CSCI 339: Distributed Systems

Syllabus for Fall 2009

General Info

Instructor:	Jeannie Albrecht
Email:	jeannie@cs.williams.edu
Office:	TCL 304
Office Hours:	MTh 1:00 - 2:30 or by appointment
Class Meetings:	TTh 11:20 - 12:35 in TCL 206
Webpage:	http://www.cs.williams.edu/~jeannie/cs339/index.html
Textbook:	*Distributed Systems, Principles and Paradigms*, by Tanenbaum and Van Steen, 2nd ed.

Course Description

This course studies the key design principles of distributed systems, which are collections of independent networked computers that function as single coherent systems. Covered topics include communication protocols, processes and threads, naming, synchronization, consistency and replication, fault tolerance, and security. We also examine some specific real-world distributed systems case studies, ranging from the Internet to file systems. Class discussion is based on readings from the textbook and research papers. The goals of this course are to understand how large-scale, distributed computational systems are built, and to provide you with the tools necessary to evaluate new technologies after the course ends.

Grading Details

Grades will be computed as follows:

5%	Class Participation
10%	Written Homework
45%	Programming Projects
20%	Midterm Exam
20%	Final Project

Each of these items are explained in detail in the following sections.

Class Participation

Lectures are mandatory and you are expected to attend regularly. One goal of this course is to promote discussion of the issues in distributed systems among all class members. As such, you are encouraged and expected to ask questions, point out weaknesses, and make observations during class.

Written Homework

The homework portion of your grade will be determined by written evaluations of a number of technical papers that we will read during the course of the semester. For each assigned paper, you will submit a 1–2 page evaluation that includes: (i) a high level summary that highlights the most important points addressed by the paper, (ii) the most glaring problem with the paper, and (iii) the conclusions you draw after reading the paper about building distributed systems. Evaluations must be turned in at the beginning of class on the day the paper is assigned. Late homework will not receive credit for any reason.

Programming Projects

There will be approximately 3 programming projects (in addition to the final project) during the semester. Details will be available on the course webpage. You get 3 flex days to use on programming projects during the semester. You can allocate these days in any way you see fit. For example, you can turn in one assignment three days late, or three assignments one day late. Beyond these three flex days, you will be penalized 25% of their value for each day late.

To complete the projects in this course, you will need the ability to develop software programs using Java and C/C++. If you have not used C/C++ recently (or at all), you may want to refresh your knowledge using one of the many good books on the topic. In particular I recommend the classic, *The C Programming Language*, by Kernighan and Ritchie, because it is short and simple. I have a few reference books in my office that you are welcome to use. There are also many online tutorials that Google can help you find. If you feel that need extra help, please come see me.

We will primarily use the Computer Science Department's Linux computers for the programming projects. If you are not familiar with the Linux/UNIX computing environment, talk to me as soon as possible so we can bring you up to speed on what you need to know.

Midterm Exam

There will a written midterm examination in this course. The exam will be closed book, closed notes, and will stress conceptual understanding of the material. Details regarding the specific format of the exam will be discussed in class.

Final Project

In lieu of a final exam (and in addition to the three programming projects described above) you will complete a final project on a topic of your choosing. I will post some sample project ideas on the course webpage later in the semester. I encourage you to be creative and pick a topic related to distributed systems that will be fun and challenging to explore. You may work with a partner on your final project. In addition to completing the programming aspect of the project, you and your partner will present your results to your classmates and write a 6–8 page paper describing your work.

Collaboration

Homework assignments and examinations are to be completed individually. You are allowed to work with a partner on programming assignments. I encourage collaboration and assistance in understanding material (especially if you have limited experience with C/C++), but not in developing solutions. Please be sure to give explicit credit for any help received. If you have any doubts about this, ask me whether or not collaboration is appropriate. Uncredited collaborations will be considered a violation of the honor code and will be handled appropriately. The computer science honor code and computer usage policy applies to all material in this class. Please review `http://www.cs.williams.edu/resources/usage.pdf`.

Calendar

The following calendar is a (very) tentative schedule of topics that we will cover in class. This schedule will likely change, so you should check the course webpage frequently for an updated calendar and the associated reading assignments. (AST is an abbreviation for your textbook.)

Date	**Topic**
Sept 10	Course Overview (AST Ch 1-2)
Sept 15	Introduction to Networks (AST Ch 3-4), *Project 1 out (AST Ch 12)*
Sept 17	HTTP, Sockets, C/C++ Overview
Sept 22	IP, TCP, UDP
Sept 24	TCP Congestion Control
Sept 29	Internet Services (AST Ch 5), *Project 1 due*
Oct 1	**NO CLASS** (Grace Hopper Conference), *Project 2 out (AST Ch 10)*
Oct 6	Resource Discovery and Search
Oct 8	RPCs and Email
Oct 13	**NO CLASS** (Fall Reading Period)
Oct 15	NTP, Logical clocks, Vector clocks (AST Ch 6), *Project 2 due*
Oct 20	Coordination and Agreement, *Project 3 out*
Oct 22	MapReduce/Hadoop, Amazon EC2, and Midterm Review, *Midterm out*
Oct 27	Semaphores and Barriers, *Midterm due*
Oct 29	Google Services: Chubby Lock Service and BigTable
Nov 3	Storage Systems and RAID
Nov 5	Distributed File Systems (AST Ch 11), *Project 3 due*
Nov 10	GFS, NFS, AFS, *Project 4 (final project) out*
Nov 12	Introduction to Wide-Area Computing, *Project 4 proposal due 11/13*
Nov 17	Security (AST Ch 9)
Nov 19	Replication (AST Ch 7)
Nov 24	Fault Tolerance (AST Ch 8), *Project 4 checkpoint due*
Nov 26	**NO CLASS** (Thanksgiving)
Dec 1	P2P Systems Overview (AST Ch 13)
Dec 3	Overlay Networks and DHTs, *Project 4 checkpoint due 12/4*
Dec 8	Project Presentations
Dec 10	Project Presentations, *Project 4 due during Reading Period*