Syllabus: CMPS 232, Distributed Systems

Computer Science Department, University of California, Santa Cruz

Winter 2010

Time: Tuesday & Thursday 10:00–11:45 AM
Location: Baskin 169
Instructor: Professor Ethan Miller
Office hours: Tue 12:00–1:00, Wed 1:30–2:30 in E2-365
Prerequisites: Undergraduate background in computer systems
Readings: Papers, available from the course home page
Home page: http://www.soe.ucsc.edu/classes/cmps232/Winter10/

Course objectives

The goal for this course is to provide an in-depth overview of research topics in distributed systems. Topics will include:

- Communication mechanisms
- Processes and mobile code
- Naming
- Synchronization
- Scalability
- Consistency, replication, and caching
- Fault tolerance
- Security and access controls

Issues in these topic areas will be illustrated through examples of real world distributed systems, including object-based systems, distributed file and storage systems, peer to peer systems, and distributed and clustered operating systems. Readings for the class will be taken from the current research literature and articles of historical significance.

Additional topics may be covered, depending on the interests of the students and the professor.

Web pages

All of the information for this class will be distributed via the Internet. The class home page is at the URL listed above, and is the starting point for lots of information about the class including readings lists, assignments, and notes. Material may be distributed either on the main class Web page, or on the course forum (linked from the home page), which is accessible from off-campus. You should create an account for the forum, and use it for discussion of general class-related topics.
Access to some of the class web pages is restricted to users on campus. In particular, papers can only be downloaded from a computer in the ucsc.edu domain. If you want to access restricted web pages or papers from off-campus, you can use the UCSC library’s off-campus access proxy (OCA), available at http://oca.ucsc.edu/. Note that OCA requires a UCSC library card/student ID.

Readings

A major component of this course will be the in-class discussion of papers on research in distributed systems. Typically, you’ll need to read two papers per class; the reading list is available online, and the papers are available as links from the reading list (though only to computers in the ucsc.edu domain). These papers should be read carefully, and a short summary of each article and a few questions or insightful comments about the material (at least 3 per paper) prepared for the class meeting in which the paper will be discussed. The summary of each article consists of brief answers to the following questions, which include three comments or questions about the paper (the last item):

- What is the problem the authors are trying to solve?
- Why is it interesting, relevant, and/or important?
- What other approaches or solutions existed at the time that this work was done?
- What was wrong with the other approaches or solutions?
- What is the authors’ approach or solution, and how does it compare to earlier approaches or solutions?
- List three comments or questions about the paper.

The optional text provides a textbook-style approach to the issues we will be covering. If your background in operating systems is less strong, you may wish to use the text as a companion to the papers we are reading. There will be no assigned readings in the text, but it’s a very useful reference on distributed systems, and many of the papers we’ll be reading are listed in the text’s bibliography.

Course requirements

Because this is a seminar, the majority of the course work will involve the reading, class discussions, and the final project.

Paper summaries

Each student is expected to write a summary of each paper, as described above, before it is covered in class. Summaries for each class are due at 10:00 PM the day before the papers will be covered in class, and must be submitted online via a form linked to the class home page. Summaries will not be graded individually; however, they will be included in the class participation grade. Students may find it useful to bring printouts of their summaries to class discussions.

Paper discussions (in-class)

We will discuss each of the papers in class; students are expected to participate actively in these discussions. In addition, each student in the class will lead the class discussion of 2–3 papers during the quarter. The exact number of papers “assigned” to each student will depend on the number of students in the class. Needless to say, you need to be in class on the that “your” paper is being discussed.
**Final project**

Students in the class must complete a research project in the general area of distributed systems. Both a paper describing the project and a poster presentation will be required. This project should be the results of experimental research (strongly preferred) or a strong survey of prior art in a focused area. Reporting work done for another course is not acceptable, unless first cleared with all instructors of the involved classes (including CMPS 232). If you wish to get an A in the class, you must do an experimental research project; a student who does a survey will receive a maximum grade of B.

There will be checkpoints every two weeks during the quarter to keep you on schedule to complete your project. The checkpoints are:

- **Jan 19** Project topic chosen
- **Jan 28** Bibliography complete
- **Feb 4** Research plan complete, implementation started
- **Feb 18** Implementation well underway
- **Mar 4** Implementation complete (or nearly so), experiments planned
- **Mar 15** Paper due (Monday at 8:00 PM)
- **TBD** Poster presentation (during finals week)

For each checkpoint, you’re expected to submit a few paragraphs describing your progress. These checkpoints won’t be graded (except for the last two, of course), but you’ll get feedback on them. The purpose of the checkpoints is to ensure that your project is on time; if you miss a checkpoint, it’s less likely that you’ll be able to finish your project by the end of the quarter. More details and a list of project suggestions will be available during the first week of class.

Your final paper should be similar to a conference / journal paper. This means that it should be about 8–12 pages long, including figures and references. Formatting instructions will be available online. I strongly recommend using \LaTeX to format your paper, though it’s not a requirement.

Posters will be presented during a 2–3 hour session to be held during final exam week (March 16–19). The final exam for CMPS 232 is scheduled for Tuesday, March 16th from 4–7 PM so, if we can’t agree on a different time in the first week or two of classes, we’ll go with that slot. During class, we’ll discuss techniques for preparing a good poster and good poster presentation skills. Posters can be printed through the School of Engineering; the cost is minimal (or free) if you submit your job far enough in advance (about a week), but goes up if you submit it closer to the time you need it. Obviously, you’re welcome to use off-campus poster printers as well, but they typically charge around $100.

**Attendance**

Class attendance is mandatory. Because this is a graduate class, I expect students to participate actively in class, and that’s hard to do if you’re not actually there. I won’t take attendance at class (except as necessary to make the registrar happy), but you cannot pass if you miss too many classes. If you need to miss a class for a good reason, such as a conference or other research-oriented commitment, please see me in advance if possible.

Much of the course material, including assignments and lecture notes, will be posted on the class web pages. However, you’re responsible for all material covered in class, whether or not it appeared on the Web site. If you miss a class, I suggest you ask either a fellow student or me to fill in any material you may have missed.
Grades

Your grades will be determined as follows:

- Final project: 50%
- Class participation and summaries: 30%
- Paper presentations: 20%

You must turn in a final project in order to pass the class.

Collaboration vs. cheating

This is a graduate seminar—I expect that you will discuss the material with other students in the class and perhaps others outside the class. This is encouraged. However, all paper summaries must be your own. If you turn in a summary written by someone else, that summary will not count towards class participation, and it will cancel one additional summary turned in over the quarter. If this happens more than once, you will fail the class and formal cheating procedures will be begun.

You’re encouraged to use any resources (code, traces, etc.) you can get for your projects, as long as you properly attribute them in your paper. Science is built on properly crediting those whose work you use; failure to do this will not be tolerated. Improper use of others’ work in your project will result in a failing grade in the class, but it’s not as bad as improperly using others’ work in your own (independent) research, which can earn you (relatively) permanent disgrace. In short, attribute everything! If you’re not sure about the ethicality of something, feel free to talk to me before you do something.