

Syllabus and Policies for CMSC 818s: Graduate Operating Systems

Neil Spring

Fall Semester 2011

Instructor Neil Spring

E-mail nspring@cs (include “818” in your subject line)

Phone x52909, only useful if you want to know whether I’m in my office.

Class TuTh 3:30pm - 4:45pm (CSIC 2107)

Office hours By appointment (email the night before) or drop by 4133

Web <http://www.cs.umd.edu/class/fall2011/cmsc818s>

1 Goals of this course

The primary goal of this course is to prepare you to collaborate with systems researchers. This means being able to have an intelligent conversation when you interview, being able to read and comment on papers written by your friends in networking and systems, and having an eye for systems research problems that appear in your own work.

Students already involved in networks and systems research should take extra care to understand not only how the topics covered in this course relate to their area of interest, but also to master the style of writing good systems research papers. That is, while I expect all students to comprehend the insight and contribution of the papers we read in this class, I expect systems students to pay particular attention to the evaluation methods, the organization of the papers, and how to mimic both.

2 General Topics

We will read research papers in the following domains.

- Prehistoric and classic operating systems: THE, hydra, Mach, etc.
- Scheduling and threads: Lottery, scheduler activations, etc.
- Influential and modern file systems: Fast file system, log-structured file system, Google file system, etc.
- Protection and virtualization: spin, multics, xen, singularity, etc.
- Security: Logic of authentication, TrInc, etc.
- Distributed systems: TreadMarks, Lamport clocks, Porcupine, MapReduce, etc.

Course background content (for the midterm) will include the following general operating system knowledge.

- General terminology: processes, threads, files, permissions, capabilities, etc.
- Synchronization primitives: semaphores, monitors, condition variables, etc.

- Processor and I/O scheduling: multi-level feedback queues, elevator algorithm, etc.
- Virtual memory: inverted page tables, hierarchical page tables, etc.
- File systems: FAT, journaling, etc.
- Interprocess communication: sockets, pipes, shared memory, etc.

3 My philosophy of a systems course

My role in this course is to facilitate your learning: lecture will not be the primary means of learning. You will learn most by reading, thinking, listening to your peers, and doing.

There are differences between graduate systems classes and undergraduate classes. First, you are grown-ups, responsible for your own learning and motivated to pursue it, so the class can spend less time on textbook-described details. Second, you are researchers, who evaluate ideas based on the problems they solve, their novelty, the completeness of the solution, etc., rather than on market success or the importance of skills. The result is that we spend more time evaluating new and potentially bad ideas rather than established solutions.

4 Background Reading

If you find yourself unfamiliar with the vocabulary of the course, have not yet taken an undergraduate networking class, or would like to make sure you understand the concepts in a textbook-like presentation, get a textbook; we use Silberschatz, Galvin and Gagne, and there are likely to be several copies floating around the department. Tanenbaum's book is well-regarded too. W. Richard Stevens "Advanced Programming in the UNIX environment" is generally useful for practical matters.

These are *recommended*, but not required (I will assign no readings and no exercises from these books). It would be reasonable to share a copy with a friend or borrow an older edition. Stevens's book is a true classic; I refer to it from time to time.

5 Grading

The grading of this course will be based on the following criteria. In this section, I present both the approximate percentage breakdown and a description of the criteria.

5.1 Class participation (10%)

I intend to conduct class time mostly in discussion. Being part of the academic community means making well-reasoned arguments about a paper's quality or lack thereof: it helps you get papers you like accepted by program committees and helps you to gain standing by asking insightful questions at workshops and conferences.

If you expect to sit in the corner and not say anything, you will miss out on these points. Do not expect an A.

5.2 Presentation (15%)

You will lead the discussion of a paper or two in a class meeting in the second half of the semester. Leading the discussion means generating some slides or notes, calling out strengths and weaknesses of the paper, providing a forum for students to discuss their on-blog comments, describing follow-on work or the relationship between the paper and related work, etc.

The criteria for evaluating your presentation will be in how well you facilitate discussion, how well you are able to explain how such a paper made the reading list, whether the main discussion topics have visual anchors (for example, you've extracted the key graphs of the paper for slides), etc. I expect you to be able to do a better job than I do with the paper you present.

5.3 Background Homework (5%)

I have four homeworks prepared. They cover material that should refresh undergraduate experience or shape self-study. These are mostly to prepare you for the midterm exams.

On a homework, if I ask a specific question, provide a specific answer. If it looks like you're paraphrasing a paragraph from wikipedia or a textbook without processing that information enough to understand, even if the answer is embedded somewhere within the long response, I will consider it incorrect. You've been warned.

I don't yet have a great plan for grading homeworks; please expect that only a subset of questions will be graded. The main goal of the homework is to prepare you for the midterm.

5.4 Reading-related "Blog" participation and exercises (20%)

In my experience, students don't read papers unless there is some graded exercise involved. However, full "reviews" of papers are too long to be evaluated, focus on minutiae, and encourage harsh evaluation of what are, on the whole, pretty good papers.

Instead we have comments on the blog. One point for a reasonable effort at commenting. Two points for something insightful. (Three points for unmatched brilliance.) The distribution of insightful comments tends to be maybe four per post. That is, very few. You may not get them often.

The papers will be listed at: <http://scriptroute.cs.umd.edu/os818-fall11/>. Please use the comment feature to ask questions or post opinions before class. I hope to accomplish two things with the blog comments. First, for students more comfortable expressing themselves in writing than vocally in class, the blog gives an opportunity to gain the respect of peers without "risking" class time. Second, I hope that it will support better use of class time by skipping directly to the interesting questions or to points of widespread confusion.

You should be able to subscribe to an RSS feed for course announcements within the blog system, though this feature is not yet ready. I will use that in lieu of a mailing list when posting new assignments, extending deadlines, etc.

The software is custom, primarily to permit authentication via UMD's service without sending my server your password. Please check that your posted comments are stored. A monster log is kept, so please don't tell me the blog ate your homework. :)

In the comments, I'm looking for is evidence of significant thought, incorporating personal experience, evaluating the paper in a new way (in light of 20 years after it was written, for example), or criticism (a fatal flaw). Don't trust authors! Make the authors convince you. For example, words like "preliminary" or "we do not address" are hints that there are really difficult problems just beyond those in the paper.

Examples:

Good This paper is awesome because the problem of sprocket wobble affects ordinary users in oceania and eurasia.

Poor The authors address sprocket wobble, which they say is a very important problem.

Connections to other research, discussion material in the class, comments by other students, etc.

Good This approach violates the end-to-end arguments because it places unnecessary functionality in the network, and it violates the 'tussles' argument by choosing an outcome in favor of repressive governments over the interests of ISPs and users.

Poor I think this is too complicated and will never be deployed.

Clarity and grammar, with simple points, supported by declarative sentences in active voice. Less is more: think silver bullet, not shotgun. For me, DON'T yell (you might enclose phrases in asterisks instead). Break paragraphs for new ideas.

Good This idea is worthwhile, but not for the reasons the authors claim. An operator could use it to plan backhaul upgrades, which they find difficult (Bolot SysChat). Traffic engineering on the backbone network is child's play.

Poor The idea described would be better applied toward backhaul links. Provisioning the capacity of backhaul links is a hard problem.

I want to see that you're starting down a path that would lead toward new research agendas of your own: finding (big, architectural) limitations, finding new applications of the idea, finding the heart of the problem the paper addresses incompletely, or proposing a new and better technique. It's very hard to make this jump without thinking about the papers for a while, and it's pretty hard to explain a new idea in the blog conversation without focusing the post around that new idea.

I know this level of comment won't always be possible. Do your best.

Finally, I'm a bit picky about style and grammar. I'm particularly annoyed by misuse of effect, putting too many o's in lose, using passive voice, writing 30-word sentences, using too many parentheses, and the word 'respectively.'

One more version of my criteria:

Some hints on comments. What I'm looking for is for you to show that you've read the paper with an eye toward future research. To that end, the questions you might answer are:

- Is this a real problem? Is this the real problem? Did they solve it? (Leading to another problem that you might want to solve, or a better solution to the problem the paper describes.) I think there are such arguments with both papers this week.
- What was it about the paper that made it worth talking about? What trick did the authors play to get the paper accepted despite its limitations? (We'll read mostly conference papers this semester; they will all have weaknesses.)
- Do you believe them? Don't tell me the simulation is unrealistic (it always is); tell me why it matters that their evaluation wasn't good enough. (An anechoic chamber is not real, but it seemed like enough to me.)
- Would you buy it? If I were a farmer looking for radio-based sensors, wouldn't I just use ISM 900 MHz and some directional antennas?
- Could you apply their solution to another, similar problem? Or is there a similar problem with a different solution? WhiteFi is about decentralized spectrum sharing, and decentralized resource sharing is the subject of lots of networking protocols. It is a little surprising that their solution is so distinct.

What I know you can do, but which I am not too interested in, are the little doubts. The "on page 3 I don't understand why they said that violets are blue when they're clearly violet" style. I don't mind if you raise them, and if you post your doubts in the comments maybe someone else can help to explain. But fixing those little issues doesn't make a paper. You might point out an invalid, substantial assumption. I want you to deny the authors their premise or conclusion, or grant them everything and suggest the next step.

5.5 Course projects (25%)

The course projects will be done in pairs. You may not use the same partner twice. You will create a user-level file system using FUSE, create a kernel module for linux, and use an instrumentation toolkit to virtually plan performance tuning upgrades.

5.6 Midterm exam (10%)

There will be a midterm exam, covering basic operating systems material that you should understand based on homeworks, readings, and lecture, if not from reading an undergraduate textbook on your own or remembering from an undergraduate class.

The midterm is scheduled for Tuesday, Oct 25. If this is a problem with your travel schedule, please inform me.

5.7 Final exam (15%)

There will be a take-home final. I will be looking for creativity and synthesis, not recall.

The final exam period for the class is Dec 21, ending at 12:30pm. The final exam will be due at that time, likely distributed around noon on December 18, giving you three days.

6 Lateness

All assignments can be turned in electronically. There will be no late turns.

There are exceptions for excused absences defined by the university. If you're sick, get a note. If you need accommodation for disability, illness, family emergency, religious observance, etc., ask me *early*.

If you're traveling to a conference and need to know what you'll miss, so that you can get a head start before you leave, ask; I can likely tell you.

7 Collaboration

Reading assignments are to be done individually. Exams are to be done individually. Individually means without collaboration with other students, without asking questions of outsiders, on mailing lists, of senior graduate students, etc. If in doubt, talk to me. If you want to bounce ideas off other students in the comments, use the reply feature or ask questions in your comment on the course blog.

Course projects are to be done in pairs; while I do not expect all group members to be present for all work, I do expect each group member to know what the others are up to. If you split up the work too coarsely (one person writes the report, another gives the presentation, etc.), I will be disappointed. You are welcome to solicit the advice of senior grad students, faculty, friends, family, pets, psychics, and whomever else might help you in your quest. I will get grouchy if I see that you've sent a FAQ to a mailing list, so be careful when you represent UMD publicly.

In code you write, if you found the skeleton elsewhere, cite your source. A url in comments is sufficient.

8 Grading

You may see your scores for individual assignments on <https://grades.cs.umd.edu/>.

Many students incorrectly interpret their progress on grades.cs relative to other students ("I'm above average, so I must pass") or relative to an absolute ("90% is an A") scale. Understand that "average" scores are often held back by students who may have abandoned the class and that I do not update the grades or their weighting in real time. In other words, the information available to you on grades is unlikely to be sufficient for you to predict your grade.

9 Fall 2011 Specifics

I may not grade every assignment. I will choose somewhat at random which ones I will grade, and will make more of an effort to grade the first week's assignments early.

I am also likely to reuse assignments. Note that consulting solution keys or the assignments as issued in previous semesters will be considered cheating.

10 Master's Comps

CS graduate students: go sign up with Fatima for your Master's comps credit. Even if you don't think you want a Master's (it is very difficult to predict what may get in your way in the next few years). If I'm asked after the semester is over, I will take a long time to answer.