The topics

Computer Systems I and Computer Systems II covered the structure of hardware and software systems, building from simple components to complete, general purpose environments. For the operating system in particular, software structures could hide the hardware from programmers, providing a simpler, abstract programming interface while handling the sharing of computing resources across processes and users. These courses provided demystification, showing how complex systems structures such as an operating system could be built.

Of course, real operating systems are more complex, and operating systems research goes far beyond the question of structure. In this course, we will address problems of optimization. Operating systems are expected to perform substantial work using little space and even less time. They must allocate CPU time, memory, disk space, disk and network bandwidth, and any other physical resource to each process. Moreover, those allocations must be dynamically recomputed, with new processes starting, old ones ending, and resource demands from each process changing. In short, the operating system must approximate the “ideal” allocation of these resources at every moment, guessing the future resource needs of each process from only their past demands.

Thus, real operating systems are complex structures with complex behavior. Worse, measuring that behavior is inherently difficult. Small changes in the system’s structure can change its performance. The best policy may be implemented badly, making an evaluation of those policies difficult. The choice of which policies and algorithms is “best” depends on the workload (the programs run and the inputs provided to them), as well as the details of the hardware. It is even difficult to describe what the desired behaviors of a system are, balancing performance, fairness, and power efficiency.

For a number of topics, we will explore not only the real policies that are applied to solve a given predictive problem, but also the methods used to evaluate those policies. Specifically, for a number of classical operating system allocation problems (e.g., CPU scheduling, main memory allocation), we will read original research literature, examine their experimental methods, and compare policies. We will examine how these idealized policies are altered for real-system implementation. Finally, we will conduct some experiments of our own, both in simulation and on real systems.

Lectures and labs

This class meets on Monday, Wednesday, and Friday of each week, from 9:00 am to 9:50 am. Lectures will be held in Seeley Mudd 202, while the occasional lab will be held in Seeley Mudd 007. I will announce labs days on the course web site when they occur.

You are expected to be present for all of the lectures and labs, and so missing either is strongly discouraged. I will not teach material twice, so if you miss a lecture or a lab, you’re on your own. If you must miss lecture or lab due to an illness or other emergency situation, contact me and we will make some kind of accommodation. If you have a conflict with a lecture or lab for any non-
emergency—an athletic event, a performance, a vacation—then the choice is yours to miss or to attend. If you choose to miss the class meeting, I do not want to know why nor even that you are missing class. You have elected, voluntarily, not to attend, and you must be prepared to obtain and learn the material that you missed on your own. I, of course, recommend that you choose to attend the class meeting when these conflicts arise. Do not underestimate the willingness of those who run extra-curricular programs to make accommodations for your academic demands.

I expect you not only to attend lectures and labs, but also to be attentive for them. The time will be best spent if it is interactive, and that requires that you be up-to-date on the class material, and that you be alert and prepared to participate.

3 Texts and materials

All materials for this class will be posted on the course web page as needed. There is no textbook, and readings will primarily be drawn from research papers.

4 Assignments, deadlines, and extensions

There will be a number of projects. The deadline for each will be stated clearly on the assignment, down to the minute. The assignment will also state the manner in which you are expected to submit or show your work. Late submissions will receive failing grades. Furthermore, failure to complete any one of the projects may result in a failing grade for the course. These assignments are too important to the course not to be completed.

5 Academic dishonesty

You will be expected to do your own work on all assignments and exams in this course except where explicitly noted on group assignments. While I encourage you to interact with your classmates and discuss the material and assignments, there is a limit to the specificity of such discussions. I seek to make that limit clear here.

It is acceptable to discuss any assignment for the class with a classmate. You may even discuss your approach to a particular problem, or review relevant material for a problem with another person. However, you may not show another student your work, nor see another student’s work. If in doubt, ask me. If you are unsure whether or not a particular kind of communication would rise to the level of academic dishonesty, then you should contact me immediately and find out.