INLS 576 - Distributed Systems

Erik L. Scott
escott at renci dot org

Office Hours - Tuesdays 10:00-11:30am, Sitterson SN239
Other times as needed.

Course Description: This class will cover topics necessary to understand and design systems of cooperating computers and other devices. The emphasis is on a breadth of topics rather than depth in any one topic - this is very much a survey class. By the time this course is complete, you should be capable of taking a systems approach to the design and engineering of network-based computing.

The general plan for the course centers around a collection of class readings and discussions of these papers and chapters by you, the students. The first half of each class will focus on the readings, and the second half will be used for lectures, demonstrations, and an introduction to the next week's reading. There will be four programming assignments over the course of the semester. Assignment Zero will be a warm-up exercise intended largely to ensure that your programming environment is installed correctly and that you have some very basic capability in the Python programming language.

Prerequisite: INLS 261 or 461. Some experience with Python will be helpful, but not essential. We'll only be using a small part of what Python can do, so you can pick it up on the fly. A very basic understanding of databases will also be useful, but again we're only looking at the rudiments here.

Office Hours: I work full time for UNC, but my office is off campus at 100 Europa Drive, Suite 540. You're welcome to meet with me there any time you want to, but I strongly suggest you email and find a time when you know I'll be there. We have plenty of free parking, and we're on the D bus line. Alternatively, my regular office hours on campus are Tuesdays from 10:00am to 11:30am in Sitterson room 239, and I can meet with people on campus fairly easily.

Grading Schema:

Programming assignments:	30%
Midterm:	25%
Final:	30%
Class Participation:	15%

Grade Assignment:

Graduate Percentage	Undergraduate Percentage
H 100-95%	A 100-90%
P 94-80%	B 89-80%
L 79-70%	C 79-70%
F Below 70%	D 69-60%
	F Below 60%

Honor Code:

The UNC Honor Code applies to all work in this course. When work or ideas are not your own, you must attribute them. Unless otherwise stated, all assignments in this class are individual assignments, meaning that the substance of the work you turn in must be your own. If you have any doubts or questions about a course of action or a specific situation, please ask for clarification. Students should NOT receive (or give) major creative assistance or ongoing minor support on individual assignments, unless permitted explicitly. For more information, see https://studentconduct.unc.edu/ . If you have any questions about this, please ask me.

Special Accommodations:

The University of North Carolina – Chapel Hill facilitates the implementation of reasonable accommodations, including resources and services, for students with disabilities, chronic medical conditions, a temporary disability or pregnancy complications resulting in difficulties with accessing learning opportunities.

All accommodations are coordinated through the Accessibility Resources and Service (ARS) Office. In the first instance please visit their website http://accessibility.unc.edu, Tel:- 919-962-8300 or Email;- accessibility@unc.edu. A student is welcome to initiate the registration process at any time, however, the process can take time. ARS is particularly busy in the run-up to Finals and during Finals. Students submitting Self-ID forms at that time are unlikely to have accommodations set until the following semester.

Please contact ARS as early in the semester as possible.

Mutability: This syllabus may be updated at any time, particularly as course reserves and other readings are established, or if adverse weather forces a change in class schedule. I will (of course) follow the schedule of the university, and the university will announce any closings via AlertCarolina. Make sure you can receive these messages.

Schedule:

Jan 13 - Welcome, Introductions, and The Introduction. (wk 1)

Introductions
Administrative Details
A whirlwind tour of what we're going to be covering this semester.

Jan 20 - How Operating Systems Work (The Short Form) (wk 2)
Caution: Basketball Traffic.

Reading and Discussion:
Silberschatz, Abraham. *Operating System Concepts Essentials*. John Wiley & Sons Hoboken NJ, 2011, 3-46.

Ritchie, D. and Thompson, K. 1978. The Unix Time-Sharing System.*The Bell System Technical Journal.* 57, 6, part 2, (July/Aug) 1905-1929.

http://plan9.bell-labs.com/sys/doc/9.html - Plan 9 from Bell Labs, an actual distributed operating system.

Lecture - History of OSes, facilities they provide, detailed looks at filesystems and network stacks.

Demos - An examination of the Unix process tree, filesystem, and the basic network facilities.

Jan 27 - Quick (Review?) of Python, Postgres, Virtualization (wk 3)

https://docs.python.org/2/tutorial/ - read first 5 sections, quickly skim the rest so you know where to find things.
http://db.cs.berkeley.edu/papers/ERL-M87-06.pdf - Stonebreaker's review paper on the design

Cafaro, M. & Aloisio, G. (2011). Grids, clouds, and virtualization. In Cafaro, M. & Aloisio, G. (eds.), Grids, clouds, and virtualization (pp. 1-21). London: Springer-Verlag.
`

Demo / hands-on time:
1) Python Programming
2) Postgres installation
3) Postgres use, and where the docs are
4) Virtualization - VirtualBox, install a Linux and a Windows instance

Assignment 0: Getting our feet wet.

Feb 3 - More on Our Friend, The Operating System (wk 4)

Reading and Discussion
- Silberschatz Ch 16.

Lecture and Demos -
OS and Resources - planning, monitoring, optimizing
Local network setup
Telnet between hosts
Wireshark Watching Over Me
Installing VirtualBox and Ubuntu. Quick tour of Ubuntu, and resources for basic linux survival.

Assignment 0 due.

Feb 10 - Networking, Top to Bottom (wk 5)

Reading and Discussion
- Appropriate Chapters from Tanenbaum and from Stevens

Assignment 1: It's like Twitter, only it does less.

Feb 17 - Network Protocols, or Lingua Franca (wk 6)
Caution: Basketball Traffic.

Ad Hoc stuff crammed through a socket - simple, but has limits
XML - sacrifice performance for ease of comprehension.
JSON - it's like XML, but even simpler, and is rapidly becoming the standard interface for anything web related (aka, "anything anyone cares about")
Packages to support this in Python.

Feb 24 - Message Queues and Distributed Filesystems (wk 7)

ftp://ftp.software.ibm.com/software/integration/wmq/WS_MQ_Messaging_Backbone_for_SOA.pdf - very much an introduction

http://www.redbooks.ibm.com/redbooks/pdfs/sg247128.pdf - redbook on MQSeries

Silberschatz Ch 17 - Distributed Filesystems
Tanenbaum on Distributed Filesystems?

Queueing in General - what it is and where AMQP, MQTT, and STOMP fit in.
MQseries
RabbitMQ
Service-oriented Architecture

- As web services - SOAP
- As queued services - MQSI

File Services

- NFS
- WebDAV
- AFS

Assignment 1 due.

Mar 2 - Web Servers and Clients (wk 8)

Reading -
W3C on HTTP (1.0, 1.1 extensions, maybe an overview summary of 2.0)
Ramakrishman and Gehrke on Web Programming

HTTP protocol
A fully worked example - old school web apps
Web 2.0 - shiny shiny

- XML Data Islands
- ...begat JQuery

Programming Assignment 2: Writing *The World's Simplest Web Server*™.

Mar 9 - Midterm and Smartphone Apps with Apache Cordova (wk 9)

Mar 16 - Spring Break.

Mar 23 - Secrets, Lies, and Commerce - Cryptography In Action (wk 10)

Readings - Schnier Ch 1.
W3C protocol spec for HTTPS
Kerberos V4 paper and dialog.

Discussion and Demos:
Foundations - First Part of Schneier.
HTTPS
Kerberos

Assignment 2 due.

Mar 30 - Databases, Distributed Databases, and Distributed Database Apps (wk 11)

Reading and Discussion

- Google Pregel paper
- And why not add a Paxos discussion? Maybe the paper? Maybe something a lot shorter and easier?
- Ramakrishnan and Gehrke chapter on ODBC and similar
- ODBC tutorial

Lecture and Demos -
ODBC & JDBC - the standard way to connect a running process to a database server
A Distributed Database - mySQL Cluster

Assignment 3: The First National Bank of Fuquay-Varina.

Apr 6 - The Internet of Things (wk 12)
The Hype Cycle
Motivating Problems
Thinking Small - Embedded Computing, and the spectrum of small
Operating system design for tiny stuff
Security - Comcast burglar alarm, Jeep p4wnage, and how to mess with uranium enrichment.

Demo - TBD.

Apr 13 - High Performance Computing (wk 13)
Motivations - Analytics, IoT and Wall Street
High Frequency Trading
Streaming Databases - System S and one or more competitors
Capacity Systems vs. Capability Systems
Paper - Condor Retrospective (a Capacity System)

Assignment 3 due.

Apr 20 - Computing at Datacenter Scale (wk 14)
Google Datacenter video
Google Map/Reduce paper
Mesos
Berkeley/Apache Spark paper (and Spark Streaming)

Field Trip to Genome Sciences -

Apr 27 - Review Day (wk 15)

Final Exam -
An open-notes, open-readings short essay test to be completed over any 24 hour period from 11:59pm Apr 27 until 9:00pm May 2nd. Any material from the semester is fair game. Expect more synthesis of material from different parts of the course and less recall of factoids.