CSCI 5105
Introduction to Distributed Systems

Instructor: Abhisheek Chandra
General Information

- Class: Tu Th 4-5.15pm
- Office Hours: Tu Th 3-4 pm
  - Location: KHKH 4-209
- Teaching Assistant: Kwangsung Oh
  - Office Hrs: TBA

- Course Web page:
  http://www-users.cselabs.umn.edu/classes/Spring-2016/csci5105/

Also a moodle page (for forum, HW submissions, etc.)
Course Structure

- **Instruction:** Primarily Lecture-based

- **Text:**
  - **Required:** “Distributed Systems: Principles and Paradigms, 2nd Ed.”, Tanenbaum and van Steen
  - Selected Research Papers

- Weekly readings from the textbook(s), lecture notes and additional reading material
Course Work

- 3 Programming Assignments (45%)
  - Work in groups of 2
  - Due in ~2 weeks
- 3 Written Assignments (15%)
  - Based on course material
  - Due in a week
- Exams (35%):
  - 1 Mid-Term (15%)
  - 1 Final (20%)
- Class Participation (5%)
Programming Assignments

- Implement specifications provided
- Systematic evaluation for performance, tradeoffs

Assignment Submission:
- Provide full code, header files, makefiles, test-files
- Report: Describe program design, and include an evaluation (of algorithm, system, etc.)

Online submission by 11:59 pm on due date
- One submission per team (of 2)
- Late penalty: 10% first day, 15% per extra day (open to change under certain circumstances)
Programming Assignments (contd.)

- **The submitted code should be original**
  - Do not copy or derive from the Web or other external sources (e.g., prior offerings, senior students, programmer friend, ...)
  - No sharing of code across teams
- Discuss and ask questions on class forum, from TAs or instructor
- Grading: Points for
  - Functionality and Correctness
  - Program Design
  - Documentation and Code readability
Written Assignments

- Based on concepts discussed in previous 3-4 lectures
- Have to be done *individually*
  - Not with your project teammates
  - All answers must be original
  - **Do not** copy or search for solutions from others, Web, etc.
- Due at beginning of lecture on due date
Exams

- Mid-Term exam would cover the material of first half of the course
- Final exam will be comprehensive
- Closed notes/closed book
- No electronic devices allowed
Class Participation

- Engage in class
- Ask questions, answer to queries, initiate and respond to discussion
- Also use the Class Forum
Class Discussion Forum

-On Moodle
-You can post questions, discuss topics, course material
-Try responding to each other as far as possible
-Instructor, TA will regularly monitor the forum
-Please avoid:
  - Irrelevant mails, flame wars
  - Posts that break the rules/spirit of honesty
**Academic Dishonesty**

- **What does it include?**
  - Copying assignments, cheating on exams, plagiarism
  - Written homework must be done by yourself – do not copy from textbook, web or others
  - Code should be original (not copied or derived from the web or other sources)
  - Providing help is also considered cheating

- **Can result in serious consequences:**
  - Can range from 0 on assignment to F in class or worse
  - U requires report to Office of Student Affairs

- See Dept. Academic Conduct Policy on class website

- If unsure, just ask!
Disability Statement

- If you have, or think you have, a disability, contact Disability Services.
- Please get a letter from DS for any special accommodation request on course work.
- I will try my best to make the required accommodations.
UNITE Mechanics

- Lecture available on streaming video
  - Live to off-campus students
  - With 10 days delay to on-campus students
- Off-campus students can phone-in
- Assignments to be handed to UNITE coordinator
  - Timestamped by due date/time
- Exam can be given on-campus or arrangement with UNITE coordinator
Course Objectives

Goals:

- Study concepts that build the foundations of large-scale systems (Web, Grids, clouds, ...)
- Learn from case studies, example systems
- Get exposure to system building and distributed systems research

At the end of the course, you should:

- Understand distributed system concepts that are used in a wide variety of systems
- Learn about tradeoffs when building large systems
- Apply learnt concepts to your job/research
Pre-requisites

- CSCI 5103 or equivalent
  - Knowledge of general operating system concepts
    - Scheduling, virtual memory, file systems, synchronization, ...
- Programming experience
- Good understanding of data structures and algorithm fundamentals
- Some knowledge of networking is useful
- **Important:** Solve HW0
Topics

- Distributed Communication and Naming
  - RPC, message-passing, group communication
  - P2P systems, Distributed Directories

- Distributed Synchronization
  - Physical and logical clocks
  - Mutual exclusion, election algorithms

- Data Consistency and Replication
  - Consistency models and protocols
  - Data replication, distribution and caching
Topics (contd.)

- Reliability and Fault Tolerance
  - Failure recovery
  - Reliable communication and Agreement
- Distributed Computing and Storage
  - Cluster and Wide-area computing
  - Data-intensive computing
  - Distributed File Systems
- Protection and Sharing
  - Virtualization and Sandboxing
  - Cloud computing
Distributed Systems
Examples of Distributed Systems

- Web
- File-sharing
- Scientific computing
Distributed System

- “A collection of independent computers that appears to its users as a single coherent system”
- Hardware view: Multiple independent but cooperating resources
- Software view: Single unified system
Benefits and Problems

- Benefits?
- Problems?
Distributed Systems: Goals

- Sharing
- Transparency
- Scalability
Sharing

- Multiple users can share and access remote resources
  - hardware, files, data
Transparency

- Hide the distributed nature of system from users
- Several types:
  - Location: Hide where a resource is located
  - Migration: Resources can be moved
  - Relocation: Resources can be moved while being used
  - Replication: Multiple copies of same resource can exist
  - Failure: Hide failures of remote resources
Scalability

- Allow the system to become bigger
- Multiple dimensions:
  - Size: Adding more resources and users
  - Geographic: Dispersed across locations
  - Administrative: Spanning multiple administrative domains