CSCI 5105
Introduction to Distributed Systems

Instructor: Abhishek Chandra

General Information

- Class: Tu Th 8.15-9.30 am
- Office Hours: Tu Th 9.30-10.30 am
  - Location: KHKH 4-209
- Teaching Assistant: Kwangsung Oh
  - Office Hrs: TBA
- Course Web page:

http://www-users.cselabs.umn.edu/classes/Spring-2019/csci5105/

Also a Canvas page (for announcements, forum, HW submissions, etc.)

2

Course Structure

- **Instruction:** Primarily Lecture-based
- **Text:**
  - **Required:** "Distributed Systems, 3rd Ed.", van Steen and Tanenbaum
  - Selected Research Papers
- Weekly readings from the textbook, lecture notes and additional reading material

3

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

4

Pre-requisites

- Good understanding of operating system concepts
  - CSCI 5103 or equivalent
  - Topics: Scheduling, virtual memory, file systems, synchronization, ...
- Basic knowledge of networking:
  - TCP/IP, sockets, protocol stack, routing, naming
- Good programming experience
- Good understanding of data structures and algorithm fundamentals

5

Course Work

- 3 Written Assignments (15%)
  - To be done ***individually***
  - Due in a week
- 3 Programming Assignments (45%)
  - Work in ***teams of 2***
  - Due in ~2 weeks
- Exams (40%):
  - 1 Mid-Term (15%)
  - 1 Final (25%)

6

Programming Assignments

- Preferred: Teams of 2
- 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
  - Late penalty: 10% for <24 hrs, +30% each extra day (open to change under certain circumstances)

7

Programming Assignments (contd.)

- ***The submitted code must 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
- Ask for clarifications on class forum, from TAs or instructor
- Grading: Points for
  - Functionality and Correctness
  - Program Design and Evaluation
  - Documentation and Code readability

8

Written Assignments

- Based on concepts discussed in previous 3-4 lectures
  - Goal is to test your understanding, practice solving problems
- Have to be done ***individually***
  - Not with your project teammates
  - All answers must be original, ***in your own words***
  - **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 Canvas
- 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
  - U requires report to Office of Student Affairs
- See Dept. Academic Conduct Policy on class website
- If unsure, just ask!

13

Disability Statement

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

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 arranged with UNITE coordinator

15

Distributed Systems

What is a Distributed System?

- "A collection of independent computers that appears to its users as a single coherent system"
- Hardware view: Collection of independent computers
- Software view: Single coherent system

17

Examples of Distributed Systems?

18

Benefits and Problems

- Benefits?
- Problems?

19

Distributed Systems: Goals

- Sharing
- Transparency
- Scalability

20

Sharing

- Multiple users can share and access remote resources
  - Computing and storage
  - Services
  - Data
- Reasons?

21

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
- Reasons?

22

Scalability

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

23

Course Topics

- Communication and Naming
  - RPC, message-passing, group communication
  - P2P systems, Directory services
- Coordination
  - Physical and logical clocks
  - Mutual exclusion, election algorithms
- Data Consistency and Replication
  - Consistency models and protocols
  - Data replication, distribution and caching

24

Course 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

25