Distributed Systems Lecture 1: Overview

CS425 / ECE428

FALL 2021

NIKITA BORISOV

Hybrid Course Logistics

Hybrid Course

- Can be done in-person or entirely online

Lectures will be:

- Presented “live” in DCL 1310
- Broadcast synchronously by Zoom
- Available for later viewing on Echo 360

Questions

- Zoom chat or in-person
- Investigating CampusWire and other methods to streamline

Objectives

Define Distributed System

Overview of distributed systems issues

Course information

Examples of Distributed Systems

Properties

Definitions

A distributed system is one in which the failure of a computer you didn't even know existed can render your own computer unusable

-- *Leslie Lamport*

Definitions

A distributed system is a system whose components are located on networked computers, which communicate and coordinate their actions by passing messages. The components interact with one another in order to achieve a common goal.

-- *Wikipedia (as of today!)*

Definitions

A collection of (probably heterogeneous) automata whose distribution is transparent to the user so that the system appears as one local machine. This is in contrast to a network, where the user is aware that there are several machines, and their location, storage replication, load balancing and functionality is not transparent. Distributed systems usually use some kind of client-server organisation.

Distributed systems are considered by some to be the “next wave” of computing.

-- Free On-Line Dictionary of Computing (FOLDOC)

Definitions

A distributed system is a collection of independent computers that appears to its users as a single coherent system

-- *Tanenbaum & Steen*

Definitions

We define a distributed system as one in which hardware or software components located at networked computers communicate and coordinate their actions only by passing messages

-- *Coulouris, Dollimore, Kindberg*

Key Properties

***Multiple* computers**

- *Concurrent* execution
- *Independent* failures
- *Autonomous* administrators
- *Heterogeneous* capacities, properties
- *Large* numbers (scalability)

***Networked* communication**

- *Asynchronous* execution
- *Unreliable* delivery
- *Insecure* medium

***Common* goal**

- *Consistency* – can discuss whole-system properties
- *Transparency* – can use the system without knowing details

Why Distributed Systems?

Scaling to large number of users and big tasks

- E.g., many of today's largest websites (Facebook, Amazon, etc.) started as a single server but grew to support billions of users
- E.g., computations that use petabytes of data and millions of CPU-hours

Collaborating across the world

- A single "system" that is accessible to users and computers around the globe
- E.g. the "Facebook system" consists of millions of servers *and* billions of apps, web browsers, etc.

Comparison – Operating Systems

***Multiple* computers**

- *Concurrent* execution
- *Independent* failures
- *Autonomous* administrators
- *Heterogeneous* capacities, properties
- *Large* numbers (scalability)

***Networked* communication**

- *Asynchronous* execution
- *Unreliable* delivery
- *Insecure* medium

***Common* goal**

- *Consistency* – can discuss whole-system properties
- *Transparency* – can use the system without knowing details

Comparison – Networking

***Multiple* computers**

- *Concurrent* execution
- *Independent* failures
- *Autonomous* administrator
- *Heterogeneous* capacities, properties
- *Large* numbers (scalability)

***Networked* communication**

- *Asynchronous* execution
- *Unreliable* delivery
- *Insecure* medium

***Common* goal**

- *Consistency* – can discuss whole-system properties
- *Transparency* – can use the system without knowing details

Note: Networks use Distributed Algorithms (DNS, BGP)

Example: WWW

***Multiple* computers – Web servers, clients**

- *Concurrent* execution
- *Independent* failures
- *Autonomous* administrators
- *Heterogeneous* capacities, properties
- *Large* numbers (scalability)

***Networked* communication – Internet (TCP/IP)**

- *Asynchronous* execution
- *Unreliable* delivery
- *Insecure* medium (HTTPS)

***Common* goal – Hyperlinked information system**

- *Consistency* – can discuss whole-system properties
- *Transparency* – can use the system without knowing details

Example: Domain Name Service

***Multiple* computers – DNS server, clients, caches**

- *Concurrent* execution
- *Independent* failures
- *Autonomous* administrators
- *Heterogeneous* capacities, properties
- *Large* numbers (scalability)

***Networked* communication – Internet (UDP + TCP/IP)**

- *Asynchronous* execution
- *Unreliable* delivery
- *Insecure* medium (DNSSEC)

***Common* goal – Hierarchical Naming System**

- *Consistency* – can discuss whole-system properties
- *Transparency* – can use the system without knowing details

Example: Bank

***Multiple* computers – ATMs, teller computers, servers, credit card scanners**

- *Concurrent* execution
- *Independent* failures
- *Autonomous* administrators
- *Heterogeneous* capacities, properties
- *Large* numbers (scalability)

***Networked* communication – Internet, local networks, modems, leased lines**

- *Asynchronous* execution
- *Unreliable* delivery
- *Insecure* medium

***Common* goal – Financial Institution**

- *Consistency* – can discuss whole-system properties
- *Transparency* – can use the system without knowing details

Course Objective

Concepts in distributed computing

- Properties
- Challenges
- Impossibility results

Designs of distributed systems

- Abstractions
- Algorithms
- Implementations

Case studies

Course Information: Staff

Instructor: Prof. Nikita Borisov

- Office: 460 Coordinated Science Lab
- Office hours: TBA

TAs:

- Dayue Bai
- Jiangran Wang
- Sanchit Vora
- Yitan Ze

Sources of Information

Course website:

- Announcements, homework, MPs,
- Lecture list, reading assignments, slides

Campuswire

- Announcements, questions, clarifications
- Can post both private & public questions
- SLA: one business day response time, hopefully faster
- Used for participation

Books

Distributed Systems: Concepts and Design, Coulouris et al., 5th ed.

- Earlier eds may be acceptable
- Your responsibility to find correct reading sections

Other texts

- *Distributed Systems: An Algorithmic Approach*, Ghosh
- *Distributed Systems: Principles and Paradigms,* Tanenbaum & Steen
- *Distributed Algorithms*, Lynch

Grade Components

ASSIGNMENTS

Homeworks

- Approx. every 2 weeks
- Must be **typed**
  - Hand-drawn diagrams OK
- Must be done **individually**

MPs (4-credit version)

- 3-4 projects
- Groups of 2
- First “warm up” project out this week

EXAMS

2 Midterms

Final

- Comprehensive
- Dates TBA

Grading Scheme

	3-credit	4-credit
Homework	33%	16%
Midterms	33%	17%
Final	33%	33%
MPs	N/A	33%
Participation	1%	1%

Grading Policy

Homework:

- ~60 points per homework, => 360 points total
- 3-credit: 330 points needed for 100%
- 4-credit: 160 points needed for 100%
- No overflow to other categories

MPs (4-credit only):

- 4 MPs worth 50, 100, 150, and 100 points, respectively => 400 points total
- 330 points needed for 100%
- No overflow to other categories

Bonus questions and assignments *may* be assigned

Academic Integrity

Academic integrity violations have serious consequences

- **At least** 0% on assignment
- Potentially additionally reduced course grade, failed course, suspension or expulsion
- All cases are reported to your department, college, and senate committee

Examples of violations:

- Collaborating on exams
- Copying homework solutions
- Sharing source code (outside group)

Cite *all* sources and use your judgment

- You *may* get 0 if you cite and use a disallowed external solution (e.g., past semester's solution)
- You *will* get 0 and a FAIR report if you use a disallowed source and don't cite it

Grading Curves

Homework and MPs are *not* curved

Each exam curved individually **up** to a B- average with 1 SD per grade

- Curved grade >= raw grade

Acknowledgments

Material borrowed from:

- Prof. Jennifer Hou
- Prof. Mehdi Harandi
- Prof. Klara Nahrstedt
- Prof. Indranil Gupta
- Prof. Nitin Vaidya
- Prof. Sayan Mitra
- Prof. Radhika Mittal

Lecture Summary

Distributed Systems properties

- Multiple computers
- Networked communication
- Common goal

Course goals

- Concepts, designs, case studies

Your responsibilities

- Read assigned sections
- Monitor Campuswire
- Participate in lectures
- Preserve academic integrity

Next Lecture

Failure Detection

- Readings: §2.4.2, §15.1