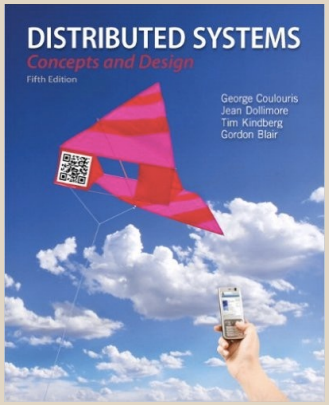
TEXAS A&M UNIVERSITY

ECEN 757: Distributed Systems and Cloud Computing

I-Hong Hou

Acknowledgement

- Materials in this course are mainly from two sources:
- Textbook



- Lectures of Prof. Indy Gupta in UIUC

Our Main Goal Today

TEXAS A&M UNIVERSITY

To Define the Term Distributed System

Can you name some examples of Distributed Systems?

Can you name some examples of Distributed Systems?

- Client-Server (NFS)
- The Web
- The Internet
- A wireless network
- DNS
- Gnutella or BitTorrent (peer to peer overlays)
- A “cloud”, e.g., Amazon EC2/S3, Microsoft Azure
- A datacenter, e.g., NCSA, a Google datacenter, The Planet

Are They Distributed Systems?

- Multi-core CPUs?
- I own a Xbox One, a PS5, and a Nintendo Switch
- Do they form a distributed system?

What is a Distributed System?

TEXAS A&M UNIVERSITY

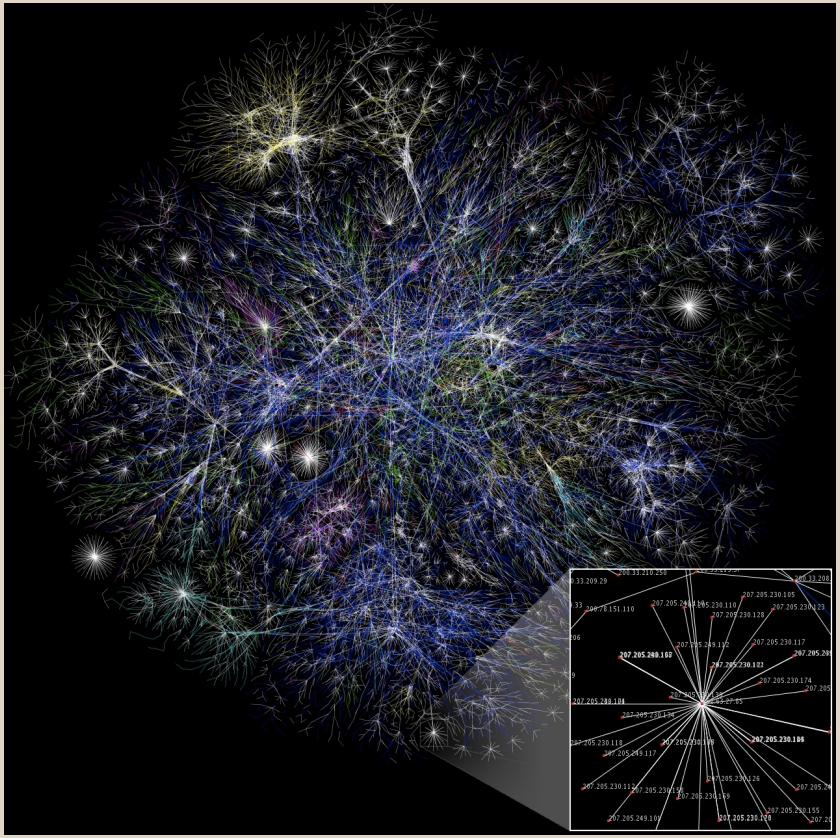
- Textbook:

- 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

Internet

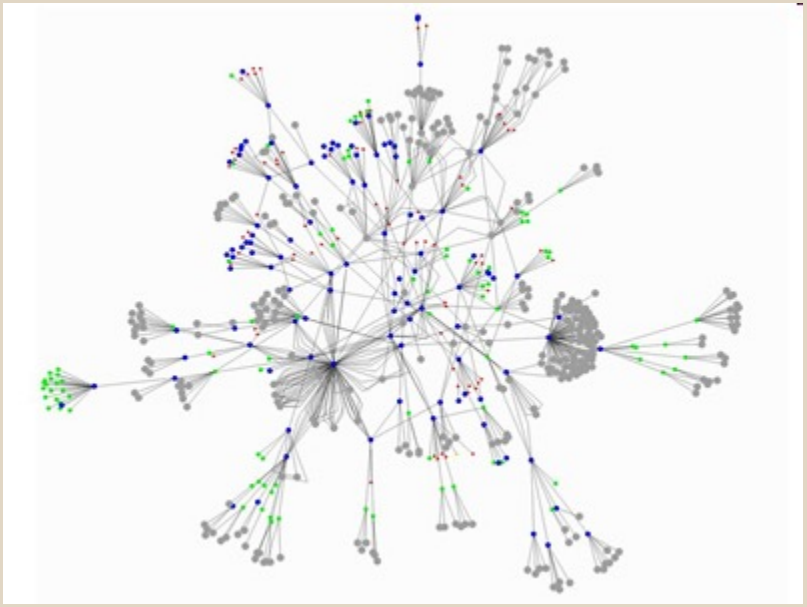
What are the "entities" (nodes)?

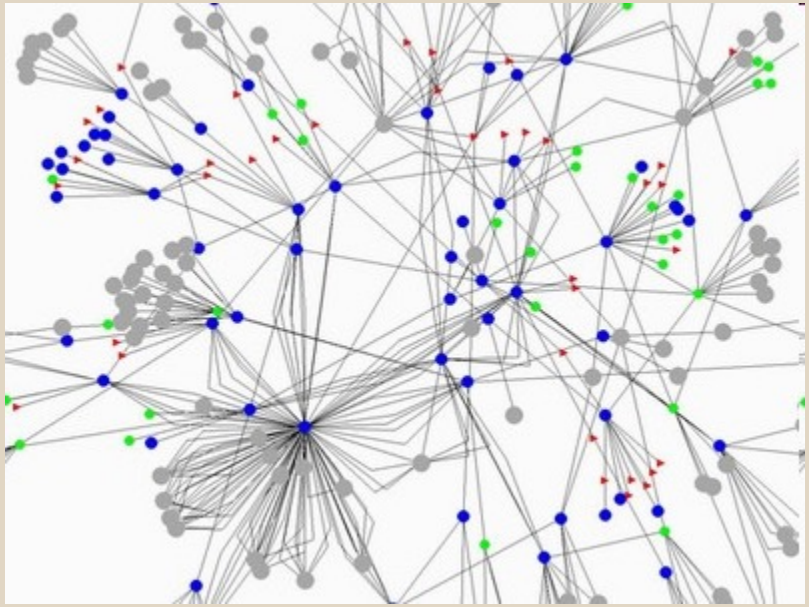
What is the communication medium (links)?



Source: https://upload.wikimedia.org/wikipedia/commons/d/d2/Internet_map_1024.jpg

Gnutella Peer to Peer System





What are the "entities" (nodes)?

Source: GnuMap Project

What is the communication medium (links)?

Datacenter

What are the “entities” (nodes)?

What is the communication medium (links)?

What are the Fundamental Challenges?

TEXAS A&M UNIVERSITY

- Concurrency: Machines work on the same job simultaneously. However, they may have different opinions about the “current state” of the job
- No global clock: When two individuals make the same stock order around the same time, how do we agree on who make the order first?
- Independent failures: When we consider a system with tens of thousands of machines, failure is a norm, not an exception. Also, messages can be arbitrarily delayed or even lost

Organization of this Course

- Part 1 (Weeks 1 and 2)

- Introduction: Some overviews about Internet, cloud computing, and MapReduce

Organization of this Course

- Part 2 (Weeks 3 – 5)
- Issues related to “Time”

- Synchronization: How to agree on current time?
- Logical Time: At least, agree on A precedes B?
- Global State: What is the “current” system state?
- Consensus: How to agree on an action?
- Concurrency Control

Organization of this Course

- Part 3 (Weeks 6 – 8)
- Issues related to “Data”
- How much data is “big data”?

Organization of this Course

- Part 3 (Weeks 6 – 8)
- Issues related to “Data”
- How much data is “big data”?

- Ex 1. I use grades in this course to study the correlations between homework scores and midterm scores
- Is this big data?

Organization of this Course

- Part 3 (Weeks 6 – 8)
- Issues related to “Data”
- How much data is “big data”?

- Ex. 2. I track students enrolled in TAMU in 2021 (~73,000) and study the relations between GPA and employment upon graduation
- Is this big data?

Organization of this Course

- Part 3 (Weeks 6 – 8)
- Issues related to “Data”
- How much data is “big data”?
- At least > 1 TB. Or, you cannot store all data in one machine

- Distributed storage: How to store big data?
- Replica: How to avoid data loss due to failure?
- P2P: Where to find data?

Lecture Mode

- Slides will be uploaded to the shared Google Drive before the lecture starts
- You are strongly encouraged to take notes during lectures
- Links to recorded lectures will be posted in the document “Links and Announcements”

Midterm

- Midterm on Mar 28

Organization of this Course

- Part 4 (Weeks 10 – 14)
- Emerging topics

- We will discuss 20 papers on 10 interesting topics

- Each paper is presented by a group of three students

Presentation

- The list of papers and group assignments will be posted on Canvas/Google Drive folder

- You need to prepare a ~25min class presentation

Topics

- Blockchain
- User behavior
- How to make the cloud greener?
- Caching
- Fault-Tolerant Design
- Job Allocation
- Internet of Things
- Crowdsourcing
- Edge/Fog computing
- Middleware

Quiz

- Every lecture in Part 4 has a quiz
- Presenters in the previous lecture need to prepare a 5min elevator pitch with at most one slide for each paper (no animation)
- The quiz contains one problem for each paper (two problems per lecture)
- The problems are supposed to be easy

Distance Learning Section

- In part 4, you need to do the following for every week:
- After watching the presentations of four groups, pick a paper to read
- Write a three-page review for the paper
- Also, write a two-page summary of the four presentations (~0.5 page per presentation)

GRADING POLICIES

READ THIS BEFORE YOU DO YOUR HW/EXAM/PRESENTATION

Academic Integrity

- “An Aggie does not lie, cheat or steal, or tolerate those who do.”
-
- “Texas A&M University students are responsible for authenticating all work submitted to an instructor. If asked, students must be able to produce proof that the item submitted is indeed the work of that student. Students must keep appropriate records at all times. The inability to authenticate one’s work, should the instructor request it, may be sufficient grounds to initiate an academic misconduct case” (Section 20.1.2.3, Student Rule 20).
-
- You can learn more about the Aggie Honor System Office Rules and Procedures, academic integrity, and your rights and responsibilities at aggiehonor.tamu.edu.

Grading Overview: In-Person

- Homework: 15%
- Midterm Exam: 40%
- Class Presentation/ Review: 25%
- Quiz: 20%

Grading Overview: DL

- Homework: 15%
- Midterm Exam: 40%
- Class Presentation/ Review: 45%

Homework

- Scan and submit your homework through Canvas
- First, try to solve the problems without external help
  - External help = anything other than lecture notes, the textbook, and your brain
- Then, you can use any external help
- If you do, you need to clearly state the sources of help
- Also, provide some reflections: “I couldn’t solve this problem independently because…”

Midterm Exam

- You can bring two sheets of letter-sized double-sided cheat sheets
- You need to scan and upload your solution through Canvas
  - Deadline = 10 mins after the exam ends
- Problems are likely to be twists of homework problems
- For DL: Please check if your company provides proctoring services

Presentation

- Roughly speaking, there are three types of papers: Theory, system, and measurement

Presentation for Theory

- What is the problem?
- What is the theoretical result?
- Why does the result make sense?
  - It is ok if you cannot fully grasp the proof, but I want you to convince me that it is intuitively correct

Presentation for System

- What is the main insight of the design?
- How to address practical issues of implementation?
- How do we evaluate the system?

Presentation for Measurement

- How do they conduct measurement?
- What are the surprising results gained from the measurements?
- What are the implications of these measurements on system design?

Tips for Presentation

- Originality: Other than simulation/experiment figures, everything should be from you, not copied from the paper
  - Do not look at the paper when making the slides
  - The best structure for a paper is not the best structure for a presentation
  - Extra points for your own examples/insights/explanations
- Your presentation should be understandable to people that have not read the paper
  - Instead of trying to cover everything, focus on conveying the two or three most important ideas
  - Be sure to do practice talks and get honest feedback

Quiz

- Quiz will be conducted after the elevator pitches
- We will conduct the quizzes through Quizizz on your smartphone
- You are not allowed to use any resource, such as the paper itself, for the quizzes
- You are strongly encouraged to take notes during presentations, and study them before quizzes
- 30 seconds for each problem

Tips for Paper Review (DL)

- Originality: Other than simulation/experiment figures, everything should be from you, not copied from the paper
- It is a “review”, not just a summary
- You should comment on the strengths, weaknesses, and potential future work of the paper
- Rule of thumb: ~1.5 page summary, ~0.5 page strengths, ~1 page weaknesses and future work

Tips for Presentation Summary (DL)

- It should be in your own words
- Take notes during the presentations
- Your summary can simply be the polished version of your notes
- Your own thoughts are welcomed, but not necessary