CS-340 Introduction to Computer Networking

Lecture 1: Introduction

Steve Tarzia

Many diagrams & slides are adapted from those by J.F Kurose and K.W. Ross

Today's lecture

- What will you learn from this class?
- How will this course be operated?
- Are you ready to take this class?

- Overview of the Internet

Course logistics

- We'll use Canvas for assignments, announcements, lecture slides.
- I will post two videos on Panopto per week.
  - I will record new videos, but last year's lectures are already on YouTube.
- Reading assignments will also be posted.
- Ask asynchronous questions on Piazza (not by email):
  - https://piazza.com/class/kels8h2i9pi2nv
- Teaching staff and office hours will be finalized/announced soon.

Discussion sections

- MWF 1:50-2:40pm class time will be used for **discussion sections**.
  - The idea is to bring together a small group of students to discuss one problem/question related to the week's topics.
  - Participation will be graded (5% of total). Two absences per quarter are allowed. We'll try to schedule one section late at night for people in Asia.
- You must attend one 25-minute session per week on Zoom:
  - Mon 1:50-2:15
  - Mon 2:15-2:40
  - Wed 1:50-2:15
  - Wed 2:15-2:40
  - Wed 9:00-9:25pm central time (for people in faraway time zones)
  - Fri 1:50-2:15
  - Fri 2:15-2:40

Making the most of remote learning

- To make remote learning work, we must learn new habits.
- When watching lectures and reading, be ready to write your questions.
  - Note the slide number.
  - Ask questions later on Piazza or during office hours.
- "Stop and Think" cues will ask you to pause the video and ponder a question I have asked. Please do it!

Office Hours

- This is the best time to get help with your projects.

Prerequisites

- Data Structures (EECS-214)
- Basic C/Unix programming (EECS-213 or EECS-205)
- All programming assignments will be in **Python**

- If you're unsure, then start the first project ASAP
- If you're going to drop, better to drop sooner than later

Grading

- 50% -- 3 Python Projects
  - Done in pairs (or individually)
  - Usually due on Mondays at Midnight.
- 20% -- 4 Homework assignments
  - Short written answers, based on reading, lectures, & Wireshark labs
- 10% -- Midterm exam
- 15% -- Final exam (cumulative) on *Monday, Dec 7th at 9am*..
  - Exams are similar to homework assignments
  - Exams are *open book* & *open notes*
  - Plan to print all the lecture slides.
- 5% -- Participation in discussion section

Why take Computer Networking?

- To learn how the Internet works!
- It's one of the greatest technical achievements in human history:
  - Somehow it's global, reliable, & scalable
  - Without centralized control, and with unreliable components.
- We will explore several layers:
  - Applications (Web, Email, DNS, etc.)
  - Encryption (TLS/SSL), *at end of class*
  - Reliable connections (TCP)
  - Routing (IP) & Network control (BGP)
  - Physical connections (Ethernet, WiFi, Cellular)
- Internet principles also apply to other networks & engineered systems in general.

Projects

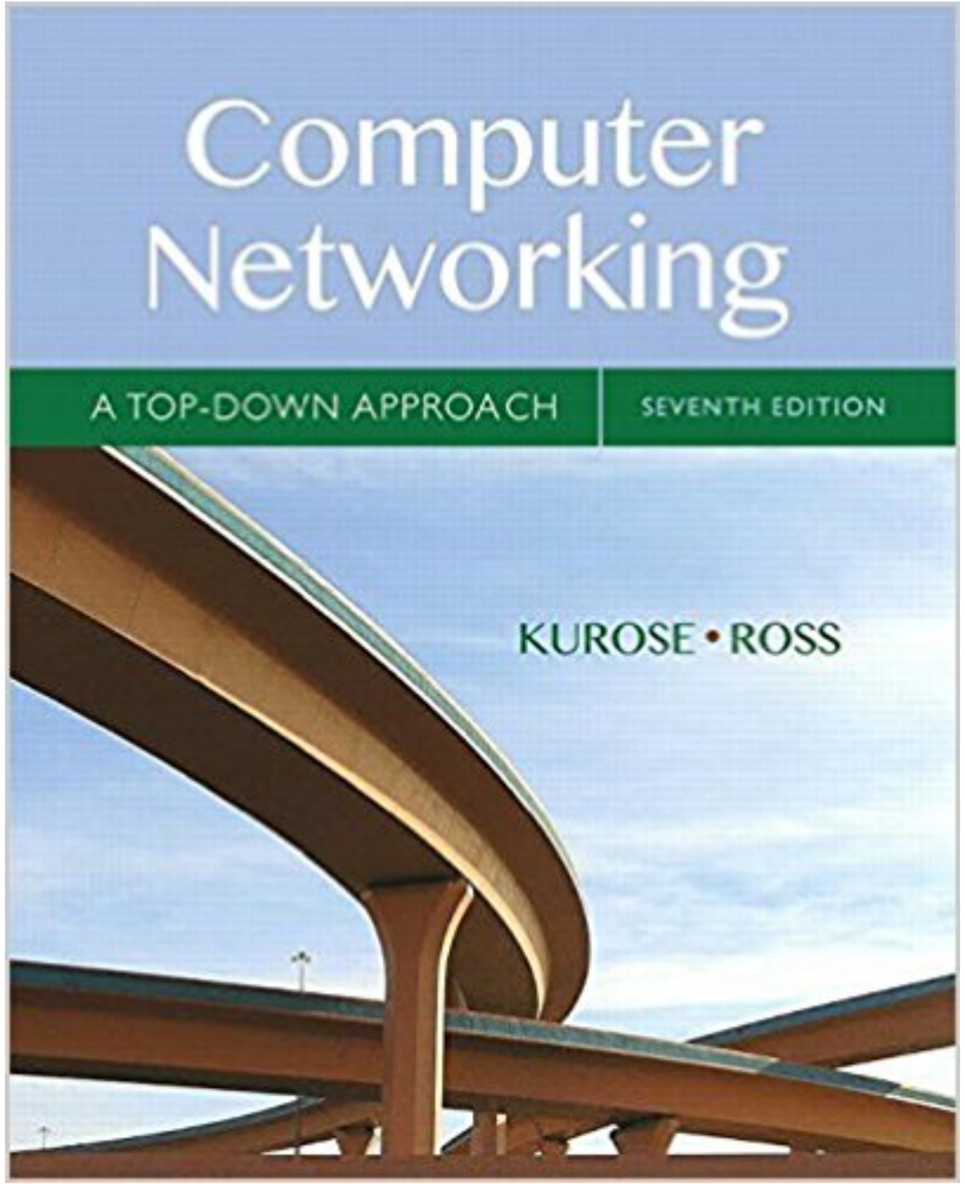
1. Build an HTTP (web) server.
2. Implement a reliable transport protocol (simplified version of TCP).
3. Implement distributed routing algorithms (BGP, etc.)

- All will be in Python.
- Will develop on your machine and test on a shared Linux server.
- Will work in pairs or individually
  - Groups of three are not allowed. Just take turns working in different pairs.

Collaboration & cheating policy

- Helping each other is OK, but you should not do anything that allows another student skip the learning process.
- You may talk to other students about the homework and projects
- You may **look** at **small parts** of each other's code (on the screen)
- You **may not** send a *copy* of your code to friends
- If you copy code from the Internet, you must add a comment explaining the source.
- You must understand what your code does
- I will use *MOSS* to compare your code to everyone else's code and prior years' code.
- If you spent fewer than 6 hours on a project, then you probably cheated!
- I will notify the Dean of blatant cheating and you may be expelled from Northwestern.
- If you're unsure about this policy, please ask me.

Required reading



Computer Networking: A Top-Down Approach (7th Edition), by Kurose and Ross.

- 864 page textbook
  - Well-written and easy to read.
  - Chapter 1-6, 8, part of 7
  - To do well on exams you must read it!
- Kind of expensive: $155
  - Feel free to buy 6th or older edition or "global edition"
- Repeats & reinforces lectures
- Exams are open-book, so please buy a paper copy of the book

About Steve

- Fourth year teaching at Northwestern. "Assistant professor of instruction"
- PhD from Northwestern in 2011, BS from Columbia in 2005,
  both in Computer Engineering
- Research expertise was acoustic sensing on mobile systems.
- Worked at a few Chicago area software startups
  First as a software engineer, later as VP of Engineering
- Have published about ten iOS apps in the app store
- Fan of Linux, AWS, Java, Python, Objective-C
- Founded nonprofit National Gun Violence Memorial (GunMemorial.org).
- Enjoys playing music (guitar, drums, trumpet), cooking, repairing things, exercise, opera, walking around Evanston.
- Married 15 years & has lived in Evanston for the past 9 years.

A Quick Overview of the Internet

To get started, I'll described the entire Internet's operation very briefly.

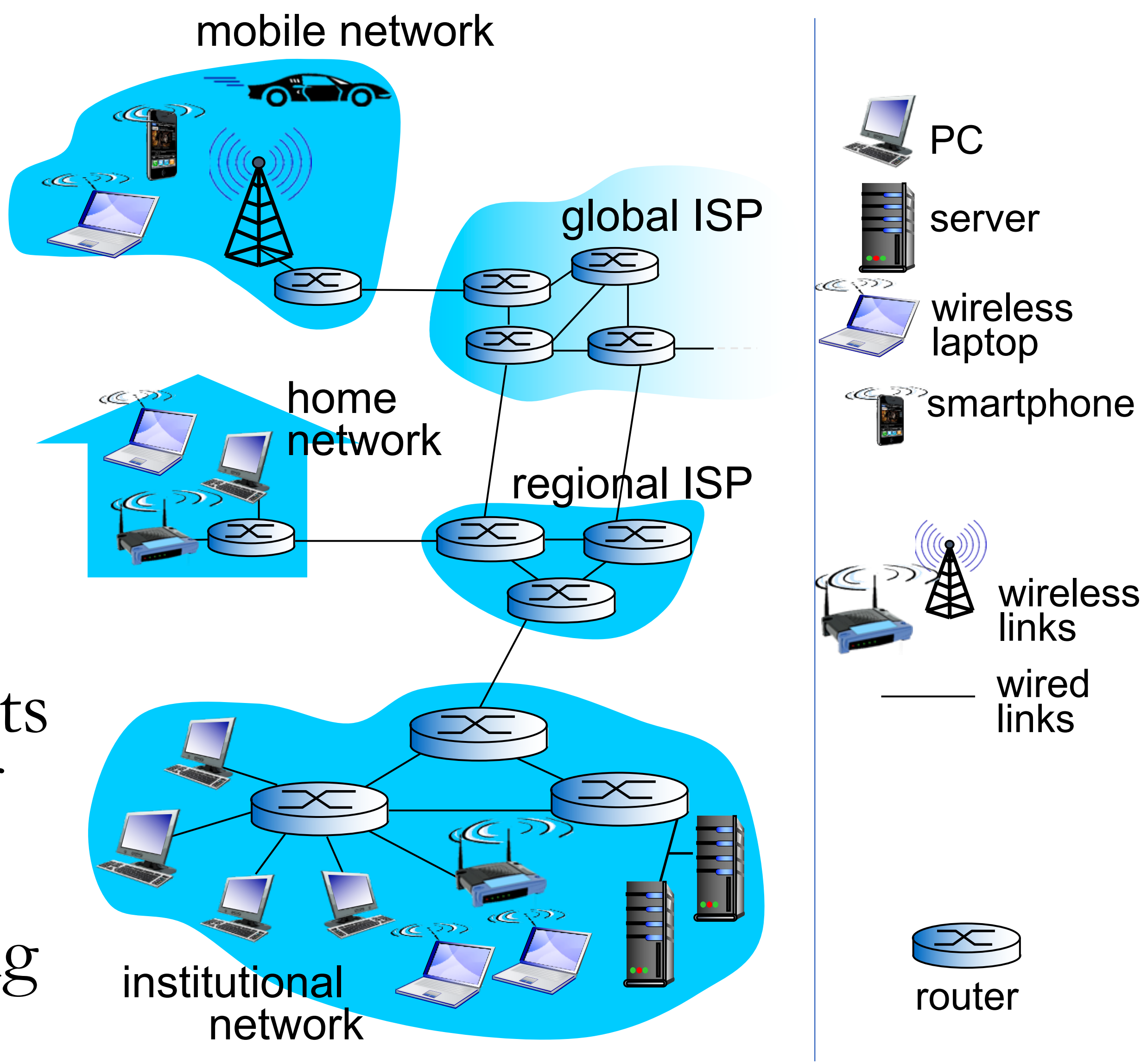
Then for the remainder of the course, we'll study it in more detail, starting at the "top" with the application layer, and moving "down" toward the physical media.

How does the Internet work?

- What steps are involved when you visit a website on your laptop?
- How is the process different on your smartphone?
- What’s the difference between the “world wide web” and the Internet?

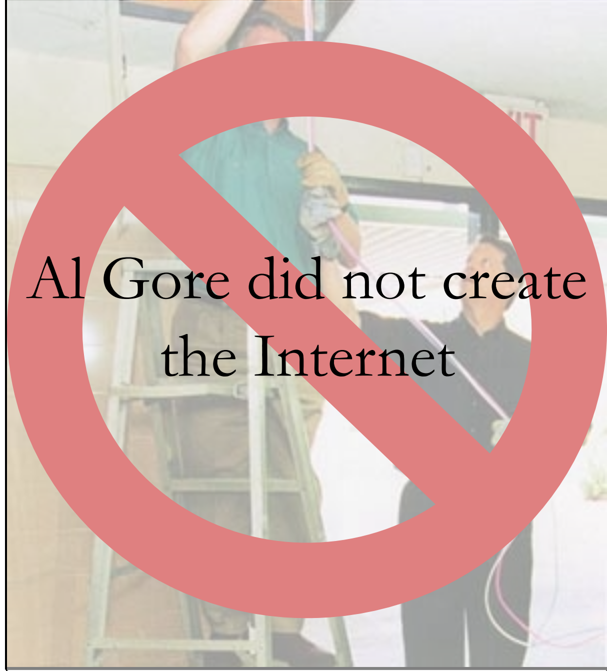
What is the Internet?

- A *network* of billions of computing devices, called *hosts* or *end systems*.
  - From thermostats to servers
  - Each has a unique, numeric *Internet Protocol (IP) address*
- Many communication links
  - copper wire, fiber optic, or radio
- Hosts send data *packets* to other hosts
  - Packets are labeled with IP address of sender and receiver
- *Routers* and *switches* direct traffic using addresses on the packets.



Who invented the Internet?

- The Internet grew out of ARPANET, a US-military-funded academic research project (in 1970s).

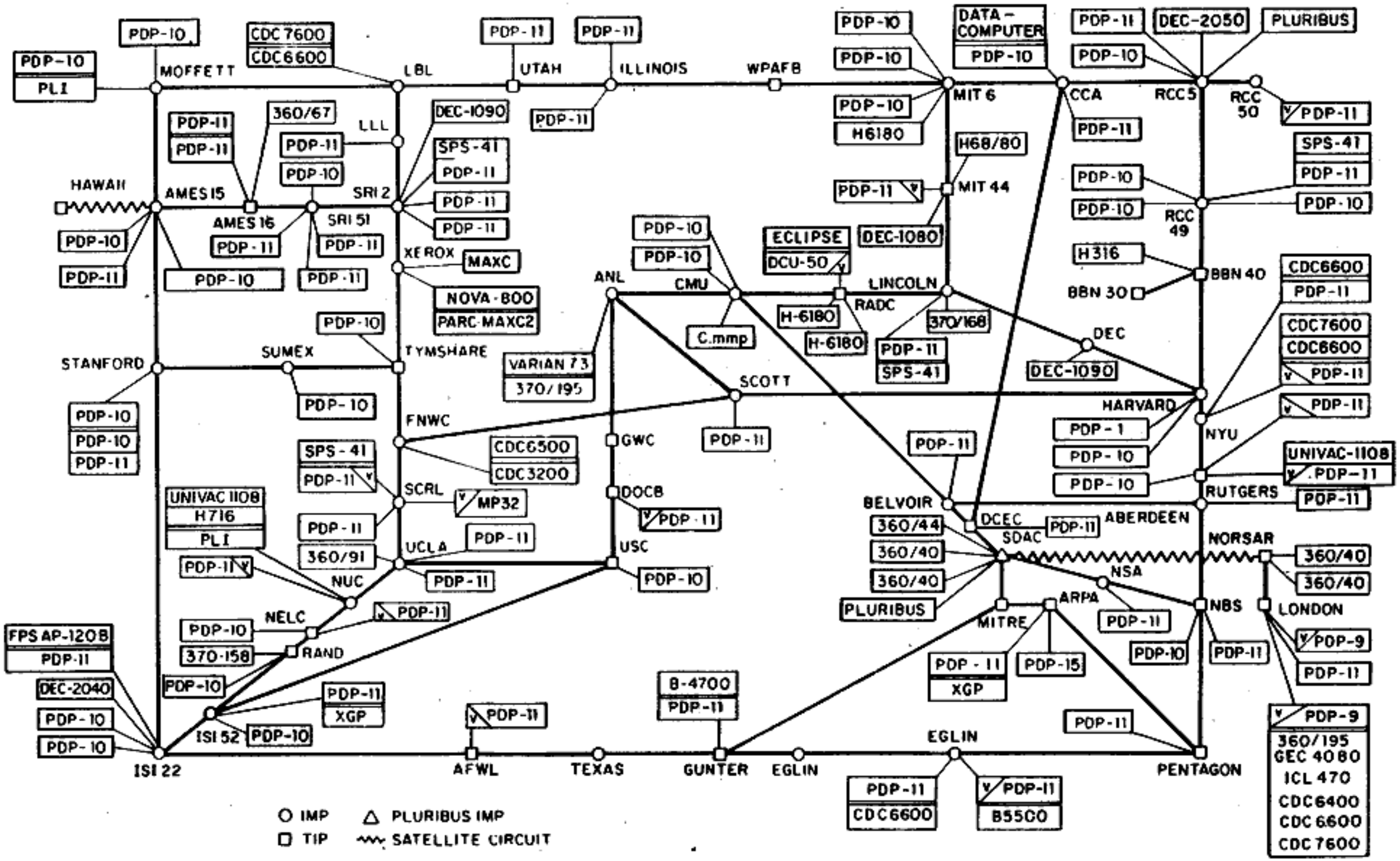


U.S. President Clinton and Vice President Gore installing Ethernet cabling in 1996.

Who controls the Internet?

- It's a loose confederation of interconnected networks.
  - A "network of networks"
- ICANN (Internet Corporation for Assigned Names and Numbers):
  - Assigns IP addresses and assigns DNS top-level domains.
  - Operates 13 root DNS servers.
- IETF (Internet Engineering Task Force):
  - Develops protocols that define how Internet devices interact.
  - RFCs (Request For Comments) are its documents defining Internet standards.

ARPANET LOGICAL MAP, MARCH 1977



(PLEASE NOTE THAT WHILE THIS MAP SHOWS THE HOST POPULATION OF THE NETWORK ACCORDING TO THE BEST INFORMATION OBTAINABLE, NO CLAIM CAN BE MADE FOR ITS ACCURACY)

NAMES SHOWN ARE IMP NAMES, NOT (NECESSARILY) HOST NAMES

The Internet != The Web

- The Internet is a worldwide network of interconnected computers.
  - ARPANET adopted TCP/IP protocol in 1983 (birth of the modern Internet)
- The Web is a worldwide network of interconnected hypertext pages.
  - Web browser app connects to web server apps.
  - Web is just one of many applications built on top of the Internet and DNS.
  - Uses HTTP protocol (first version, 0.9, was released in 1991)
  - Invented at CERN on the France/Switzerland border.
- Apart from the Web, many other applications use the Internet:
  - Email, SSH, BitTorrent, Voice-over-IP telephony, FTP, Remote Desktop, VNC, Skype, Telnet, Snapchat, FaceTime, Netflix … you name it!
  - Any smartphone app that communicate with a backend service

The Internet is **distributed** and **loosely coupled**

- No one entity is manages all the hosts, routers, and links on the Internet.
- Hosts join by building a physical connection to an existing host.
- Hardware and software come from many different vendors.
- Standard Internet protocols define how devices should interoperate.

- A network **protocol** defines proper communication patterns between devices:
  - types of messages that can be sent,
  - structure of the messages (bit-level representation),
  - possible responses and actions to take.
- Clearly-defined protocols allow us to create new devices that are compatible with the existing ones, and thus expand the Internet *ad-hoc*.

Human protocols

Uncle Owen: "You, I suppose you're programmed for etiquette and protocol."
C-3PO: "Protocol? Why, it's my primary function, sir. …"

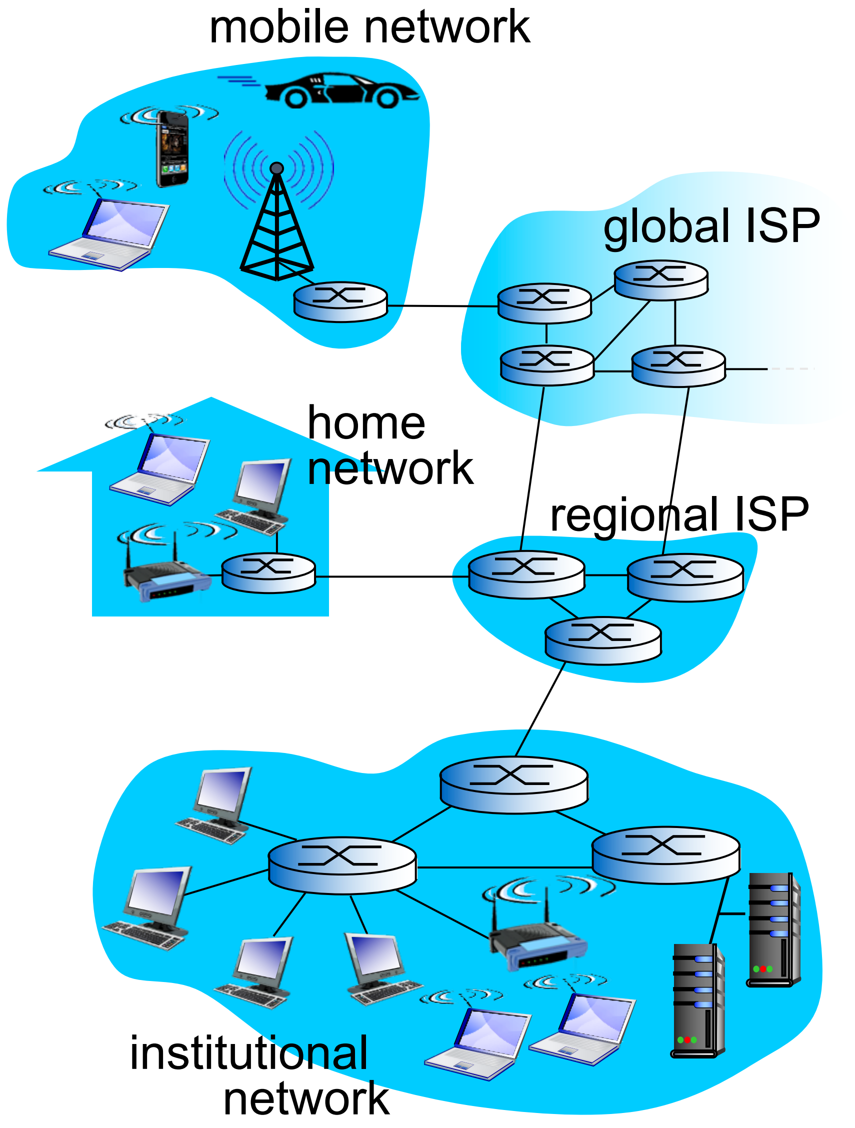
- Human societies also have widely known protocols (etiquette) to allow millions of strangers to live harmoniously together.
- It would be very inefficient for each pair of people to establish their own rules of interaction, so we follow certain standards:
  - Shake hands when meeting to signal friendliness.
  - Stand at the end of the checkout line at a grocery store to make a purchase.
  - Drive on the right side of the road, stop at stop signs, green means go…
  - Ask a person to see a movie or go to dinner to signal romantic interest.
  - … but do not ask a third person along on the date.
  - Do not say "hello" to passing strangers in a big city.
  - Do say "hello" to passing strangers in rural areas.
- Humans are intelligent and adaptable, so human protocols can be flexible.
- Computers need very **strict** protocols to allow "strangers" to interact.

The Internet – in practice

- All machines on the Internet can send and receive messages, but there are two distinct categories of participants:
- Network edge:
  - **Hosts**: clients and servers
  - Have only one link to the Internet
- Network core:
  - **Routers** connecting various Internet parts
  - These are like hosts with multiple links and living only to relay others' packets.

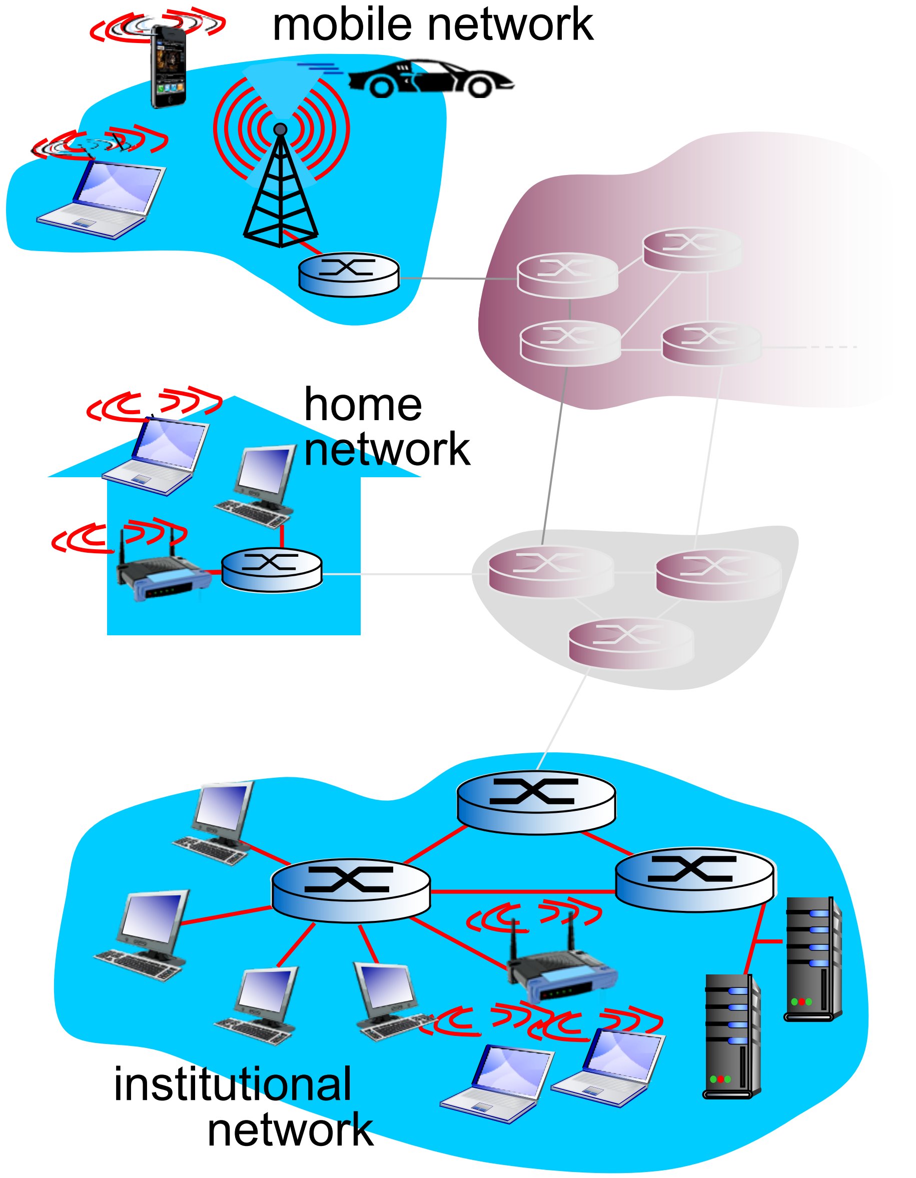
Machines are connected by physical media:

- Wired and wireless communication links



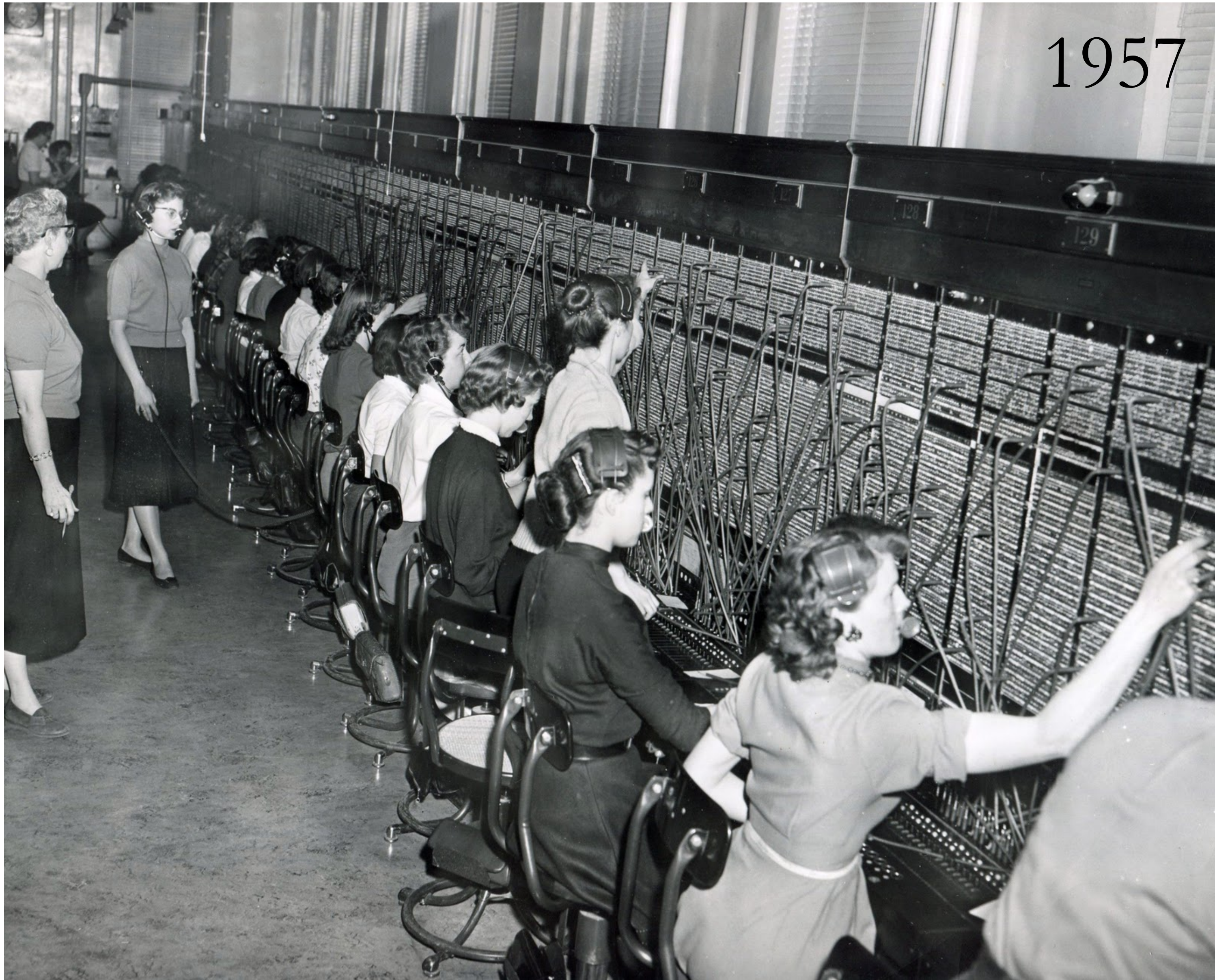
Access networks and local-area networks

- Many different technologies are used to connect devices at the edge
  - WiFi, Ethernet, Cellular
  - Speed varies depending on the link technology
  - Links may be shared or dedicated
- Some access networks assign public IP addresses to hosts, and others use private IP addresses and network address translation (NAT).



Public Switched Telephone Network (PSTN)

- Uses *circuit switching* -- a very basic design.
- An electrical connection is established end-to-end for each call.
- Originally done by moving cables plugged into switchboards.
- Later, pulse and tone dialing automated switching.



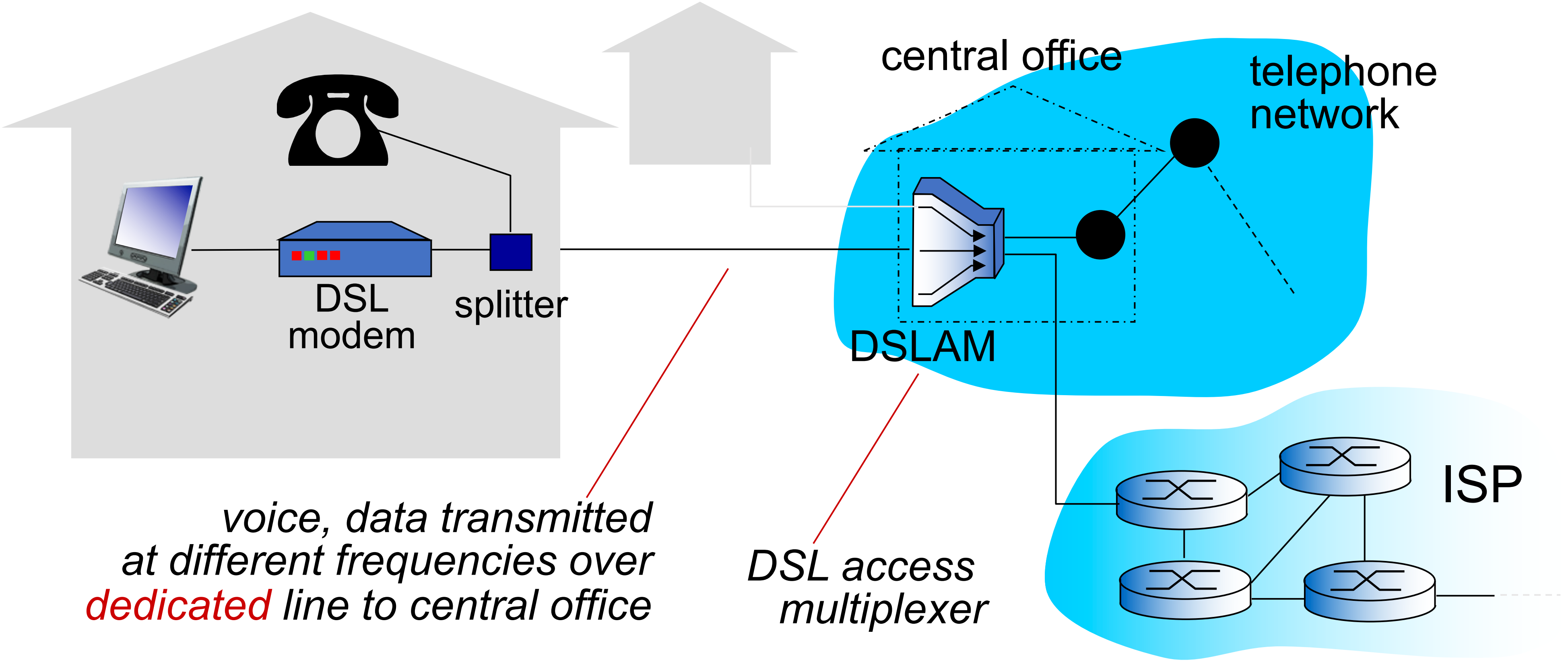
Home Internet access uses old networks

- Public Switched Telephone Network (PSTN) still uses analog signals on twisted pair cables (for the "last mile" anyway).
  - Early access used telephone modems, up to 56kbps, over the voice channel.
  - Modern DSL access uses higher frequencies than voice channel.
- Cable Television Networks use coaxial cable.
  - Coaxial cable is a much better signal carrier, much less sensitive to noise, and thus can carry much more data.
  - It was designed to carry video, which is more complex than just audio.
- Both PSTN and Cable have a modern fiber optic network at their core, and these days it's the same infrastructure that runs the Internet.

Connecting to the Internet in the 1990s

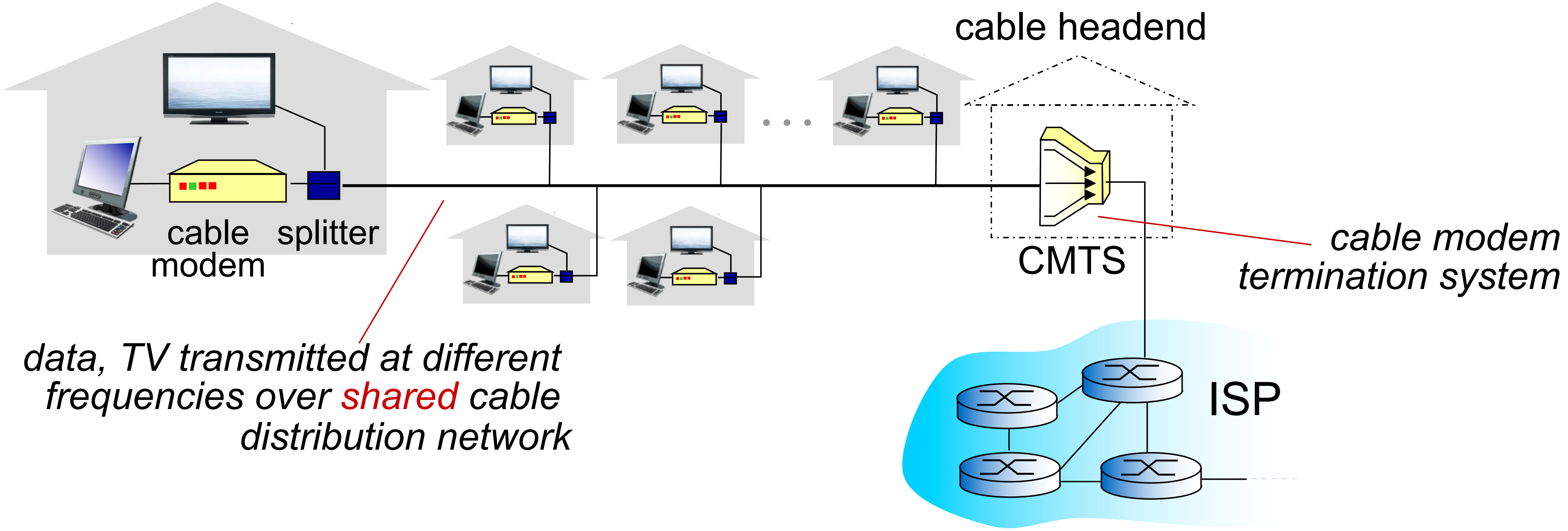
- Connect your computer's fax/modem to a regular telephone line.
- Computer dials a *local* telephone number owned by your Internet Service Provider (ISP).
- Data is encoded as sounds in the audio range
- ISP has its own connection to the Internet, shared by everyone who is dialed in.
- Hope that your mom doesn't pick up the telephone while you're surfing and interrupt your session.
- Hang up when you're done so that your household can use the telephone again.
- https://www.youtube.com/watch?v=D1UY7eDRXrs

Digital Subscriber Line (DSL)



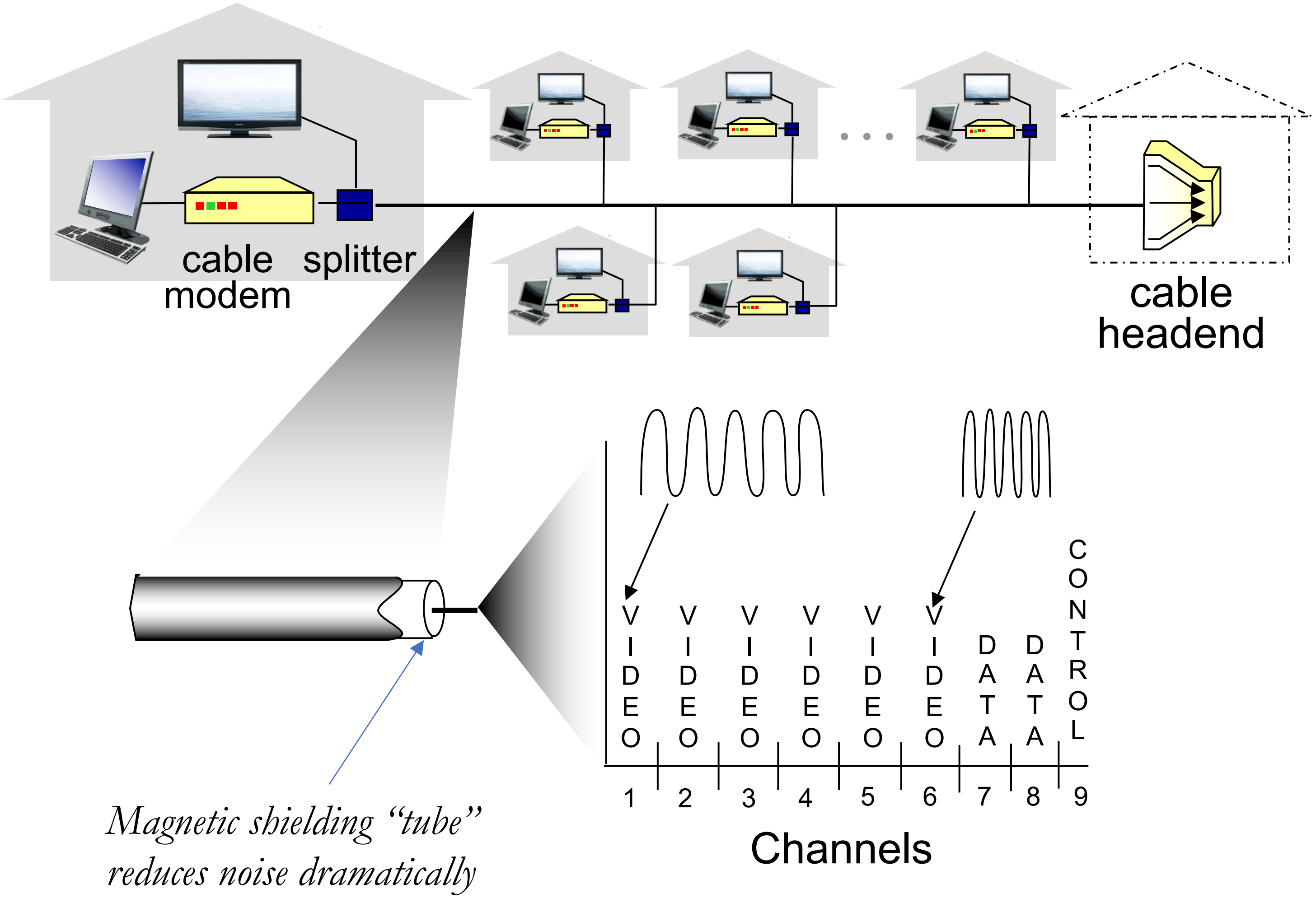
- Telephone calls do not interfere with data exchange.
- Up to ~40MBps
  - Limited by noise in “last mile” cabling originally designed for just voice.

Cable Networks



- Network was originally designed for one-way broadcast of TV
  - Entire neighborhood shares the same cable/signal.
- Up to ~200 MBps

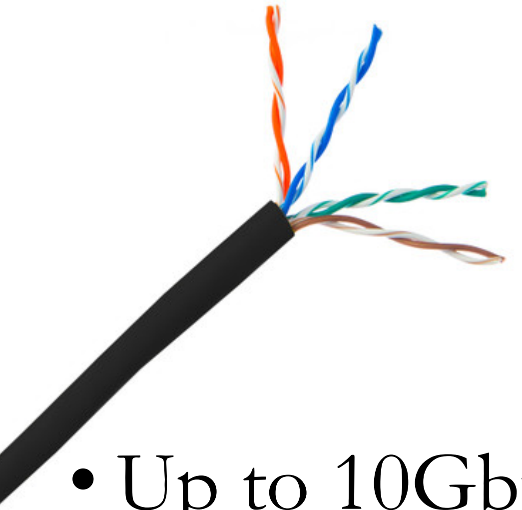
Frequency Division Multiplexing



- A way to share a single communication medium
- Different channels are simultaneously transmitted on different frequency bands.
- Different transmitters don't have to "take turns."
- Coaxial cable is a very low-noise medium, so high frequencies can be supported and channels can be narrow.

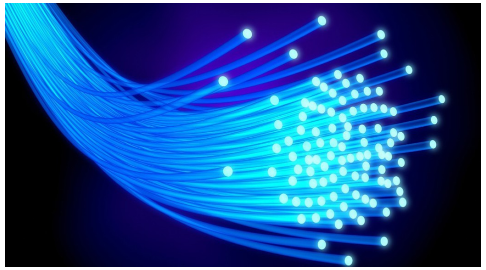
Modern wired/guided media

Twisted Pair (Cat5e, Cat6)



- Up to 10Gbps over short distances.
- Signals are transmitted electrically.
- Cheap and easy to install.

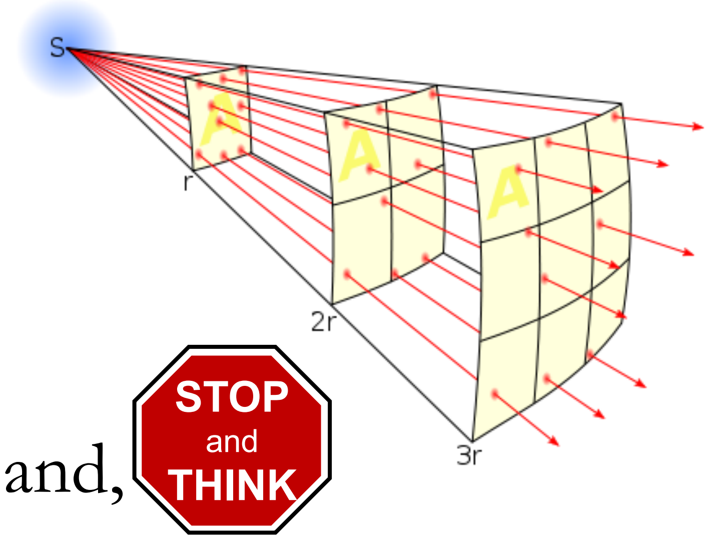
Fiber Optic



- Up to 100Gbps over long distances.
- Signals encoded as pulses of light.
- Immune to electromagnetic interference.
- More expensive.

Radio is a wireless/unguided medium

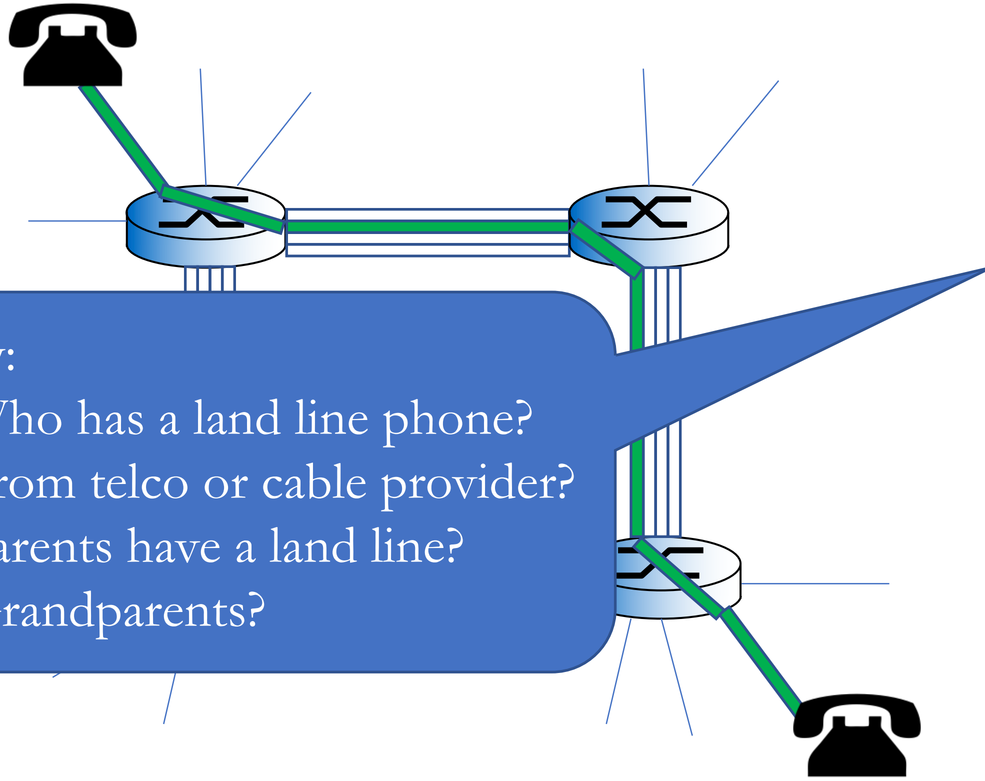
- Signal flows through open space
- (+) No need to wire-up every device
- (-) Lots of noise, interference, and competition for the spectrum.
- Objects can block/reflect signal
- Signal's power is proportional to the square root of distance
- FCC licenses different frequency bands in the U.S. to reduce interference between competing devices
- Includes: WiFi, Cellular, Satellite
- Why is WiFi range much smaller than Cellular?
  - Higher Freq., Lower xmit power, Noise in unlicensed band,



A look at the network core

Circuit Switching was used in the PSTN

- If two people wanted to talk, a dedicated electrical path (circuit) is established.
- The circuit remains dedicated to the call until the parties hang up.
- (+) Performance is guaranteed.
  - Competing calls will not interrupt yours.
- (-) You may have to wait for a free link before establishing a call. (*Congestion* tone)
- (-) It's wasteful because there is much **silence** during a call
  - Both parties rarely talk simultaneously.

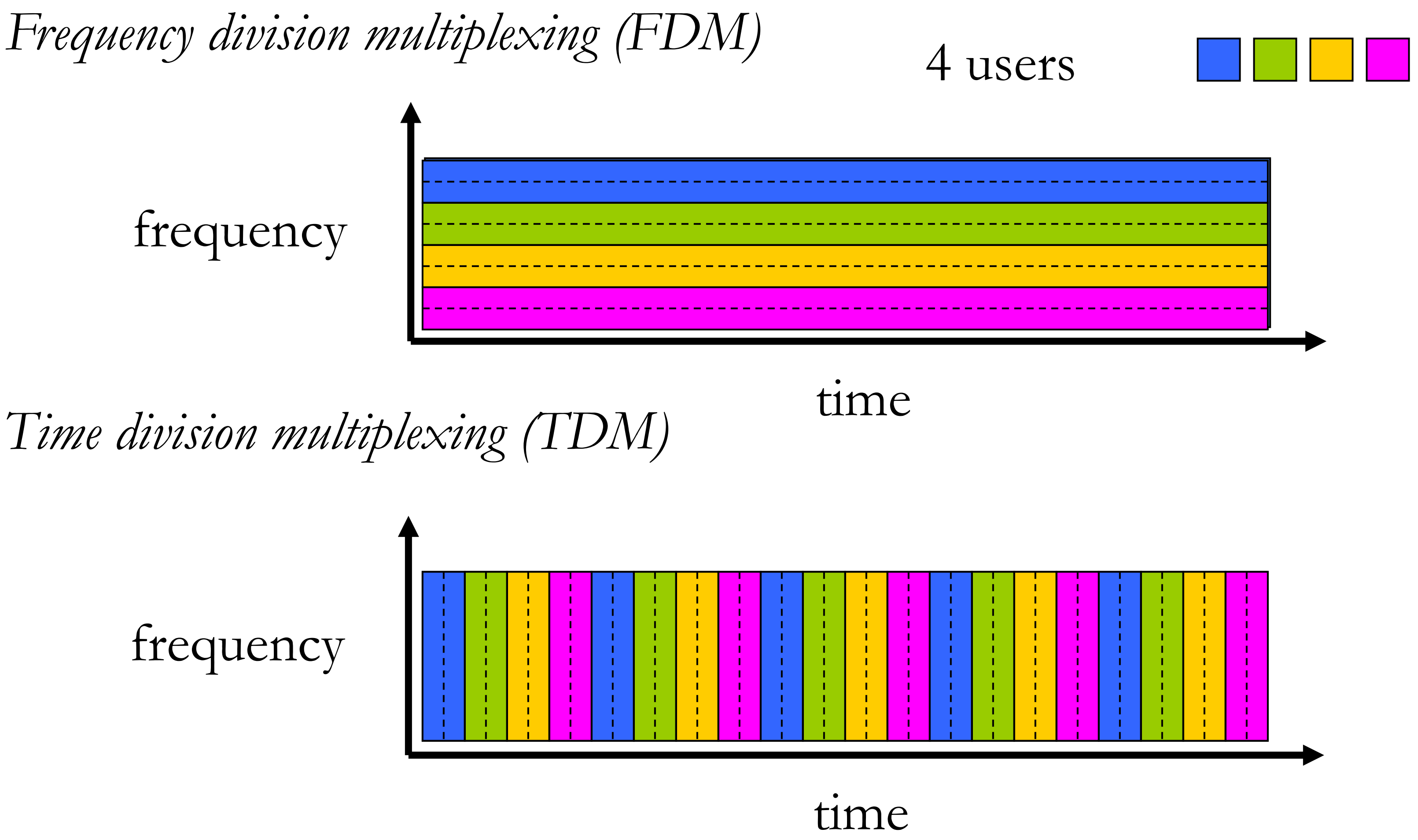


Survey:
- Who has a land line phone?
- From telco or cable provider?
- Parents have a land line?
- Grandparents?

An "all or nothing" approach. Service does not **degrade** gracefully when network is crowded, but this is good for telephony.

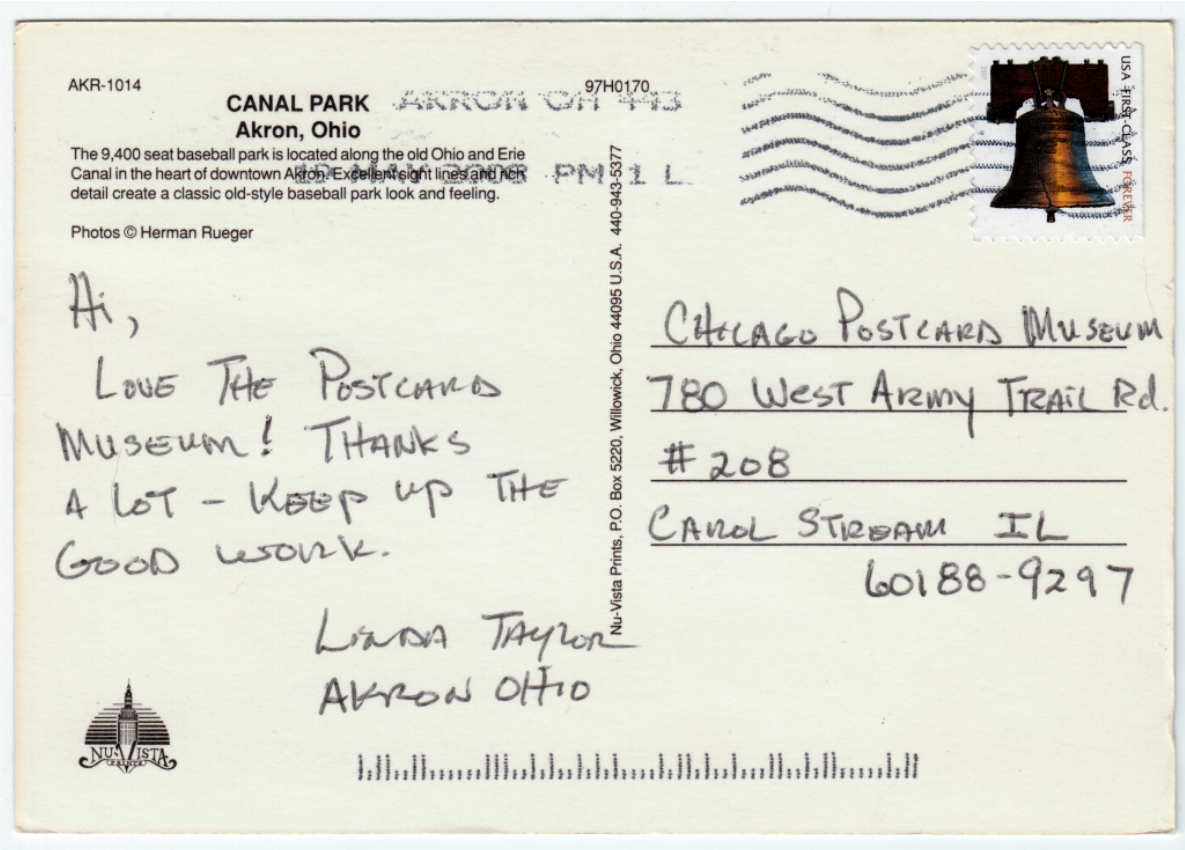
Circuit switched backbone links are multiplexed

- We don't literally use a reserved electrical path (wire) for each call.
- Reserve a channel on a multiplexed (shared) link.
- Allows one link to support many simultaneous channels (calls).
- Covered in Chapter 6.



Computer networks use **Packet Switching**

- This is an alternative to circuit switching, and a key Internet innovation.
- Data is sent as packets – small chunks (~1.5kb)
- A packet is like a postcard.
- Contains:
  - Destination address
  - Sender address
  - Limited space for a message
- Packets create challenges:
  - Large message must be split into multiple postcards (packets)
  - May arrive in any order
  - May take different paths to destination
  - Delivery is "best effort," not guaranteed.



Packet vs Circuit switching

- "Switching" here means the way a link is shared by competing users.
- **Packet** benefits:
  - Simpler:
    - No connection setup is necessary.
    - Just address packet correctly and send it.
    - Sender need not know full path to endpoint, just the "next hop."
  - Very efficient for sharing "bursty" customers.
- **Circuit** benefits:
  - Guaranteed performance.
  - Very effective for applications requiring constant, steady transfer (streaming audio and video).

Network performance metrics

There are many *ways to measure* network speed.

- **Throughput** – is the rate of data transfer. (a.k.a. **bandwidth**)
  - Usually measured in bits per second (bps) or bytes per second (Bps).
  - 1 byte = 8 bits
  - Data arrives in discrete chunks (packets) so throughput is calculated as an *average* over some time window.
- **Latency** – is the delay of an action. Network latency is the time it takes for a bit of data to arrive at the destination.
  - Often measure packet round trip time (**RTT**): time to get a response.
- Throughput and Latency are totally independent metrics, but both are important and affect applications and user experience.

Network performance is constantly changing!

- The Internet is a huge, complex, **dynamic** system.
- Eg., Latency/Throughput will vary over time and space.

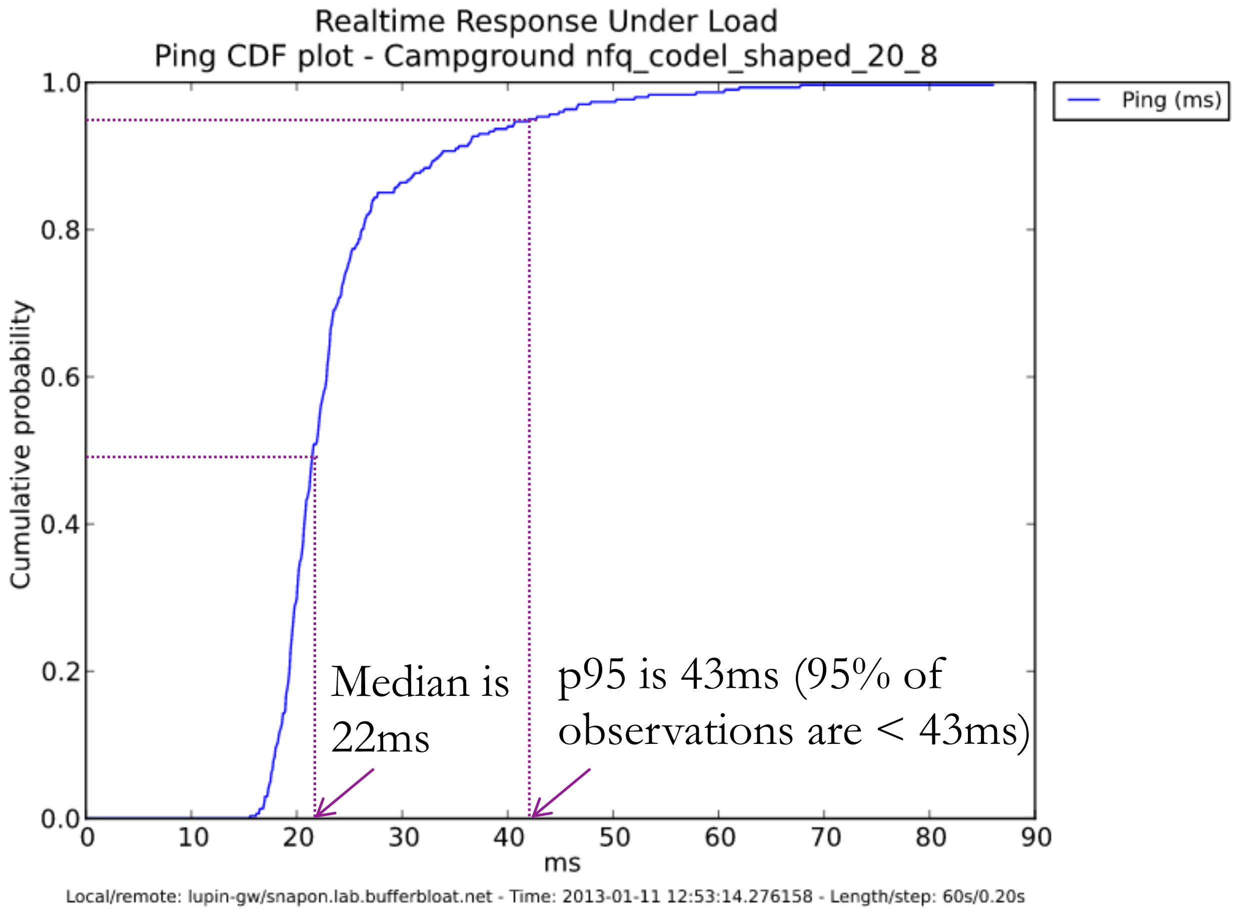
We can report:

- **Mean** average
- **Median** average
  - But these ignore variation. We want *all* users to be happy, not just "average" user.
- **Minimum**, **Maxium**
- 5^{th} percentile (**p05**), 95^{th} percentile (**p95**), 99^{th} percentile (**p99**)
  - These are like minimum or maximum but ignore outliers.

Cumulative distribution function (CDF)

- Instead of using summary statistics (mean, median, etc.), you can present the full distribution of measurements.

Y-axis gives the percentage of samples below the X-axis value



Network performance experiment

- Go to https://fast.com or https://speedtest.net
- What did you get? Now try again… what did you get the 2nd time?

(Mbps = Megabits per second)

- Why did you all get different results?
- There are dozens of reasons, and we'll learn about them in this class!

- In 2018, Charter-Spectrum ISP paid a $172M settlement for speed fraud. Speeds were 80% slower than advertised!

Recap: Internet Overview

- Various physical media and access network types are used
- It's loosely-couple and de-centralized
  - Everyone's invited to the Internet, assuming you follow the protocols and can somehow connect.
- A ***protocol*** is a set of rules for communication.
- Telephone network used circuit switching (*reserves* end-to-end path).
- Internet uses packet switching (store-and-forwarded along path)
  - Delivery is "best effort," not guaranteed.

TODO

- Find a partner for projects (you may ask around on Piazza).
- Read Chapter 1 of the Kurose book.
- Check out HW1 and Project 1, to be posted soon.