DS 593: Privacy in Practice

Data in motion and the Internet

Instructor: Alishah Chator

News?

Last time

- Computers!

Today

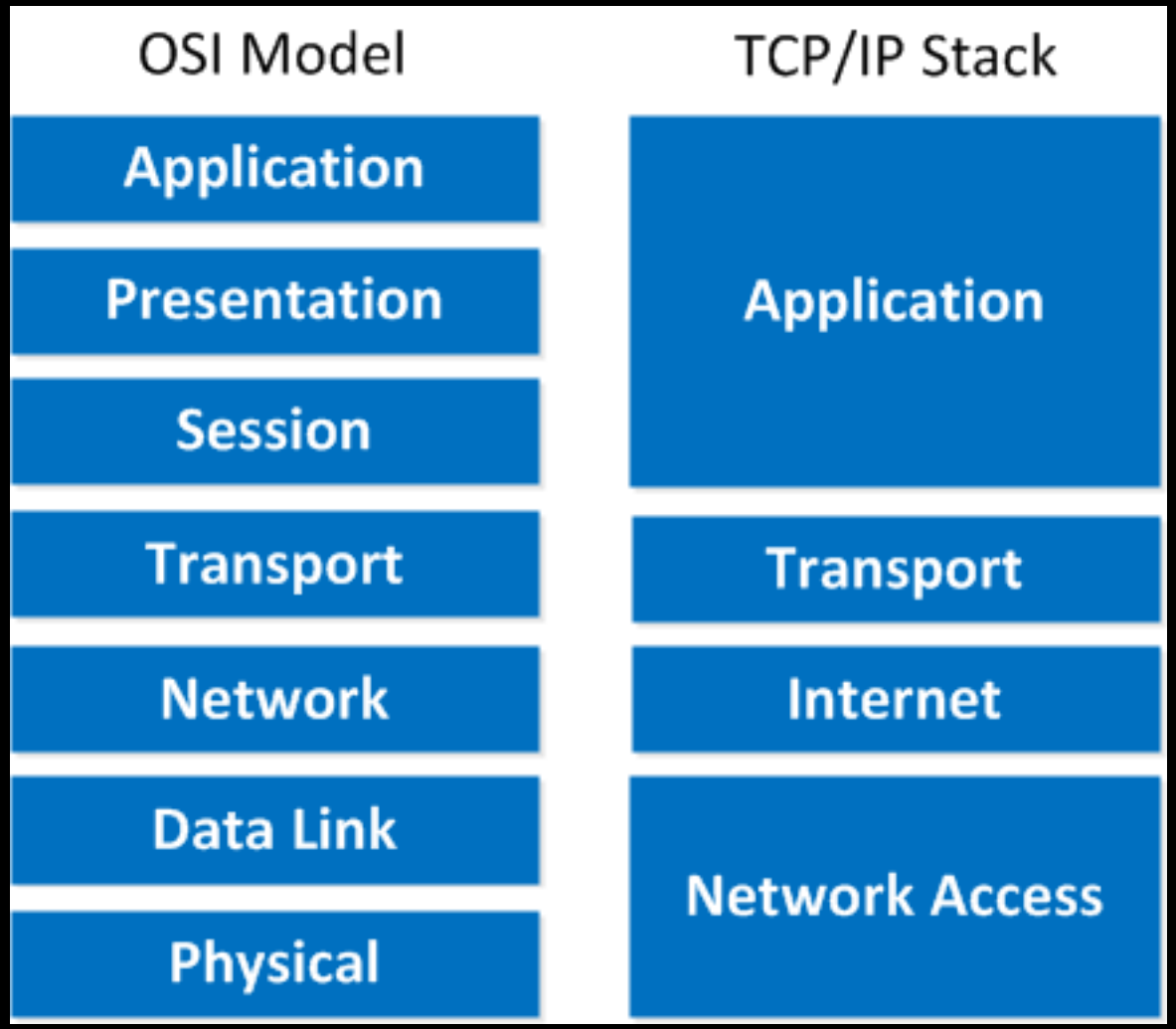
- Connecting computers together!
- The internet

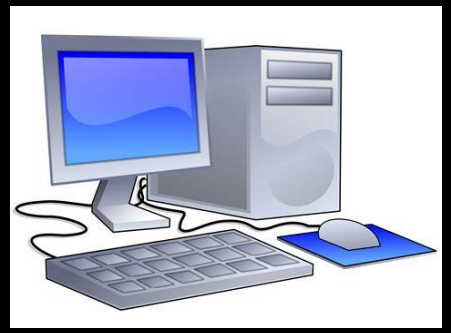
Goals

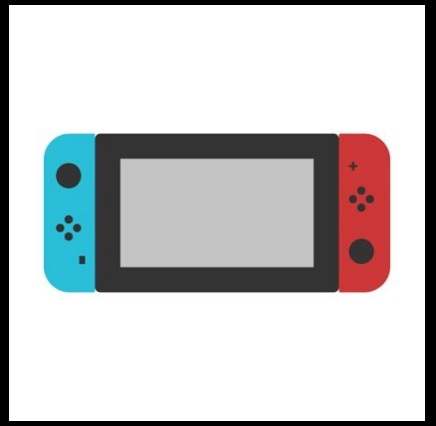
- Have a sense of how information travels between devices
- Know the difference between the internet and the world wide web

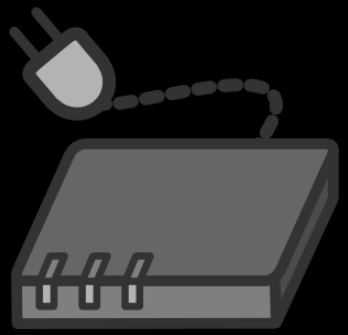
“The internet is just a world passing notes in a classroom.”- Jon Stewart

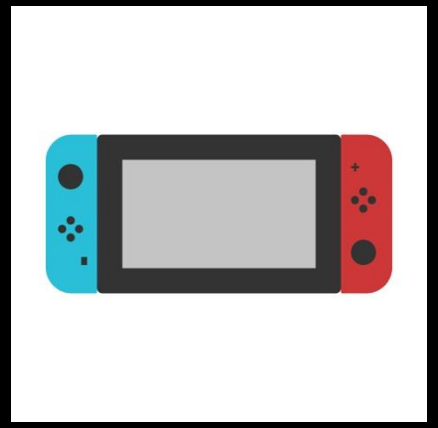
What exactly is the internet?


OSI Model
TCP/IP Stack
Application
Presentation
Session
Transport
Network
Data Link
Physical
Application
Transport
Internet
Network Access

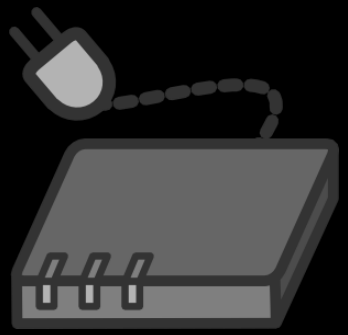


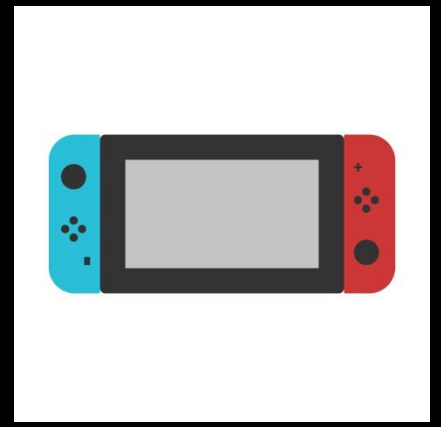




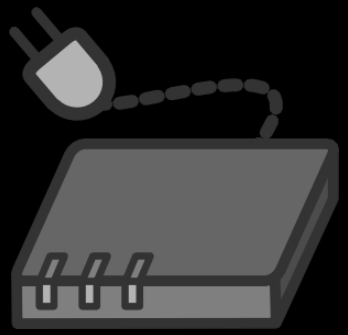


Local Area Network (LAN)

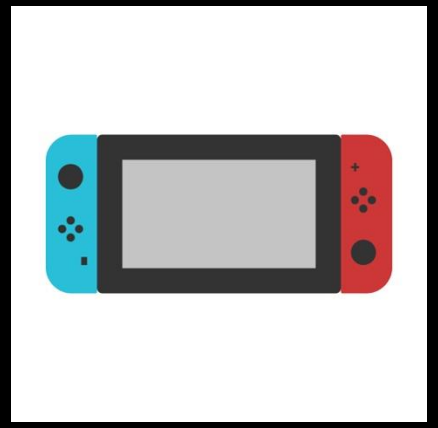




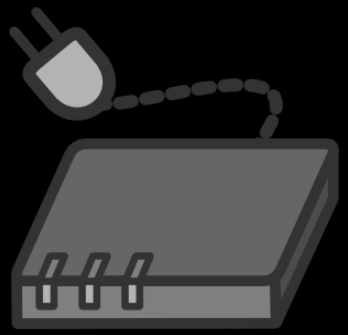
Local Area Network (LAN)



Devices can be connected by ethernet cables

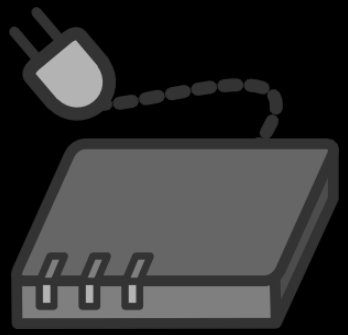


Local Area Network (LAN)

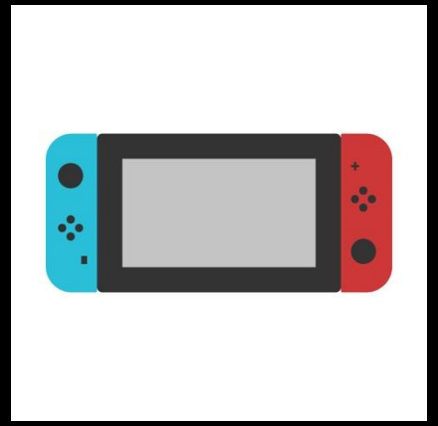


They can also be wirelessly linked

Local Area Network (LAN)



Router



Router

- The router handles how devices communicate with each other
- Each networked device has a unique identifier called a MAC address
- The router maps these to an IP address and informs devices what else is on the network and how to find it
  - More on IP address in a bit

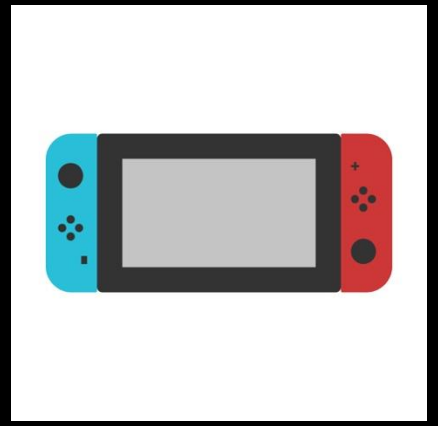
Local Area Network (LAN)

Router

192.169.0.1

192.169.0.24

192.169.0.5



192.169.0.20

Local Area Network (LAN)

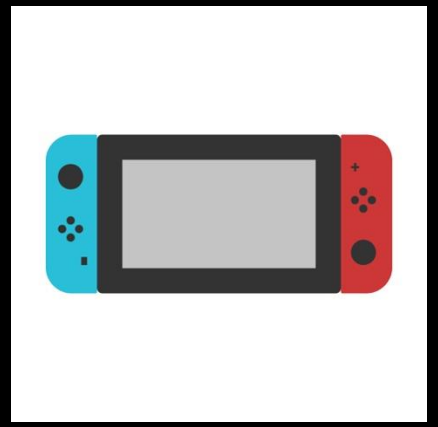
Router

192.169.0.1

Now, what if we wanted to communicate beyond this network?

192.169.0.24

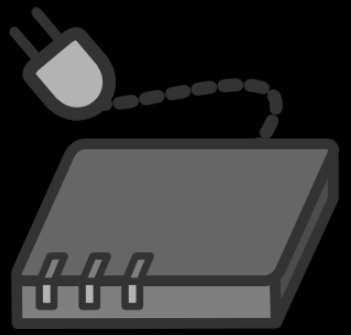
192.169.0.5



192.169.0.20

Local Area Network (LAN)

Modem



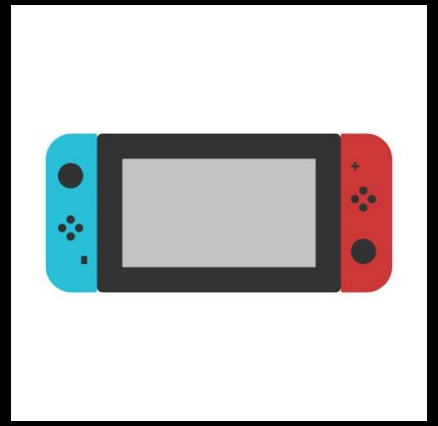
Router

192.169.0.1

Now, what if we wanted to communicate beyond this network?

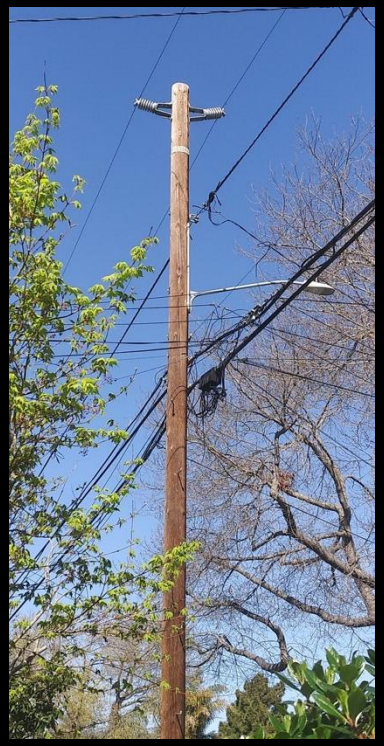
192.169.0.24

192.169.0.5



192.169.0.20

Internet Service Provider (ISP)



Modem

- Short for modulator-demodulator

- Takes the signals from the router and converts them into formats that can be transmitted over different types of transmission medium

Internet Service Provider

- These ISPs can connect a large number of smaller networks together
- Routers share information about these different networks and how they are connected
- Different ISPs connect to network access points
  - Public infrastructure that serves as the backbone to the internet


Map
Satellite
Greenland
Iceland
Finland
Sweden
Norway
Russia
United Kingdom
Poland
Germany
Ukraine
Kazakhstan
Mongolia
Canada
France
Italy
Spain
Turkey
China
Japan
South Korea
United States
North Atlantic Ocean
Afghanistan
Iraq
Iran
Pakistan
Algeria
Libya
Egypt
Saudi Arabia
India
Thailand
México
Mauritania
Mali
Niger
Chad
Sudan
Nigeria
Ethiopia
Venezuela
Colombia
Kenya
DR Congo
Tanzania
Indonesia
Papua New Guinea
Brazil
Perú
Angola
Bolivia
Namibia
Madagascar
Botswana
Indian Ocean
Australia
South Pacific Ocean
South Atlantic Ocean
Chile
South Africa
Argentina
New Zealand

The Internet Protocol (IP)

- Once we connect all of these devices together how do we actually do things?

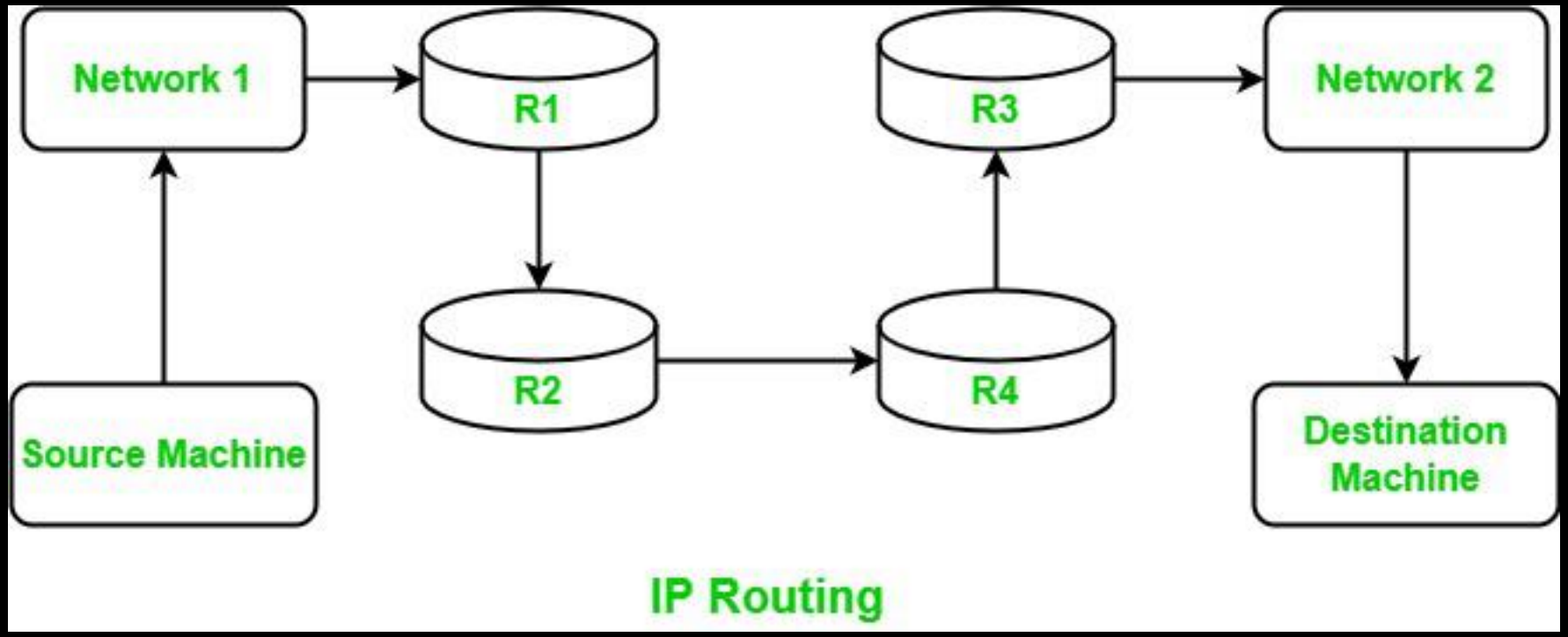
- The IP protocol is a special protocol that routers use to deliver messages from on machine to another
  - A *packet* is the common term for the smallest unit of data traveling this way

IP Addresses

- Come in either IPv4 or IPv6 formats
  - IPv4: 33.21.75.2
  - IPv6: *2001:0db8:85a3:0000:0000:8a2e:0370:7334*
- Hierarchical structure
  - ISPs, Countries, ETC get a whole chunk of addresss to give out
  - Left hand terms more coarse grained, right hand side more fine grained
- Used to local a machine connected to the internet

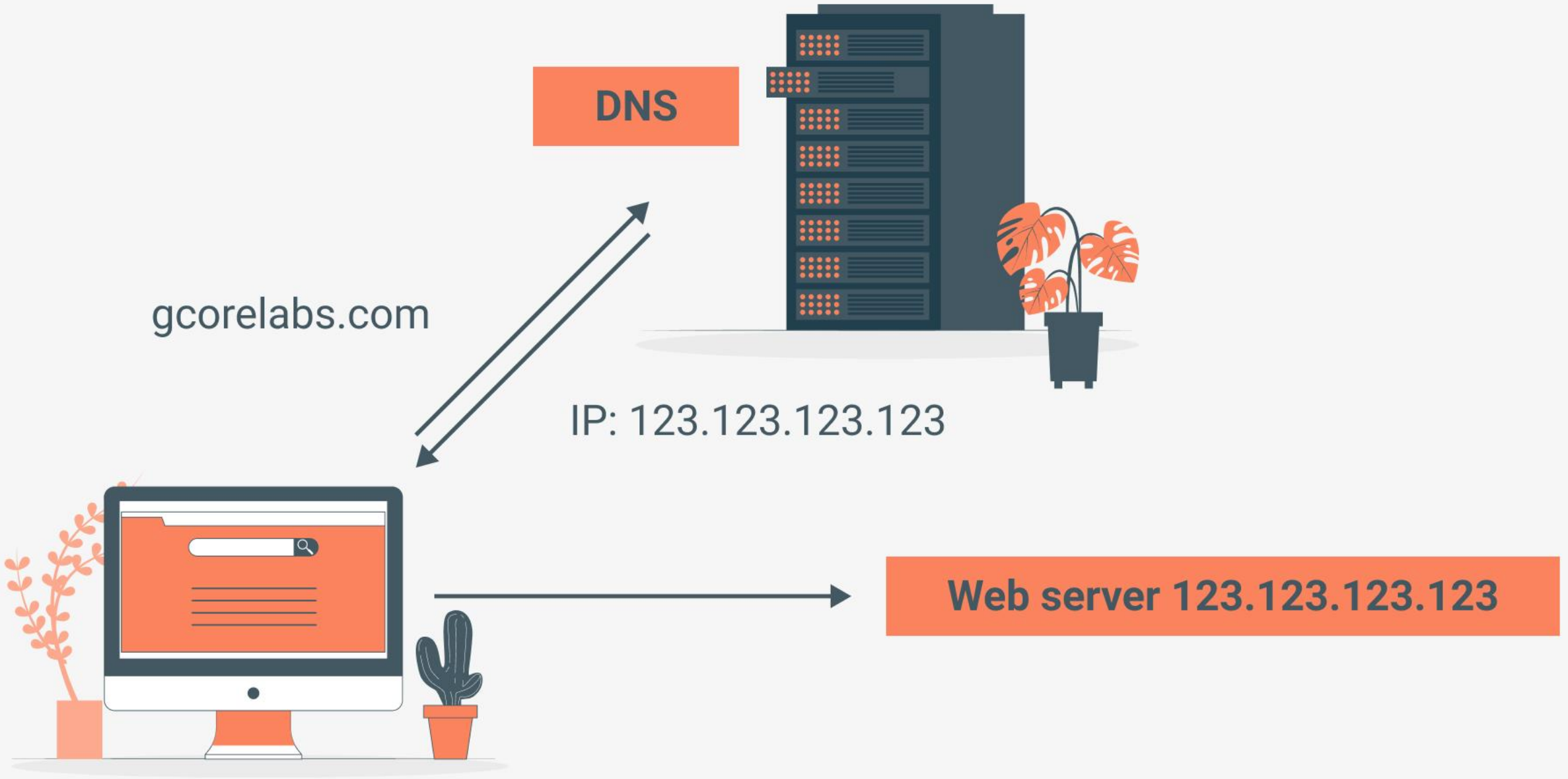
Routing

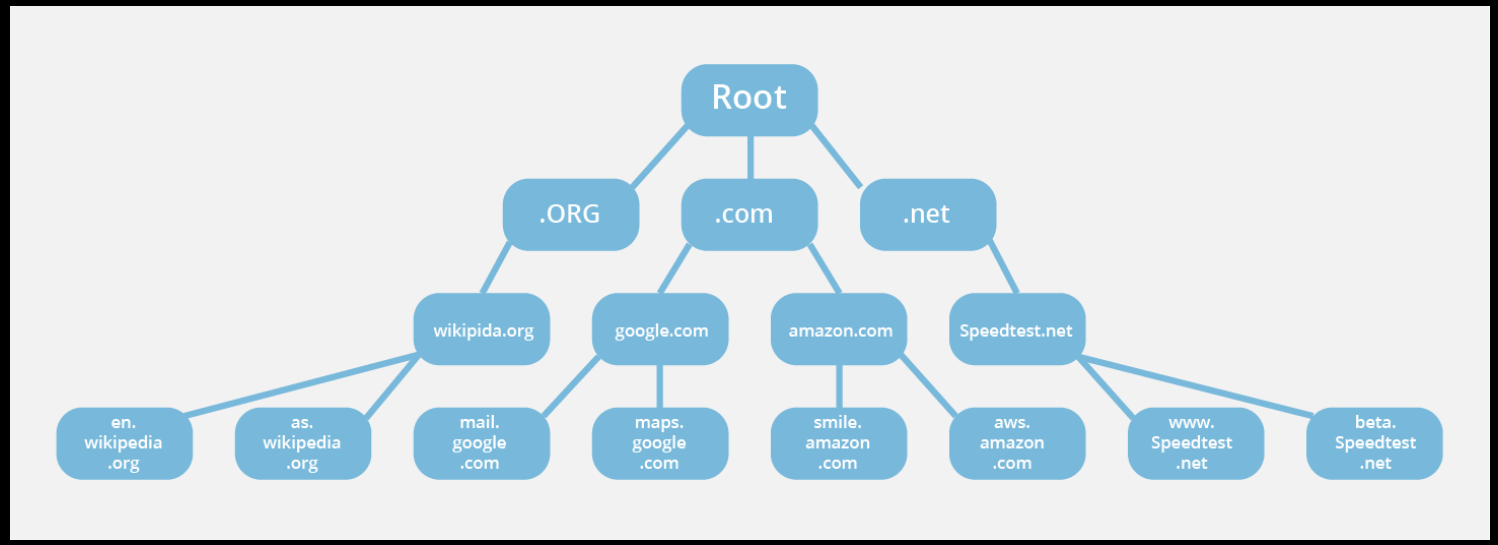
- The IP protocol lets routers learn about the IP addresses of devices closely connected to them
- They can use this to determine the best path to the destination device
- Usually more than one path -> Allows for a flexible and robust internet


Network 1
R1
R3
Network 2
R2
R4
Source Machine
Destination Machine
IP Routing

Domain Name Service (DNS)

- Easy to connect to a device if you know its IP address
  - Not exactly intuitive for humans
- Need a way to associate these addresses with easy to understand names
- DNS is a special protocol that does this mapping


DNS
gcorelabs.com
IP: 123.123.123.123
Web server 123.123.123.123


Root
.ORG
.com
.net
wikipida.org
google.com
amazon.com
Speedtest.net
en.
wikipedia
.org
as.
wikipedia
.org
mail.
google
.com
maps.
google
.com
smile.
amazon
.com
aws.
amazon
.com
www.
Speedtest
.net
beta.
Speedtest
.net


+
−
L
K
40
310
5
66
6
121
312
224
105
90
61
133
2
26
34
Leaflet | Map data © OpenStreetMap contributors

Other aspects of the Internet

- Many other protocols allowing for things to work
  - BGP
  - DHCP
  - TCP
  - UDP

- Firewalls
  - Often router level
  - Block IPs
  - Also do more fancy things

World Wide Web (WWW)

- Not enough to have a bunch of computers connected together, we want to do something fun with that

- The web comprises of the services, applications, interconnected files, etc that are accessed via the internet
  - The internet is the infrastructure, the web is the service

- Uniform Resource Locators (URLs) specify a location on the web
  - Comprises of a protocol, a hostname, and a file name
  - I.e. https://alishahc.com/papers/mixing.pdf

Other protocols

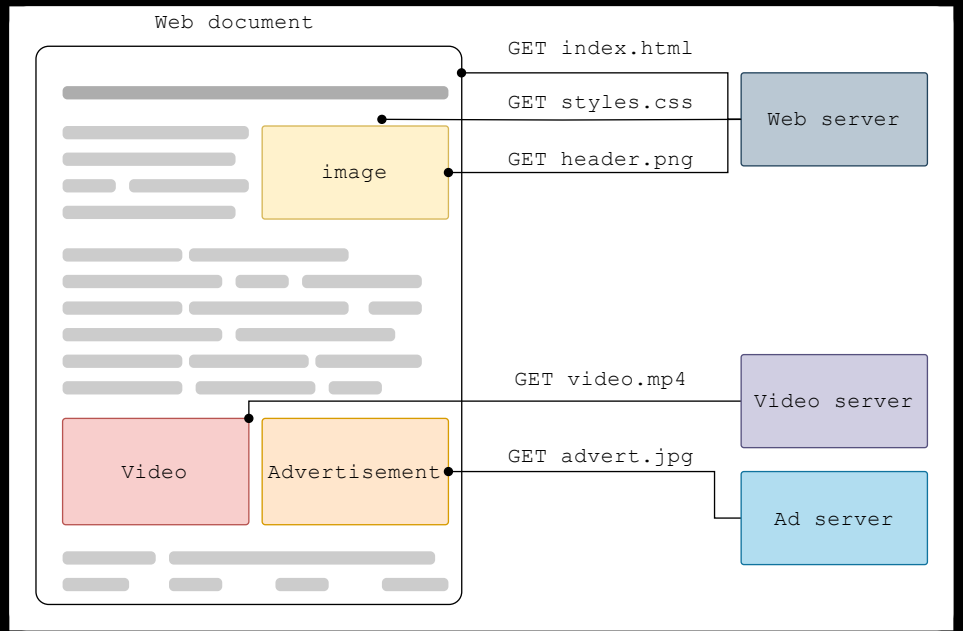
- There are many different protocols used on the Internet

- Each has a different application
  - FTP
  - POP/IMAP
  - SSH
  - VoIP

Hypertext Transfer Protocol (HTTP)

- The core of the web

- A protocol that connects webpages (HTML documents) using hyperlinks that specify URLs



Web Browsers

- Special applications designed for browser the web
- Primary purpose is fetching and rendering HTML pages
  - This involves fetching any additional resources needed by the web page
  - Running any client-side code needed by the web page
- Often is characterized by a *User Agent*
  - This is information shared with web pages about the version and other details about the web browser

Search Engines

- Search Engines are often conflated with web browsers but they have differing purposes
- A search engine is a web application that indexes web pages and in able to provide the user with a list of links
- A web browser is necessary to actually click these links and load the resulting pages
- Google Chrome is a browser, google search is a search engine

Extensions

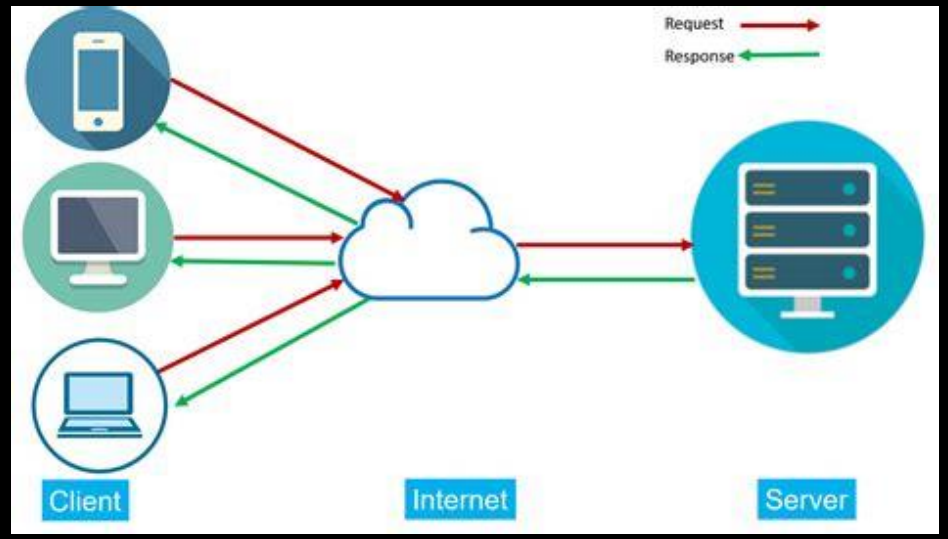
- Browsers can be customized with mini applications that change their functionality, often known as plugins or extensions
- Unlike client-side code (i.e. javascript) which runs for a specific web page and can only access the contents of that page, extensions run globally and can often access all aspects of the web browser to perform their role.
- It is this added power that allows for extensions such as adblockers

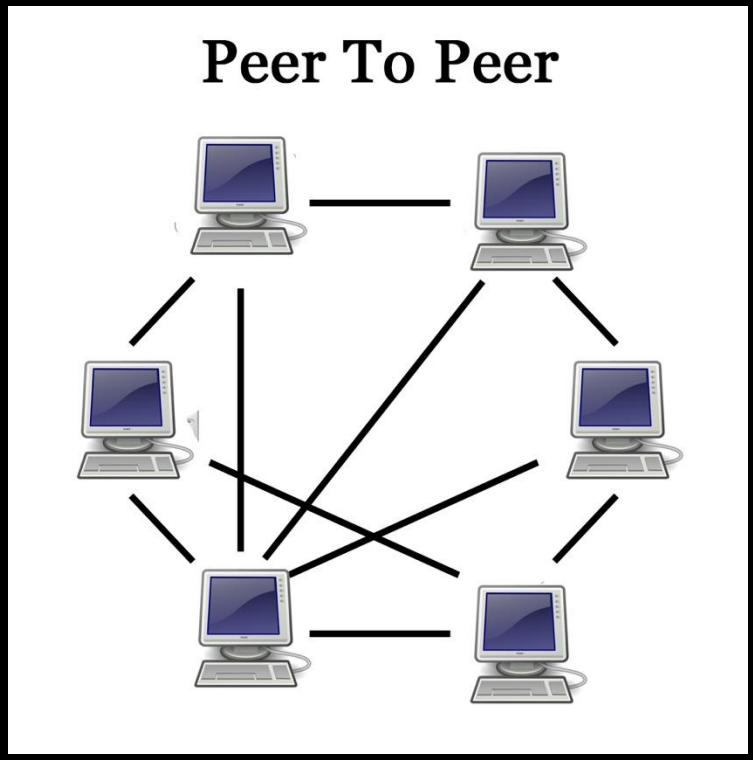
Cookies

- By default, web pages have limited information about who is visiting their sites.
  - They could try and record IP address and map information to that but that is imprecise and hard to scale

- A cookie is a small amount of data that websites save on a user's local browser in order to recall information about them or their actions on the site
  - Adding items to a cart
  - Being logged in
  - Visited pages

Different networking models





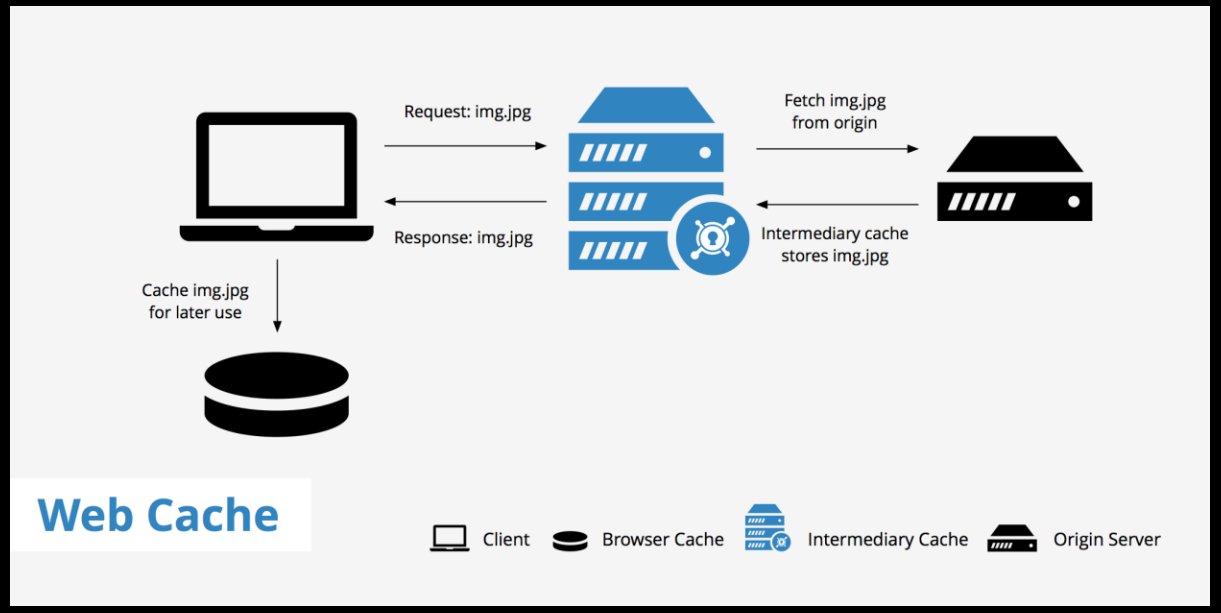
Data centers

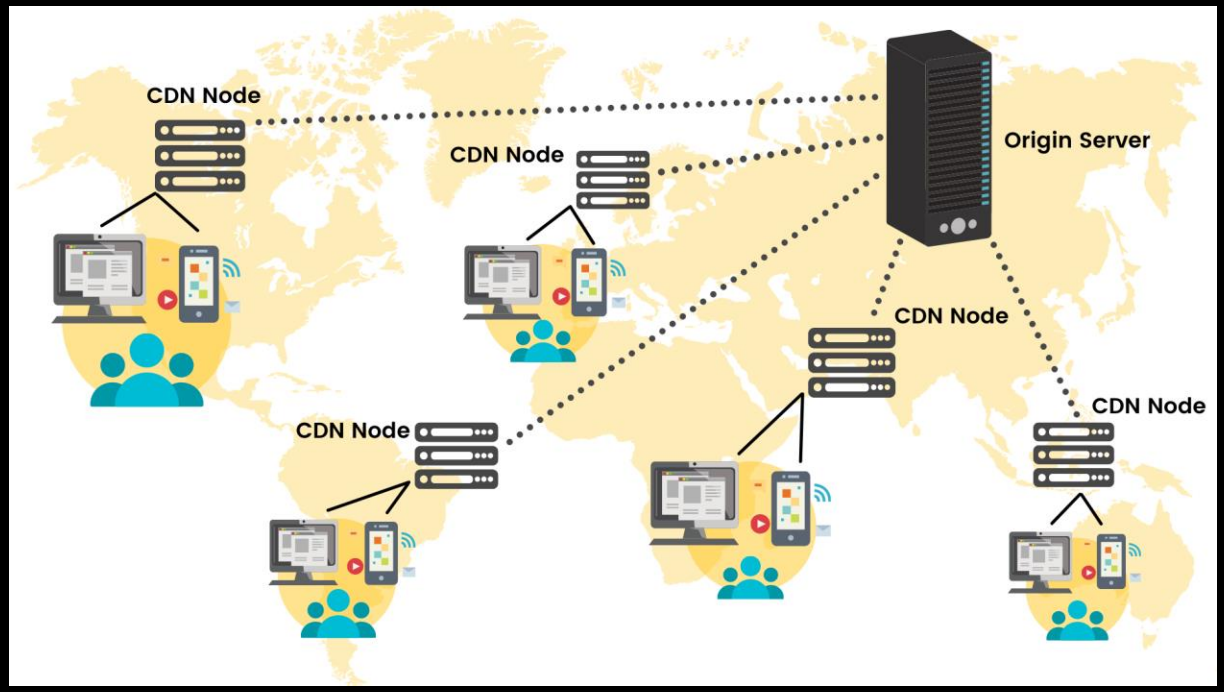
Cloud Computing

- With the vast resources available to service providers, the benefits began to outweigh network latency

- This led to a new model where users would perform computing tasks through web applications instead of locally
  - Word processing
  - Software development
  - Hosting

Caching and Content Delivery Networks (CDNs)





Cellular networks

- Modern cellular networks work quite similarly to the rest of the internet, with the cellular network infrastructure (i.e. cell towers) acting as the transmission medium

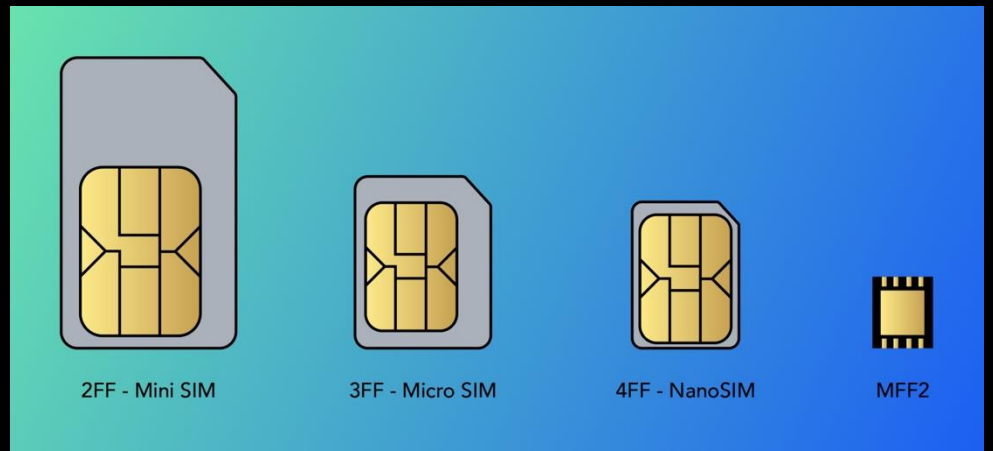
- One core difference is that phones are constantly moving
  - How to find the phone?
  - How to stay connected to the network?

- Special “handoff” procedure

SIM cards

- As the phone is moving physically around, it needs to continually reconnect to the cellular network

- A subscriber identity module (SIM) card, is used to identify that the phone should be allowed to access this network as it travels around



New forms of networked devices

- Internet of Things (IoT) devices
  - Smart Appliances
  - Security Cameras
- Wearables
  - Smart watches
  - Fitness trackers
- Public Infrastructure

Next Time

- Mechanisms for Privacy