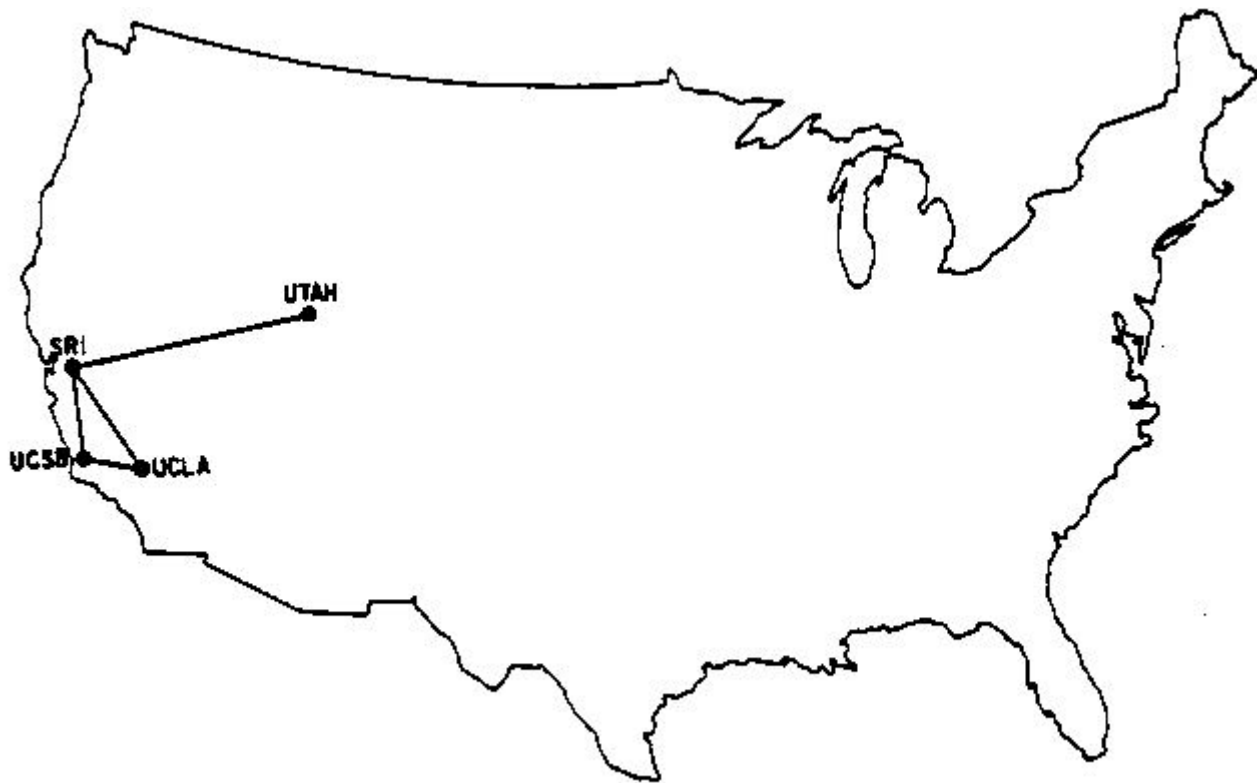


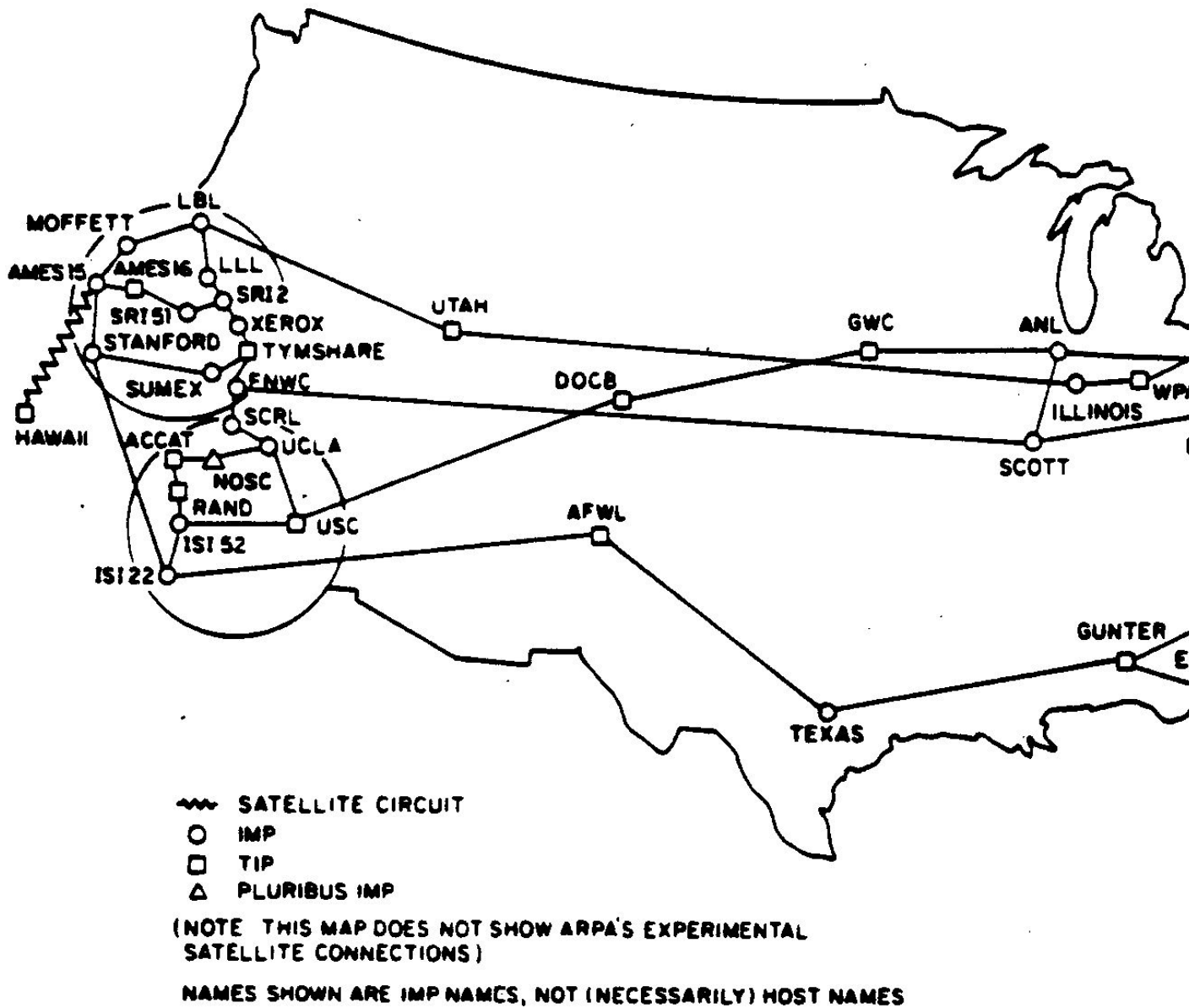
The Internet

History and Overview

The Internet is a communications network that spans billions of devices. It started as a project funded by the U.S. defense department and was originally known as the ARPANET. This is what it looked like in 1969:



and this is what it looked like in 1977:



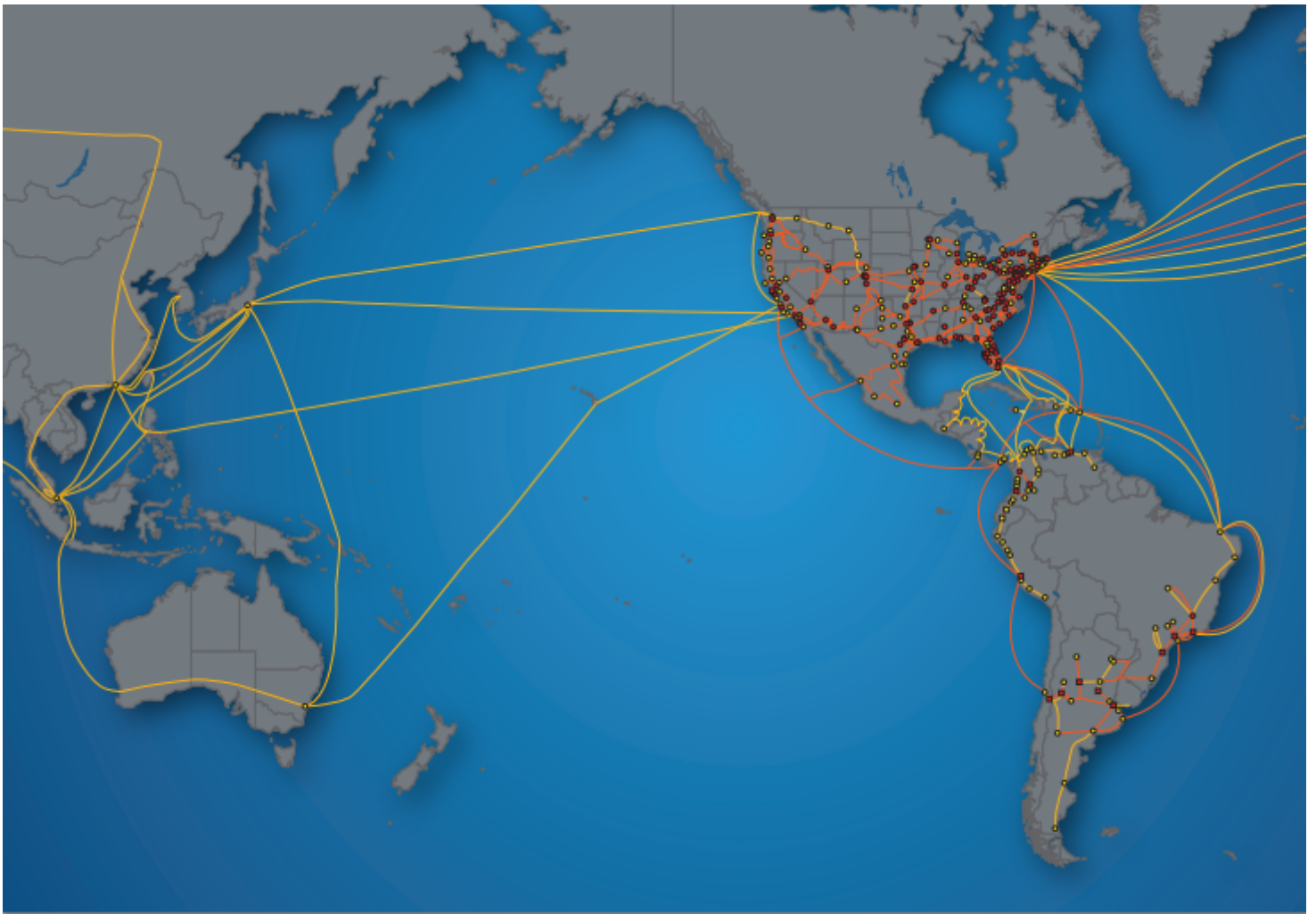
It's nearly impossible to visualize what the Internet looks like today, but here is a good try:



There are billions of devices connected to the Internet. These include consumer devices, such as computers, smartphones, tablets, and smart home devices (speakers, thermostats, refrigerators). The Internet also includes a lot of infrastructure needed for these devices to communicate, including routers, switches, and cell phone towers.

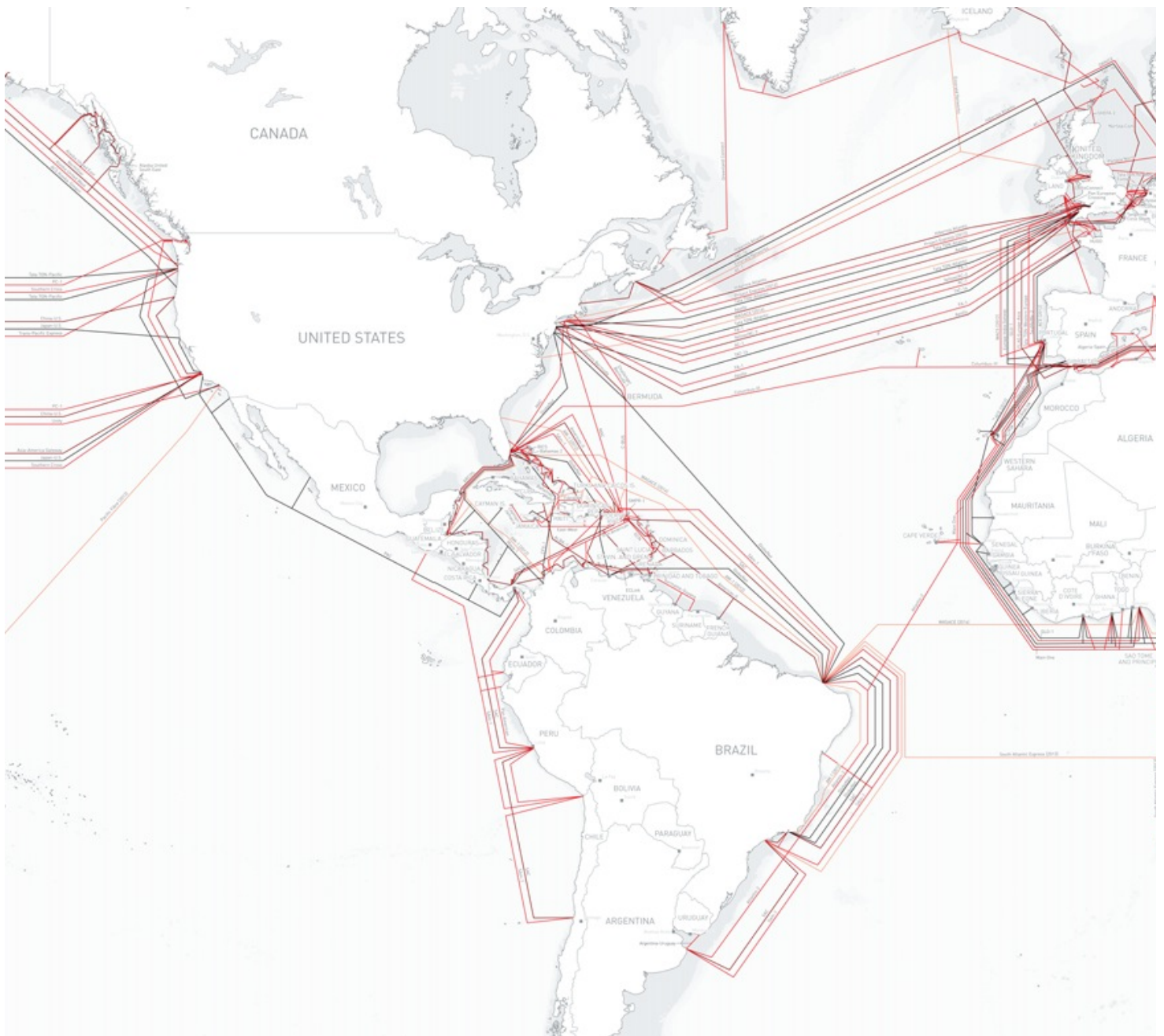
Since its early development as a U.S. government project, the Internet has since been both internationalized and commercialized, so that different parts of the network are owned by different companies and reside in different countries. Some countries (such as China) have nationalized infrastructure, whereas many others rely on the private sector to operate the network.

A single company can run its own global network. For example, here is what Level 3's network looked like recently:



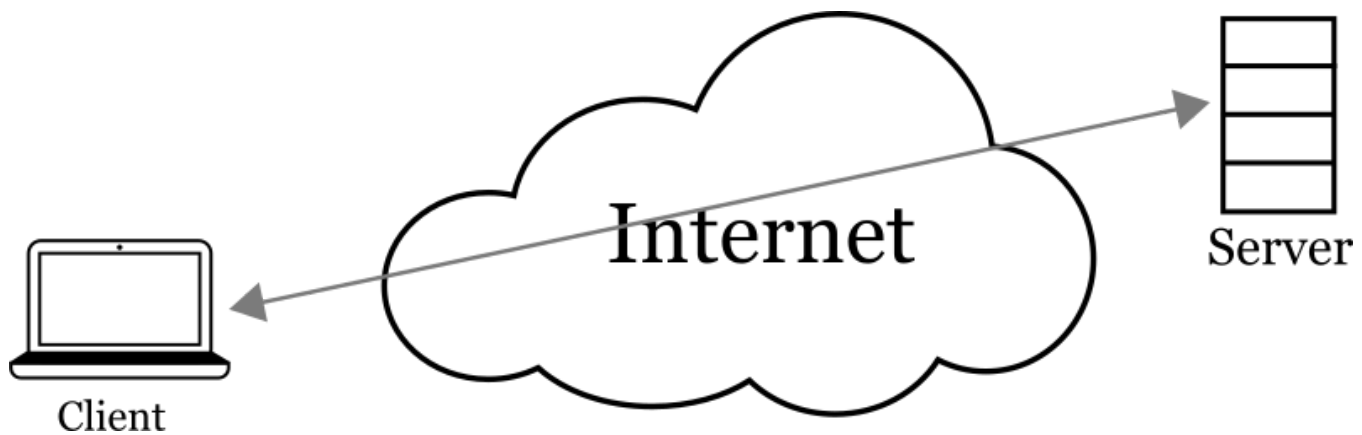
The Internet contains many of these networks, all interconnected to each other.

Another fascinating part of the Internet is that there are many cables running oceans and seas that connect parts of the network. Here is a map of some undersea cables:



How the Internet Works, Briefly

Most applications on the Internet use client-server communication. As shown below, a client connects to a server, sends one or more requests to the server, and receives responses for each request. Examples of this kind of communication include a web browser connecting to a web server (e.g. amazon.com), the Twitter app connecting to the Twitter service, an instant messaging application connecting to the instant messaging server, and so on.



DNS

The first step for a client usually involves figuring out what server to connect to. Out of all the billions of devices on the Internet, the client needs to specify *which* other device it wants to connect to and which service it wants to use when connecting to that server (e.g. web, email, or some other application).

To do this, clients use the Domain Name Service (DNS). The client specifies the *name* of the server it wants to connect to (e.g. amazon.com), and asks DNS to return the *IP address* of this server (e.g. 176.32.98.166).

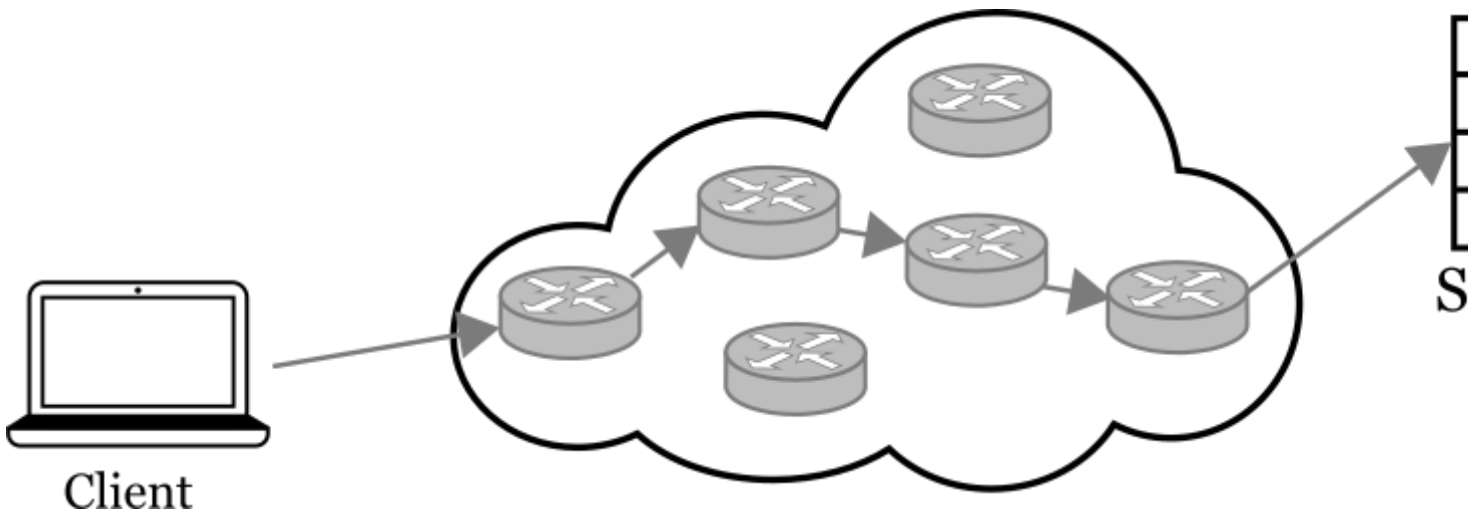
TCP

The next step for a client is to use the Transmission Control Protocol (TCP) to make a connection to the server. To do this, the client specifies the IP address it learned from DNS, plus the *port* number of the service it wants to connect to. For example, the port number for a web server is 80 (HTTP) or 443 (HTTPS), for email it is 25 or 465 or various other ports, and for SSH it is 22.

Once TCP makes a connection between the client and the server, the client can send requests to the server ("send this web page", "retrieve this email message"), and the server can send responses. The format of these messages depends on the application being used. TCP ensures that the connection between the client is *reliable*---any message the client sends is received by the server and vice versa. The only reason messages will not go through is if the network is not working. TCP does *not* guarantee anything about performance. The messages exchanged by the client and server will be delivered, but they might be delivered slowly or quickly, depending on a variety of factors.

Routing and IP

Whenever TCP sends a message from a client to a server, it divides the message into a series of packets and sends each one individually. Each packet must find a way across the Internet from the client to the server. As shown below, the connection between the client and server traverses a particular *path* through the Internet. This path consists of a series of routers that are determined by a collection of routing protocols.



At each step in this path, the packet is sent to the next hop using the Internet Protocol (IP). At each hop, IP sends packets as fast as it can. If a router receives packets faster than it can send them, packets can be delayed (because they have to wait in a buffer before being sent to the next router) or dropped (because the buffer is full). The effects of delay and dropping are often called *congestion*. TCP will re-send any packets that are dropped, but it will also slow down when loss occurs.

Layers and Standards

TCP and IP are the primary *protocols* used to send data across the Internet, although there are many more. The protocols that are used are often arranged in *layers*, from top to bottom, as shown below. The application layer is at the top (e.g. HTTP for a web browser), followed by the transport layer (e.g. TCP), the network layer (e.g. IP), and the link and physical layers (e.g. WiFi or Ethernet).

Application
Transport
Network
Link
Physical

These protocols are standardized by a variety of organizations:

- W3C: standardizes web protocols, such as HTML and CSS.
- IETF: standardizes Internet protocols, such as HTTP, TCP, and IP.
- IEEE: standardizes link protocols, such as WiFi and Ethernet.

It's important to standardize protocols so that all the different devices on the Internet can interoperate. The alternative is to use proprietary protocols, which then only work with devices or software from one or several manufacturers.

Learn More

If you want to learn more about how the Internet works, [*Computer Networking: A Top-Down Approach \(7th Edition\)*](https://www.amazon.com/Computer-Networking-Top-Down-Approach-7th/dp/0133594149) [\(https://www.amazon.com/Computer-Networking-Top-Down-Approach-7th/dp/0133594149\)](https://www.amazon.com/Computer-Networking-Top-Down-Approach-7th/dp/0133594149), by Kurose and Ross, is an excellent networking textbook. This is the book currently used for CS 460.