How the Internet Works

Learning Goals

- Recognize core terms related to the internet, including: browsers, routers, ISPs, IP addresses, DNS servers, protocols, packets, and cloud

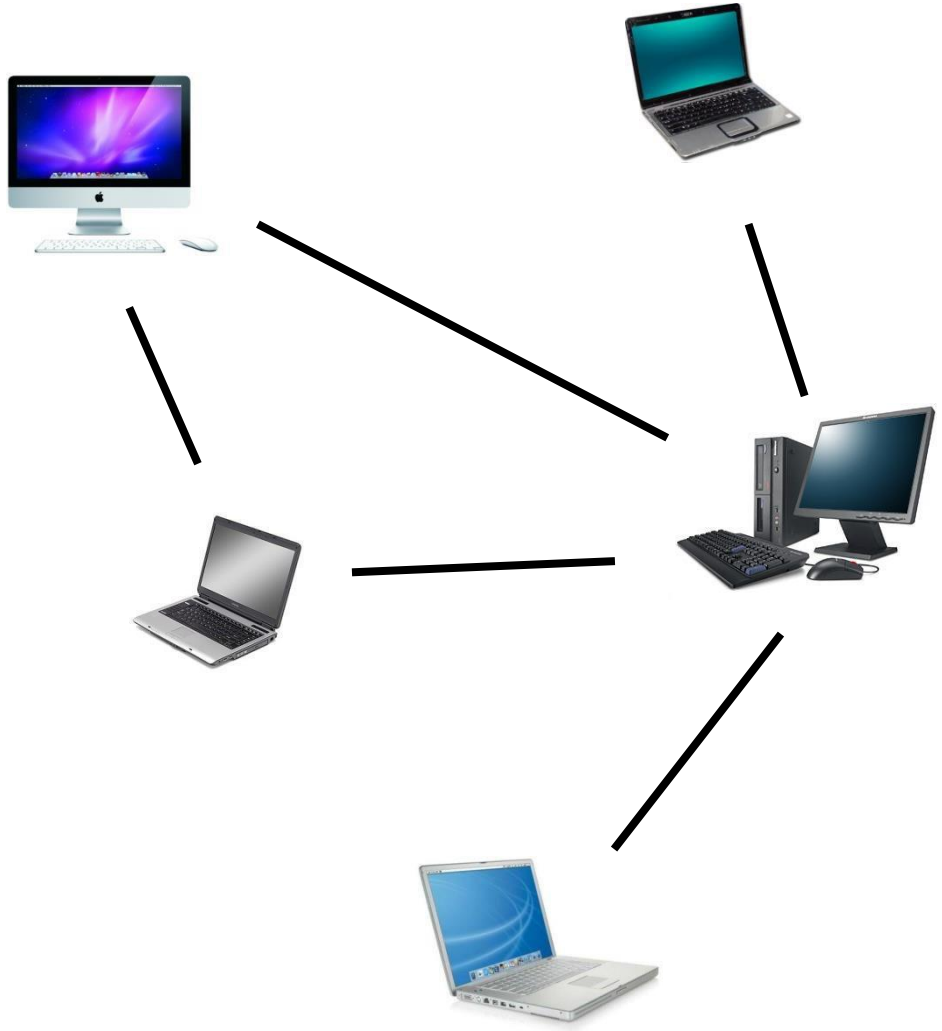
- Understand at a high level the internet communication process that happens when you click on a link to a website in your browser.

- Understand at a high level that the internet is fault tolerant due to being distributed

What is the Internet?

The Internet is a network of computer networks all across the world that are connected. The purpose of the internet is to send data between different computers in a manner that is decentralized--no one person has control over the whole thing.

It's like a graph where the nodes are computers and the edges are different methods of transmitting information.

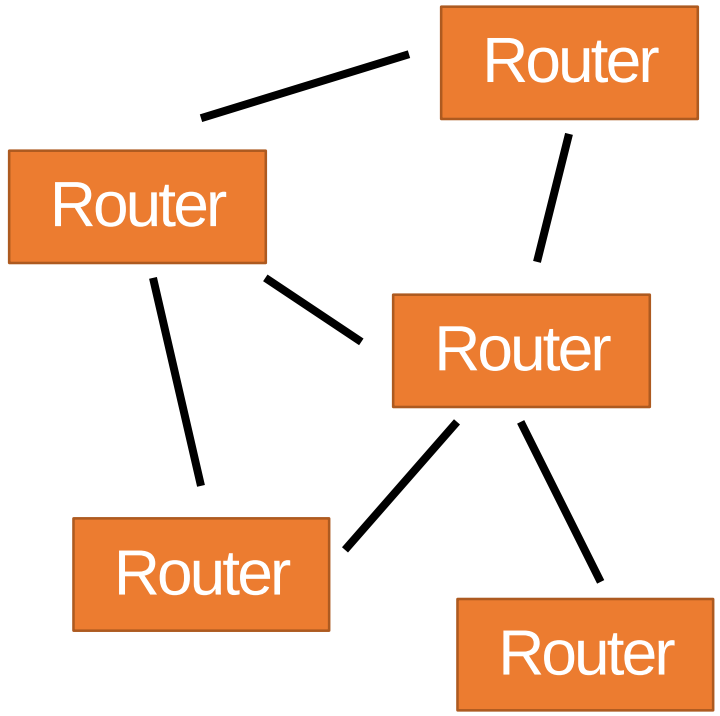


Routers are the Core

The core of the internet is a collection of devices called routers.

These devices act like switches – they take in data and send it to one of many possible locations based on the end destination of the data and the current connections on the internet.

There are thousands of routers spread across the world to help move data around from one place to another.

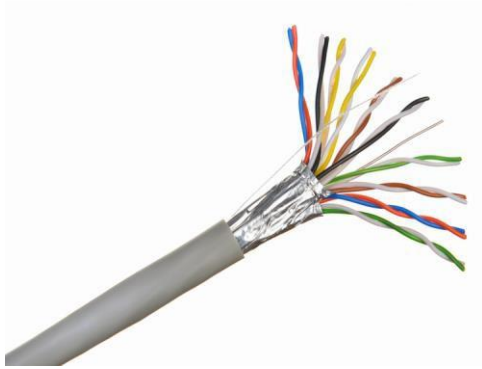


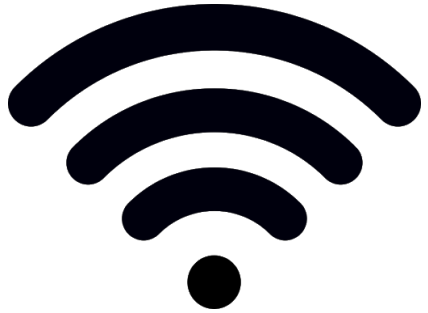
Connections Between Routers

Routers are commonly connected by cables, which are used to send data across a long distance. That data is usually represented using bits.

Cables range from telephone wires to coaxial cable to fiberoptic cable. All of these systems convert bits to different real-world representations (analog signal, electricity, light, etc.).

Computers can also send data to routers over Wi-Fi. In this connection, data is sent over a short distance via radio waves.





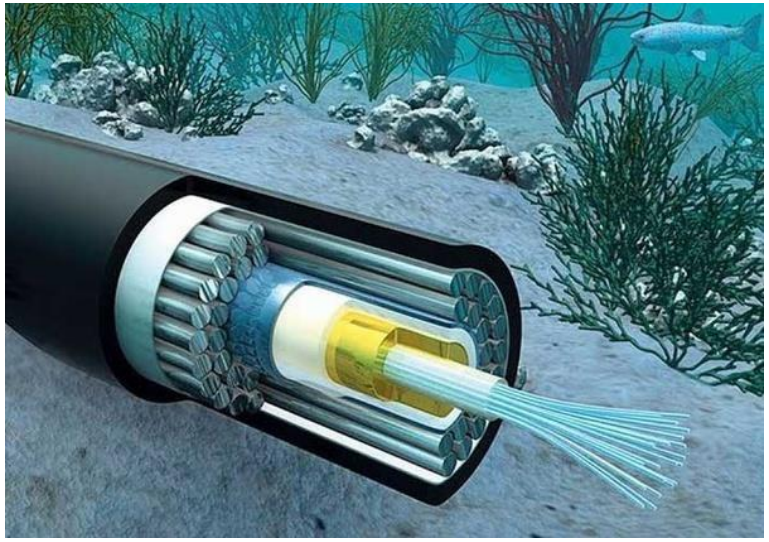
International Internet

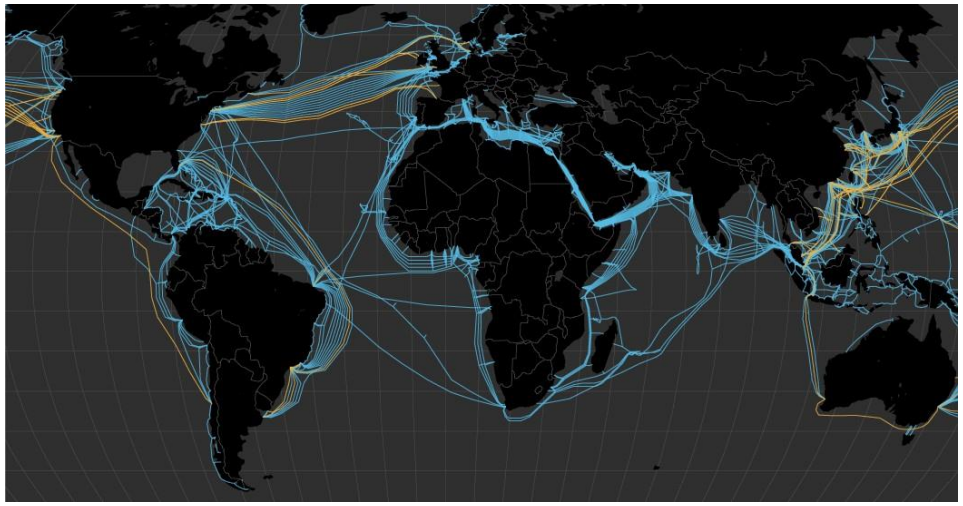
How does the internet connect across continents?

Giant fiberoptic cables have been laid on the ocean floor. Most international internet traffic is transmitted through these cables.

Read more:

https://www.nytimes.com/interactive/2019/03/10/technology/internet-cables-oceans.html



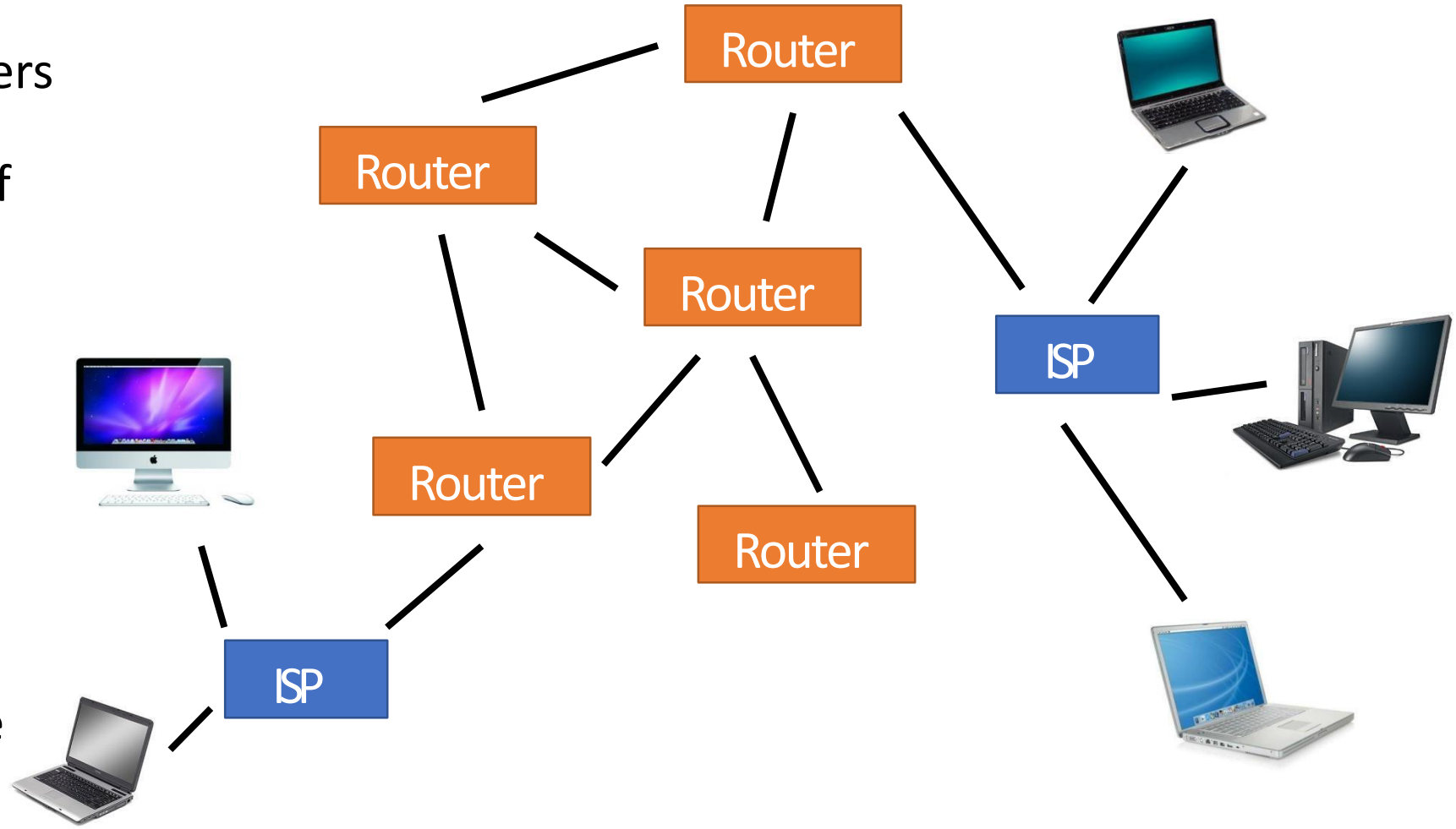


ISPs Manage Local Connections

Internet Service Providers (ISPs) connect a user's computer to the core of the internet.

Verizon, Comcast, etc. are all ISPs.

Organizations can be their own ISP – Duke is its own ISP, for example

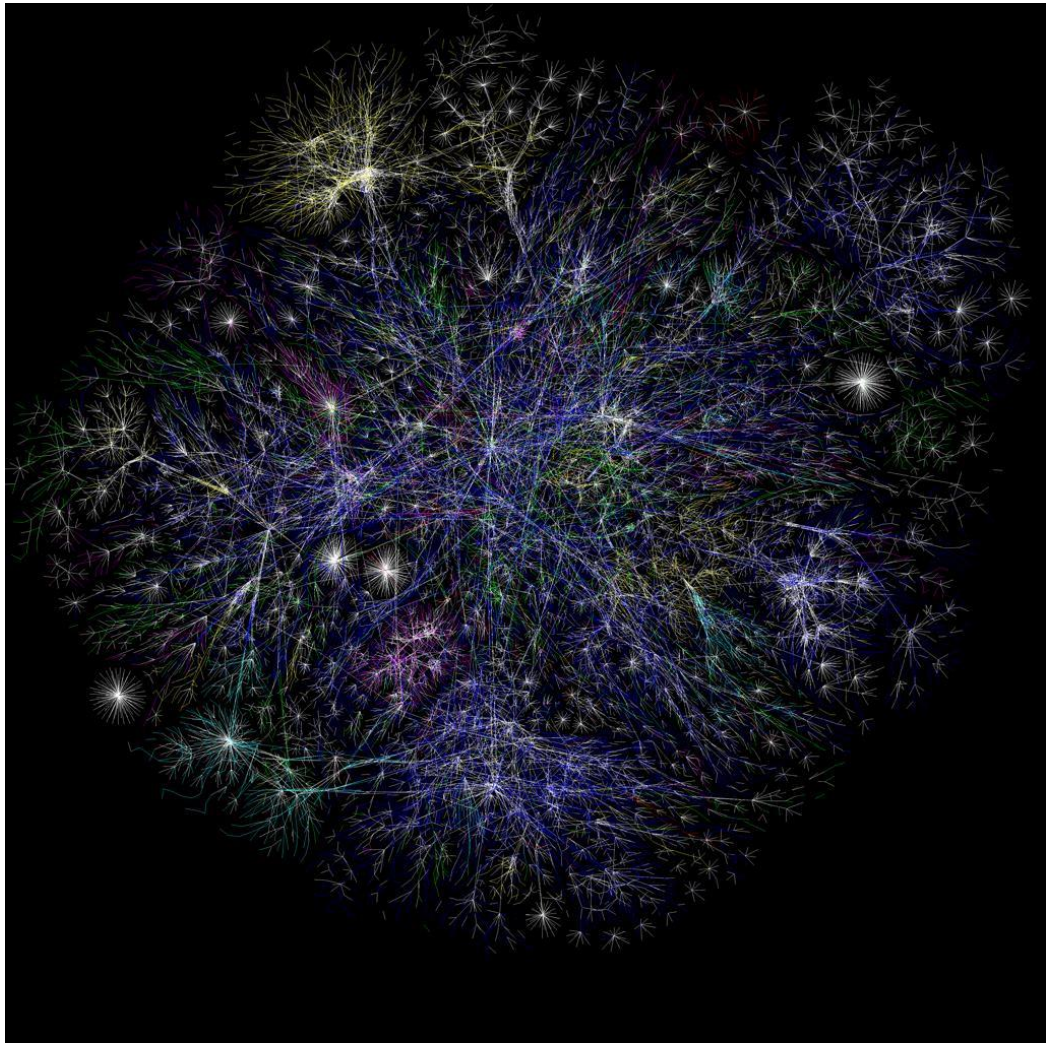


The Modern Internet is Huge!

The internet today is used widely across the world and contains millions of computers and connections.

The picture to the right (from the Opte Project) illustrates the connections of the internet in 2005 – it's even more widely connected now!

How is it possible for us to make a request for a specific website in this massive web and get the result back so quickly?



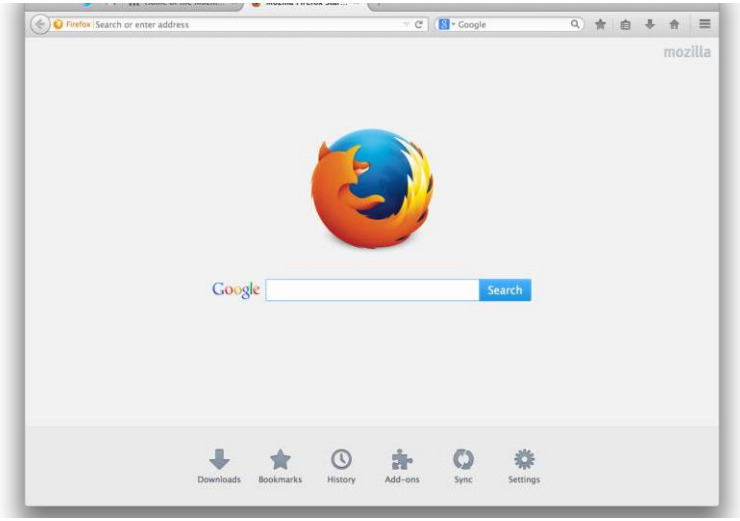
Journey of a Website

Browsers Display Data

Your browser (Firefox, Chrome, Safari, etc.) is an application that receives data from the internet and organizes it into a webpage that you can read.

Browsers receive webpages as text and turn that text into visual content using a protocol called HTML (HyperText Markup Language).

You can view the HTML of any webpage by right-clicking and selecting 'View Page Source'.




```
</nav>

<div class="container">

<h2>15-110: Principles of Computing</h2>

<br>
<p><b><font color="#990099">Due to the COVID-19 Epidemic, all classes and office hours from 03/16 onwards will be conducted remotely. Please
refer to the class Piazza for links to the Zoom class sessions.</font></b></p><br>

Principles of Computing (15110) is a course in fundamental computing principles for students with little to no computing background. Programming
constructs: sequencing, selection, iteration, and recursion. Data organization: arrays and lists. Use of abstraction in computing: data
representation, computer organization, computer networks, functional decomposition, and application programming interfaces for graphics. Use of
computational principles in problem-solving: divide and conquer, randomness, and concurrency. Classification of computational problems based on
complexity, non-computable functions, and using heuristics to find reasonable solutions to complex problems. Social, ethical and legal issues
associated with the development of new computational artifacts will also be discussed. Prerequisites: none. <br><br>

<h4>Meeting Times</h4>
<table class="table table-striped">
  <thead><td><b>Session</b></td><td><b>Instructor(s)</b></td><td><b>Time</b></td><td><b>Location</b></td></thead>
  <tr><td><b>Lecture 1</b></td><td><b>Kelly Rivers (krivers)</b></td><td><b>MWF 2:30-3:20pm</b></td><td><b>DH 2210</b></td></tr>
  <tr><td>Recitation A</td><td>Diaj (dtoussai) and Enock (emaburi)</td><td>R  9:30-10:20am</td><td>GHC 5207</td></tr>
  <tr><td>Recitation B</td><td>Amanda (lianglij) and Neeraj (neerajsa)</td><td>R 10:30-11:20am</td><td>GHC 5207</td></tr>
  <tr><td>Recitation C</td><td>Mahima A. (mahimaa) and Rachel (rachelt1)</td><td>R 11:30-12:20pm</td><td>GHC 5207</td></tr>
  <tr><td>Recitation D/J</td><td>Frank (frankh) and Mahima S. (mshanwar)</td><td>R 12:30-1:20pm</td><td>GHC 5207</td></tr>
  <tr><td>Recitation E</td><td>Andrea (arestrad) and Emily (eding)</td><td>R 1:30-2:20pm</td><td>GHC 5207</td></tr>
  <tr><td>Recitation F</td><td>Meghan (mamcgraw) and Rishabh (rishabhc)</td><td>R 2:30-3:20pm</td><td>GHC 5207</td></tr>
  <tr><td>Recitation G</td><td>Elyana (erhurst) and Iris (ilu1)</td><td>R 3:30-4:20pm</td><td>GHC 5207</td></tr>
  <tr><td><b>Lecture 2</b></td><td><b>Margaret Reid-Miller (mr54)</b></td><td><b>MWF 3:30-4:20pm</b></td><td><b>DH 2210</b></td></tr>
  <tr><td>Recitation H</td><td>Jonan (jseeley) and Laura (lkoye)</td><td>R  9:30-10:20pm</td><td>GHC 5210</td></tr>
  <tr><td>Recitation I</td><td>Lauren (leheller) and Rhea (rkudtar1)</td><td>R 10:30-11:20am</td><td>GHC 5210</td></tr>
  <tr><td>Recitation D/J</td><td>Frank (frankh) and Mahima S. (mshanwar)</td><td>R 12:30-1:20pm</td><td><b>GHC 5207</b></td></tr>
```

URLs are Website Nicknames

Find www.google.com

At the beginning of the process, you have to make a request to access a specific website.

You generally do this by clicking on a link on a webpage or typing out a URL (Uniform Resource Locator). The URL is like a nickname for the website you want to access.

IP Addresses are Real Names

Find www.google.com

If a URL is a nickname for a website, an IP Address is its real name.

Every computer on the internet is assigned a series of numbers, like 172.217.9.206. That series of numbers uniquely identifies the computer that hosts a website.

The first step in finding a website is to translate the URL into the equivalent IP Address.

Google

172.217.9.206

IP Address Assignment

IP Addresses aren't a core part of a computer; they aren't built into the hardware or software. But they aren't entirely random either.

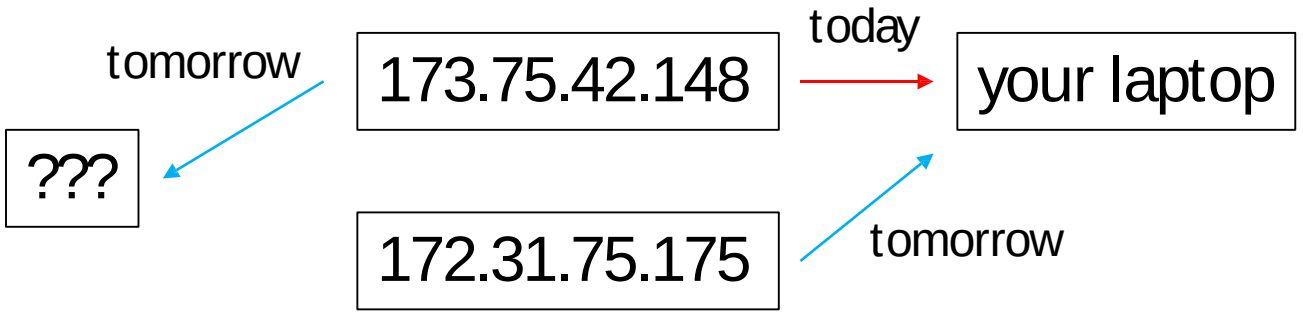
An organization called ICANN (Internet Corporation for Assigned Names and Numbers) assigns groups of addresses to different organizations (like ISPs and companies). The organizations then assign their numbers to individual computers when they connect to the internet.

IP Addresses are Static or Dynamic

Some IP Addresses are static. Many of these are the addresses of specific websites (like Google, or Duke).

152.3.72.197 -> Duke

Other IP Addresses are dynamic. They get assigned to different computers at different times. This is used for computers that go online and offline regularly (like your computer).



IP Address Meaning

The core standard for IP Addresses consists of four numbers, each between 1-255. In other words, each number is a byte.

Some numbers contain geographic information (country); some numbers contain information about the organization that owns the address.

But an IP Address does not say who owns the associated computer, or what kind of machine it is. This makes it possible for internet communication to be private to outside observers (though it is not private to the ISP).

IP Addresses at Scale

Question: if every computer needs to have a unique IP Address and an IP Address is 4 bytes long, how many computers can be on the internet at the same time?

Answer: $(2^8)^4$, or about 4.3 billion.

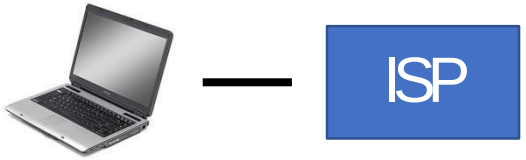
This seems like a lot, but it's still less than the world's population, and that doesn't count all the websites on the internet.

ICANN now has a new system for IP Addresses that contains 16 bytes; that can handle 1038 addresses. We're good for now!

Finding IP Addresses

Find www.google.com

How do we go from a URL to an IP Address?



Google

172.217.9.206

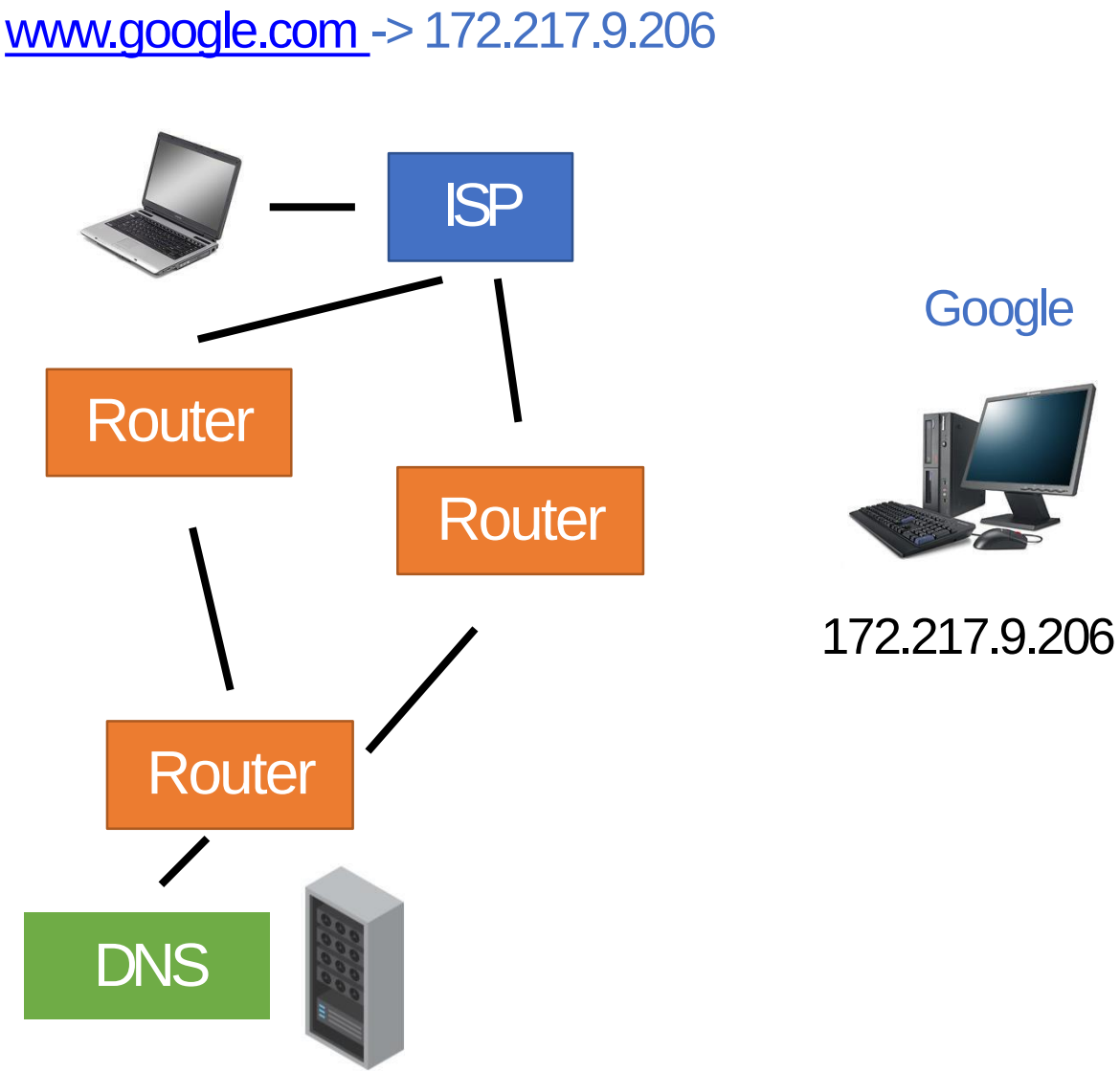
First, your computer checks with your ISP. It keeps a list of frequently-requested websites, so if someone has requested the same website recently, it can send back the IP Address immediately.

Finding IP Addresses

If your ISP doesn't know the IP Address, it sends your request on to the nearest Domain Name System Server (DNS server).

Your request may need to pass through several routers to get to a DNS server.

A DNS server is a computer that maintains a mapping of all URLs to IP Addresses. It will be able to find the correct address and send it back through the routers to you.

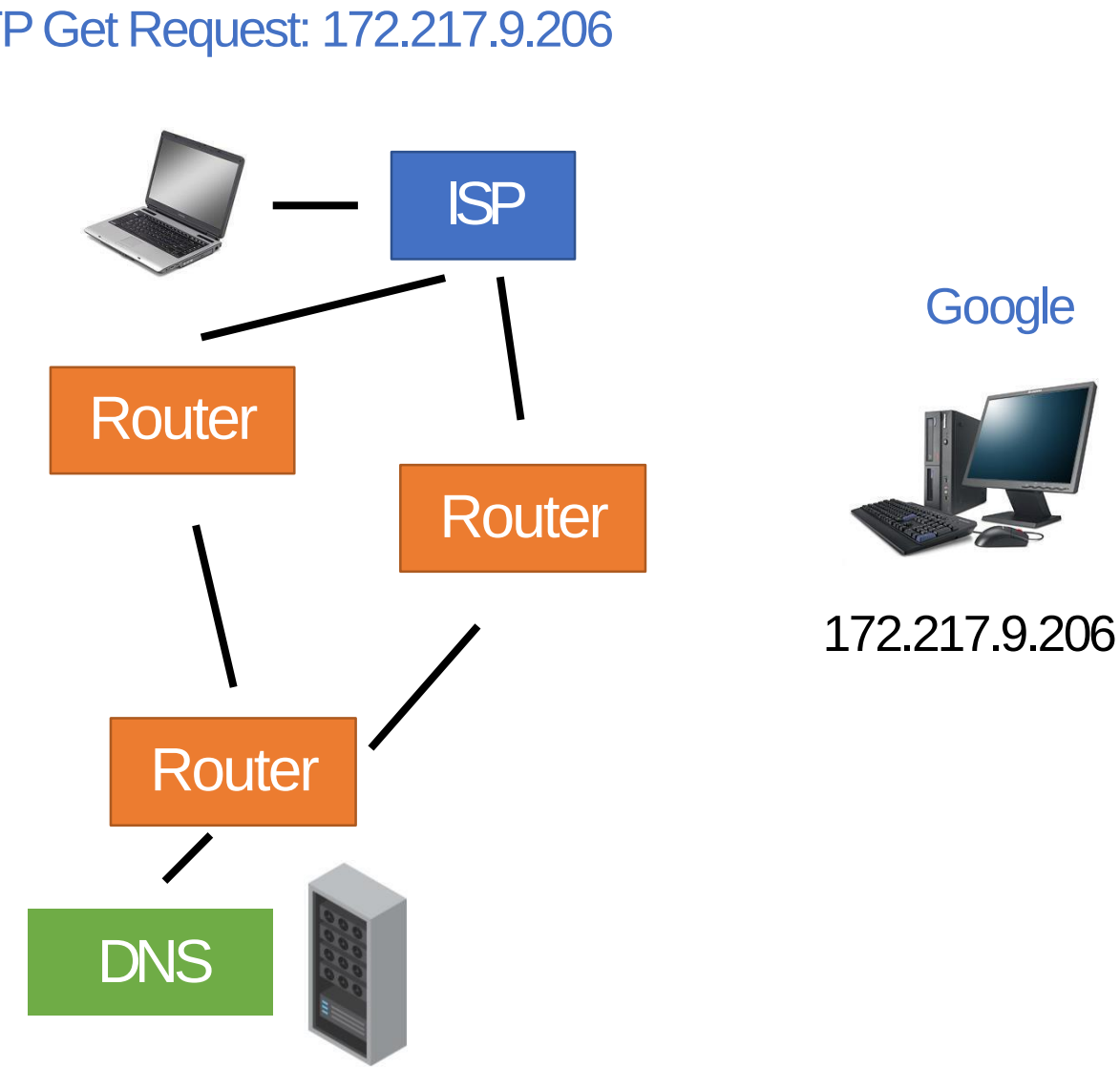


Requesting a Website

Once your computer knows what the IP Address is, it sends a request for a specific page to the IP Address.

The request is structured to match a certain protocol. For example, HTTP (HyperText Transfer Protocol) is a standard that describes how to request information from a website.

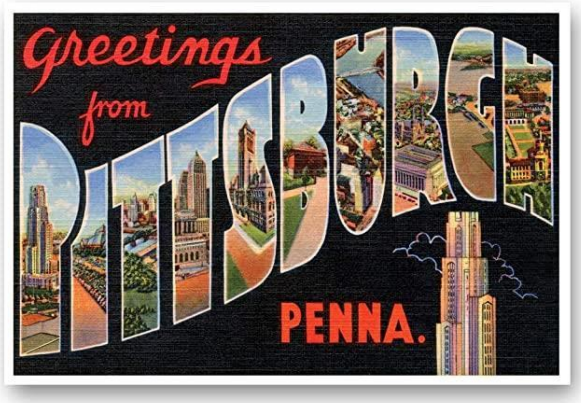
This request is sent using something called a packet.

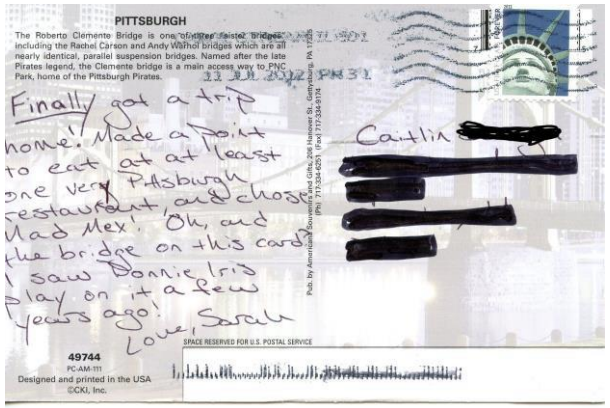


Packets Store Data

A packet is a small message that is sent to a particular IP Address.

It's similar to a postcard – it has a message (the data), a destination address (IP address), and a return/sender address (IP address).





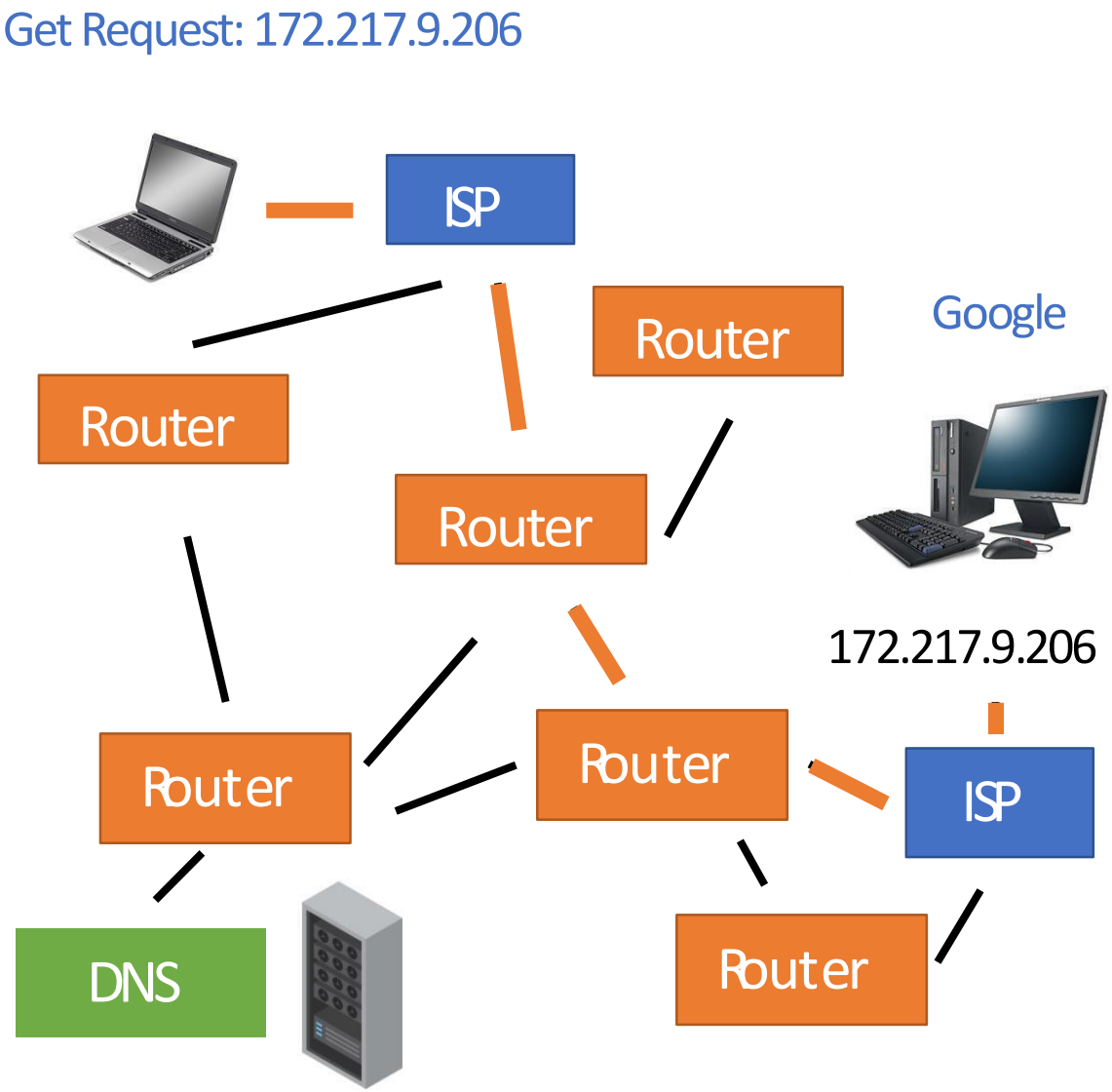
Because a packet is small, it can be sent along a wire very quickly.

A Packet Can Take Many Paths

Sending a packet across the internet is like sending a postcard through the mail.

You don't tell the post office which roads to take; you just tell it the destination, and the post finds a route to get it there.

Similarly, you don't tell the internet which routers to visit; the internet figures it out.

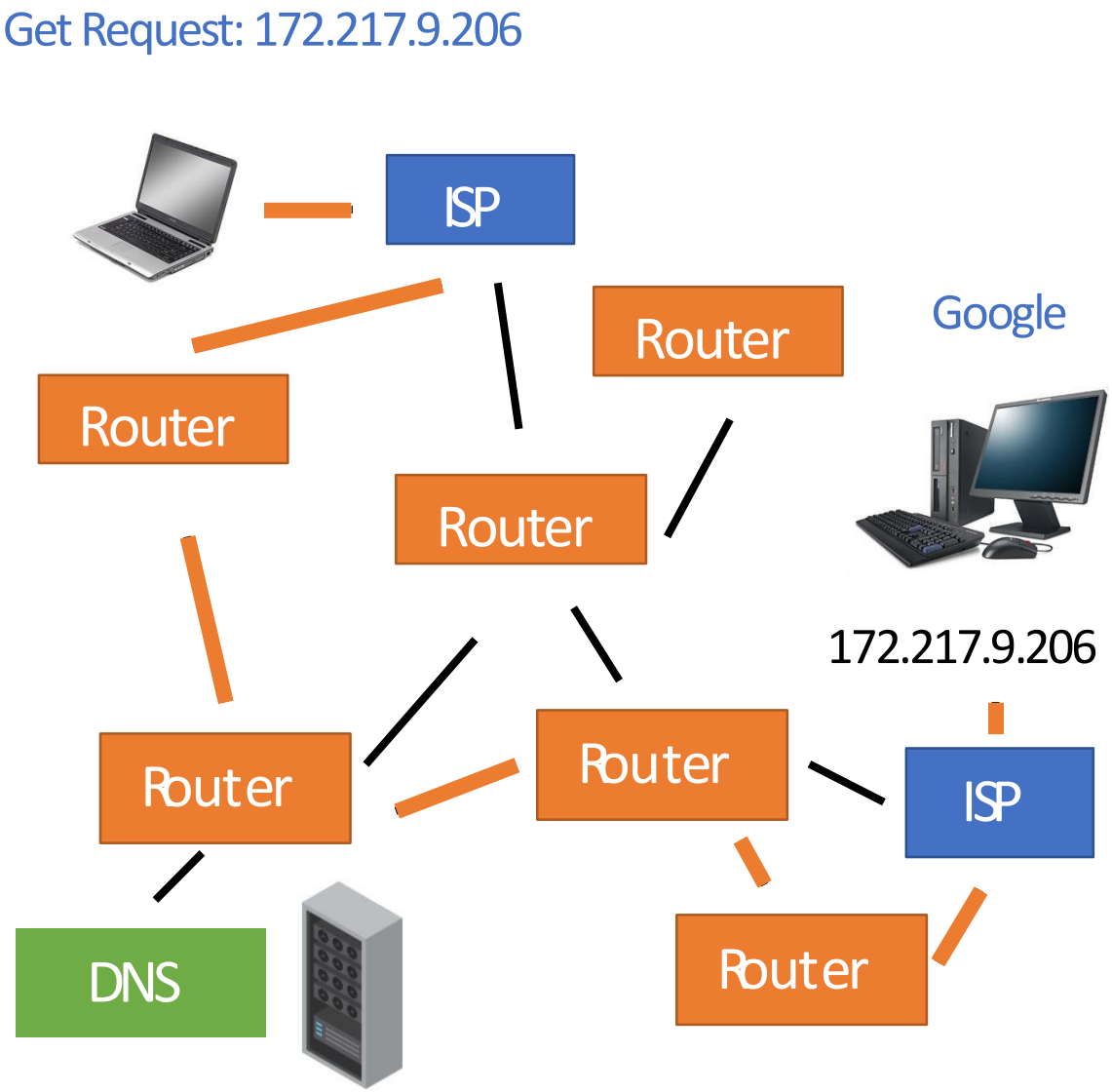


A Packet Can Take Many Paths

Sending a packet across the internet is like sending a postcard through the mail.

You don't tell the post office which roads to take; you just tell it the destination, and the post finds a route to get it there.

Similarly, you don't tell the internet which routers to visit; the internet figures it out.

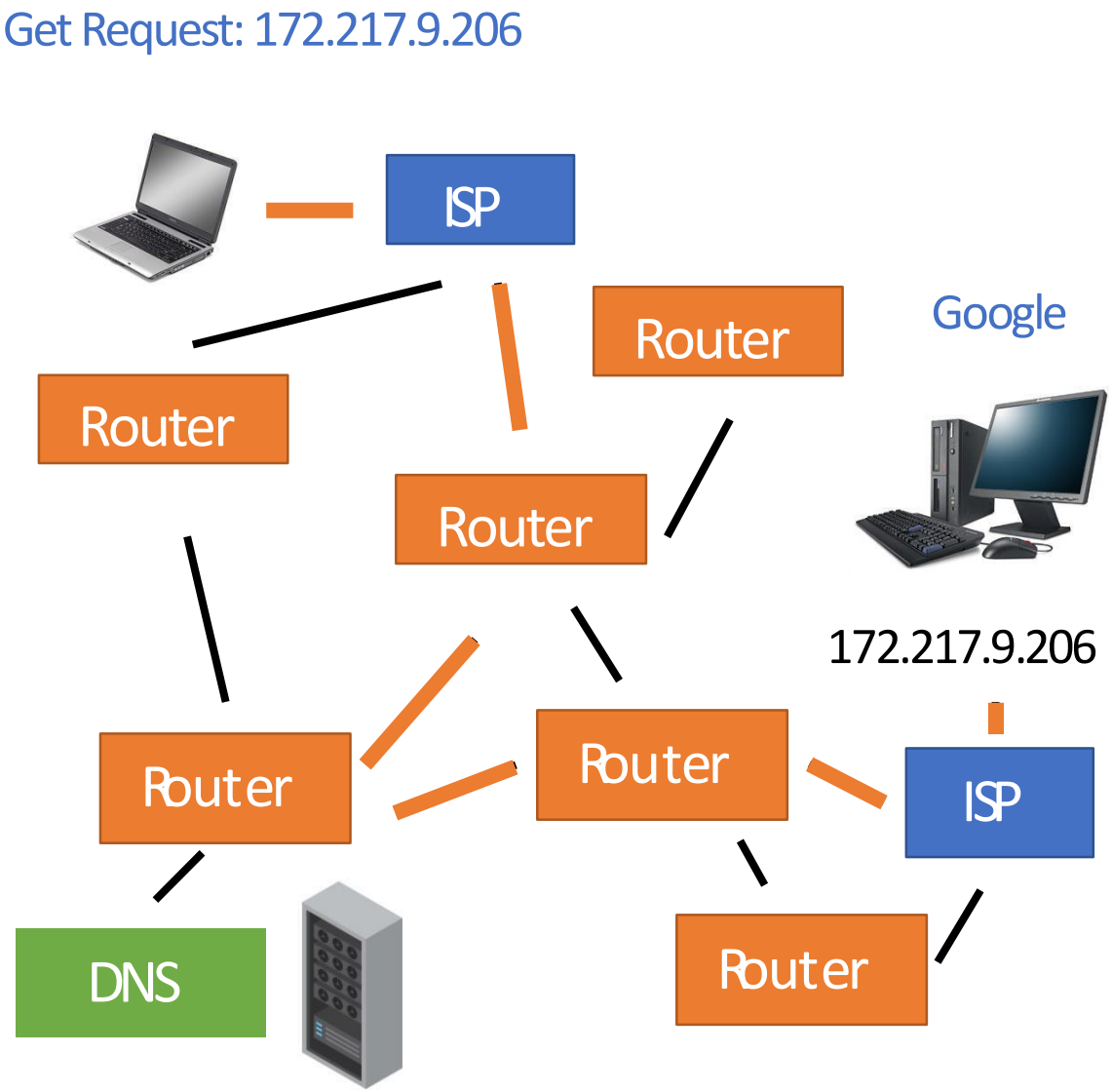


A Packet Can Take Many Paths

Sending a packet across the internet is like sending a postcard through the mail.

You don't tell the post office which roads to take; you just tell it the destination, and the post finds a route to get it there.

Similarly, you don't tell the internet which routers to visit; the internet figures it out.

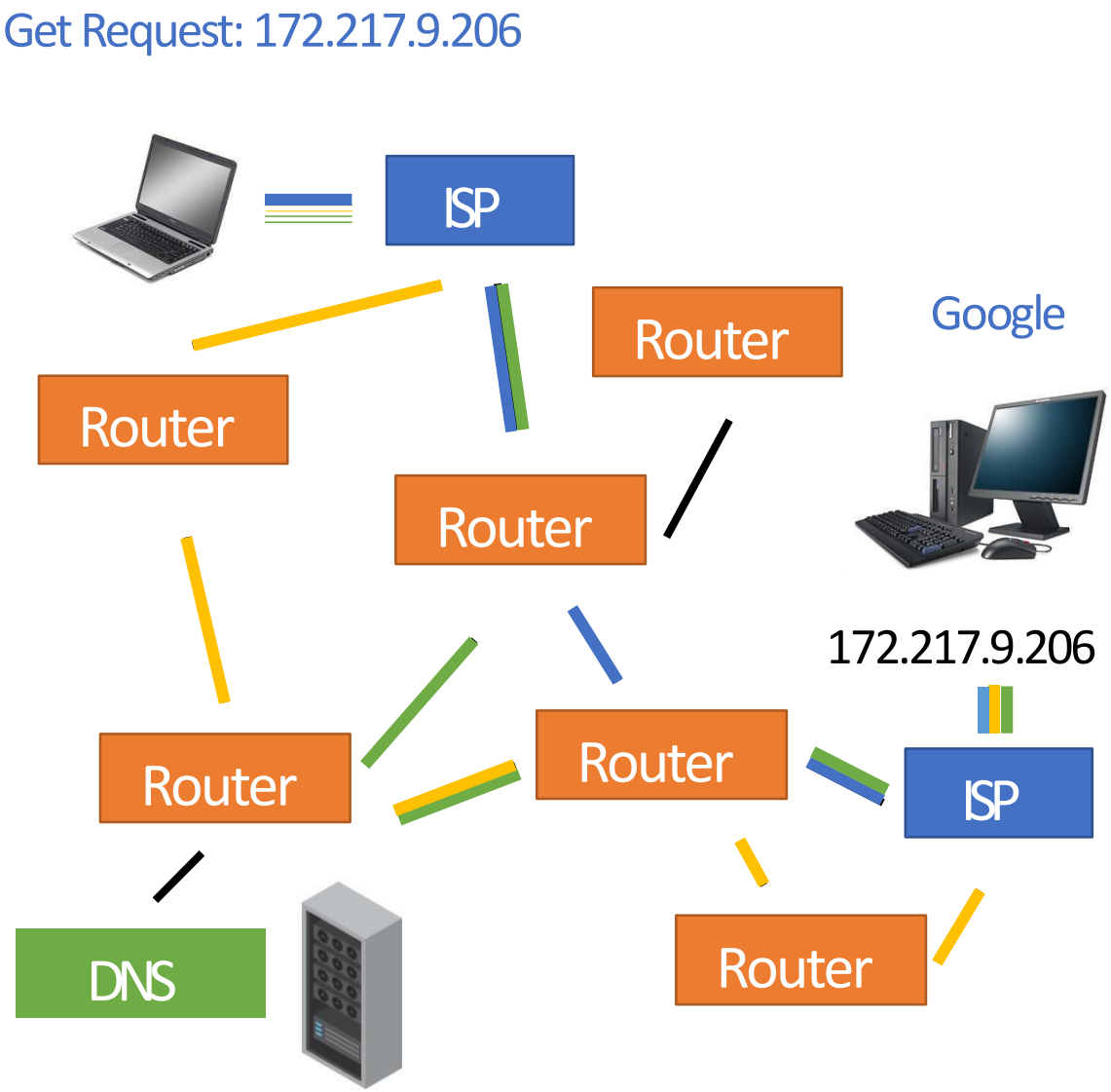


Webpages are also Packets

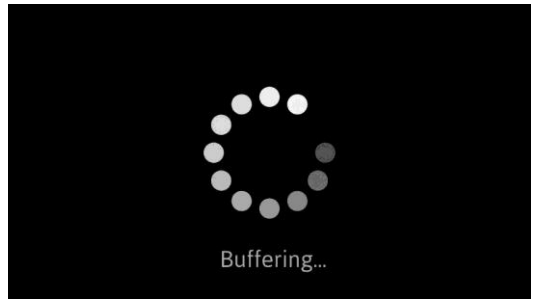
When the website gets your request, it might need to send back a response (like a webpage).

Since webpages are generally large, the page is split into multiple packets and the packets are sent back through the routers to your computer.

When all the packets get to your computer, the browser assembles them to produce the HTML of a website. It's like putting together a jigsaw puzzle.



Buffering



Some webpages need a lot of packets. For example, a video takes a lot of data to render. Packets may take a long time to reach the browser, which can cause lag.

Your browser uses buffering to show you part of a website while the rest of it loads. Buffering occurs when the browser receives enough of the early packets to pre-load the initial content onto your computer. While you read or watch the content, the browser silently loads the rest of the content as it arrives.

If a buffer pauses for a long time, your browser is probably waiting for a few packets that are still missing.

Packets Are Not Reliable

Can we rely on packets to show up properly? Not really...

- There are no guarantees that all packets will use the same route
- There are no guarantees that computer will receive the packets in the intended order
- There are no guarantees that all the packets will arrive at your computer
- There are no guarantees that the packets will not be corrupted

So how does the internet function? It's designed to be incredibly fault tolerant.

Fault Tolerance

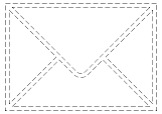
The Internet is Fault Tolerant

The internet is designed so that when things go wrong (which they do), there are always plenty of backups and checks in place to make it right.

This is true both in how packets are delivered, and in how computers are networked to each other.

Packet Fault Tolerance

Q: What happens if a packet goes missing?



A: Your computer knows how to put packets back together based on the data they carry. Most protocols can tell if a packet is missing. If it is, the browser simply sends another request for a new set of packets.

Q: What happens if a packet is corrupted?

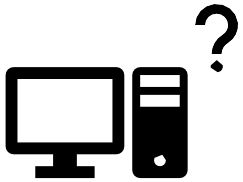
A: Every packet contains a checksum that the computer can check to make sure it's not corrupted. If it is corrupted, the computer just sends a request for a new set of packets.

Network Fault Tolerance

Q: What happens if your computer goes down?

A: This isn't great for you, but it's fixable and happens all the time! When your ISP sees that your computer has gone offline, it holds any data you've received until you come back.

Q: What happens if a company's website (a server) goes down?



A: Most companies have many servers that can all handle traffic to the same website, so traffic to the server that is down gets re-routed.

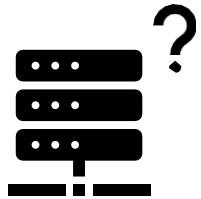
If all of a company's servers go down, then the website goes down too.

Network Fault Tolerance

Q: What happens if a router goes down?

A: This is fine – traffic will just be sent to other routers instead. The core of the internet is heavily connected and decentralized, so this will not disturb traffic.

Q: What happens if a DNS Server goes down?



A: There are lots of DNS Servers spread across the world. If one goes down, your request gets sent to a different one.

Network Fault Tolerance

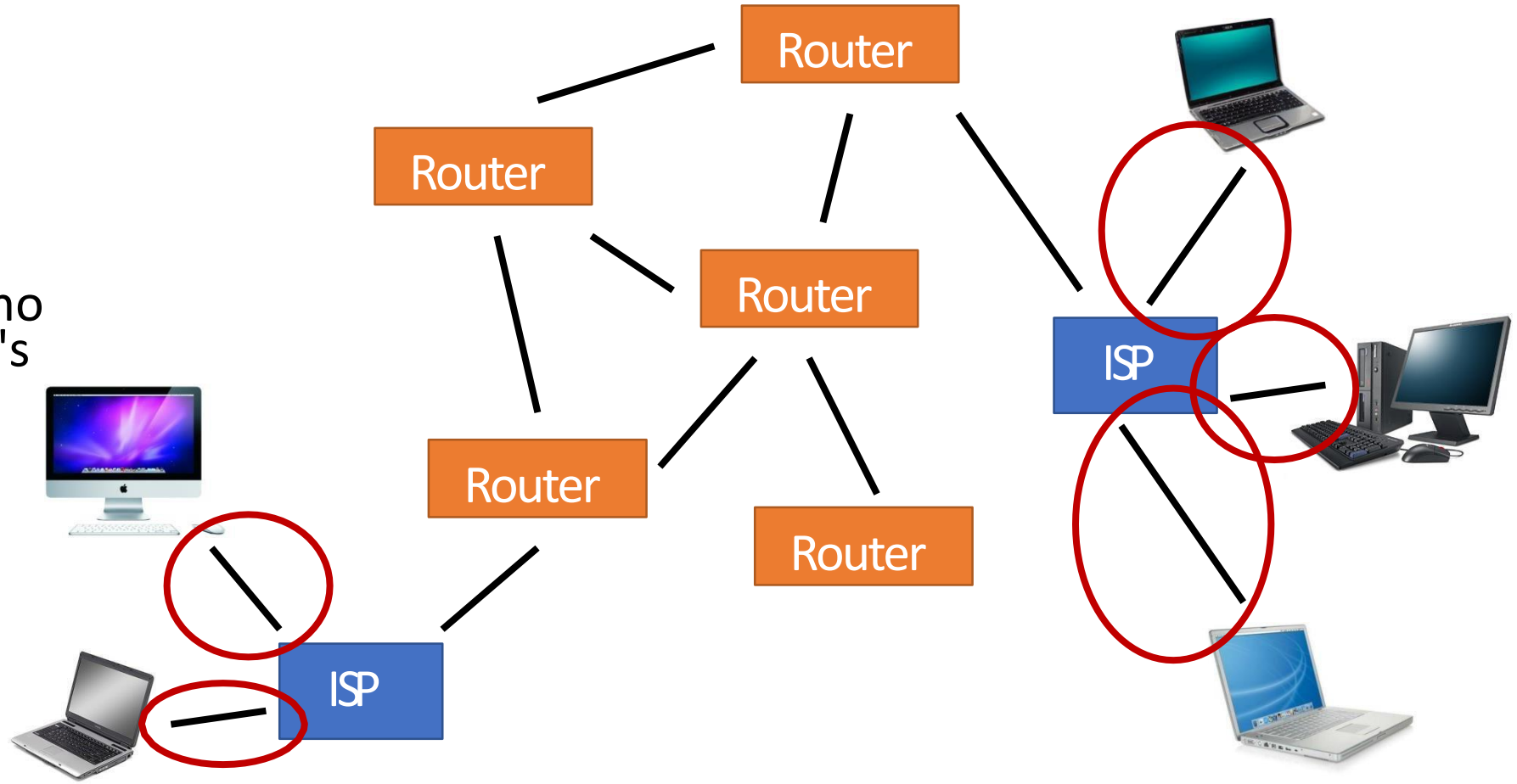
Q: What happens if your ISP goes down?

A: This is finally the place where you get into trouble. If your ISP goes down, you lose your connection to the entire internet because the ISP is the only place you can connect to.

The Internet is Hard To Control

It's hard for one organization to control the whole internet, because the core of the internet contains no bottlenecks. There's no one point of control.

That's not true for your local connection – your ISP is a bottleneck to the rest of the internet.



How Do Governments Turn Off Internet?

When a government shuts off the internet for an entire country, it's generally possible because they control the ISPs.

If all the ISPs shut down traffic, local computers have no way to access the broader internet. It's still there – it's just not connected.

If there is only one main connection between the core of the internet and a country (like a single router that serves as the general entry point), the government can also shut down the internet if they shut down that router.

Miscellaneous Internet

Other Internet Buzzwords

Finally, let's go over a few internet buzzwords you've probably heard before.

We'll talk about three big ideas: net neutrality, the Cloud, and IoT.

Net Neutrality

You may have heard the term net neutrality used in various political debates.

Net Neutrality is a principle which states that ISPs must treat all internet traffic equally. Packets should not be prioritized or de- prioritized based on who sent them, who is receiving them, or what is in them.

In terms of policy, Net Neutrality states that internet access should be considered a utility, like phoneline connections.

Net Neutrality Effects

Without Net Neutrality, an ISP could ask a website that sends a lot of packets (like YouTube) to pay them for the extra work. If the company refused, the ISP could de-prioritize that website's traffic to make it appear slower to the user. This is called throttling. On the other hand, if the company pays, maybe they get prioritized instead.

The ISP could also offer deals to their customers based on the websites they visit. For example, Verizon might make your monthly bill cheaper if you only visit websites on a Verizon-approved list. In an extreme example, an ISP could entirely block your access to a website it doesn't approve of.

Net Neutrality is currently not a law in the United States. It is law in some other countries, like India.

The Cloud is Other People's Computers

When a company says that they store things in "the cloud", they're referring to other computers that are connected to the internet. This is just distributed computing!

Companies use the cloud because it makes storage cheap, can scale at need, and is available on demand. It's generally easier to access computers that are provided by another company than to maintain a set of servers yourself.

You probably use the cloud too. If you store data online (like in Google Drive, or Instagram), you're storing data in the cloud.

IoT is Objects with Computers

If you have a 'smart' device of some kind in your home (like a smart thermostat, or watch, or a device like Alexa), you already own an IoT device. IoT stands for Internet of Things.

A smart device is an everyday object (like a thermostat) that has a computer inside of it. That computer connects to your home's Wi-Fi network. It can then send data to another machine in your home network, or back to its company's server.

Smart Devices Use Sensors and Actuators

Smart devices are different from passive devices because they tend to have sensors that collect data about the real world, like the temperature. They can use that sensed data to make decisions.

They also have actuators that let them do real-world actions, like adjusting the AC in your house. They use a combination of sensed data and requests received over the internet to decide which actions to take.

By combining sensors, actuators, the internet, and programming, we can make devices that make their own decisions without us telling them what to do.

IoT devices are really interesting, but they can also have a large number of security flaws.