# The Internet

## Introduction to Networking

# The Internet

The Internet is many computers. Some are small, like your cell phone. Some are even smaller, like a router or a Roomba floor vacuum robot or an Amazon Echo.

All of these computers work together to form the Internet.

The Internet is many computers communicating with one another. They might communicate over radio waves (like WiFi or Bluetooth) or over physical cables (like Ethernet or Cable TV).

One of the remarkable features of the Internet is that the computers connected are not really aware how they're connected. They are able to communicate over many different kinds of connections.

# Protocols

Once they’re connected, how do they communicate? Through network protocols.

Communication between connections on the Internet are governed by rules known as **protocols,** or formal standards for transmitting information.

Protocols on the Internet work the same way. In order for every system to communicate with each other, every computer on the Internet has agreed to communicate using the exact same protocols.

These protocols and standards are what constitute the “Network of Autonomous Systems” that is the Internet. The word *autonomous* here is used to mean that the Internet works without any human intervention.

# LAN

**local area network**(LAN): a group of interconnected computers that share resources in a small geographic area.

- Smallest example of a LAN is two computers connected by an CAT6(ethernet cable).

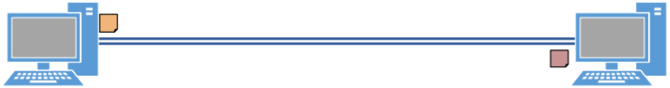


Image:Michael Lamont; www.slideshare.net/MichaelLamont

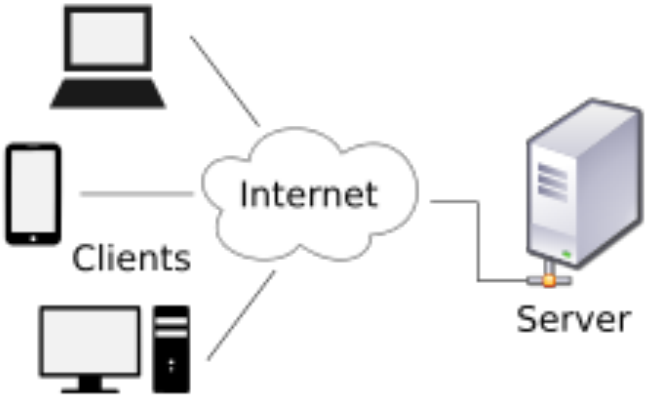
- A **host** is a computer on a network.

# Server-Client

A **client** is a computer that is requesting a resource or service.

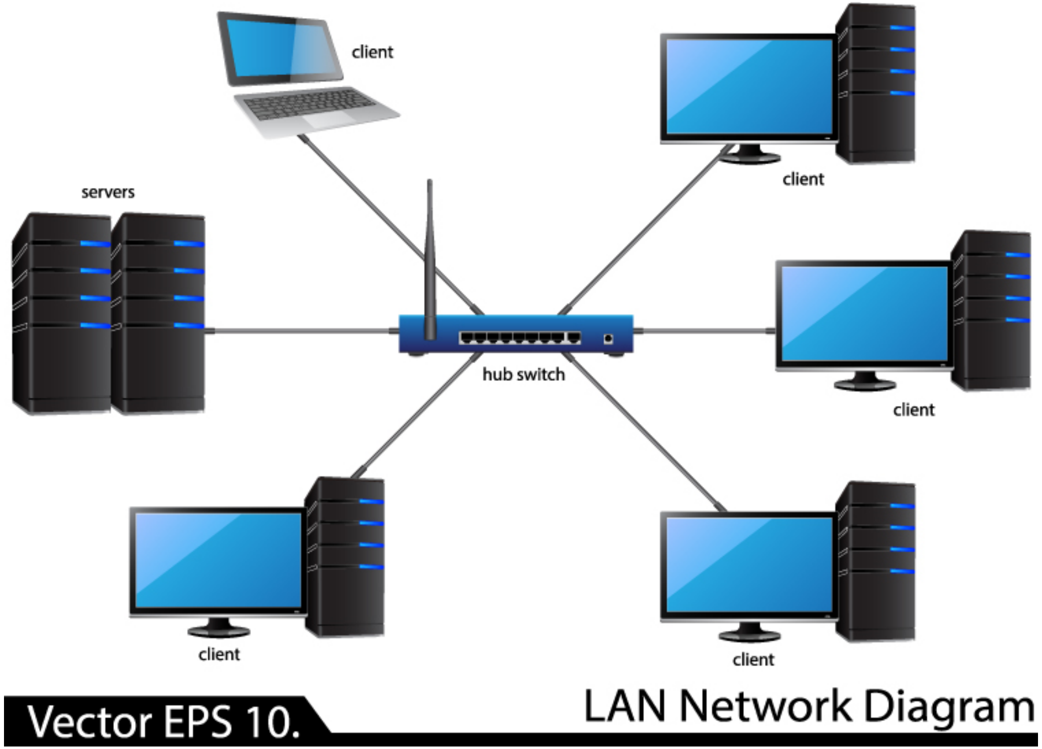
A **server** is a computer that provides a resource or service.

- Print servers provides printing service, web servers serves webpages, email servers, file servers, etc…



# Hubs and Switches

**Switches** allow computers on a LAN to communicate with one another.



# Switches

A **switch** is a networking device can forward data directly to computers connected to its ports using the destination physical or **hardware address**(MAC address).

- A port is a physical input/output interface to a networking device.
- E.g. a computer sent a request to a printer via a switch; only the printer sees the request.

# MAC Address

A **network interface card**(NIC) on a computer contains a unique hardware address or physical address called the **MAC(media access control) address**.

- Every device that can make Wifi, ethernet or bluetooth connections has a MAC address, e.g., laptops, phones, bluetooth speakers, smart fridge, tablets, etc…

A MAC address is 6 bytes or 48 bits in length. It is displayed in 12 hexadecimal digits.

- Example of a MAC address: 00-AA-00-B6-7A-50. The first six digits identify the vendor; in this case 00-AA-00 belongs to Intel.

# IP Address

The MAC address provides the physical address for the NIC but provides no information as to its network location, LAN or in which building, city or country the network resides.

The **Internet Protocol**(IP) address provide worldwide addressing that identifies the computer's local network.

- An IP address is a network address or logical address of a computer.
- Can be IPv4 or IPv6. Although IPv6 is becoming more popular, IPv4 is still the addressing technique of the internet.
- This is like a street address for your home or school, except, on the Internet, every computer has a unique IP address that only that computer can have.

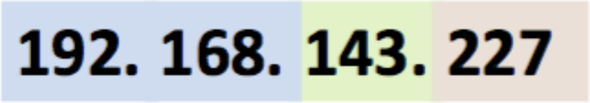
A google search like "what's my ip" will likely display the IP address of your computer.

# IP Address(IPv4)

The **IP address** is a unique 32-bit that identifies on which network the computer is located as well as differentiates the computer from all other devices on the same network.

The address is divided into four 8-bit parts. The format is A.B.C.D, where A,B,C,D are decimal equivalent of the 8-bit binary value. Each value is in the range 0-255.

IP address has three parts: network number(identifies the network), subnet and host ID number(identifies the particular computer or host). For example:



192. 168. 143. 227

# IP Address(IPv4)

For IPv4, with 32-bits, there are a total of 4,294,967,296 addresses.

- Not a lot of addresses considering the number of tablets, smartphones, smart TVs, Rokus, laptops in each home!

To conserve the public IP address space, **private IP addresses** are used within private networks. These private addresses are not valid addresses for Internet use.

The three address blocks for private IP addresses are:

10.0.0.0 – 10.255.255.255

172.16.0.0 – 172.31.255.255

192.168.0.0 – 192.168.255.255

# Private IP Addresses

In a small LAN, computers/hosts are assigned private IP addresses that can identified them locally.

- These private IP addresses are not unique; two hosts from two different LAN can have the same private IP address.

When a computer make a request for a resource outside of its LAN, e.g. CNN’s webpage, the request passes through the router that is responsible for data packets leaving/entering the network.

- This router is known as the **default gateway.**
- The **Internet Service Provider**(ISP) assigns one public IP address to the entire network. Any outside traffic to and from computers inside the network is routed to this public IP address.

# Private IP Addresses

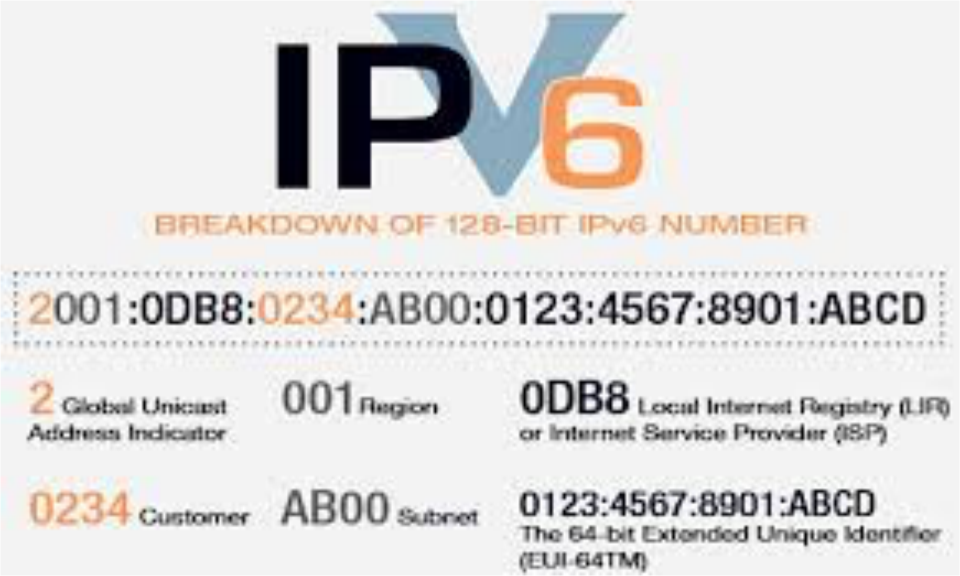
When you try opening a website from your computer, the request is sent from your computer with a private IP address to your router, after which your router requests the website from your ISP using the public IP address assigned to your network.

Once the request has been made, the operations are reversed - the ISP sends the contents of the website to the public IP address of your router, which forwards the address to the computer that asked for it.

# IP Address(IPv6)

For **IPv6** uses 128-bits for a total of approximately 3.4x10^{38} addresses.

The address is divided into eight 16-bit parts. The format is 8 groups of four hexadecimal digits.


IPV6
BREAKDOWN OF 128-BIT IPv6 NUMBER
2001:0DB8:0234:AB00:0123:4567:8901:ABCD
2 Global Unicast Address Indicator
001 Region
0DB8 Local Internet Registry (LIR) or Internet Service Provider (ISP)
0234 Customer
AB00 Subnet
0123:4567:8901:ABCD The 64-bit Extended Unique Identifier (EUI-64TM)

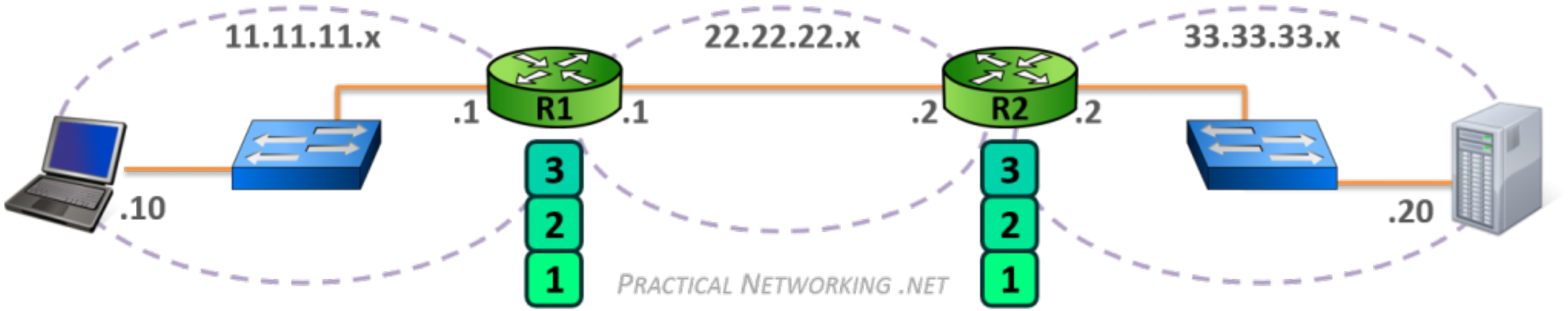
# Switch vs. Router

A **switch** is a networking device that facilitates communication within a local area network using source and destination MAC addresses.

- A switch is responsible for hop-to-hop delivery
- A data unit sent by a switch is a frame.

A **router** is a networking device that facilitates communication between networks using source and destination IP addresses.

- A router is responsible for end-to-end delivery
- A data unit sent by a router is a **packet**.



# Network Applications

A **network application** lets a computer interact with other computers by performing a specific set of tasks.

- Browsers, email clients, etc…

The application is responsible for managing the transmitting and receiving of data required to perform its tasks .

The application has to be able to communicate with applications on other networked computers for it to be useful.

# Network Applications

A **network protocol** is a set of rules for how applications intercommunicate.

Common protocols/applications include:

- **Simple Mail Transfer Protocol**(SMTP)
- **Hypertext Transfer Protocol**(HTTP) for sending/receiving webpages
- **HTTPS(HTTP Secure)**
  - Encrypted with SSL(Secure Socket Layer) or the new TLS(Transport Layer Security)
- Secure Shell Access(SSH)
- File Transfer Protocol(FTP)

# TCP/IP

**Transmission Control Protocol**(TCP) is the protocol that is responsible for sending data packets across the internet.

TCP is a host-to-host protocol that provides **reliable**, **connection-oriented** communication over IP networks between two endpoints.

- A TCP packet is sent using IP addresses, hence the term TCP/IP.
- **reliable**: data received is intact, error-free and assembled in the correct order. Any lost data is retransmitted.
- **Connection-oriented**: Using a system of acknowledgements called the Three-Way-Handshake, two computers confirm the connection before data transmission.

.

# TCP/IP

The suite of protocols that make up **TCP/IP(Transmission Control Protocol/Internet Protocol)** define:

- How data is transmitted across a network
- How data should be formatted so other networked systems can understand it

TCP/IP is the standard for modern data communications across all networks.

Another protocol similar to TCP is **User Datagram Protocol(UDP)**, which is a connectionless protocol.

- Unlike TCP, UDP has no guarantee of delivery or correct assembly. It requires less overhead and is faster. (Skype uses UDP)

.

# TCP/IP

TCP/IP came out of ARPAnet the Department Of Defense(DoD)'s attempt to create a decentralized, no single point of failure network.

- There were many network protocols in the 1990's. The "Protocol Wars" ensued and TCP/IP won the war.

Two key features TCP/IP features that support decentralization:

- **End node verification**: the two endpoints of any data transfer are responsible for making sure it was successful – no centralized control scheme
- **Dynamic routing**: End nodes can transfer data over multiple paths, and the network chooses the best (fastest, most reliable) path for each individual data transfer

# HTTP

In addition to lower-level protocols such as **TCP/IP**, there are higher-level protocols (protocols that operate on top of TCP/IP) such as **HTTP** and **SMTP**.

Computers on the Internet need to know not only how to send data to each other but also need to know what to do with the data. For example, when viewing websites, your computer is communicating using **HTTP (Hypertext Transfer Protocol).**

This is also the reason that you sometimes see web addresses begin with an http:// or an https://.

This indicates to the browser to use the HTTP protocol to communicate.

# SMTP

This provides the computer with enough context to be able to handle the data and communicate in a way that is best for websites.

Similarly, there is also a protocol for sending email known as **SMTP (Simple Mail Transfer Protocol).**

HTTP and SMTP were designed by the **Internet Engineering Task Force(IETF)**. This is an organization that actively promotes the use of standardized protocols on the Internet.

They are also involved in developing and creating resources (such as standards and documentation) to make the adoption of these standards easier.

# Name Resolution

Logical IP addresses are “friendlier” than physical MAC addresses, but still aren’t really human readable.

**Domain Names**: structured, user friendly system names provided by TCP/IP. Examples of **top-level domain names**:

- .com, .org, .edu, .gov, etc…

**Name Resolution**: the process of mapping logical addresses back and forth into domain names

- 172.217.10.46 is mapped to www.google.com
- Special name servers store the mapping information in databases
- TCP/IP’s **Domain Name Service** (DNS) provides a **hierarchy** of name servers that handle name resolution for the Internet.

# Domain Names

When you visit a particular website you are most likely accessing it by typing something into your browser such as:

google.com

This **domain name** can be broken down into two parts. The TLD, or **top-level domain**, (in this case "com") is the first indicator of where to go to find the website.

The "google" part of the domain name actually specifies which computer to point to.

# Domain Names

Domain names are organized **hierarchically** from right to left, where the right is more general and left is more specific.

For example, if you wanted to visit the English Wikipedia page, you would have to type in “en.wikipedia.org”.

This indicates that you would like to visit a website within the top-level domain “org”, a website specifically called “wikipedia”, and the “en” sub-domain of Wikipedia that will give you the English-language version of the website.

# Port Numbers

An IP address identifies a computer on the internet; a **port number** is a 16-bit number identifies the application or service on a computer/server.

A **port** is a logical connection used by programs to exchange information. There are 65536 TCP ports on a computer; each with a 16-bit port number from 0 to 65535. (There are also 65536 UDP ports)

- This allows a computer to use different network applications simultaneously. Data packets from different applications use different ports.
- E.g. if I have a browser and an email client such as Apple Mail or Microsoft Outlook, then at least two ports are needed for data transfer.
- Even within the same webpage request, multiple ports can be used to download the html, images, javascript files, etc… simultaneously.

# Port Numbers

The port numbers in the range from 0 to 1023 are the well-known ports. They are used to identify widely used types of network services.

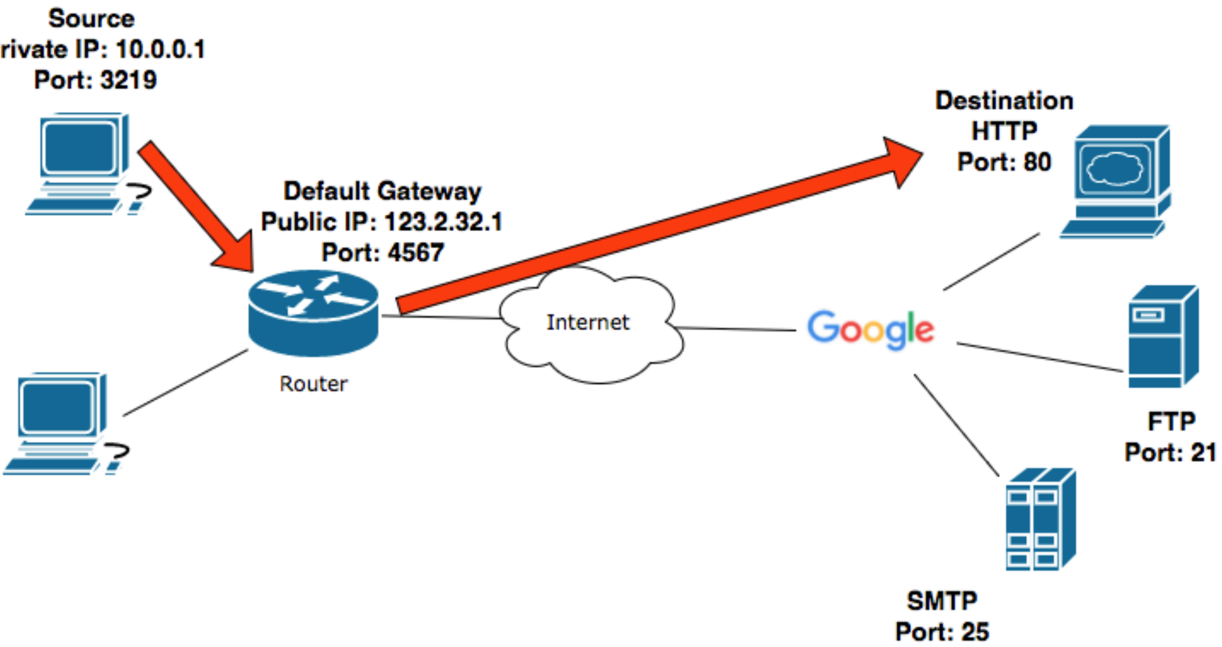
- Port 80 for HTTP, Port 21 for FTP, port 25 SMTP, port 23 for Telnet.

To make a HTTP request for a website, a TCP connection is established.

- This connection includes among other things the IP address and port number of the client and the IP address and the port number of the server.
- The port number of the client is dynamically assigned to a random port above 1023(ephemeral ports). The port number of the server is one of the well-known port.

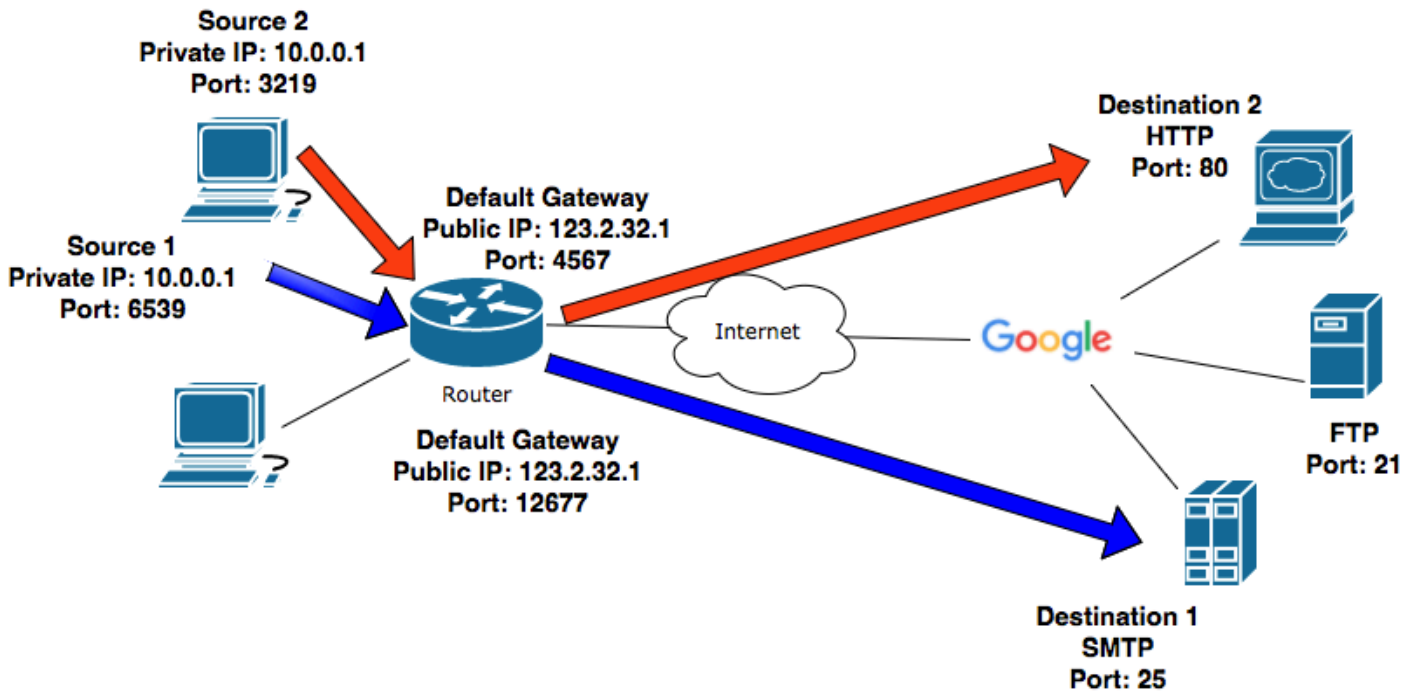
# An HTTP Request

- A webpage(HTTP) request from a computer.



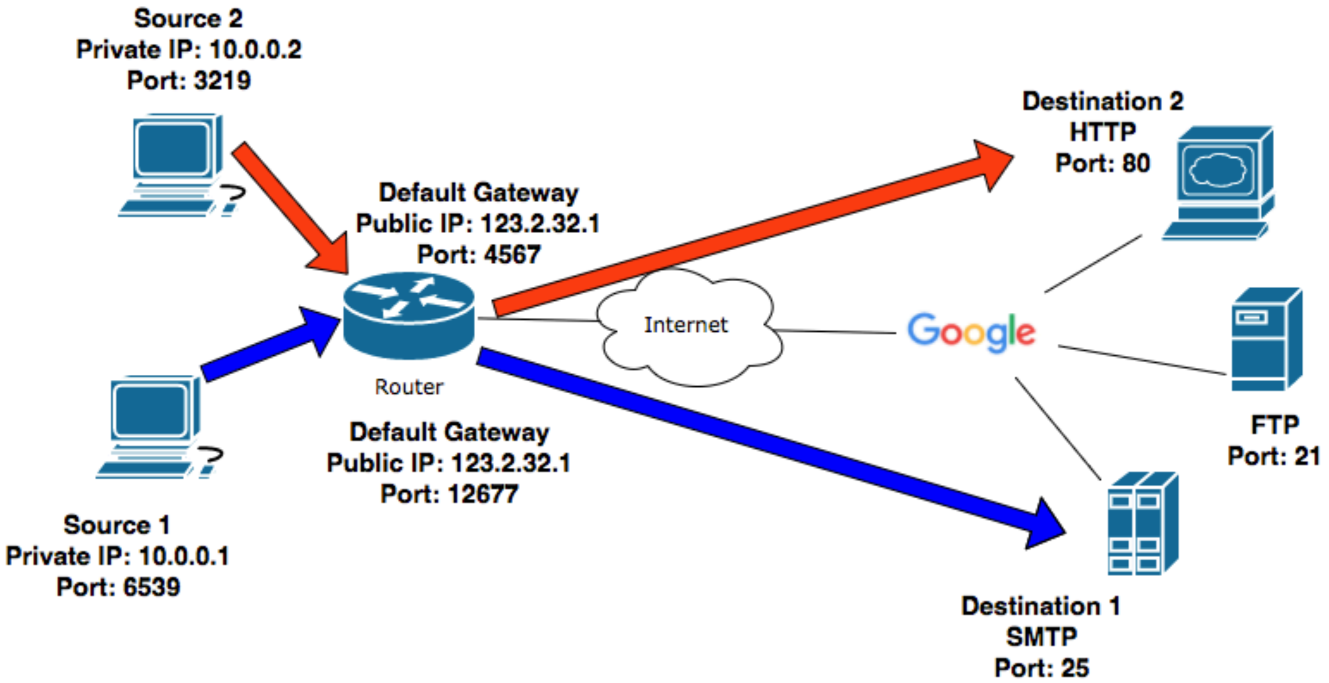
# An HTTP Request

- A webpage request and sending an email from the same computer.



# Port Numbers

- Computer 1 sending an email. Computer 2 requesting a webpage. Both computers are in the same local area network.



# Packets

The Internet transmits data from one computer to another by breaking the data down into small manageable units that are known as **packets**. This system is called a **packet-switching system**.

A packet is defined as a small chunk of digital data (like an email or a picture) that contains both a sender address and an address to where the packet is supposed to go.

A single image could contain anywhere from just a handful of packets to upwards of 10,000 depending on a variety of factors.

The process of determining where packets go is called **routing**, or selecting a path for traffic in a network.

# Packets

Packet-switching system is similar to traffic routing.

If a road is becoming too crowded, traffic lights that allow cars to go onto that road might turn red in order to prevent any more cars from coming on.

A road might even be under construction, so no cars are able to enter that road.

Detour signs might route the cars to other roads that are more accessible in order for the cars to get to their destination.

# Packets

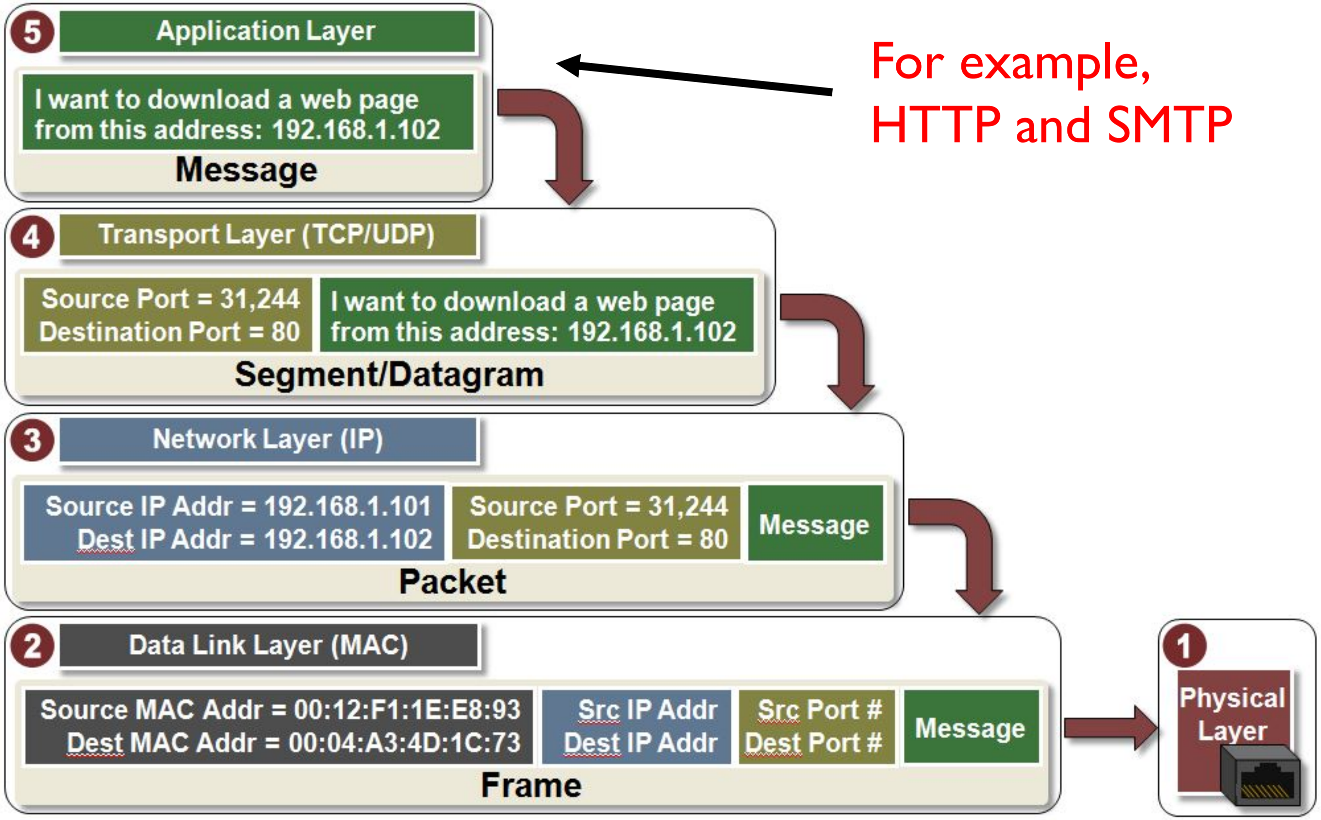
Similarly, in packet routing, a certain connection might only be able to transfer a certain number of packets per minute, so if the connection isn't able to transfer them quickly enough, a new path for the packets to go will need to be found by a router. Thus, routing is **redundant**(more than one way to route data).

Redundancy increases the reliability of the internet and helps it scale to more devices and people.

If some connection is broken or doesn't function, the router needs to find a new path for the data to get to its destination, otherwise the packets will never be delivered. Thus, routing is dynamic and **fault tolerant.**

# Packets

A message that needs to be sent across the internet receives headers from each layer as it travels down the protocol stack. The headers provide routing/addressing information, e.g., port numbers, IP addresses, MAC addresses, etc…



5 Application Layer

I want to download a web page from this address: 192.168.1.102

Message

4 Transport Layer (TCP/UDP)

| | |
|---|---|
| Source Port = 31,244<br>Destination Port = 80 | I want to download a web page from this address: 192.168.1.102 |

Segment/Datagram

3 Network Layer (IP)

| | | |
|---|---|---|
| Source IP Addr = 192.168.1.101<br>Dest IP Addr = 192.168.1.102 | Source Port = 31,244<br>Destination Port = 80 | Message |

Packet

2 Data Link Layer (MAC)

| | | | |
|---|---|---|---|
| Source MAC Addr = 00:12:F1:1E:E8:93<br>Dest MAC Addr = 00:04:A3:4D:1C:73 | Src IP Addr<br>Dest IP Addr | Src Port #<br>Dest Port # | Message |

Frame

1 Physical Layer

For example,
HTTP and SMTP

# Hierarchy

The Internet is built hierarchically.

Domain names are broken into parts that provide general to specific information about the website you are trying to access.

IP addresses are broken up into several numbers that delineate the different sections of the Internet from each other.

Routing, especially, which uses the DNS to jump around from network to network to find its way to the destination, is built on the hierarchical system of networks that are all interconnected.

This hierarchy and connectedness allowed the Internet to grow to the global scale that it is today.

# Bandwidth and Latency

When you have a large system there is going to be **latency**, however. The latency of a system is the measure of the time it takes for a message to get to its intended recipient from the time it was sent.

Another way to measure a systems effectiveness is with **bandwidth**. Bandwidth is the amount of data that can be transferred on a system in a certain amount of time.

The amount of data is measured in bits. These days, an average Internet connection speed might be anywhere from a megabit (about a million bits) per second to a gigabit (about a billion bits) per second.

Downloading an email requires less bandwidth than watching a movie on Netflix.

# References

1) Runestone Academy. CS Principles.
2) InetDaemon. (2014, September). Networking Tutorials. Retrieved from http://www.inetdaemon.com/tutorials/networking/
3) Allen, Scott. (2012, January). A Software Developer's Guide to HTTP. Retrieved from

   https://odetocode.com/Articles/741.aspx
3) Microchip Technology.(2018). Introduction to TCP/IP. Retrieved from

   http://microchipdeveloper.com/tcpip:tcpip-intro-video