**Domain Names and ICANN**

The internet as we know it today, and oftentimes referred to as the world wide web, was invented by Tim Berners-Lee. While inter-computer and inter-network communication already existed thanks to the ARPANET, Berners-Lee dreamt up HTTP (Hypertext Transfer Protocol) communication across the internet. Berners-Lee also founded and is still the current director of the W3C (World Wide Web Consortium). Both of which will be discussed in greater detail in later sections.

The innovation that Berners-Lee brought to the forefront, was simplicity and ease of access. As we will go into detail later, HTTP is the language that is used for communication between web browsers and DNS (domain name servers). These domain name servers will also be described in greater detail later, but they are extremely important to making the internet as we know it accessible. They are so important that there is a regulatory group specifically in charge of managing domain namespaces as well as their databases. This regulatory group is known as ICANN (Internet Corporation for Assigned Names and Numbers)

ICANN “coordinates the allocation and assignment of names in the root zone of the Domain Name System (**DNS**) and coordinates the development and implementation of policies concerning the registration of second-level domain names” (ICANN.org). Put in similar terms, the ICANN is responsible for allocating and maintaining the name spaces for the internet. This is both the “website name” (i.e. Facebook, Amazon, Google, etc.) as it is referred to by many people as well as the “dot com” segment.

The “dot com” segments can more accurately be called top-level domains. The most generic of these is “com”, but several others exist. “Com” was initially used in order to be the commercial top-level domain, but it has since become the generic one to use. There is also .gov for government and it’s agencies, .mil for US military use, .edu for educational use (mainly colleges and universities), and then .org and .net for additional “generic” top level domains.

The domain names listed above are by no means exhaustive and several other have been added and maintained by ICANN. This can be international and country specific codes like .uk, .fr, .de, etc. as well as even some city specific domains like .tokyo, .london, .nyc, etc. ICANN has also expanded and allowed for the use of several other generic top level domains in the form of domains like .xyz, .online, and several others that can be seen below.





**IP Address, Packets, and Routing**

The internet has always been a network of networks (some smaller than others). Local networks (i.e. devices like laptops and cellphones connected to a home Wi-Fi enabled or wired router) are then connected to a greater network through an ISP (internet service provider) like Comcast or Verizon. The ISP then connects you to other ISPs and other networks that are all interconnected. While this may all seem simple and intuitive now, it was not always the case.

Instead of letters, the messages being sent are comprised of 0's and 1's called bits. Bits are assembled into groups of 8 called bytes. Bits can be thought of atoms that make up data. Today we send bits through 3 main mediums (electricity, light, radio). The transmission capacity of these three mediums is referred to as bandwidth. This can be thought of how many “letters” (bits) can fit into a cable or wave at a time. How fast these “letters” (bits) travel is known as the bitrate and the time it takes them to get from origin to destination is known as latency.

As mentioned above, we have three main transmission mediums, electricity, light, and radio. Each of these transmission mediums has a specific material, speed, and set of pros and cons. The first medium we look at is ethernet. Ethernet cables are made up of copper wires that allow for cheap data transfer of electrical signals. They suffer data loss and interference very quickly over long distances.

The next medium is fiber optic cables. Fiber optical cables allow for fast data transfer of light signals. They are also extremely efficient and suffer little too no loss even at large distances. Expensive cables and very hard to work with.

Lastly, we have radio signals. Radio signals (wireless) allow for transfer of data without physical connections. Bits are translated into different wave lengths based on different frequencies and then translated back into data at the receiving end. Signal loss is great at distances (distance varies depending on frequency i.e. AM vs FM). Wi-Fi takes local radio waves and converts them into copper or fiber messages to be sent over the physical network.

During the initial days of the internet, there was no standard in packet sending or receiving. Internetworking Protocol (IP) was developed by Vint Cerf and Bob Kahn. They thought that the internet would be better served with a design philosophy and architecture expressed in a set of uniform protocols. A protocol is well known set of rules and standards that all parties agree to use to communicate between machines. This has allowed the expansion of the internet and internet capable devices due to uniform standards (i.e. which protocols to use and work with).

All devices on the internet have a unique address (just a number) within a given network. These unique IP addresses work similar to, and can even be thought of as, mailing street addresses of the physical world. Visiting a website can be broken down / thought of as your computer sending mail with the IP address of the website as the recipient address and a return address of your computer. This allows the website to return information to you in the same manner.

Traditional IP addresses are 32 bits long and can be broken down into four 8-bit segments. While this is no longer always the case, IP addresses used to be segmented into country/region/subnetwork/device. This previous standard of segmenting IP addresses was known as IPv4. It was originally designed in 1973 and widely adopted in the 1980s.

While the original IPv4 address allocation allowed up to 4 billion unique addresses (devices) to be connected to the internet, the world and its engineers realized that this would quickly no longer be enough. Due to this, the world is currently in transition to a new standard of IP address allocation called IPv6. This new IP addressing schema uses 128 bits per address (8 segments of 4 hexadecimal characters).

Even with the implementation of these new, more complex, IP addresses, browsing the internet will not change. This is due to the fact that people no longer need to rely on IP addresses to browse the internet like the early days. In today’s world, most people never know or see actual IP addresses of websites due to DNS.

Data transfers on the internet are not point to point but very indirect. Paths can change extremely often, even during a transmission. Packets are data traveling along the internet and work more like cars than trains. Packets are allowed to take any available road (whether it is the most efficient route or not).

Packets can contain almost any type of data and depending on size, messages may be broken up into several packets. Groups of packets can take diverging paths and arrive at the destination at different times and potentially out of order. Packets simply carry the data as well as headers containing to and from addresses and are moved through the networks via Routers.

Keeping with the analogy of IP addresses being similar to mailing street addresses, routers can be thought of as mail sorting machines located within a post office. Routers act like traffic managers to keep packets moving smoothly through the network. Routers attempt to push packets through the "cheapest" available path at any given moment. Having multiple routes allows the network to be fault tolerant (i.e. multiple roads can be closed or blocked but packages can still get to their end destination).

TCP (transmission control protocol) manages the sending and receiving of all your data packets (i.e. guaranteed mail service where TCP validates all packets have been received) so that when all packets are received, they can be reassembled. TCP and Routing were built with scalability in mind. The hope and design philosophy is more routers = more redundancy.

**Domain Name Server (DNS)**

Most people never know or see actual IP addresses due to DNS. DNS stands for domain name server. The purpose of these servers is to act like a lookup table or phonebook. First, a user from their web browser, will enter the name of a website they want to go to. For example, a user can type in www.google.com. The web browser then puts out a request to a DNS server that attempts to correlate that name to an actual address.

One example of a public DNS server is Google. If all else fails, users can attempt to configure their internet capable device to communicate first with Google at the address 8.8.8.8 which will then look up the appropriate forwarding address to send to the user’s device.

DNS servers are broken up in a manner that dedicates certain servers to handle certain website types. For example, some servers only handle .com while others handle .org domains. DNS was originally designed to be an open and public communication protocol. Due to its openness, however, it is susceptible to cyber-attacks like DNS spoofing. This allows malicious users to reroute unknowing users to fake or incorrect websites. Both DNS and IP addresses were designed with scalability in mind.

**HTTP and HTTPs**

To access the internet, you use a web browser to enter the URL (uniform resource locator) of a webpage. Your computer then reaches out to a DNS which sends HTTP responses back to your computer. HTTP stands for hypertext transfer protocol and is the language used to communicate between web browsers and servers. These are relatively simple messages mainly made up of “Get” requests. Get requests are simple lookup requests (i.e. get document-name or get /login).

Obviously, the more image, video, other HTTP requests, the slower a webpage will load. When sending information to a server, your computer sends “Post” requests to servers. One simple and common example is a post requests to websites with login information. Web browsers then take this back and forth of Gets and Ports and stores any relevant information in cookies. Cookies are temporary files that allow websites to save data and remember things like who you are.

HTTP communications are handled via plain text requests and are susceptible to hackers and other security vulnerabilities. This led to creation of SSL (secure socket layer) and TSL (transport layer security) incorporated into HTTPS. This is done via digital certificates that provided by verified websites. If the certificates being sent by websites have not been verified, you web browser will warn you.

Encryption (scrambling or changing a message to hide it) allows us to safely transfer private data over the open and unsecured internet. Decryption is the process of unscrambling am encrypted message to make it readable. Caesar's Cypher was one of the first encryption methods.

Based on Caesar's idea of "encrypting" messages so that intercepted messages were unreadable to his enemies, letters in the messages were shifted a certain number of letters up or down in the alphabet, for example, all letters were shifted up or down 3 letters. Keys are the number of letters that a sender and receiver have agreed upon so that the messages can be unlocked. HELLO with a key of 5 would be MJQQT.

This cypher is extremely simple and easy to crack. So much so, that using brute force would only take at most 26 different shift combinations. A 10-digit encryption key on the other hand would create 10 billion possible solutions. While this would take a human being several centuries to decrypt, it would still only take a computer seconds to do the same. Today's encryption keys are made up of 256 bits.

Symmetric encryption is when the same key is used to encrypt and decrypt messages. Asymmetric encryption is what is used by communication on the internet. There is a public key for encrypting and a private key for decrypting. The public key is shared so that anyone can encrypt messages (normally sent by websites to clients). This way messages can only be decrypted once received back to website servers via their private keys.

One real world analogy for this would be that public keys can open delivery slots on mailboxes, private keys can open the mailbox itself. This pubic / private key encryption method is the basis for SSL and TLS encryption protocols. Web browsers indicate this with the lock icon and https.

**W3C and HTML and CSS**

As mentioned in the first section of the project, Tim Berners-Lee brought about the core requirements and building blocks to make the internet accessible to the average person. With the invention of the World Wide Web via the standardization of HTTP and other protocols, Berners-Lee created a standardized platform that allowed the internet to become the huge, ever growing monster it is today.

Tim Berners-Lee not only helped to pioneer the World Wide Web, all associated protocols, and the very first versions of HTML in 1989, but in 1994 he also helped found the World Wide Web Consortium (W3C). Berners-Lee is still actively involved in the internet and its inner workings. He currently sits on the Board as its Director.

HTML stands for hypertext markup language and is the language used to tell a web browser how a page should look. HTML is made up of tags for fonts, links, images, formatting, alignment, etc. Images and videos are not included in HTML but see separate files that are linked to via separate URLs. These requests are handled via separate HTTP requests from the main page and are displayed as they come in.

CSS, which stands for Cascading Style Sheets, expands on HTML and other markup languages. While HTML can be thought of as the frame of a house, the CSS can be thought of as the siding, paint, and furniture. Instead of being limited by the defaults and sometimes rigid formatting of HTML, CSS enables the separation of presentation and content, including layout, colors, and fonts. CSS and its standards are also maintained by W3C.