Unit 7 - The Internet - Study Guide

Hardware of the Internet

- The Internet is a huge system of interconnected networks. Since it's not feasible for every device to have a direct connection to every single other device, we use a network of **routers** that are all connected to one another - eventually.
  - This system has messages sent from one router to the next, with the path determined by the lowest **cost**. There are a few factors that go into this decision:
    - Distance from the destination
    - Reliability of recipient router
    - Type of connection (see below)
  - This network of routers is both **scalable** and **fault tolerant** because of the way that it's built.
    - More routers can always be added to the network, increasing reliability and potential paths.
    - If a single router goes down, the entire Internet doesn't fail. Information can still get from one point to another - a different path will just have to be taken.
- There are 3 primary ways in which digital data is transmitted between devices:
  - Electricity via Ethernet
    - Cheap, middling in range
  - Light via Fiber Optic
    - Expensive, incredibly fast and long-range with very low latency
  - Radio Wave via WiFi/Cell Towers
    - No wires! Lowest speed and range of the 3
- Every device that is connected to the Internet must have a unique identifier, so that it can be located across the network. This is called the **IP Address**.
  - IPv4
    - This is the old standard, and it uses **32 bits** to store addresses. This means that there are ~4 billion distinct addresses, which is no longer enough to support all the devices connected to the Internet.
  - IPv6
    - This is the standard being shifted to. It uses **128 bits** to store its addresses. This means that there are 2^{128} distinct addresses - that's 340 **undecillion**.

Rendering a Webpage

- There are many steps involved in getting a web page loaded on your computer. They are as follows:
- **URL**
  - URL stands for **U**niform **R**esource **L**ocator. They are used to locate the resources we're looking for on the Internet.
  - Every URL has 2 components:
    - **Domain**
      - The domain for a URL determines **where** on the Internet you need to look to find the resource
    - **Path**
      - The path determines the actual resource being requested
        - The resource could be an HTML file, an image, a video, or any number of other things!
- **HTTP Request**
  - HTTP stands for **H**yper**T**ext **T**ransfer **P**rotocol. It is the standardized language that computers use to send and receive Internet resources.
  - HTTP requests contain all of the information about what exactly is being requested - which domain is hosting the resource, what resource is being requested, the format for the resource being requested, among many other things.
- **DNS Lookup**
  - DNS stands for **D**omain **N**ame **S**ystem. This is a sequence of servers containing information about the IP Addresses for every website in existence.
  - Your computer will use the DNS if it does not have the IP Address for the **domain** in the URL in its **cache**.
  - The DNS works from more general to more specific domains.
    - Top-Level Domain
      - The top-level domain includes things like **.com**, **.org**, and **.net**.
      - A **Root Server** contains information about where information on each of these top-level domain servers can be found.
    - Second-Level Domain
      - The second-level domain is going to be the name of the website you're trying to access. Things like **example**.com, **vcs**.net, **wikipedia**.org.
      - Information about these can be found on the servers a Root server will point you towards.
    - Sub-Domain
      - Sub-domains include things like **www**.example.com and **learn**.vcs.net.
      - A Hosting server will provide you with the IP Address you're looking for.

- **Create Request Packets**
  - Once the destination IP Address has been squared away, the HTTP Request is ready to be sent. The entire message isn't all sent at once, though - it's broken down into **packets**. Each packet contains a small portion of the request being sent.
  - The formatting for packets is determined by TCP/IP. This stands for **T**ransmission **C**ontrol **P**rotocol/**I**nternet **P**rotocol. IP includes information about the sender, recipient, and size of the packet, where TCP includes information about the way in which the packets should be ordered upon receipt.
  - The information added by TCP/IP is called **metadata** - it's data *about* the data that's actually being transmitted.
- **Route Request Packets**
  - The packets containing the request are then sent out into the Internet. Each one can take its own path through the massive network of routers in order to reach its final destination.
- **Request Received, Processed**
  - Once the server for the domain receives the packets containing the request, it will ensure that it has received all of them, using the information added by TCP. If it's missing any packets, it will request that the sender re-send them.
  - If all the packets have been received properly, the server will reassemble them into the original request. It knows which order to do this because of TCP!
  - With the request in hand, the server can read through it and determine what resource is being requested.
- **HTTP Response**
  - Similarly to the HTTP Request created by our computer, the server will create an HTTP Response to be sent across the Internet.
  - This response will include a status code, the resource being requested, and much of the same information as the original request (format of data, language, etc.).
- **Response Packets Created**
  - This process is identical to the creation of the request packets.
  - Both devices know how to structure packets, because they are both using the same set of protocols!
- **Response Packets Routed**
  - Similarly to when the request was sent, the response packets are sent through the Internet to reach the original requester's device.
- **Response Received, Processed**
  - Your computer receives the packets, then re-assembles them into a coherent message. Once again, it knows the correct order because of TCP's metadata!
- **Webpage Rendered**
  - Your Internet browser then reads the HTML response it got from the server and renders it into a webpage for you to view.

Sequential, Parallel, and Distributed Computing

- Each of these styles of computing determines the method/order in which tasks are completed by a computing system.
- **Sequential**
  - Tasks are completed **one at a time**. A single processor is used to accomplish tasks in the order they were assigned.
  - This is the simplest computing style, since it requires no additional setup. It is, however, slower than the alternatives.
- **Parallel**
  - Multiple tasks can be completed **simultaneously**. Multiple processors, sharing memory, are used to accomplish tasks.
  - This will usually be faster than sequential computing, because more than one task can be completed at once. It does require additional setup and maintenance, though, and can be more difficult to troubleshoot.
- **Distributed**
  - Tasks are divided up between **multiple computers**. Similar to Parallel Computing, but using the processors from different computers rather than multiple within the same machine.
  - Very fault tolerant, since tasks can continue to be run even if a single device completing tasks goes down.
  - Scalable, since more machines can be added to the network and thus increase the processing power of the collective.
  - Requires lots of setup and can be expensive to maintain.

Cybersecurity

- Because all information on the Internet travels over shared wires, it's possible for our messages to be intercepted by nefarious ne'er-do-wells.
    - There are many types of attacks that hackers can undertake in an effort to gain access to information, including **D**istributed **D**enial of **S**ervice (DDoS) and **DNS Spoofing**.
    - Phishing is a type of attack against the Internet *user*, in an effort to get them to divulge important information about themselves.
- In order to prevent these dastardly villains from doing anything heinous with our information, there are several types of precautions we take.
    - **Encryption**
        - SSL/TLS encryption is used to ensure that messages cannot be read, even if they are intercepted!
        - This type of encryption uses the Public Key Encryption method.
        - Certificate Authorities (CAs) help us ensure that the public keys we receive from websites are valid.
    - **Security Software**
        - We can install anti-virus software to keep our computers safe from malware.
    - **Following Best Practices**
        - Often, **we** are the greatest hole in our Internet security. There are several things that we can do to protect ourselves:
            - Use strong passwords that are difficult for computers to guess
            - Use a different password for each website you log into
            - Keep your software updated
            - Don't click suspicious links or download fishy files

Impact of the Internet

- The Internet is a philosophy of making information available to any and everyone.
- This huge technological marvel has had enormous impacts on the world, both positive and negative.
- **Communication**
  - The Internet enables communication of information in a capacity that is unprecedented throughout history. People can instantly send messages with others across the world!
  - With information so freely available, anyone can learn just about anything that they want.
  - There is a dark side to the coin - some information *shouldn't* be shared with the public, like classified government secrets! If those are made public, they could put people in danger.
- **E-commerce**
  - The Internet has allowed people to purchase things from their homes, without the need to get up and go anywhere! This enables them to do research about the quality of the products they want, as well as makes sure people aren't limited by their location.
- **Digital Divide**
  - Roughly 60% of the global population is currently connected to the Internet.
  - Access to the Internet and the knowledge it provides varies greatly based on geography and socioeconomic status. There are many people working all over the world to solve this issue, but it's still far from fixed!

Creative Credit and Copyright

- Whenever someone creates a work of art, they become the owner of the **copyright** for that work. They can choose from many different types of licenses for that work:
  - **All Rights Reserved**
    - The owner makes all decisions on how the work can be used, distributed, and modified
  - **Some Rights Reserved**
    - The owner allows the work to be used, distributed, or modified, but is still required to be credited for their work
  - **Public Domain**
    - The owner has relinquished control of how the work is distributed, used, and modified.