**The Internet**

Table of Contents

[Introduction 2](#_Toc65538527)

[IP Addresses 3](#_Toc65538528)

[DNS Servers 4](#_Toc65538529)

[Browsers 5](#_Toc65538530)

[TCP/IP 6](#_Toc65538531)

[Internet Architecture 9](#_Toc65538532)

[FTP 10](#_Toc65538533)

[Hardware Connections 12](#_Toc65538534)

## Introduction

The internet was initially a connection of major networks belonging to different universities, research centres and larger institutions. The World Wide Web was created in 1991 as a method to organize the information that was shared through these connections. It consisted of hyperlinks, that linked pieces of text that contained similar information. This process is called the Hyper Text Transfer Protocol, or HTTP. Later, hypermedia was created, which did the same thing using images and other graphical objects. The ‘websites’ we use today are exactly the same. They are just information that resides on a computer with multiple pages of hyperlinks and hypermedia, with a single homepage.

## IP Addresses

Every computer connected to the internet is given a unique IP address, so that other computers are able to identify it. This is also how websites function. The website resides on a computer with a specific, fixed IP address, and a connection is made between the IP address of that computer and that of the user when the website is visited. For the typical user, the IP address is supplied automatically by their Internet Service Provider (ISP) of from a Dynamic Host Configuration Protocol (DHCP) server.

A DHCP is just once computer on a network that has the job of providing all other computers on that network with an IP address. It has multiple IP addresses at its disposal, and usually assigns IP addresses dynamically, meaning they are temporary and not fixed.

## DNS Servers

The IP address belonging to each computer and website is just a long series of numbers. A user cannot be expected to remember the exact IP address of every website they wish to visit, since there is nothing special about the IP address. Instead, each website is also accessible via a unique name, called its domain name. A DNS server is like a directory that maintains a list of domain names and their associated IP addresses. When the user enters a domain name into their browser, the local DNS cache on their system is first checked. If the domain was previously visited, the IP address has been stored on the computer. If it is not found, then the browser sends a request to the DNS server. The DHCP server or the ISP connects to a DNS server to check which IP address is associated with that domain name, and then connects the user to the specified IP address. Usually, more than one DNS server is available, so that if one is down, another can be used. ISPs tend to use their own DNS servers, but public ones are also available and are usually more reliable. This is because public DNS servers, like the one provided by Google, tend to be exposed to a much larger range of IP addresses making them more reliable, and are also maintained by much larger companies so they are less likely to fail. If the IP is not found even there, the root DNS server is sent a request. This is a system of 13 servers distributed across the planet that drive the entire internet. The root DNS server however, does not return the address of the actual domain requested, since it does not always know. It does however, return the address of the top-level domain (TLD) DNS server. For example, a domain name ending in .com belongs to a .com TLD server. The address of the TLD server is stored locally so that the root DNS server does not have to be asked again. The TLD server returns the required IP address.

## Browsers

When a user types in a URL or a search term into their browsers address bar, modern browsers check to see if it is an actual URL. If it is typed in correctly, the browser will build the complete URL. For example, the user may type in ‘google.com’, but the browser will automatically add an http:// to the beginning of this by default.

Next the browser starts a DNS lookup to retrieve the IP address. After the IP address has been found, a request is sent by the browser to the IP address. The request consists, of the request line, which contains which HTTP method to use, the resource location and the protocol version, the request header, which must contain at least the host name and connection type, and the request body, which is optional.

Once the request is sent, the server returns a response. If the request was successful, the server returns a code ‘200 OK’. Otherwise, it returns an error code. Common error codes include ‘404 Not Found’, ‘403 Forbidden’ and ‘500 Internal Server Error’.

With a successful request, the browser receives an HTML body, along with other resources required by the page such as graphics files, CSS files and JavaScript files. The browser then renders the page as required.

## TCP/IP

The huge amount of data that is transferred through the internet every day can make things very difficult to manage and keep organized. To do this, the Transmission Control Protocol (TCP) and Internet Protocol (IP) were developed by the United States Department of Defence. The TCP/IP protocol is a set of rules that must be followed while transferring data over the internet.

In accordance with this protocol, data is first broken into smaller parts called packets. This is to make it easier to handle data from a large number of users. If a user were to send a huge piece of data at once, it would be difficult to manage and would cause delays since all the other users would essentially have to wait in line for their data to be transferred.

The protocol also requires that each packet have a source IP address and a destination IP address attached to it. The destination IP address is needed to know where to send the data. The source address is mostly used by the ISP at the receiving end, to make sure that the receiving IP has permission to access the data from that source. The process of attaching the required IP addresses to the data packet is called ‘encapsulation’.

Other than this, there are some applications that connect users directly through the internet, such as chat applications. These are called network applications. These require a further layer of identification, so that the particular application requesting the data can be identified and the data is not sent to any other application. They are thus given an additional address, called a port.

The source IP address, and if required its port, and the destination IP address, and if required its port, can thus form a connection. This connection and its four components combined is called a socket.

The logical connection for communication between two systems is called a session. Every time a website is visited, the browser starts an HTTP session. The session is a limited time communication and has a specific ID that is generated randomly, for additional security. The server will remember the browser for the duration of the session.

The networks themselves are connected to each other through routers. Routers send data from the source network to the destination network. This means that another layer of addressing is needed. The data, along with its encapsulated source and destination IP addresses, is first sent to its local network. The network encapsulates the packet in another layer of addressing, which identifies its router. The router is identified using a ‘gateway’ address. The router now decapsulates the packet, removing the previous layer of addressing, and encapsulates it again so as to send it to another network. The movement of data from one network to another is called a network hop. This process continues until the destination IP address is reached. There, the packet is completely decapsulated so as to retrieve the data.

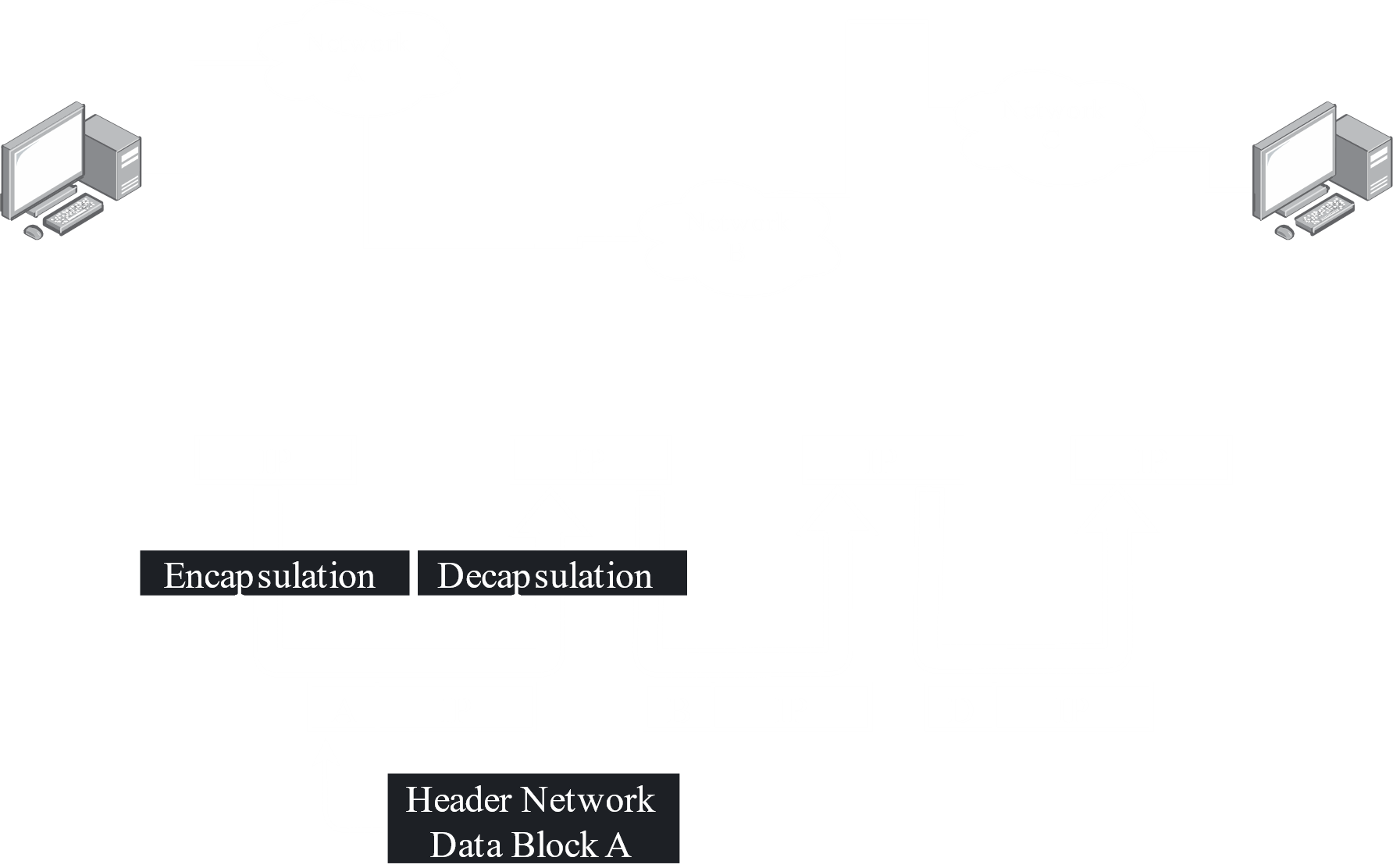
Computers on the same network however, can disregard all of this. They do not require access to other networks, and so do not need to follow the process outlined by TCP/IP, since they are essentially not connecting to the internet. Data can be directly transferred from one computer to another. However, some operating systems fail to realize that data is being transferred within the same network. This results in the data first being sent to the router instead of the destination computer. The router then sends the data back to the local network, from where it is sent to the destination computer. The whole process of data transfer and encapsulation and decapsulation of data packets by the network and router, is much slower than if the data had just been sent directly to the destination computer.

IP packets are independent of each other and are routed individually in the network. This means that there is no way to detect if any packets have been lost. TCP solves this. It allows recovery of a lost fragment by retransmitting the data.

TCP/IP allows two types of connections, open and closed. HTTP is a closed connection.

## Internet Architecture

The entire structure of data transfer from source to destination, through multiple networks and routers is called the Internet Architecture. The structure, if visualized, would look something like this:



Internet Architecture consists of three layers, the IP layers, the TCP layer and the Application layer. The Application layer includes different protocols on which to build internet services like E-mail, File Transfer Protocol (FTP), transfer of hypermedia pages and the transfer of distributed databases (like in the World Wide Web).

## FTP

The File Transfer Protocol (FTP) is a protocol computer on the internet follow to transfer files to and from one another. It is the internet standard for file transfer between computers using TCP or IP networks. FTP enables a computer to upload files to a hosted web server so that the files can be viewed and downloaded on the internet. An FTP client can make this job easier. It is a software application that allows the user to log into a server and manage files on the server. FTPs may be authenticated, using usernames and passwords, or anonymous, which do not require any authentication.

FTP uses one connection for commands and another for transferring data. The FTP server looks for connections on a standard port number. A port is a logical connection point for communication using the Internet Protocol. Usually, the port 21 is used only for sending commands and is called the command port. For example, the FTP client may ask for the list of files and folders available, and the FTP server will send back this list. Data is transferred on the data port, whose number can vary depending on the mode of connection.

There are two types of modes of connection, active and passive. In an active FTP connection, the client opens a port and searches for incoming connection requests from servers. They generally use port 20 as the data port. In a passive FTP connection, the server opens a port and searches for clients that are trying to connect to it. Usually, passive connections are made since active connections may pose security risks. Firewalls tend to block connection requests from the outside, while allowing outgoing connection requests.

There are some problems that come with such a convenient service however. The FTP server is just a computer that provides the FTP service. It is very lucrative for hackers, and thus necessitates security software and utilization of usernames, passwords and file access control. Furthermore, since if the FTP is publicly accessible and anyone can upload files to it, there is a big possibility that some of those files will contain viruses since it is not possible to check the possibly huge number of files on the FTP. This means that users who download files need to be extra careful, using only trustworthy FTPs and checking any files they download for viruses.

## Hardware Connections

The connections between networks, routers and computers are done with a multitude of cables. ISPs connect users to their network using co-axial cables. The cables are connected to the router using a RJ45 connector, and the cables themselves are known as Category 5 (CAT5) and Category 6 (CAT6) cables, which are twisted pair cables. Connections between computers on the same network are made with CAT5 cables as well. These connections are called Local Area Network (LAN) connections.

Twisted Pair Cables consist of 4 pairs of wires. Each pair is spirally twisted in order to reduce electromagnetic interference. 1 pair sends data, 1 pair receives data and 2 pairs help reduce electromagnetic interference so that data transfer is not hampered. The wires are twisted to reduce magnetic interference so data can travel further. The maximum distance a twisted pair cable can cover is 100 meters. A repeated is needed for larger distances.

Connections from the ISPs network to other networks and routers are made through fibre optic cables. The underwater submarine cables that make connections between different countries use these as well. These work with the help of total internal reflection. They consist of 2 wires, one to receive data and one to send data. Fibre optic cables can cover a few thousand miles. If a part of the cable gets damaged, the entire part must be replaced. Repairing it is not an option since it may damage the cable’s data transfer capabilities.

An alternative to submarine cables is the Very Small Aperture Terminal, or VSAT. These are satellite dishes that send data directly to a satellite in order to access the internet. Communication is very slow however, and this is only used as a backup plan by high-level government institutions that cannot afford to be disconnected from the internet.

Another form of internet communication widely used is that provided by mobile operators. They provide internet access through mobile technology. Communication here is also very slow.

At the user end, the computer sends data to the modem as electrical signals. The modem works as a converter. It converts electrical signals to optical signals, which it sends to the cables, and coverts the optical signals it receives to electrical signals, which it sends to the computer. The modem also has the responsibility of maintaining serial communication, which means it ensures data is sent one bit at a time, in the correct order.

As established, the data must first go to the ISP. The ISP regulates the speed and quality of the connection. The more the user pays, the higher speeds and better quality they receive. This means that no connections is actually ‘slow’. Data is transferred at very high speeds between networks through the submarine cables. The data is simply held back by the ISP, and is transferred depending on the bandwidth the user has paid for.

The ISP sends the data to a regional network, which connects to the terminal where connections to the submarine cables are managed. This is needed because the submarine cables already have a very large number of channels. It would become very messy if every ISP had direct access to them.