HTTP 5110: how the Internet works (more or less).

You might have noticed this thing called the Internet. Other than being a rather time consuming and engaging technology it is also the modern equivalent of moveable type and mass production: a new technology that will have a profound impact on society.

As you enter this course you might ask yourself: how does the Internet work? Most of us would feel that the answer would be extremely dense in its jargon and technological mumbo jumbo: a good cure for insomnia or, when printed out, a handy way of levelling the dining room table. And you would be right but my job is to help you understand how the Internet works. In a very simplified format

The Internet is a large collection, probably billions, of computers that are able to use a telephone in order to transfer data in binary format. Binary is a series of discrete electronic signals that, for our convenience, translates into zeroes and ones. The computers utilize a protocol that tells other computers what to do with those zeroes and ones. The two main protocols in this case are TCP IP.

TCP IP are the main protocols of the Internet.

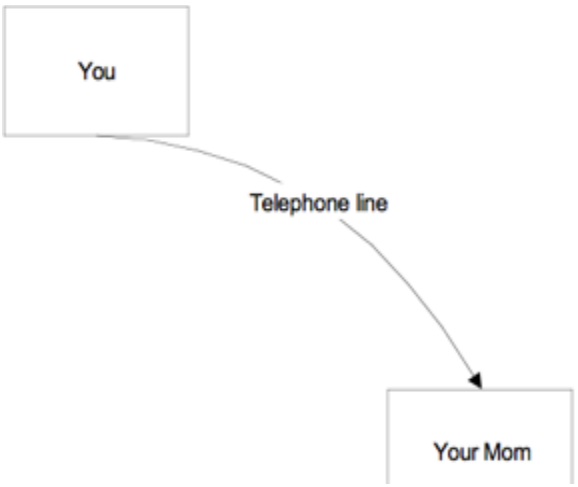
TCP is the Transmission Control Protocol
IP is the Internet Protocol.

How does it work?

A way back before you were born, say 1964, researchers at Stanford University had a moment. They realized that within the computer different components; say the keyboard and the CPU, talk to each other through wires. Now what other technology uses wires to communicate? If you said: telephone you would be right. If wires run the telephone system and wires run computers then why could not computers talk to each other over the telephone?

They wrote the programs to get the computers to talk to one another and hooked them up and…it worked. They could transfer data from one computer to another. One computer could collect that data while another computer could analyse it. They could avoid having to send a research student across the campus with a huge stack of punch cards and then wait for the report to come back.

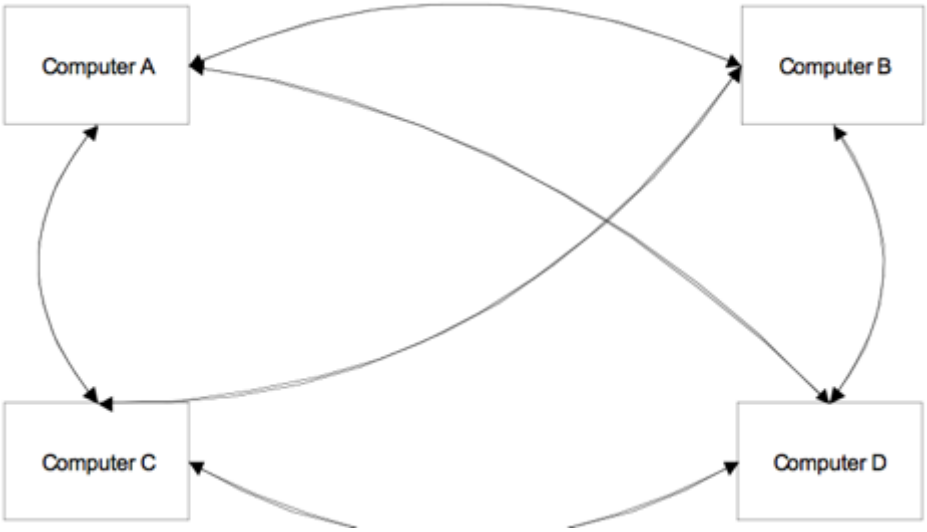
Data transfer over the telephone has several problems. The main one being that while you are the telephone with your Mom no one else can use these two devices or the wire they are attached to. This is not usually a problem but what happens in an emergency? How easy is it to overload the system?



Initially the network was slow, 5 words per second, and was plagued by a telephone system that was imperfect. Moreover, if for some reason the transfer was interrupted, someone picked up the phone, then the transfer would have to start over again. So they needed another means of creating the data and getting it ready for transfer and they chose packet switching.

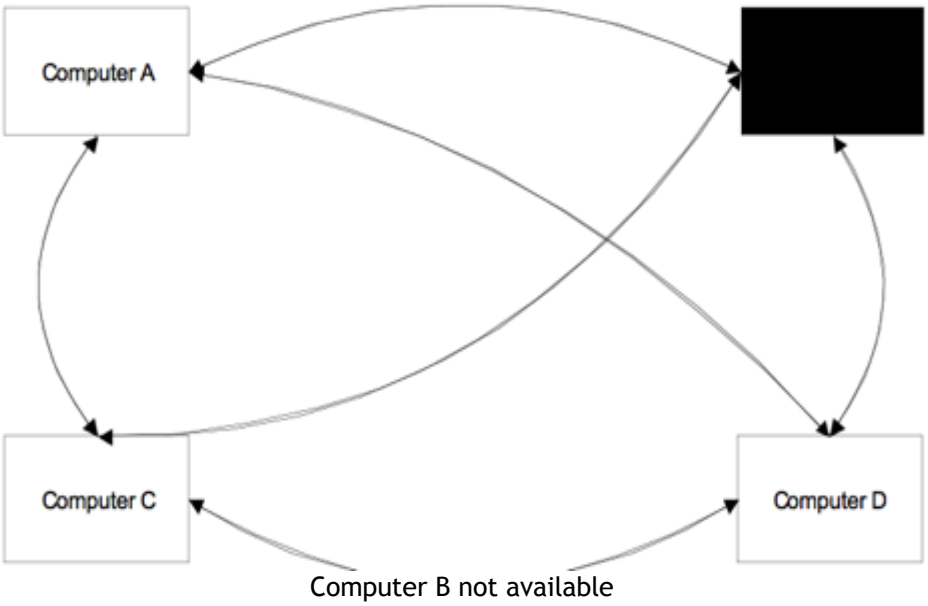
Packet Switching is another term for TCP.

The essence of TCP is that it separates the data into smaller packets and then sends those packets over the network where they are collected by the receiving computer and reassembled into the original sequence. The packets are ‘switched’, control is transferred, because individual packets can take any route from the sending computer to the receiving computer.



If the arrows are data then they can take any route from Computer A to Computer D: ABCD, ACBD, AD, ABCD, and so on. So if one computer hangs up, or goes down, or is too busy, there are other computers in the network that can take on the job.

The real ingenuity of this system is: what happens if a major computer centre goes down?



Considering this is the 1960's: can you imagine a situation where a major computer centre would suddenly not be available? Think Cold War think nuclear meltdown. You get what I mean – the Internet was designed to keep working even though everything else stopped working.

This is not a problem: TCP finds another available computer to send the information to. Moreover, TCP is smart enough to spot bad packets and to know if there are dropped packets: these are simply asked for again. It is for this reason that most downloads can be disrupted and restarted.

TCP is the content and IP, Internet protocol, is where to send the content. If you look at the address bar in your browser you have seen numbers that might look like this: 123.456.789.123, or 401.703.921.412. These numbers represent the address of a particular computer connected to the Internet. Each computer connected to the Internet has a unique number that allows it (allows you the user) to send and receive data. In the diagrams above the computers marked ABCD would each have an IP address consisting of four octets or sets of numbers. Every device connected to the Internet gets a number like this and,

like a telephone number, this is how you are able to get the data. Trying to remember all of these numbers would be quite hard – to make it easier we instead use domain names.

Domain names have two parts – the name and the top-level domain. The name can be anything but the top-level domain has to be a recognized word or a two letter country code. The original top-level domains were .com, .net, .org, .edu, .gov, .int, and .mil. You are familiar with .com, .net, and .org. Country codes are similar, .ca for Canada, .uk for the United Kingdom and so on. The Domain Name System (DNS) maps the domain name to the IP address. When you register a domain your domain is placed into a database that is accessed again and again as your domain is visited. Each access will eventually propagate your domain into more and more databases serving domain names. When you request a document in your browser your request is checked against your ISP's cached addresses and if it is not there your request is checked against a regional database and so on until your request is handled by a root name server. There are thirteen root name servers in the world. All supporting this highly effective and highly redundant names database.

The Internet is based on TCP-IP and the Web is based on HTTP or the Hyper Text Transfer Protocol. HTTP works on TCP-IP like an application, Word, works on an operating system, Windows. HTTP is a client server system where the client requests a document and the server retrieves it and the browser displays it. You make your request by asking for a URL, a Uniform Resource Locator, and your ISP, Internet Service Provider, looks up the domain name, translates it to an IP address and finds that resource or computer. Once found then copy the information and send it back to you. This is another interesting part of using the Internet – every time you see something on your screen you have taken a copy of it.

TCP slices the data into small parts and transmit them through the network wires and IP makes sure the information arrives in the right spot. There are other protocols that work in a similar fashion. FTP, File Transfer Protocol, moves files from your desktop to the server, SMTP, Simple Mail Transfer Protocol, handles your email.

Overall, it is well designed and reliable system. The beauty is in the redundancy that ensures the system works regardless of the state of any one access point or computer. For example, the World Trade Center, was a key access point in the northeast US. When it collapsed in 2001 the Internet traffic was almost stopped for about 15 minutes before it recovered.

For further reading:

HTTP 5110: how the Internet works (more or less).

https://www.hp.com/us-en/shop/tech-takes/how-does-the-internet-work This article has a nice graphic.

https://developer.mozilla.org/en-US/docs/Learn/Common_questions/How_does_the_Internet_work From Mozilla one of the older groups on the Internet.

https://developer.mozilla.org/en-US/docs/Learn/Getting_started_with_the_web/How_the_Web_works and some more from Mozilla

https://datatracker.ietf.org/doc/html/rfc793
https://datatracker.ietf.org/doc/html/rfc1180
For those of you who want to get their geek on – these are the original specifications for TCP IP.

http://www.walthowe.com/navnet/history.html A nice series of the non-technical side of building the Internet.