Networking

"Latency, not bandwidth, is the performance bottleneck for most websites! To understand why we need to understand the mechanics of TCP and HTTP protocols." – High Performance Browser Networking

Measuring Performance

When measuring the performance on an application that communicates with other end points over a network, we will consider three important terms.

	Description
Bandwidth	Maximum amount of data per unit of time e.g. gigabytes per second
Throughput	Actual amount of data per unit of time
Latency	The delay between the sender sending and the receiver reiving information.

We can think of bandwidth as the width of the pipe and latency as the delay incurred by packets traveling from one end of the pipe to the other. If I use a high bandwidth, high latency system I would expect to wait a while for my file to start downloading but from the point I see the download start it would be fast. If I were using a low latency, low bandwidth channel the file would start downloading immediately but the download would take a while.

> **ACCEPTABLE DELAY**
>
> Most people notice a delay of 100-200ms and any delay greater than 300ms is considered sluggish. 1second starts to be a bad delay. To provide a great use experience we need to aim for response times of 100-200ms. This is one of the driving factors behind the adoption of content delivery networks to locate servers close to users. See the following link

We will now look at the three elements of networking performance.

Latency

Latency describes the time taken by a packet to travel from sender to receiver. Latency is impacted by the following factors.

	Description
Propagation Delay	Length of channel dividend by the speed of the signal travelling down it. Typically, the speed of the signal is close to the speed of light.
Transmission Delay	Time required to push a packet's bits onto the wire. Simply a function of the number of bits in the packet and the rate of transmission in bits per second. If we have a 100Mb file then it will take 100 seconds to push it onto the wire using a 1Mb/s link and 1 second to push it onto the wire using a 100Mb/s link
Processing Delay	The delay incurred by a router processing the packet header, handing errors and deciding on the next destination
Queuing Delay	If incoming packets arrive faster than a router can deal with them, they get queued inside a buffer

Typically, a packet traveling across a network will encounter many intermediate routers and will hence be impacted by multiple processing and transmission delays. Where load is heavy, we have increased risk of incurring queueing delays.

Very roughly we can assume that a signal in fibre travels at about 200,000,000 metres per second. The distance between London and New York is 5500km so the round trip in seconds is 0.055 seconds or 55ms. The world circumference is 40075 km, so the time taken to circumnavigate the globe is then 0.2 seconds or 200ms. In practice the route is not the minimum distance and there will be multiple transmission, processing and queueing delays.

Another problem is that there is a delay between the home and the ISP of between 10ms and 70ms. This is known as the last mile latency.

These might add 100% giving us 400ms to circumvent the globe and 110ms for round trip between New York and London. This latency is affected by the ISP, technology used and time of day.

Bandwidth

A single fibre optic link support multiple channels via wavelength multiplexing. The capacity of a fibre is 171Gb/s so across all channels we get 70Tb/s. A single cable often has four strands of fibre giving a bandwidth of 280Tb/s. The ends of these fibres, however, are connected to much lower bandwidth technologies such as DSL, cable and wireless technologies. The available bandwidth is driven by the weakest link which is typically these low-capacity technologies. In 2018 the worlds average download speed was 46Mbs and upload speed was 22Mbs. My own sky gave figures of 16Mbs and upload 0.9Mbs.

The following online resource provides details on networking.

https://hpbn.co