The Internet

The internet and World Wide Web

There is a common misconception around the definition of the internet and WWW. People tend to use the terms interchangeably but there is a distinction.

The **internet** is a global network of interconnected computers. It allows communication between computers and devices using TCP/IP. It has globally unique IP addresses. Individual devices are attached to a local area network, which in turn are connected to wide area networks. The internet is a network of wide area networks and is itself is a wide are network. The internet refers to the physical hardware components such as computers, cables, routers, gateways and so on. The internet has be around since the 1960s.

The **World Wide Web (WWW)** is a service that has web pages and other content that runs on the internet. It is made up of interlinked hypertext documents and uses mainly the HTTP protocol. The WWW was developed in the early 1990s by Tim Berners Lee.

Local Area Network (LAN)

- Individual devices are attached to a local area network.
- A local area network is a network that covers a relatively small geographical area typically extends over the range of a single organisation such as a university campus, school site.
- LANs are typically managed by a single organisation.

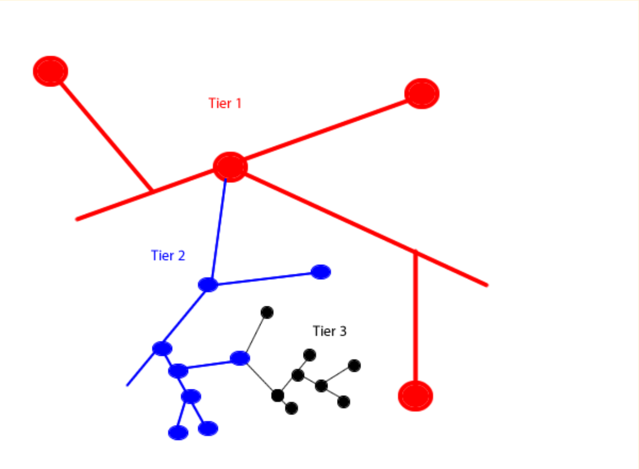
Wide Area Network (WAN)

- A wide area network is made up of many local area networks and covers a much wider geographical area.
- The internet the ultimate wide area of networks. It is a network of networks and allows billions of devices to be interconnected.

Structure of the internet

- **Tier 1** networks are considered the backbone of the internet. These are country wide networks that are connected to other networks by fibre optic cables that link different parts of the world and include cables that cross ocean floors including across the Atlantic ocean.
- **Tier 2** networks are regionally based and allow connectivity between local area networks
- **Tier 3** networks and LANs and enable internet access for homes and small businesses and individual organisations.

Internet tiers



Packet switching

Transfer of data across a network relies on the principal of packet switching. Packet switching is the process of data being broken into packets before being sent over the network and then reassembled at the other end. Packets are forwarded through a series of routers between their source and destination. This allows data to be transfer across a network efficiently. Large files would otherwise clog up the network. Small packets can choose different routes through the network, however packets do get lost during transfer. A typical packet size is 1500 bytes.

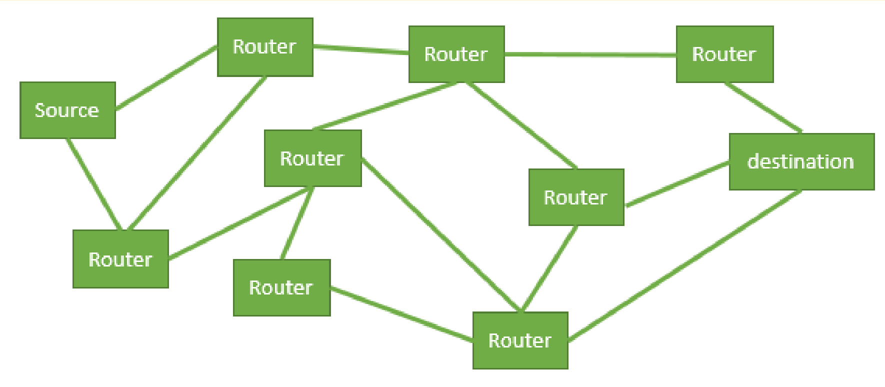
Packet contents

Header	Source IP address	Where the packet has come from
	Destination IP address	Where to send the packet
	Packet identification	Necessary so the computer knows how and in what order to reassemble the packets
	Destination and source MAC Address	This is the address of the network card
	Destination and source port numbers	
	Protocol	Which protocol is being used (eg HTTP, FTP)
Body	Data/Payload	Part of the file that you want to send
Footer	Error control bits	Check for errors in the packets to make sure they have not been corrupted in transport

Gateways and routers

- The purpose of a router is to connect networks together. The router directs internet traffic to destination along the quickest and least congested routes. Normally a packet will be routed through multiple routers before it reaches its destination. Routers have at least two network cards so that traffic can be directed in different directions.
- A gateway, like a router, connects networks together. A gateway provides a single access point to a network. Networks may have different protocols so gateways allow communication to be translated between different protocols.

Possible routes for packets



Uniform Resource Locator

The Uniform resource locator is the address of a resource on the world wide web. The resource can be a web page or other file such as mp3, pdf for instance. It contains both the protocol and domain name and takes the form of:

Protocol://Fully qualified Domain Name/Path

- Protocol: These tell the browser what to do with the web address (eg HTTP, HTTPS)
- FQDN: This is the name of the website
- Path: Points to where the specific page is on the website.

The URL for http://www.bbc.co.uk/news/world-middle-east-23691571

Protocol:	http
FQDN:	www.bbc.co.uk
Path:	news/world-middle-east-23691571

IP Address

Every device on the internet needs to have a unique IP (internet protocol) address. Packets contain the sender's and receiver's IP address so that routers know where to direct the packets. Just like every house in the country has a unique postal address.

Domain name

Each web server has an IP address, and the IP address can be used to request a web page. However, IP addresses are hard to remember so domain names are used to identify IP addresses and are much easier to remember. Multiple IP addresses can be associated with a single domain name.

Example

Domain name	www.google.co.uk
IP address	216.239.238.120

The Domain name identifies the location of the resource on the internet. It is structured hierarchically with domains and sub-domains.

Generic top level domains	.com	.org	.net	.gov		
Country top level domains	.fr	.nz	.au	.uk	.tv	de
Second level domains for UK	.co.uk	.org.uk	.sch.uk	.ac.uk	.gov.uk	.nhs.uk

Fully qualified domain name (FQDN)

The fully qualified domain name contains the complete domain name and hostname of the web server.

Examples of FQDN

access a webserver	www.bbc.co.uk
access a webmail server	mail.google.com

The domain name hierarchy for www.bbc.co.uk.

root	.
Top level domain	uk
Second level domain	Co
Local domain name	bbc
Hostname of server	www

Domain Name server (DNS)

When a domain name is requested the domain name server searches through huge databases to find the corresponding IP address. If the server cannot find the corresponding IP address it then requests the IP address from the DNS system (other DNS servers).

Internet registries

Internet registries allocate domain name to one or more IP addresses. This is a complex task that is overseen by an organisation called ICANN (Internet Corporation for Assigned Names and Numbers). Regional internet registries are responsible (e.g. RIPE NCC in Europe) for allocating a set of IP addresses to domain names. This ensures that domain names and IP addresses are globally unique.

Internet Security

Why do we need network security?

- To prevent unauthorised access to our electronic devices
- To protect our data eg to prevent sensitive data being stolen, to prevent personal data from being stolen.

Firewall

A server is dedicated to acting as the firewall. It has two network cards one for the LAN and one for the internet. A firewall prevents unauthorised access to a network and represent the first line of defence for a network. Networked computers have lots of incoming and outgoing data packets. Whilst most data packets are harmless, some data packets may be harmful and contain malware. It is the role of the firewall software to identify and prevent these packets getting on to the LAN/computer in the first place.

Static packet filtering Incoming packets via the internet are monitored and inspected. The information in the packet headers including the source and destination IP addresses, ports and protocols are checked. Any packets that do not meet the filtering criteria (eg only accept packets with specified source IP addresses) are blocked, otherwise packets will be passed onto the LAN.

Stateful inspection – dynamic packet filtering

- Stateful inspection is form of packet filtering that monitors both outgoing and incoming packets.
- In static packet filtering only information in the header is examined, but in dynamic packet filtering the contents of the packet are also examined.
- Stateful inspection monitors the state of the connection for a particular communication. Static packet filtering is based in a set of predefined rules, whereas dynamic packet filtering considers the context of the connection based on previous packets. For instance if a request to a web server is made, then response packets from that server will be expected and allowed to pass to the LAN.
- Stateful inspection offers a better level of protection than static packet filtering.

Proxy server

When a proxy server is used there is no direct connection between the LAN and internet. Traffic is routed through the proxy server. The proxy server will have a different IP address to the devices on the LAN allowing the IP address to be hidden outside the network. The proxy server then acts as the firewall as required.

Encryption

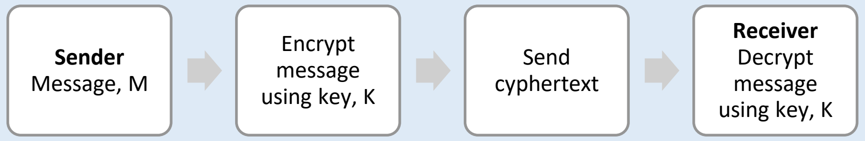
When passing sensitive data over the internet such as credit card numbers you need to ensure that the data are encrypted. This means that the message is garbled up so if the message gets intercepted as it is being transmitted to its destination it will be almost impossible for anyone without the key to read the original message. Ensure that you are using the secure hypertext transfer protocol (https) and also on most web browsers a little green padlock should appear on the URL bar.

Symmetric encryption

In symmetric encryption both the sender and receiver use the same key to encrypt the decrypt the data.

$$C = f[M, K]$$
$$M = f^{-1}[C, K]$$

where C is the cyphertext, M is the message, K is the key and f can be any symmetric encryption algorithm


Sender Message, M
Encrypt message using key, K
Send cyphertext
Receiver Decrypt message using key, K

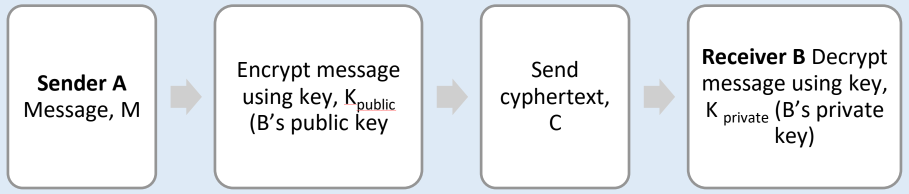
Asymmetric encryption

In asymmetric encryption a different key is used to encrypt and decrypt the data. This is a one way function, you cannot use the same key to encrypt and decrypt the data.

$$C = f[M, K_{public}]$$
$$M = f^{-1}[C, K_{private}]$$

where C is the cyphertext, M is the message, K_{public} is the public key, $K_{private}$ is the private key and f can be any symmetric encryption algorithm.

The sender A encrypts the message using the receiver B's public key. The receiver then decrypts the message using the private key that is not shared with anyone.


Sender A Message, M
Encrypt message using key, K_{public} (B's public key
Send cyphertext, C
Receiver B Decrypt message using key, $K_{private}$ (B's private key)

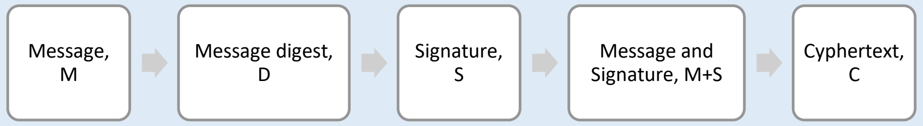
Supposing Alice wants to send a message to Bob. Alice uses Bob's public key that is made available to all in order to encrypt the message. She then send the message and Bob decrypts the message using his own private key that is known only to him. To reply Bob uses Alice's public key to encrypt his message. Bob sends the message and Alice decrypts the message using Alice's private key.

Digital signature- Sending

A digital signature is a way of verifying that a message has been sent from the correct source. To send a message from Alice to Bob:

1) Apply a hash (one way encryption), H to the message, M to produce the message digest, D such that:
$$D = H[M]$$
2) Encrypt the message digest D, using the Alice's (the sender's) private key ($A_{private}$) and an asymmetric encryption algorithm E to produce the digital signature, S.
$$\text{S} = E(H[M], A_{private})$$
3) Append the signature to the original message and encrypt using Bob's (the receiver's) public key (B_{public}) to produce the cyphertext C to be sent.
$$\text{C} = E(M + S, B_{public})$$

Hash Encrypt using $A_{private}$ Encrypt using B_{public}

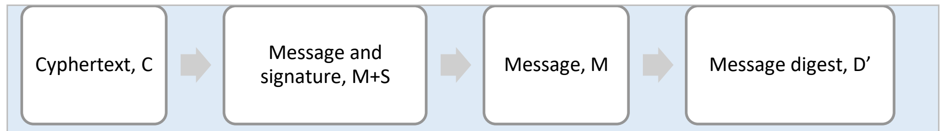


Digital signature - Receiving

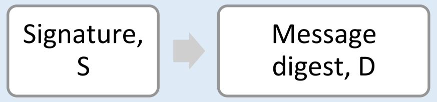
To receive a digital signature sent from Alice to Bob:

1) Decrypt the cyphertext using Bob's private key to extract the message and signature:
$$M + S = E\,[C, B_{private}]$$
2) Apply the hash, H to the message, M to produce the new message digest, D' such that:
$$D' = H[M]$$
3) Decrypt the signature to get the message digest D using Alice's public key, A_{public}
$$\text{D} = H[M] = (S, A_{public})$$
4) Evaluate D versus D'. If they are the same then the message is authentic and has not been tampered with

Decrypt using $B_{private}$ Hash



Decrypt using A_{public}



Digital certificate

Digital certificates are another way of verifying internet communications. Digital certificates are issued by certification authorities. The certificates contain information about the owner the public key and also the digital signature of the issuing body. Digital certificates are typically used by banks and e-commerce websites.

Malware

Malware is short for malicious software. Malware is software that has been purposely developed to damage, disrupt or take control of computer systems.

Types of malware

Trojan software is malware that gains access to a computer by pretending to be legitimate software. The Trojan allows hackers unauthorised remote access to your computer without the user being aware. From there the hacker can control your computer and use the machine for nefarious purposes such as installing key loggers to record passwords and pin numbers or launch attacks on other computers thereby obfuscating the original source of the attack.

Computer viruses are software that replicates themselves and can transfer from one computer to another. They are activated by a user often as email attachments and attachment to other files and programs. Once a virus is on a computer system it can make undesirable and unauthorised changes to a computer system. Viruses require human agency to be activated, eg opening an email attachment.

Worms spread like viruses but do not require human intervention, and attached themselves to network tools to spread themselves. In that sense they are more destructive than computer viruses because they can spread automatically from one computer to another very quickly and can affect whole networks.

Code quality, monitoring and protection

- Common areas of exploitation are buffer overflow and SQL injections.
- SQL injections are a way of hacking into databases. The hacker enters input that is able to access data in the database that they should not have access too (eg passwords of other people) or modify the database.
- Ensuring that code is built with a high degree of protection in the first place will avoid many of the potential ways of exploitation.
- With buffer overflow the hacker deliberately overwrites memory locations using existing code in RAM. In those locations the hacker can place malware.
- It is important that any security patches that come out for software that these are installed as soon as possible.

Antivirus software

- Antivirus software is software that scans the computer and is able to detect known malware against a database and quarantine and remove them from the system.
- Quarantine means that affected files are isolated and are unable to infect a computer system.
- Antivirus software needs to be regularly updated to keep up with new viruses that are continuously being developed.