AQA Computer Science A-Level

4.9.3 The Internet

Advanced Notes

Specification:

4.9.3.1 The Internet and how it works:

Understand the structure of the Internet.

Understand the role of packet switching and routers.

Know the main components of a packet.

Define:

- router
- gateway

Consider where and why they are used.

Explain how routing is achieved across the Internet.

Describe the term 'uniform resource locator' (URL) in the context of internetworking.

Explain the terms ‘fully qualified domain name’ (FQDN), ‘domain name’ and ‘IP address’.

Describe how domain names are organised.

Understand the purpose and function of the domain service and its reliance on the Domain

Name Server (DNS) system.

Explain the service provided by Internet registries and why they are needed.

4.9.3.2 Internet security:

Understand how a firewall works (packet filtering, proxy server, stateful inspection).

Explain symmetric and asymmetric (private/public key) encryption and key exchange.

Explain how digital certificates and digital signatures are obtained and used.

Discuss worms, trojans and viruses, and the vulnerabilities that they exploit.

Discuss how improved code quality, monitoring and protection can be used to address worms, trojans and viruses.

The structure of the Internet

The Internet is defined as a network of interconnected computer networks which uses an end-to-end communication protocol.

The Internet is mostly a wired network, with cables that pass under oceans to connect different continents.

Internet service providers

An internet service provider (or ISP) is a company that provides its customers with access to the Internet. The largest internet service providers are national companies and are referred to as national internet service providers.

National internet service providers provide internet access to smaller regional and local ISPs, from whom homes and businesses can buy access to the Internet.

Synoptic Link

The protocol used by the Internet is TCP/IP.

TCP/IP is covered in much more detail in the notes for **the transmission control protocol / internet protocol.**

Protocol

A set of rules that allows for communication between devices.

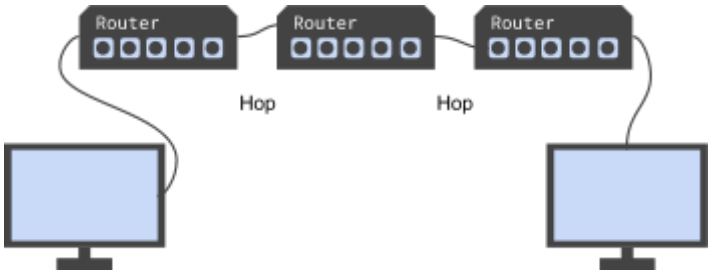
Packet switching and routers

A packet is a container in which data is transmitted over networks. You can think of a packet as an envelope in the postal system, they're labelled with addresses for their sender and recipient and contain information intended for the recipient.

Sender: 56.133.21.19	Sender: 56.133.21.19
Recipient: 158.66.12.3	Recipient: 158.66.12.3
Hello,	world!
TTL: 6	TTL: 6
1 of 2	2 of 2

A packet switched network is one in which data is sent in packets. One message is frequently split into multiple packets, each of which is sent to its recipient via the best possible route before being reassembled with other packets by its recipient.

When a packet is sent through a network, it usually has to pass through a number of routers before reaching its destination. A router uses the recipient address on a packet to determine where to send the packet. Every time that a packet passes through a router, a hop is said to occur.



Each packet can only pass through a finite number of hops. A packet's time to live (or TTL) is a number that indicates how many hops the packet can partake in and is reduced by one with each hop. When a packet's TTL expires, the packet is said to be dropped, meaning that the packet is deleted. The recipient will notice a missing packet and request that the sender transmits the missing packet again.

The primary components of a packet

Component of packet	Purpose
Sender's address	Identifies where the packet was sent from, and therefore where the response should be sent to.
Receiver's address	Identifies the packet's intended recipient, allowing it to be routed to the correct device.
Packet contents	Where the packet holds the data that is being transferred.
Time to live (TTL)	Holds the number of hops a packet can go through before being dropped.
Sequence number	Contains the number of packets in a message and identifies a packet's position in relation to others. This allows packets to be reassembled in the correct order and allows missing packets to be identified.

Routers and gateways

Two types of network device, routers and gateways, both perform essentially the same job. They connect different networks, allowing packets to reach their destination.

Routers send packets to their recipient via the fastest possible route. This might be the route that involves the lowest number of hops or the route that is least congested at the time. Routers hold tables with information relating to the fastest routes to certain devices which they frequently update so as to enable maximum performance.

Synoptic Link

Routers can use **Dijkstra's algorithm** to find the shortest route between two devices on a network.

Dijkstra's algorithm is covered under **optimisation algorithms** in **fundamentals of algorithms**.

Where two networks use different protocols, packets must be modified so as to conform to both protocols. This is where gateways come in, they strip away most of the packet's details, leaving just the packet's contents. The gateway will then give packets new sender and receiver addresses which comply to the new protocol.

Uniform resource locators

A uniform resource locator (or URL) is an address assigned to files on the Internet. Different protocols can be used in URLs to access different types of files in different ways.

`https://www.bbc.co.uk/news/technology/index.html`

Take for example the URL above. Each part of URL is broken down in the table below.

Part of URL	Purpose
`https://`	The protocol being used to access a file. HTTPS stands for hypertext transfer protocol secure.
`www`	Subdomain for world wide web. This will usually point to the web server hosted at the following domain. Other subdomains can be used to point to specific directories, for example: `news.bbc.co.uk`
`bbc.co.uk`	Domain. BBC is the name of the organisation. `.uk` is a top-level domain (TLD) and `.co` is a second-level domain (2LD)
`/news`	Directory of the file being requested.
`/technology`	Subdirectory of the file being requested.
`/index`	Name of the file being requested
`.html`	The file's extension. Hypertext markup language (HTML) is frequently used for creating web pages.

There is a wide variety of top level domains (TLDs) available for use. While `.com` is the most frequently used, TLDs like `.org` and `.net` are also common.

Domain names

A domain name identifies an organisation or individual on the Internet. They use alphanumeric characters which make them easy for humans to remember.

Fully qualified domain names

A fully qualified domain name (or FQDN) is a domain that specifies an exact resource and can be interpreted in only one way. An FQDN will always include the server's host name.

`https://bbc.co.uk/news/index.html` FQDN ✗

`https://www.bbc.co.uk/news/index.html` FQDN ✓

IP Addresses

An internet protocol address (IP address) is assigned to every computer on the Internet and every device that communicates on a network.

IP addresses are not easy for humans to remember, which is why domain names are used. Domain names map to IP addresses, meaning they are essentially a human-friendly representation of an IP address.

Synoptic Link

IP addresses are covered in more detail in notes for **The Transmission Control Protocol / Internet Protocol (TCP/IP).**

The domain name server (DNS) system

As mentioned previously, each domain name has a direct relationship with an IP address. When you enter a domain name into your browser's address bar, a domain name server is used to translate the domain name into its corresponding IP address.

A domain name server stores a table of domain names and their corresponding IP address. If a domain name server does not have a record of the domain that you are trying to access, your request will be passed to another domain name server. Some very small and rarely visited websites will require numerous changes of servers before a record can be found. This makes those websites slightly slower to access than large websites.

DNS Table	
Domain Name	**IP address**
turing.me	41.47.142.208
lovelace.com	16.57.142.88
babbage.net	84.88.49.3

Internet registries

An internet registry is an organisation responsible for the allocation of IP addresses. There are only five internet registries in operation, each serving a different geographical area.

An important part of an internet registry's work is to protect the world's depleting pool of unallocated IP addresses. When a new IP address is requested, an internet registry will first look for a previously allocated IP address that has become unused rather than allocate a brand new IP address straight away.

Synoptic Link

IP addresses and the decline in unassigned addresses are covered in more detail in notes for **The Transmission Control Protocol / Internet Protocol (TCP/IP).**

Internet security

Firewalls

A firewall sits between a device and the Internet and regulates the packets that pass through it. Firewalls can be either software or hardware and work as a proxy server which can perform both packet filtering and stateful inspection.

Packet filtering

Firewalls use packet filtering to accept and block packets based on their source IP address or the protocol that they are using (determined by their port number). A network's administrator can specify particular IP addresses or protocols to block or use automatic filtering software that can block suspicious packets.

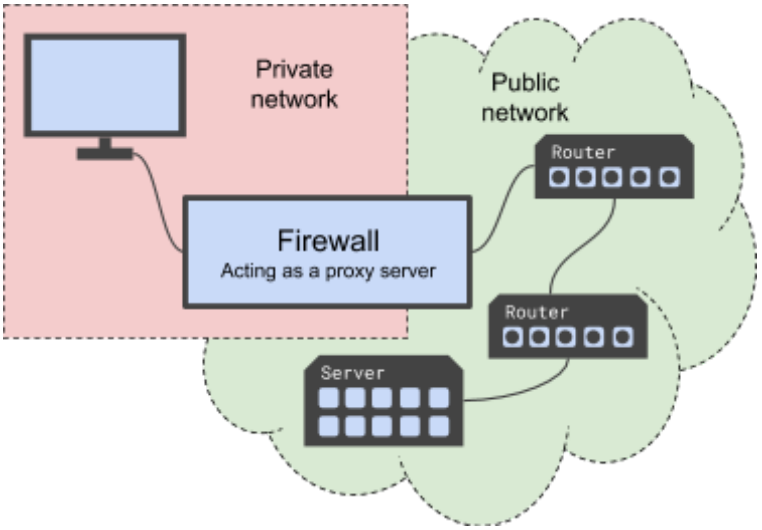
Stateful inspection

Stateful inspection actually examines the contents of a packet before deciding whether to allow it through the firewall. Some firewalls keep a record of current connections in a network, allowing them to filter out packets that aren't related to activity on the network.

Proxy server

A server that sits between a public network and a private network is called a proxy server. These devices manage every packet that passes between the two networks. Firewalls can be said to act as proxy servers when they control the movement of packets between public and private networks.

When a device in a private network sends a packet through a firewall and into a public network, the packet's "sender" address is that of the firewall, rather than the device's private IP address. This provides some degree of anonymity to devices on private networks as their private address is never sent beyond the private network.



Symmetric and asymmetric encryption and key exchange

When information needs to be transmitted securely over a network, encryption is used.

Symmetric encryption

In symmetric encryption, both the sender and receiver share the same private key. This key is used to both encrypt and decrypt data sent between the two parties.

Before sending any information, the sender and receiver must participate in a key exchange to ensure that they both have a copy of their shared key. If the key is exchanged over a network, it is vulnerable to interception. This is a major flaw in symmetric encryption that asymmetric encryption overcomes.

Synoptic Link

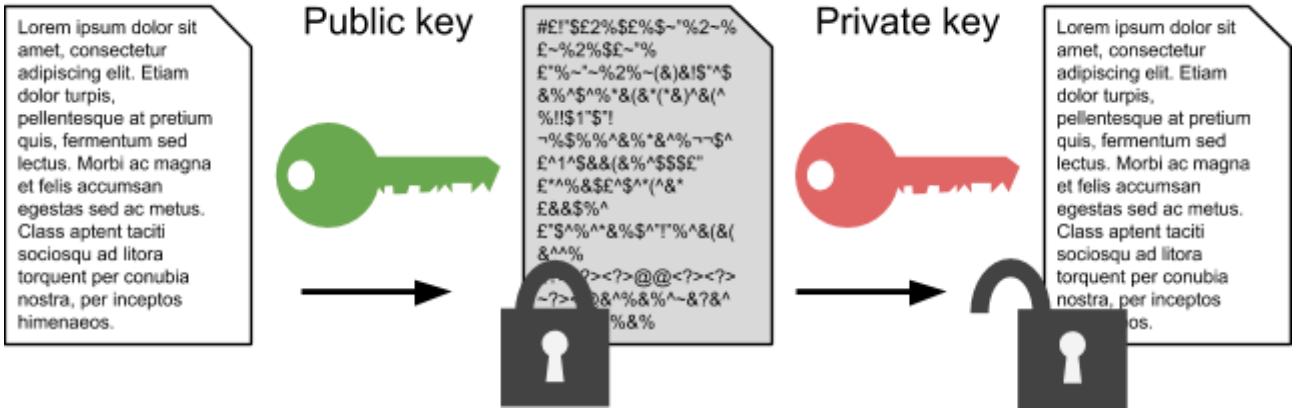
Encryption is the process of scrambling data so that it cannot be understood if intercepted.

Encryption is covered in the notes for **representing images, sound and other data** under **fundamentals of data representation**.

Asymmetric encryption

When two devices communicate using asymmetric encryption, four different keys are used. Each device has a pair of mathematically related keys, one of which is kept secret (the private key) and the other shared on the Internet (the public key).

When a message is encrypted with a public key, only the corresponding private key can decrypt it and vice versa.



Before a message is sent, it is encrypted by the sender using the recipient's public key. This means that the message can only be decrypted by the corresponding private key (as explained earlier), the recipient's private key, which only the recipient has access to. This means that the recipient is the only person who can decrypt the message.

Digital certificates & digital signatures

Synoptic Link

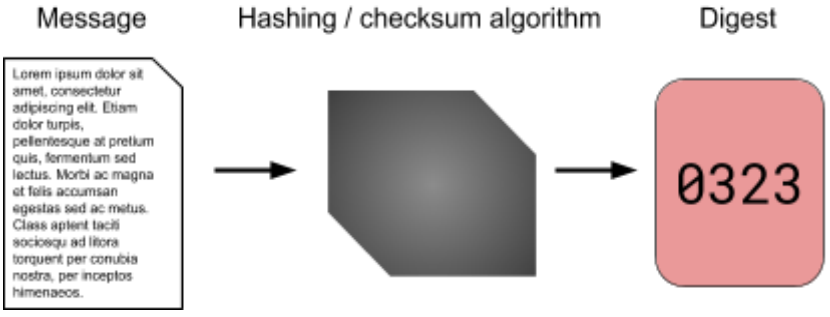
Hashing and **checksums** are both ways of producing a value from a piece of data.

They are covered in more detail in the notes for **information coding systems** under **fundamentals of data representation**.

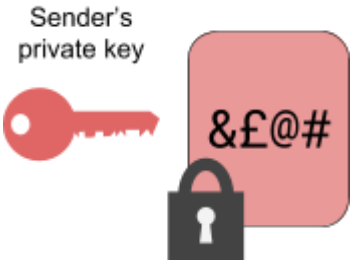
Digital signatures

When using asymmetric encryption, a digital signature can be used to verify the sender of a message and to verify that a message has not been tampered with during transmission. The process involves a number of stages, which are detailed below.

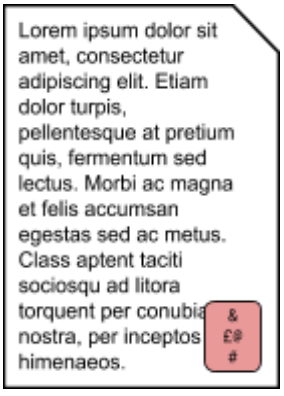
1. A digest of the message is created, perhaps by a hashing or checksum algorithm. The value of the digest depends on the content of the message and will not be the same if the message is changed.



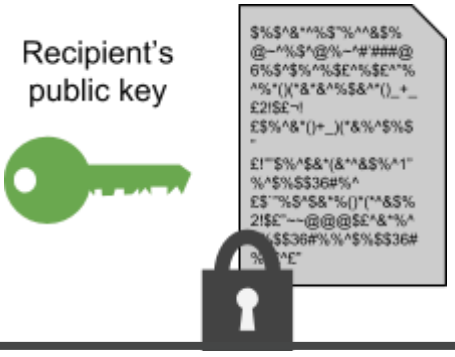
2. This digest is encrypted with the sender's private key (which anyone can decrypt with the sender's public key)



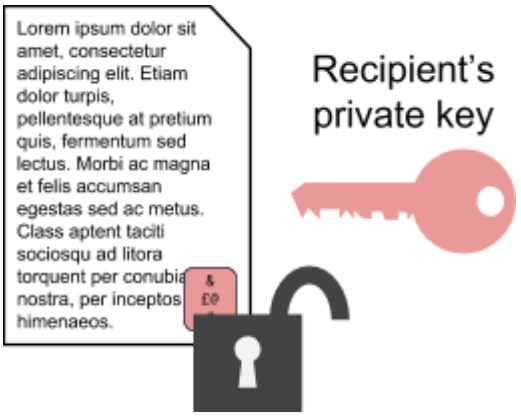
3. The encrypted digest is appended to the message



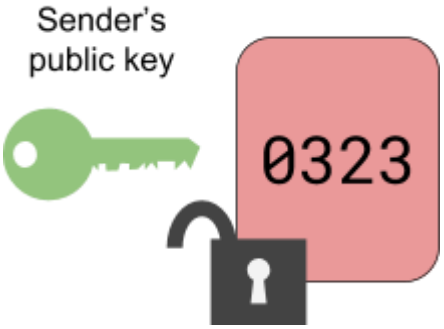
4. The message and appended digest are encrypted with the recipient's public key, meaning that only the recipient can decrypt the information.



When the recipient receives the message, they first decrypt it using their private key; leaving them with the decrypted message and an encrypted digest, as shown in the image on the right.



As the digest was encrypted using the sender's private key in stage 2, it can be decrypted using the sender's public key as shown in the image below. This verifies that the message was really sent by the sender as only they have access to their private key.



The recipient then carries out the same hashing or checksum algorithm on the message and checks whether their result matches the now decrypted digest. If everything matches, the recipient can be certain that the message was sent by the sender and hasn't been tampered with or corrupted during transmission.

Digital certificates

A digital certificate verifies ownership of a key pair used in asymmetric encryption and can be used to check that a fake key pair isn't being used by an imposter. Issued by certificate authorities, these files contain: a serial number, the owner's name, an expiry date, the owner's public key and the certificate authority's digital signature.

Worms, trojans and viruses

Worms, trojans and viruses are all types of malware that can infect computers.

Worms

Worms are pieces of malicious software that can self-replicate between computers, either within a network or by users downloading and running a malicious file.

Trojans

A Trojan is a type of malware that is disguised as a benign file that users can be tricked into opening. These are often spread as email attachments or downloaded from malicious websites.

Viruses

If you've studied viruses in Biology, you may know that viruses require a host cell. This is the same with computer viruses, which require a host file in which to reside. These files are typically executable files, meaning that viruses can lie dormant in a computer until their host file is opened or run.

Viruses can spread between computers over a private network, the Internet or even through the use of physical media like hard drives, flash drives and optical disks.

Preventing malware

Although it is difficult to avoid malware completely, there are a number of precautions that can be taken in order to protect computers from malicious software.

Malware often exploit bugs in code that enable them to take hold of a computer system. Good code quality is an important factor in preventing malware and small oversights by developers can have devastating consequences. Other vulnerabilities that malware exploit include a lack of antivirus software, out-of-date software and poor security.

One general rule for preventing malware is to install antivirus software. Antivirus programs are specialist pieces of software that scan the files on a computer and remove any suspicious files. Many modern operating systems come with some level of antivirus installed as a default.

In organisations, employees can be trained about the risks of opening suspicious email attachments in order to reduce the risk posed by malware.