Network Characteristics and Protocols

- A single computer not connected to anything else is known as a standalone machine.
- Once you connect one computing device to another, either via cable or wireless, you have a network.

Advantages and Disadvantages of Networks

Advantages

- Users can share files.
- Users can share peripherals and connections to other networks such as the internet.
- Users can access files from any computer on the network.
- Servers can control security, software updates and backup of data.
- Communication with other people - e.g. email and social networking.

Disadvantages

- Increased security risks to data.
- Malware and viruses spread very easily between computers.
- If a server fails, the computers connected to it may not work.
- Computers may run slower if there is a lot of data travelling on the network.

The Need for Standards

- Standards are a set of hardware and software specifications that allow manufacturers to create products and services that are compatible with each other.
- Without most standards, most devices wouldn't be able to interact or communicate with one another.

What is a Protocol

- A key way of ensuring technology-based standards are established and adhered to is to use protocols.
- Much like spoken languages, if two devices have different protocols, they cannot communicate.
- However, if two devices share the same protocol, they can exchange information - in other words, they speak the same language.

Common Protocols

Protocol Acronym	Protocol Full Name	Area / Purpose
TCP	Transmission Control Protocol	Communication over LAN / WAN
IP	Internet Protocol	Communication over LAN / WAN
UDP	User Datagram Protocol	Communication over LAN / WAN
HTTP	Hypertext Transfer Protocol	Web page requests
HTTPS	Hypertext Transfer Protocol Secure	Web page requests
FTP	File Transfer Protocol	File transfers
POP	Post Office Protocol	Email
IMAP	Internet Message Access Protocol	Email
SMTP	Simple Mail Transfer Protocol	Email

TCP/IP

- The Transmission Control Protocol (TCP) provides error-free transmission between two routers.
- The Internet Protocol (IP) routes packets across a wide area network (WAN).
- Together, they make up the TCP/IP protocol stack, the foundation of communication over the internet.

UDP

- The User Datagram Protocol (UDP) uses a simple, connectionless transmission model.
- It is an alternative to TCP but has no error checking.
- It is used to send short messages using datagrams, where speed is more important than accuracy.
- It maintains an open, two-way connection ideal for online gaming.
- It is largely obsolete now, as it is significantly less reliable than TCP.

HTTP/HTTPS

- The Hypertext Transfer Protocol (HTTP) is a way for a client and server to send and receive requests and deliver HTML web pages.
- It is the fundamental protocol of the World Wide Web.
- The Hypertext Transfer Protocol Secure (HTTPS) is effectively the same as HTTP, except it adds in encryption and authentication.
- HTTPS should be used whenever a website deals with sensitive information such as passwords or bank account details.

FTP

- The File Transmission Protocol (FTP) is a protocol used for sending files between computers, normally on a wide are network (WAN).
- Often people use FTP clients - software that sits on top of the actual FTP protocol.
- When you interact with the program, the client generates and sends the appropriate FTP commands.

POP/IMAP/SMTP

- Three popular protocols are used in conjunction with mail servers to deal with email.
- Mail servers can be thought of as a virtual post office handling incoming and outgoing email.
- Simple Mail Transfer Protocol (SMTP) transfers outgoing emails between servers and from email clients to servers.
- Post Office Protocol (POP) retrieves emails from a mail server and transfers them to your device, removing them from the server in the process.
- Internet Message Access Protocol (IMAP) keeps email on the mail server, maintaining synchronicity between devices.

TCP IP, DNS and Protocol Layers

The Internet

- A router is connected to an Internet Service Provider (ISP), typically via a telephone connection or fibre optic cable.
- The ISP is connected to a Domain Name Service (DNS) and other routers that make up the backbone of the internet.
- Those routers are also connected to:
  - Devices on their own LANs
  - Other routers on the WAN
  - Servers

The Concept of Layers

- The concept of layering is to divide the complex task of networking into smaller, simpler tasks that work in tandem with each other.
- The hardware and / or software for each layer has a defined responsibility, and each one provides a service to the layer above it.
- The advantages of layering include:
  - Reducing the complex problem into smaller sub-problems.
  - Devices can be manufactured to operate at a particular layer.
  - Products form different vendors will work together.

TCP/IP Protocols and the use of Layers.

- TCP/IP is one of the most important protocol stacks in use today.
- The TCP/IP stack refers to a set of networking protocols, consisting of four layers working together.
- All incoming and outgoing data packets pass up and down through the various layers.

TCP/IP Layers

- The original TCP/IP model had four layers. The updated model has five, breaking the link layer in two - data link and physical.
- There were very few physical connection options when TCP/IP was originally conceived. Now, we have twisted pair, Wi-Fi, fibre optic etc.
- Originally, it didn’t make sense to split physical connections from data delivery. Today, it makes more sense to do so.
- As the four-layer model is perfectly sufficient for A level, this is the one that is used for the specification.

The Four Layer TCP/IP Stack

Layer	Description	Examples
Application Layer	Network applications such as web browsers or email programs operate at this layer	FTP, HTTP, HTTPS, SMTP, IMAP
Transport Layer	Set up communication between two hosts - they agree settings such as language and packet size	TCP, UDP
Internet Layer	Addresses and packages for data transmission. Routes packets across the network	IP
Link Layer	Network hardware and connection port standards. Operating system device drivers also sit here. Facilitates the transmission of binary via any media.	Copper, Twisted Pair

TCP/IP - Application Layer

- The application layer uses an appropriate protocol relating to whatever application is being used to transmit data.
- If a message was being sent via a web browser, then the list appropriate protocols would include HTTP, HTTPS, FTP etc

TCP/IP - Transport Layer

- The transport layer uses the TCP part of the stack as well as other conversation protocols like UDP.
- It is responsible for establishing an end-to-end connection and maintaining conversations between application processors.
- These protocols use port numbers to track sessions and add this information to the header.
- Once the connection is made, the transport layer splits the data into packets. It adds to each packet:
  - Its number / sequence.
  - The total number of packets.
  - The port number that the packet should use.
- Packets are numbered so they can be reassembled in the correct order.

TCP/IP - Internet Layer

- The internet layer uses the IP part of the stack.
- It adds to each packet:

    - The source IP address
    - Destination IP address
- All routers operate at this layer. They use the IP address to find out where the packets are heading. We now have what is known as a socket:
    - socket = IP address + port
    - E.g. 127.56.87.2:80
- We now know:
    - The device the packet is being sent to (IP address).
    - The application on that device that needs the packet (port).

TCP/IP - Link Layer

- The link layer represents the actual physical connection between network devices.
- It is responsible for adding the unique Media Access Control (MAC) address of the:
    - Source device
    - Destination device
- When transmitting data between routers over a wide area network (WAN), the MAC address is changed at each hop on the route.

MAC vs IP Address

- Communication between two devices on the same local area network (LAN) only requires the link layer, which creates a frame using MAC addresses.
- Communication between two devices on different networks (WAN) requires both the network layer, which uses IP addresses to create a packet, and the link layer frame.
- In practice, communications via modern LANs also makes use of IP addresses, treating the local network as though it were a wide network.
- Every physical device should have a unique MAC address. However, for a router, storing references to every MAC address in existence would be unmanageable - it would take too long to find a particular address to decide which connection to route traffic down.
- Therefore, switches learn and store MAC addresses for connected LAN devices only, while routers cache some IP addresses.
- The MAC address tells who you are and the IP address tells where you are.
- MAC addresses need to be unique.

The WWW and Domain Name System (DNS)

- The internet offers many services - one of the most popular is the World Wide Web.
- The WWW is a collection of files with information stored in hypertext (known as web pages), and other associated files, hosted on web servers.
- These web servers host (store) the files and handle client requests - for example, a HTTP GET request for a page or resource.
- A web page, stored as text (HTML, CSS, JAVASCRIPT), is sent to a web browser application, which uses rules to render it correctly.

Domain Name System

1. A user requests a URL via a browser. e.g. google.com
2. The browser sends the domain name to a Domain Name System (DNS).
3. The DNS maps the domain name to an IP address and returns it to the browser.
4. A GET request for the web page or resource is sent to the web server using the IP address.
5. The requested web page or resource is returned to the client's web browser.

Example

1. First, the human-readable URL is received by a DNS resolver server. e.g. google.com
2. The server then queries a DNS root name server.
3. The root server responds with the address of the top-level domain server (TLD) e.g. for .com
4. The resolver makes a request to the .com TLD server.
5. The TLD server then responds with the IP address of the domain name's server, e.g. google.com
6. The recursive resolver sends a query to the domain's name server.
7. The IP address of google.com (8.8.8.8) is then returned to the resolver from the name server.
8. Finally, the DNS resolver responds to the web browser with the IP address of google.com.

LANs and WANs

Local Area Network (LAN)

- Any network that covers a small geographical area, typically located on a single site.
- All the hardware for a LAN tends to be owned and controlled by the organisation using it.
- LANs are typically connected using:
  - UTP cable
  - Fibre Optic
  - Wi-Fi

Wide Area Network (WANs)

- A wide area network (WAN) is any network that covers a large geographical area.
- When multiple LANs physically located in different areas are connected, they form a WAN.
- The infrastructure that connects LANs to form a WAN is leased from telecommunication companies who own and manage it.
- WANs are typically connected by:
  - Telephone lines
  - Fibre Optic Cables
  - Satellite Links

Client Server & Peer to Peer

Client-Server Model

- With a client-server network model, a server:
  - Controls access and security for a shared file store.
  - Manages access to the internet.
  - Manages printing jobs.
  - Provides email services.
  - Runs regular backups of data.
- A client makes requests to the server for data, connections and other services.

Advantages

- Easier to manage file security.
- Easier to back up shared data.
- Easier to install software updates to all computers.

Disadvantages

- Can be expensive to set up and maintain.
- Requires IT specialists to maintain.
- The server is a single point of failure.
- Users will lose network access if the server fails.

Client-server networks are best suited to organisations with a large number of computers or situations where many computers need access to the same information.

Peer-to-Peer Model

- A Peer is a computer connected to the network.
- A peer is equal to all other peers.
- Peers serve their own files to each other.
- Each peer is responsible for its own security and data backup.
- Peers usually have their own printers.
- Peers can send print jobs to another peer to process, but that peer must be switched on to communicate with the printer.

Advantages

- Very easy to maintain.
- Specialist staff are not required.
- No dependency on a single computer.
- Cheaper to set up.
- No expensive hardware.

Disadvantages

- The network is less secure.
- Users will need to manage their own backups.
- Can be difficult to maintain a well-ordered file store.

Peer-to-peer networks are best suited to smaller organisations with fewer computers or where fewer users need access to the same data.

Packet and Circuit Switching

Circuit Switching

- Circuit Switching provides the basis for traditional telephone networks.
- Circuit switching creates a temporary and dedicated link of fixed bandwidth between the source and destination that only lasts until the transmission is complete.
- Using this method guarantees the quality of the transmission through dedicated bandwidth, making circuit switching excellent for data that needs a constant link end-end such as real-time video.
- The downside is that a lot of the potential bandwidth can be wasted.

Packet Switching

- Packet switching breaks streams of data into smaller blocks, each sent independently of one another.
- At each node, packets are sent via whichever route the node decides is the least congested - this maximises bandwidth but doesn't guarantee the quality of the transmission.
- It also means that packets can take different routes to their destination and may arrive out of order.
- Packet switching is more affordable and efficient than circuit switching as all bandwidth can be used at once. It also requires less complex infrastructure and can easily respond if parts of the network fail.

Circuit Switching vs Packet Switching

Circuit Switching

- Physical path between source and destination.
- All packets use the same path.
- Reserves bandwidth in advance.
- Can cause a lot of bandwidth wastage.
- No store and forward transmission.

Packet Switching

- No physical path.
- Packets travel independently.
- Does not reserve bandwidth.
- No bandwidth wastage.
- Supports store and forward transmission.

Network Security Threats

The Need for Network Security

- When we connect computing devices together via a network, allow them to share data and subsequently connect them to the internet, the become vulnerable to attack or misuse.
- All networks need protection from unwanted intrusion and hacking.
- The aim of network security is to:
  - Only allow authorised users to access what they need.
  - Prevent unauthorised access.
  - Minimise the potential damage caused by unauthorised access.

Hackers

- A hacker is a person who attempts to gain access to a computer system with the intent of damaging data or somehow harming that system.
- Hackers come in many forms - some not malicious.

Black-hat

- These are your traditional hackers, as often portrayed in TV and films.
- They attempt to gain access via nefarious means, typically to steal company secrets or cause damage.

White-hat

- These are security experts (often ex-hackers) employed by a company.
- Also known as ethical hackers, they use their expertise to try and find vulnerabilities and fix them.

Grey-hat

- These hackers sit somewhere between the other two.
- They are not employed by a company, but they still attempt to locate flaws in company-wide computer systems as a hobby.
- What they do it technically illegal, but they then inform the company of the flaw so the company can fix it.

Malware

- Malware is an umbrella term that covers any computer code written with the intent to frustrate or harm.
- Malware can have a wide range of effects depending on its type. These typically include:
  - Deleting, corrupting or encrypting files.
  - Causing computers to crash, reboot or slow down.
  - Reducing network speeds.
  - Logging keyboard inputs and sending them to hackers.

Virus

- The computer virus is probably the most well-known and popularised form of malware.
- Viruses are pieces of code capable of copying themselves and spreading throughout a system.
- They are typically designed to have a detrimental effect like corrupting a file system or destroying data.

Spyware

- Spyware is a form of malware that covertly obtains information about a user's computer activities by transmitting data from their device.
- It can be used in a variety of ways to harvest all sorts of sensitive and personal data from a device:
  - Internet surfing habits
  - Email addresses
  - Visited web pages
  - Downloads / downloading habits
  - Passwords
  - Credit card numbers
  - Keystrokes
  - Cookies

Denial-of-Service Attack

- A denial-of-service (DoS) attack is when an attacker floods a server with useless traffic, causing the server to become overloaded.
- Many past DoS attacks exploited limitations of the TCP/IP stack.
- DoS attacks often target web servers of high-profile organisations such as banks, large scale e-commerce businesses and the

government.

- Though DoS attacks do not typically result in the theft or loss of data or other assets, they can cost a great deal of time and money to handle.
- An additional type of DoS attack is the **Distributed Denial of Service (DDoS)** attack.
- A DDoS attack occurs when multiple systems orchestrate a synchronised DoS attack against a single target.
- There are several ways this can be achieved, which go beyond the specification. Essentially, instead of being attacked from one location, the target is attacked from many locations at once.
- In this example, several devices have unknowingly become infected and are not acting as so-called zombies, carrying out independent DoS attacks - often without the device owner's knowledge.

SQL Injection

- SQL injection is a code injection technique used to attack data-driven applications.
- This kind of attack is designed to exploit vulnerabilities in poorly coded database applications.
- Code is entered into input text boxes and is then executed by the server.

Social Engineering

- Social engineering is an umbrella term covering several different manipulation techniques that exploit human error, with a view to obtaining private information, access to restricted system or money.
- Social engineering scams lure users into exposing data, spreading malware or providing access to a system. These scams include:
  - Baiting
  - Scareware
  - Pretexting / blagging
  - Phishing
  - Pharming
  - Shoulder-surfing
  - Quid pro quo
  - Vishing

Phishing

- Phishing is an online fraud technique used by cybercriminals to trick users into giving out personal information:
  - Usernames
  - Passwords
  - Credit card details
- Perpetrators disguise themselves as a trustworthy source in an electronic communication such as an email or a fake website.

Pharming

- Pharming is another online fraud technique used by cybercriminals.
- Malicious code installed on a PC or server misdirects users to fraudulent websites without their knowledge.
- Pharming is sometimes referred to as "Phishing without a lure."

Preventing and Minimising Threats

Firewalls

- A firewall is a piece of hardware or software configured to let only certain types of traffic through it.
- It can be set up to prevent:
  - Unauthorised internet traffic from outside a LAN.
  - Users in LAN from a accessing parts of the internet prohibited by the company they work for.
- A firewall can block certain ports and types of traffic. It can also inspect data travelling across it to see if it looks suspicious.
- Operating systems and home routers come with built-in firewalls. More sophisticated ones are also available for purchase separately.

Secure Passwords

- A common method of preventing unauthorised access to a system is requiring users to log in with predetermined credentials such as a username and password.
- These systems can often be made more secure by implementing password rules.
- However, it is well-documented that enforcing overly strict rules often causes users to write their password down so they do not forget it - defeating the object entirely.

Up-to-Date and Anti-malware Software

- A common method of preventing and minimising threats is to install anti-malware software.
- These applications often come pre-installed with your operating system, but you can purchase others from dedicated vendors.

- It is also crucial to ensure software - especially operating system and anti-malware software - is always updated with the latest patches.
- When software is first released, it is often far from perfect. Early builds can contain bugs or flaws, which can then be exploited by malware.
- As flaws are discovered, companies develop fixes, known as patches.

Proxies

- A proxy is simply a physical device placed between a network and a remote source.
- All traffic travels through the device on its way in and out of a network.
- For example, a proxy could handle web page requests. If the page is not on the proxy's banned list, it will pass on the request.
- It can also look at the page and choose whether to pass it back to the user or not.
- A proxy ensures there is no direct, physical connection between a single user and a remote source.

Encryption

- Encryption is the process of turning plain text into an unreadable form.
- The plain text is encrypted using an algorithm and a unique key.
- Only someone with the appropriate key will be able to translate the cyphertext back into readable form.
- Encryption won't stop you from being hacked, but it will make any stolen data very difficult to read.

Search Engine Indexing

- To make the job of finding information on the web easier, we use search engines.
- When you submit a search via a search engine, it isn't actually searching every single available web page on the internet.
- Instead, the search engine runs your request against its index.

- Indexing is the process of a search engine collecting, sorting and storing data in its index.
- Therefore, the index is the place where all the data the search engine has gathered is located.
- When you see search results appear on your screen, it is the search engine index that provides these.
- Searching the index is very fast. However, the index must be constantly updated to ensure that:
    - New sites/pages are added.
    - Old sites/pages are removed.
    - Broken links are updated.
- Search engines use programs known as spiders or crawlers that travel the world wide web.
- They index any pages, content and metadata they find and map links between pages by following all:
    - Internal links
    - External links
- In doing so, they continuously add to and update the index.

Server and Client-side Processing

- When we use the internet, the majority of our interactions involve two types of connected entities:
    - The client or user.
    - The server holding the web page or resource we want.
        - The client requests the page.
        - The client’s request is received by the web server.
        - The web server returns the page to the client’s browser.

Client-side Processing

- Initial data validation (JavaScript)
- Manipulates user interface elements
- Applies website styles (CSS)
- Reduces load on the server
- Reduces the amount of web traffic

Server-side Processing

- Provides further validation
- Queries and updates the server database
- Encodes data into readable HTML
- Keeps organisation data secure
- Performs complex calculations

Lossy vs Lossless

- The purpose of compression is to:
    - Reduce the size of files.
    - Reduce download times.
    - Reduce storage requirements.
    - Make best use of bandwidth.
- Given the vast amounts of data sent and streamed over the internet every day, making efficient use of bandwidth can be critical.
- By reducing a file's size as much as possible, you can considerably speed up the time it takes to transmit.

Methods of Compression

- Compression achieves its goals by reducing the overall size of a file as much as possible.
- There are two different methods of compression:
    - Lossy
    - Lossless
- Both reduce the overall size of files but in quite different ways.
- When a compressed file arrives at its destination, it needs to be uncompressed so it can be read.

Lossy Compression

- With images, audio and video, a small reduction in quality is not very noticeable.
- Therefore, lossy compression is considered an acceptable compromise of quality vs file size - and download time.

Lossless Compression

- Another approach doesn't sacrifice any quality during compression - known as lossless compression.
- With lossless compression, we can reduce the size of an image file, but we are able to store the image in its full, original quality when it is uncompressed.
- However, this method is only effective on images with large areas of continuous colours.
- Therefore, lossless compression is ideal for vector-style images such as logos, cartoons and icons but far less so for full-colour photographs where there are very few blocks of continuous, repeating colours.

Suitability of Lossy and Lossless Compression

- File type often determines which method of compression is best - some files are simply not suitable for lossy compression.
- With text documents and executable program, we must not lose any of the data during compression.
- For these files, we must use lossless compression, as we need to be able to restore the file in its entirety.

Lossy vs Lossless

Compression

- Reduces the size of a file.
- Makes files quicker to transfer.
- Files take up less space in storage.

Lossy Compression

- Some data is lost when the file is compressed.
- Slightly reduces quality but significantly reduces file size.
- Suitable for image, audio and video.

Lossless Compression

- None of the original data is lost.
- The original file can be recreated when it is uncompressed.
- Suitable for executable files and documents.

Run-length Encoding and Dictionary Coding

- Dictionary coding is ideal for the compression of text-based documents.
- Run-length encoding is more suited to the compression of images.
- Being lossless compression techniques, both utilise a method of encoding data that allows us to recreate the file in its original quality.

Dictionary coding

- Dictionary encoding works by building an index that we visualise as a table.
- Every data item or entry in the file is recorded, along with an indexed reference or unique code.
- The compressed file now consists of:
    - The dictionary index.
    - The sequence of occurrences needed to recreate the original file.

Run-Length Encoding

- While dictionary coding is great for compression of text-based files, run-length encoding is ideal for compressing bitmap images.
- Bitmap images are made up of discrete pixels.
- To recreate this image faithfully using lossless compression, we need to be able to recreate every pixel.

PageRank Algorithm

- PageRank is a trademark algorithm developed by one of Google's founders, Larry Page.
- It is used to help compile and rank website pages and the list of results returned by a search engine.
- It is not the only algorithm used by Google or other search engines to rank search results, but it is the first, the most famous and the one you need to know for the exam.
- At a high level, it works by checking the number and quality of links to a page in order to determine roughly how important that page is.
- The assumption is that websites of greater importance are more likely to be linked to from other websites.

PR(A) = (1-d) + d (PR(T1)/C(T1) + … + PR(Tn) /C(Tn))

- **PR(A)** is the PageRank of page A.
- **C(Tn)** is the total count of outbound links from web page n, including the inbound link to page A.
- Each web page has a notational vote of 1, shared between all the web pages it links to.
- **PR(Tn)/C(Tn)** is the share of the vote page A gets from pages T1 through Tn.
- Each of these vote fractions are added together and multiplied by d.
- d is the damping factor that prevents **PR(Tn)/C(Tn)** from having too much influence.
- It is notionally set to 0.85, which equates to roughly six clickthrough links. The average user will then end their browsing session or enter a new web address rather than following another link.

Applying the PageRank Algorithm

- A web page's PageRank is determined in part by the PageRank of other pages linking to it.
- The algorithm works without needing to know the PageRank of any back-linked pages.
- In the first instance, the algorithm makes and informed guess. After several further iterations, the algorithm begins to home in on the correct PageRank.
- The number of iterations required for the final PageRank number to stop moving can be hundreds - if not millions.
- Once the final PageRank is achieved, the average PageRank of all pages will be 1.