This document is part of an ongoing educational journey by Ben Foster and is not to be distributed without appropriate credit.

Network plus textbook

Contents

Section 1: Network Fundamentals

1.1 Network components

Understanding the core components of a network is key to both troubleshooting and building reliable systems. Each device or system plays a specific role in how data is managed, transferred and secured.

- **Clients**: this is any device, like a PC or phone, that connects to a network to request services or resources from a server
- **Servers**: these are a type of computer that provides services or resources (like files, websites or authentication) to clients over the network. They have specialised hardware layouts, software and operating systems for their roles
- **Hubs**: simple devices that forward data to all connected devices, regardless of the intended destination – now mostly obsolete
- **Switches**: these are networking devices that connect devices within a LAN and intelligently forward data only to a device it's intended for
- **WAPs (wireless access points):** this is a networking device that allows wireless devices to connect to a wired network using Wi-Fi. They plug into switches or routers and broadcast Wi-Fi signals
- **Routers**: these direct data between different networks and provide connectivity to the internet
- **Firewalls**: these control incoming and outgoing traffic based on security rules, protecting networks from unauthorized access

- **Load balancers**: these are devices that distribute incoming traffic across multiple servers to improve performance and reliability
- **Proxy servers**: these act as an intermediary between a client and the internet, often used to cache content or filter traffic
- **IDS (intrusion detection system):** this is a device which monitors network traffic for suspicious activity and alerts administrators to potential threats
- **IPS (intrusion prevention system**): this is a device which actively blocks detected threats in real time, stopping malicious traffic before it causes harm
- **Controllers:** these manage multiple wireless access points centrally, allowing easier configuration and monitoring in large networks
- **SAN (storage area network):** this is a dedicated high-speed network that provides access to consolidated storage devices
- **Media**: media refers to physical or wireless means by which network signals travel
- **NAS (network attached storage):** NAS is a dedicated file storage device connected to a network that allows multiple users and devices to access shared files over the LAN

1.2 Network Resources

In a network, resources refer to shared assets like files, printers, storage devices and applications. How these are accessed and managed depends on the network model used: either client server or peer to peer (P2P).

Client server model

In this model, resources are centralised on one or more dedicated servers. These servers manage file storage, print services, authentication, databases, and more. The clients (user devices) send requests to the server, which responds accordingly.

Benefits are:

- Centralised management and security
- Its scalable as you can easily add more clients
- Ideal for business and enterprise networks
- Better control over backups, updates and access permissions (easier management)

Drawbacks are:

- Server hardware and setup can be expensive
- If the server fails, many services may go down
- Requires regular maintenance and skilled administration
- More expensive as it requires dedicated hardware and management expertise

P2P model

In the P2P model, each device (peer) can act as both a client and server. Resources are shared directly between devices without a centralised authority.

Benefits are:

- Easy and inexpensive to set up
- No need for dedicated server hardware
- Good for small networks

Drawbacks are:

- Poor scalability
- No centralised control over resources or security, so admin and backup is difficult due to having to manage so many devices
- File sharing may lead to version conflicts or duplication

1.3 Network Geography

Network geography refers to the physical size and scope of a network – from a single room to global communications. The type of network used depends on how far data needs to travel and how many devices are involved.

PAN – Personal Area Network

This is the smallest network. It connects devices within the range of a person – typically a few meters. Example would include a connection between a phone and earbuds or smart watch.

Range: under 10 meters

Use case: Personal convenience and mobility

LAN – Local Area Network

A LAN connects devices in a single building or location, such as a home, office, or school. Examples would include computers and printers in an office sharing a router an file server.

Range: up to a few hundred meters

Use case: fast, secure communication within one organisation or home.

CAN – Campus Area Network

A CAN convers multiple buildings in close proximity, such as a university campus, hospital, or corporate complex. Examples would include several departments of a university connected under one private network.

Range: up to a few kilometres

Use case: unified control and services across a tight cluster of buildings

MAN – Metropolitan Area Network

A MAN spans a city of large town, often linking multiple LANs or CANs together. It may be owned by a company of service provider. Examples would include city government or uni buildings spread across the city sharing a common infrastructure each.

Range: Tens of kilometres

Use case: Public utility networks or city-wide infrastructure

WAN – Wide Area Network

A WAN connects devices across long distances, from city to city or even globally. The internet is the most well-known example. Other examples would include a company with offices around the world connected through leased lines or VPNs.

Range: Countrywide to Worldwide

Use case: Enabling communications across remote locations

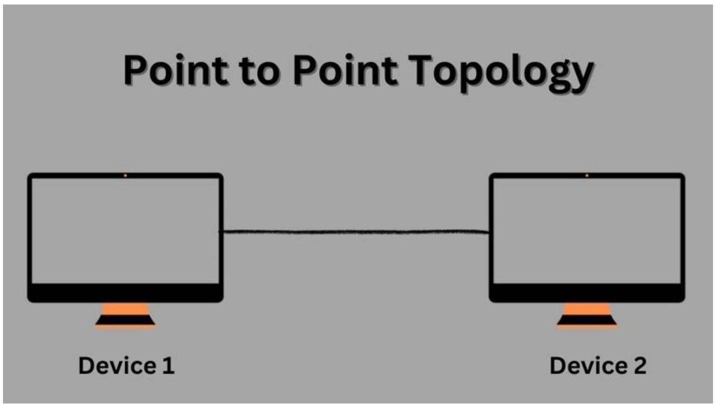
1.4 Network Topologies

A network topology describes how devices (or nodes) are physically or logically connected. Each layout affects the speed, cost, fault tolerance and performance of the network. Below are the most common topologies you'll come across.

Wired topologies

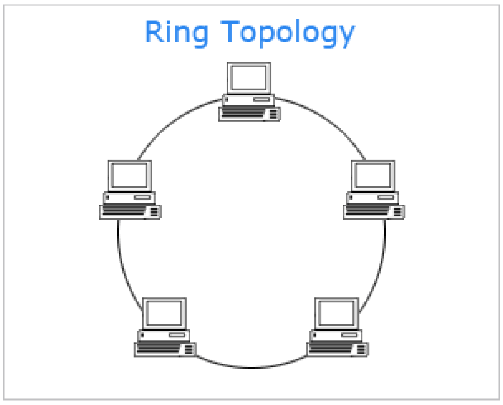
P2P (point to point)

A P2P topology connects two devices directly. Its simple and effective, often used when only two nodes need to exchange data – such as a computer linked to a printer or a dedicated server connection. This type of topology is reliable, as there's no extra hardware or paths to fail, but its not scalable, as adding more devices means rewiring the setup entirely.



Ring

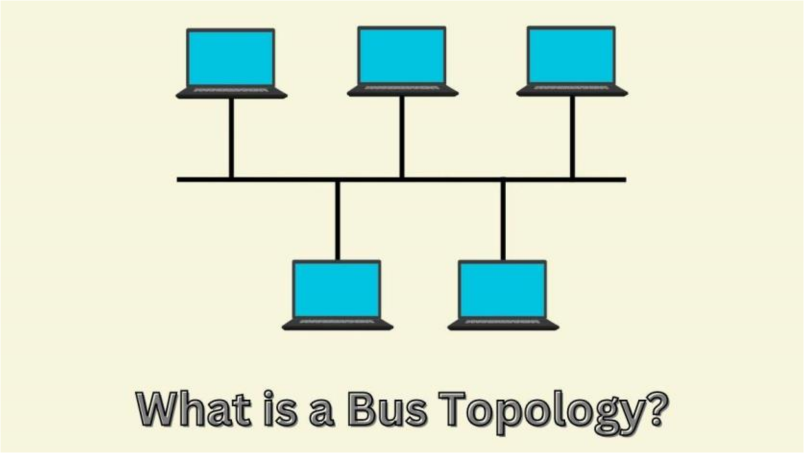
In a ring topology, each device is connected to two others, forming a closed loop. Data travels around the ring in one direction (or both, in dual ring set ups). This is orderly and avoids collisions by using tokens to pass data, but if any one device or cable in the ring fails it can break the entire loop. That makes fault tolerance a major concern unless it's a dual ring, which adds cost and complexity.



ComputerHope.com

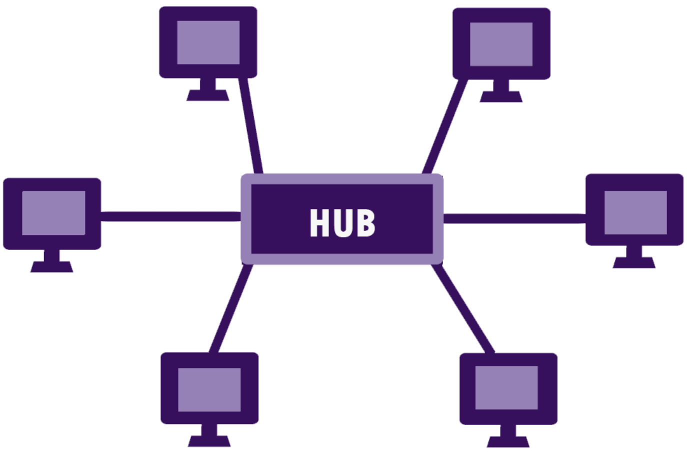
Bus

All devices share one main cable, which is known the bus. They communicate using this cable like a spine. Each device taps into the bus to send or receive data. This is cheap and easy to set up, especially for small networks, but because all data travels along the same backbone, there's a risk of data collision, and if the main cable fails, everything goes down. Its not ideal for modern networks.



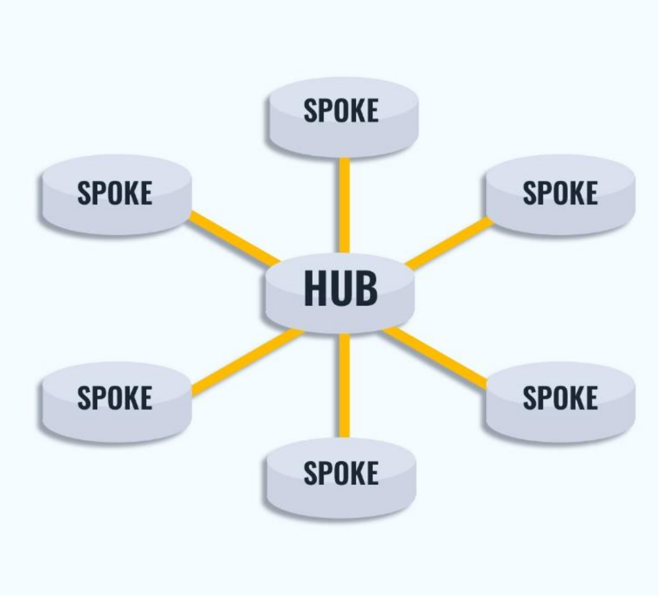
Star

Each device connects to a central hub or switch. The hub manages traffic between devices. This is easy to manage and troubleshoot; if one device fails, the rest of the network stays online. But if the central hub or switch goes down, the entire network is affected, making that point a single point of failure.



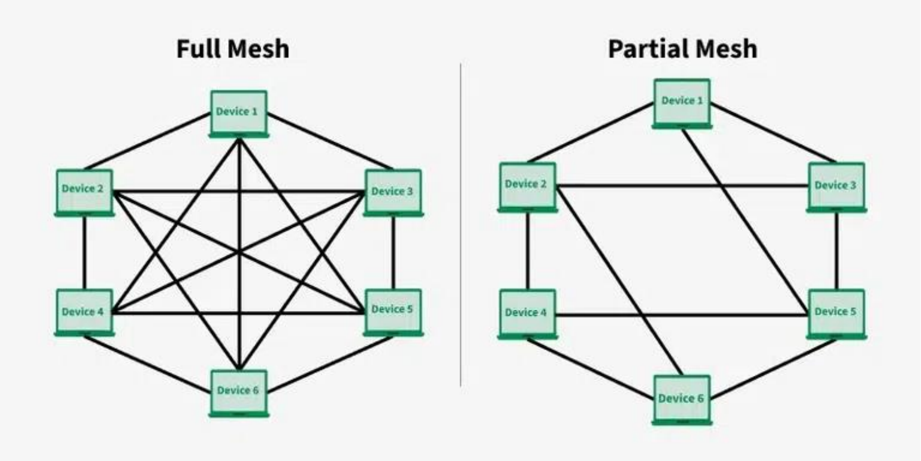
Hub and spoke

This is essentially a variation of the star, often seen in WANs. Branch offices (spokes) connect back to a central office (hub) using routers or private lines. This centralises control and is great for hierarchical networks, but again, each hub becomes a point of central.



Mesh

In a mesh, every device connects to every other device. This creates multiple paths for data to travel, offering extreme redundancy and high fault tolerance. This is ideal for critical systems where uptime is vital, but all that wiring is expensive and complex, especially in full mesh where every node has a direct link to every other node. Partial mesh can be a compromise, where only the most important nodes are connected together.



- Every node is connected to every other (full mesh)

Wireless topologies

Unlike wired networks that rely on physical cabling, wireless networks use radio signals to connect devices. But just like wired networks, they still follow topological principles –

the layout of the connections still matters. Here are the three most common wireless topologies you'll come across.

Infrastructure mode

This is the most common wireless setup – the one you use at home, in cafes or in offices. Every device (laptop, phone, tablet) connects through a central device, usually a wireless access point (WAP) or wireless router. This infrastructure allows centralised control, providing security, standardised range, and connection stability. It's also ideal for managing lots of devices at once, because the access point coordinates all the traffic. However, the WAP is a single point of failure, meaning if the WAP goes offline, every wireless device loses its connection.

Ad hoc mode

Ad hoc is a more spontaneous setup. Devices connect directly to each other without needing an access point. It's like creating a mini-network on the fly – perfect for file transfers between laptops or setting up temporary group chat in a room. The main benefit of this is the flexibility; you can create a network anywhere, instantly, with no extra hardware. But without a centralised controller, things can get messy – there's limited range, less security, and poor performance with more than a few devices. It's not built for long term or large-scale use, but it can be a quick and useful solution in the right moment.

Wireless mesh

In a wireless mesh, every node (like a WAP or mesh extender) talks to multiple others, creating a web of connections, similar to a wired mesh. If one link breaks, data automatically reroutes through another path – a built-in self-healing design. These are highly reliable networks, especially over large areas like campuses or cities. They also offer excellent coverage and fault tolerance, since there's no single device everything depends on. The trade-off is complexity. Mesh setups require more planning, and they're more expensive due to the number of device and the way they communicate.

1.4 Data Centre Topologies

A data centre is a facility that hosts critical systems – servers, storage, network equipment – and supports the delivery of applications, storage and services.

To manage this large, complex environment efficiently, we design it with a specific topology.

Three tired hierarchy

The traditional approach to building a data centre network is called the three-tiered hierarchy. Think of it like a pyramid made of three layers, each one with a different responsibility.

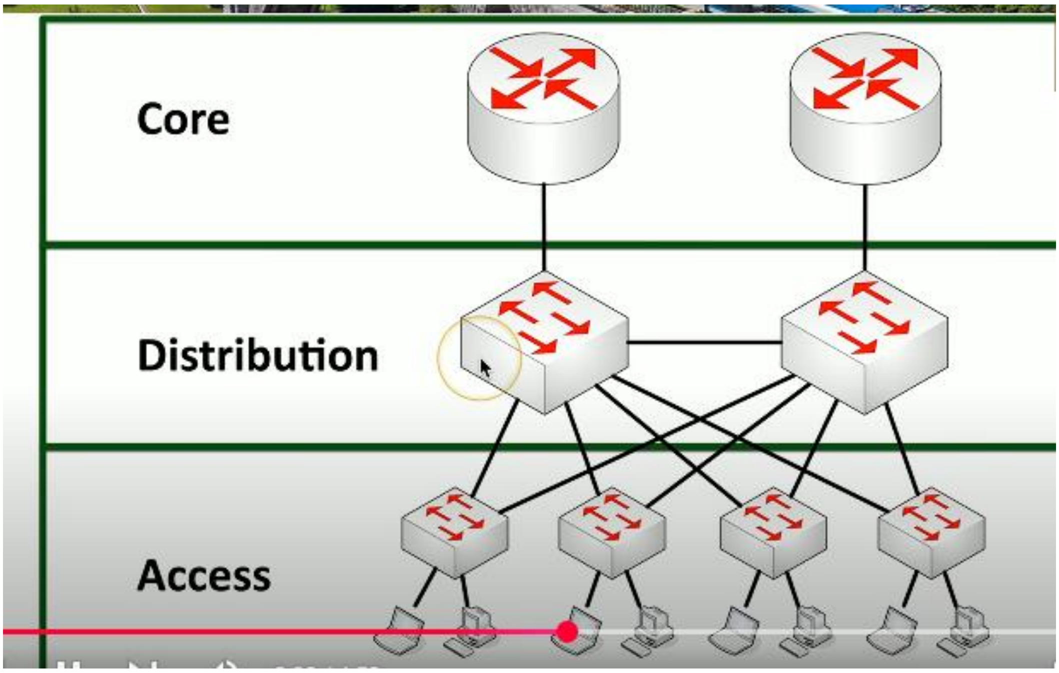
At the top is the core layer. This layer doesn't deal with individual devices or users. Instead, its role is fast and efficient transport. It connects different parts of the network together – sort of like a high-speed highway that links cities.

Beneath that, **in the middle, is the distribution layer**, sometimes called the aggregation layer. This is the decision-making hub. It connects smaller networks together and often applies rules, like ACLs, routing policies, or load balancing. If the core is the highway, the distribution layer is where you find roundabouts and traffic signals that decide which road you should take next.

You might be wondering – why build it this way? Why not just connect everything in one big web?

The answer lies in scalability, performance, and control. Each layer has a focused role, so its easier to manage. If you need to scale, you can just add more access layer switches without affecting the core. If something breaks, it's isolated to a layer, which makes troubleshooting faster. And if you need to apply security or routing policies, the distribution layer gives you a central place to do it.

This model looks like this, where each core device only connects to one distribution device beneath it. The distribution devices connect to one another, and multiple access devices beneath the distribution layer.

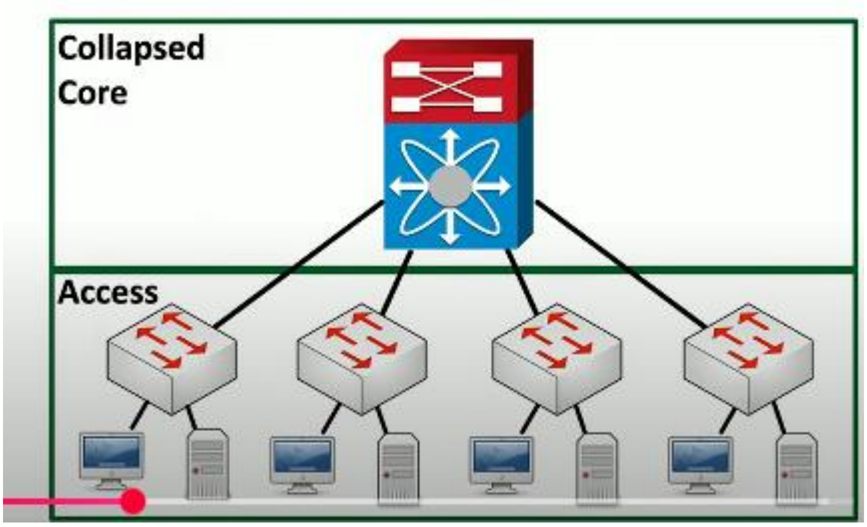


Collapsed core

While the three-tiered model is common in large environments, smaller organisations often don't need all three layers. In those cases, the core and distribution layers are merged together. This is called a collapsed core design.

It reduces the amount of hardware and simplifies management. But it comes at a cost. Still, for a small data centre or branch office, it makes perfect sense. You keep the basic benefits of a layered design while cutting down on complexity and cost.

This model looks like a three tiered architecture with only two layers:



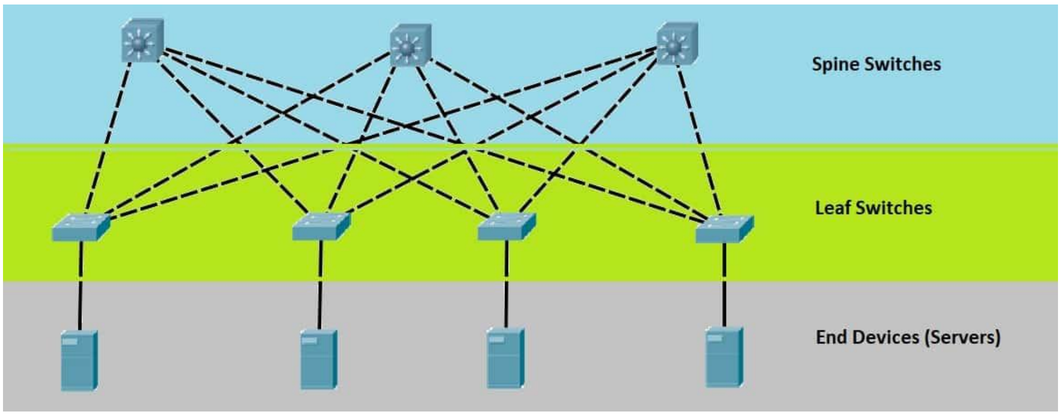
This is simpler, easier to manage and troubleshoot, but offers less redundancy as there's less routes for data to travel on, so unlike in a three-tiered architecture when one route goes down there's others available, in this layout that may not be the case.

Spine and leaf architecture

As modern data centres shifted toward virtualisation, cloud computing and containers, the old hierarchical model started to show its limitations – especially when it came to internal communication between devices.

A new design emerged called spine and leaf. Imagine it like a web where every point has a direct, equal connection to every other point.

This model looks like this



This model has the end devices connecting a switch above them in the leaf layer. The switches don't connect to one another, only the devices above them, and these switches connect to multiple routing devices above them. The spine switches (or routers) don't connect to each other, just as the leaf switches don't. the spine switches however connect to multiple leaf switches, and the leaf switches connect to multiple spine switches.

In this model, the leaf switches are at the bottom. These are equivalent to the access layer – they connect directly to servers, storage and firewalls. But instead of connecting upward into a traditional distribution layer, each leaf switch connects to every spine switch above it. The spine switches are high performance switches whose only job is to interconnect the leaves.

This setup creates a kind of mesh where any device connected to one leaf switch can reach any other device through just one spine switch. You get consistent latency, high bandwidth, and scalability – because all traffic has multiple possible paths.

This is perfect for what's known as east west traffic – communication that stays within the data centre, like between virtual machines or microservices. The older three-tiered model was optimized for north-south traffic - from users outside the data centre in toward the servers. But in modern architectures, most traffic actually moves sideways, not up and down. That's where spine and leaf shines.

Traffic flows: north-south vs east-west

Understanding traffic flow is key to know why these topologies matter.

When someone from the outside world makes a request – like loading a website or accessing cloud storage – that traffic travels from the outside, through firewalls, routers, and into the data centre. This is north-south traffic. It moves from the top (north) of the network into the bottom (south).

But once inside the data centre, traffic often travels laterally between services. For example, a web sever might talk to a database server. Or a load balancer might redirect a request to one of several app servers. This internal movement is called east-west traffic, and it now accounts for the majority of activity in modern data centres.

This is why the spine and leaf model has become so popular. It handles the east-west flows more efficiently than the old hierarchy.

Comparing designs

While both the three tiered and spine and leaf architectures are valid, the choice depends on the needs of the business.

A three-tiered topology is excellent for control, segmentation, and clear traffic flows in large enterprises or networks where north-south traffic dominates.

A collapsed core works well in smaller organisations with fewer devices and simpler needs.

A spine and leaf architecture is ideal for high-speed, scalable environments where performance and east-west traffic are critical – like cloud platforms or data centres serving many virtual machines.

Each design represents a balance of cost, complexity, performance and scalability. And as a network engineer or architect, understanding the strengths of each will help you design networks that are robust, efficient and future proof.

Section 2 – The OSI Model

2.1 Layer 1- Physical

The physical layer of the OSI model serves as the foundational stage of network communication. It is responsible for the transmission of raw data bits across physical media. This layer defines the electrical, mechanical, and procedural standards for activating, maintaining, and deactivating the physical connection between network devices. Unlike higher layers that deal with data formats or addressing, the physical layer is concerned only with how binary data is converted into signals and transmitted through mediums such as copper wires, fibre optics, or radio waves.

Electrical signalling and bit transmission

At its core, the physical layer transmits information in the form of bits, the smallest unit of data, either a 1 or a 0. These bits are not meaningful at this stage; they are simply voltage changes, light pulses, or radio frequencies, depending on the transmission medium. To send these signals effectively, devices must modulate their physical characteristics in a way that represents the intended data. One basic modulation technique is transition modulation, where the state change (such as a shift in voltage) represents a bit. The specifics of modulation schemes can vary, but at this layer, understanding the existence of such transformations is more important than the underlying mathematics. This modulation allows bits to move from one end of a cable to another in a predictable and recoverable manner.

Cables and media types

The method of bit transmission is highly dependent on the cabling and medium used. Common cable types include:

- **Straight through cables**, typically used to connect different types of devices (such as a pc to a switch)
- **Crossover cables**, used to connect the same devices directly (switch to switch or router to router)

The choice of cabling determines not just how signals travel, but also how they interact at connection endpoints. While modern auto-sensing ports can often detect and adjust to the cable type, understanding this distinction is crucial for troubleshooting layer 1 issues.

Timing and synchronisation

Communication over a physical medium may follow different timing models

- **Asynchronous communication** involves devices sending data without needing to sync their clocks, relying instead on start and stop bits to mark data frames.
- **Synchronous communication** requires devices to share a synchronised time source, enabling continuous streams of data without start/stop delimiters. This model supports higher efficiency but demands greater precision in timing.

The presence or absence of timing coordination directly affects how the bits are interpreted on the receiver's end.

Bandwidth usage: baseband vs broadband

Another key layer 1 distinction involves how the physical medium is used to carry signals:

- **Baseband transmission** dedicates the entire channel to a single signal. Ethernet commonly uses this method – sending one signal at a time over the medium
- **Broadband transmission** divides the channel into multiple frequency ranges, allowing multiple signals to coexist. This approach is much more common in cable television or DSL technologies

Both forms operate at the physical layer but differ in how they utilise the transmission medium's capacity.

Multiplexing techniques

When bandwidth must be shared among multiple signals or users, multiplexing techniques are employed to allow simultaneous data transmission:

- **Time division multiplexing (TDM)** assigns time slots to each signal in a fixed rotation
- **Statistical TDM** improves on this by allocating slots dynamically, based on active traffic
- **Frequency division multiplexing (FDM)** divides the available bandwidth into multiple frequency bands, with each carrying a separate signal simultaneously

These strategies optimize channel usage and are foundational in wide area network design and telecommunications systems.

Layer 1 devices

Devices that operate strictly at layer 1 do not examine or alter the data they transmit; they merely regenerate or relay electrical or optical signals. Examples include:

- **Repeaters**, which regenerate and amplify signals to extend the range of a network without modifying the content
- **Hubs**, which function as multi-port repeaters, broadcasting incoming signals to all other ports. Their simplicity reflects the lack of intelligence at this layer – there is no addressing or filtering

Because these devices do not interpret any data beyond the signal level, they exemplify the minimalism of layer 1's responsibilities.

Conclusion

The physical layer forms the basis upon which all higher layer functions depend. Its role in converting binary data into transmission signals, managing electrical and timing properties, and handling raw connectivity is essential to the functioning of any network. While its functions may appear simple in contrast to those of higher layers, a failure at this level – whether due to faulty cabling, incorrect timing, or media incompatibility – renders all higher layer irrelevant until resolved.

2.2 Layer 2 -Data Link

The data link layer occupies the second position in the OSI model and is responsible for node-to-node data transfer. It ensures that frames of data are reliably transmitted across the physical medium established by layer 1, and that they reach the correct destination device on a local network. While the physical layer concerns itself only with transmitting raw bits, the data link layer adds order; it groups bits into frames, applies addressing, detects errors, and introduces mechanisms for flow control. It marks the beginning of logical networking.

Framing and MAC addressing

At this layer data is organised into frames. A frame is a structured package of data that includes not only the payload (the actual data) but also metadata required for successful delivery. This includes:

- A destination MAC address, identifying the receiving device
- A source MAC address, identifying the sender
- A frame check sequence, often using a checksum to detect errors in transmission

The MAC address (Media Access Control) is a hardware level identifier assigned to each network interface card (NIC). It is globally unique and functions as the address used on the local network. While higher layers may deal with logical addressing (like IP), layer 2 operates exclusively using the physical addresses to distinguish between connected hosts.

Logical view of local networks

Although physical cabling connects devices at layer 1, it is the data link layer that provides a logical view of how devices communicate on a local network segment. This logical structuring defines what constitutes a single broadcast domain and determines how frames are switched, forwarded or filtered. By using MAC addressing and logical segmentation, this layer creates an abstract sense of network neighbourhood, where only relevant devices receive and process frames.

Logical link control and flow regulation

The data link layer is traditionally divided into two sublayers:

- **Media Access Control (MAC):** handles addressing and access control for medium
- **Logical Link Control (LLC):** provides mechanisms for identifying protocols, error detection, and in some cases, flow control

Flow control refers to techniques that manage the rate at which data is sent, preventing a faster sender from overwhelming a slower receiver. While more prominent in higher layer protocols like TCP, basic forms can exist here as part of the LLC sublayer’s responsibilities.

Error detection: checksums and reliability

To prevent the forwarding of corrupted data, the data link layer includes basic error detection using checksums, often calculated using algorithms such as CRC (Cyclic

Redundancy Check). When a frame is received, the checksum is recalculated and compared. If the result differs, the frame is considered damaged and discarded. This process improves the integrity of data at the local level, although recovery mechanisms typically exist in upper layers if retransmission is necessary. The result is stored in the frame check suquence (FCS) field, which is part of the trailer of the frame.

Timing mechanisms: Asynchronous, synchronous and isochronous

Like layer 1, the data link layer concerns itself with timing models – particularly when data is transmitted over time sensitive or coordinated systems.

- Asynchronous communication continues to rely on start and stop indicators to distinguish bits, without clock synchronisation
- Synchronous communication uses a shared clock or synchronised signal to time the transmission and reception of data
- Isochronous communication is designed for time sensitive streams (such as audio or video), ensuring consistent timing between frames. This method prioritises timing consistency over delivery certainty

While all three can be implemented across various technologies, they are relevant in layer 2 in protocols and systems that manage the timing of data transmission between two locally connected devices.

Layer 2 devices and their intelligence

Devices operating at layer 2 are inherently more intelligent than those at layer 1. They are capable of reading MAC addresses and making forwarding decisions based on them. The primary example is the network switch.

Switches maintain a MAC address table, which maps MAC addresses to specific ports. When a frame arrives, the switch inspects the destination MAC address and consults its table to forward the frame only to the appropriate port, rather than broadcasting it to all connected devices as a hub would. This targeted delivery not only improves efficiency but also enhances security and reduces collisions.

The increased intelligence of switches, compared to repeaters and hubs, reflects the functional advancement of the data link layer. In introduces selective forwarding, segmentation of collision domains, and support for full duplex communication.

Conclusion

The data link layer is the first stage at which logical decisions are made. It transforms the stream of bits passed from the physical layer into structured frames, identifies devices using MAC addresses, and enforces local delivery rules. With mechanisms for timing, flow control and error detection, it ensures that communication between devices on the same network segment is orderly and reliable. This layer, through devices like switches, bridges the gap between raw transmission and meaningful networking.

2.3 Layer 3 - Network

The network layer governs how data is transferred between devices across multiple networks. Unlike layer 2, which is limited local segments, layer 3 introduces logical addressing and routing, allowing data to travel across intermediate networks toward distant destinations.

IP and logical addressing

At the core of this layer is the internet protocol (IP). IP assigns each device a logical address that reflects its position within a broader network hierarchy. Unlike MAC addresses, which are tied to hardware, IP addresses are reassign able and reflect the structure of the network. This allows routers to make decisions based on destination networks rather than physical devices.

Switching methods: Packet, Circuit, and message

The network layer can be associated with several switching methods

Switching is how data is moved across a network – from one device to another – by passing through routers and other network devices. At layer 3, switching determines the path that packets take between networks.

- Packet switching; the standard on modern networks. Data is split into packets, each sent separately and possibly along different routes. Efficient and resilient
- Circuit switching; a dedicated path is set up before communication begins. All data follows that path. Reliable but inefficient. Used in old telephone systems
- Message switching; entire messages are stored at each switch before being forwarded. Flexible but slow and memory heavy. Now obsolete

Routing: discovery and selection

Layer 3 is responsible for route discovery – identifying available paths through the network – and route selection, choosing the best one based on metrics such as hop count, bandwidth, or administrative cost. This process is managed by dynamic routing protocols (like OSPF or RIP) or through static configuration. It ensures that packets follow an optimal or administratively preferred path.

Connection services

Though IP itself is connectionless, the network layer supports basic connection related functions:

- Flow control; managing how quickly data is routed toward a destination

- Packet reordering, ensuring data can be properly reconstructed at the receiving end, particularly when paths of different lengths are involved

These features provide support for reliable delivery, although full reliability is typically handled at higher layers like transport.

ICMP: diagnostics and reporting

The internet message protocol (ICMP) operates alongside IP to report errors and perform diagnostics. It is essential for functions such as:

- **Ping**; a simple test of connection between devices
- **Traceroute;** a recording of the steps along the journey of a ping

Layer 3 devices

Devices operating at this layer include routers and layer 3 switches. These systems read IP headers too forward packets between networks. They maintain routing tables, make decisions about path selection, and enforce access policies based on IP level rules

Summary

The network layer abstracts the idea of physical proximity. It introduces IP addressing and routing, allowing devices on different networks to communicate. Through packet switching, route selection, and control protocols like ICMP, it lays the foundation for internetworking – ensuring that data can find its way across the globe.

2.4 Layer 4 - Transport

The transport layer marks the boundary between the lower, network focused layers and the upper layers that serve applications. Its primary role is to manage end to end communication between devices – breaking data into smaller units, tracking it, ensuring it's delivered correctly (if required), and reassembling it at the destination.

Two core protocols define this layer: TCP and UDP. Each offers a distinct approach to communication depending on the needs of the application.

TCP: Reliable, connection-oriented delivery

TCP is a connection-oriented protocol. Before any data is sent, a three-way handshake sets up a session between devices, synchronizing sequence numbers and confirming readiness. This process enables TCP to guarantee delivery, order and integrity.

Data is sent as segments; each tracked with sequence numbers. The receiver sends acknowledgements, confirming receipt. If data is lost, TCP triggers retransmission. To optimize flow, TCP uses windowing, which adjusts the volume of unacknowledged data based in current network conditions.

TCP is suited for cases like file transfers, web browsing, and email – where completeness and accuracy are essential.

UDP: Fast, connectionless transmission

UDP operates without connection setup. Each datagram is sent independently, with no guarantee of delivery, order, or duplication handling. It's a connectionless protocol, and therefore much lighter.

Because it lacks acknowledgements, windows, and sequencing, UDP sacrifices reliability for speed. It's ideal for real time applications like streaming, VoIP and gaming, where delays (such as those that come with retransmission of dropped packets, as would happen with TCP) are more disruptive than occasional packet loss.

In some contexts, dropped packets aren't a problem, and handling reliability at the application level is preferable. In other words, UDP is suitable for the situations when speed of delivery is more important than coherence of delivery, such as in a phone call when its essential for latency to be low, but users can understand what's being said despite a few dropped packets.

Key differences between UDP and TCP:

TCP ensures:

- Delivery via **acknowledgements**
- Correct order via **sequencing**
- Adjusted flow via **windowing**

UDP omits these features, trading control for speed and simplicity.

Buffering and flow control

To accommodate mismatched speeds between sender and receiver, the transport layer relies on buffering-temporarily holding data in memory-and windowing, which manages how much can be sent before needing acknowledgment. Together, they smooth data flow and respond to changing network conditions.

Layer 4 devices

While most functions are handles in software, some devices operate with transport layer awareness. These include load balancers and stateful firewalls, which inspect TCP/UDP headers to manage sessions or enforce port-based rules.

Summary

Layer 4 ensures that data flows efficiently and appropriately between endpoints. It supports both precise, reliable communication through TCP and fast, low overhead delivery through UDP-depending on what the situation demands.

2.5 Layer 5 - Session

The session layer serves as the manager of conversations between networked applications. Positioned above the transport layer, it takes on the responsibility of organising and controlling dialogues between systems – essentially defining when communication starts, how it proceeds, and when it ends. Unlike the transport layer, which focuses on data transport and reliability, the session layer introduces structure into multi-packet exchanges, particularly when the dialogue must be resumed or kept in sync.

What is a session?

A session is a coordinated exchange of information between two endpoints. This could be a phone call over VoIP, a video stream, or a persistent connection between client and server during a file transfer. The session layer ensures these exchanges are well regulated through three distinct stages:

1. **Session establishment**

A connection is formally initiated between the two systems. During this stage, both parties negotiate parameters such as authentication requirements or resource allocation. This prepares both ends for a sustained exchange.

2. **Data transfer and maintenance**

Once established, the session handles data exchange. This includes mechanisms for checkpointing (saving progress), synchronisation (ensuring communication remains in order) and -if needed- reestablishing a session after disruption. It can also manage token passing or half duplex communication where only one party can speak at a time.

3. **Session teardown**

When communication is complete, the session layer coordinates the termination of the session, ensuring all data is accounted for and resources are released cleanly. This process avoids data loss or dangling processes.

Functions of the session layer

While layer 4 (transport) ensures the reliability of data delivery, the session layer introduces the concept of managing the dialogue itself. This includes:

- **Session management**: tracks sessions and can resume them after interruption
- **Acknowledgement of receipts**: while layer 4 handles segment level acknowledgement, the session layer may provide high level application acknowledgment – for example, confirming that a streamed video segment has been played

- **Dialogue control**: coordinates which side is allowed to communicate and when. This is especially relevant in protocols requiring strict turn taking
- **Session recovery**: in the event of a dropped connection, the session layer can use checkpoints to resume the session from the last successful data exchange

Common protocols at the session layer

Several protocols operate within or rely on session layer functions, especially those related to real time media and persistent communication.

- **H.323**: a protocol suite for voice, video, and data conferencing over packet-based networks. It handles call signalling, session setup, and media negotiation
- **RTP (real time transport protocol):** although RTP technically operates at the transport layer, its use in real-time audio/video streaming requires coordination from the session layer to manage timing and session continuity
- **NetBIOS (network basic input/output system):** often associated with older windows-based networking, NetBIOS provides session establishment and teardown, especially in local area networks

Connection to previous layers

The session layer builds directly on layer 4's mechanisms. While TCP provides a connection-oriented transport with features like sequencing and retransmission, the session layer adds meaning and coordination to that flow. It treats TCP streams not just as raw data but as conversations that need to be structured, resumed or closed.

This structuring becomes essential in multimedia applications, client server sessions, or any long-lived process where disconnection risks or turn taking are significant concerns.

2.6 Layer 6 - Presentation

The presentation layer sits between the session layer and the application layer (layer 7), acting like a translator between the two. Its job is to ensure that the data being sent by one device can be understood by the receiving device, even if their formats or encoding systems are different.

Imaging two people speaking different languages. The presentation layer is the interpreter, making sure both sides understand the same message – even if it's wrapped differently.

Core responsibilities

1. Data formatting

When raw data travels across a network, it needs to be structured in a way both ends can understand. This is where formatting comes in. the presentation layer handles things like:

- Test formats (ASCII, EBCDIC)
- Image formats (JPEG, PNG, GIF)
- Video formats (MPEG, MP4)
- Data structure encoding (XML, JSON)

So when a JPEG is sent in an email or a JSON file is returned from a web server, this layer ensures its structured correctly and recognizable to the receiving application.

2. Compression

Though not always mentioned, compression often happens at this layer. This helps reduce the amount of data sent across the wire – making communication faster and more efficient.

Think of how a zIP file shrinks content. The presentation layer may do similar work for network traffic.

3. Encryption and decryption

This is one of its most important roles in modern networking. It handles encryption before transmission and decryption on arrival. This keeps sensitive data protected while it moves over potentially insecure networks.

Common encryption methods seen here include:

- TLS (transport layer security) – often works alongside layer 7 apps like HTTPS, but encryption happens at this layer
- SSL (now deprecated, but historically relevant)
- Encryption algorithms themselves (like AES, RSA) operate here as well

Protocols and standards

While this layer isn't as protocol heavy as others. Key technologies associated with it include:

- TLS 0 for encrypting sessions securely
- MIME – used in emails to support text, images, audio etc
- SSL (legacy) still seen in older systems

Summary

The presentation layer is all about how data is packaged:

- Translation and formatting of data

- Encryption and decryption of data
- Compression and decompression of data
- Ensuring structure and encoding are compatible on both ends

While if often works silently behind the scenes, it plays a crucial role in making communication seamless and secure.

2.6 Layer 7 - Application

The application layer is the final stop in the OSI model – the layer that interacts directly with software on your device. But don't be misled – this isn't about the applications themselves (like chrome or outlook). It's about the network services those applications use to communicate.

Think of layer 7 as the front desk of the network. It doesn't move packets or format data – it simply offers services to the applications running on your machine.

What happens at layer 7?

- Advertisement of services that are available on the network – like email, file transfer, or web access
- Application layer protocols are used to manage these services
- This layer is responsible for network related parts of applications, not the full applications themselves

So if a user opens a browser and visits a website, layer 7 is where the HTTP/HTTPS protocol kicks in to ask, "Hey server, can I see this webpage?"

Or if an email app is syncing, layer 7 is where SMTP, POP3 or IMAP are quietly doing the work under the hood.

Key services at layer 7

Here are some examples of services and protocols that live here:

- HTTP/HTTPS – browsing websites
- FTP – Transferring files
- STMP/POP3/IMAP – sending and receiving emails
- DNS – translating domain names into IP (though often seen as both layer 7 and layer 5/6 depending on context
- DHCP – assigning IP addresses (application-level service for networking)
- SNMP – monitoring and managing devices

These protocols are like waiters at a restaurant – each one knows how to handle a certain kind of request and delivers it to the correct kitchen (service).

Key insight

Layer 7 is all about communication between machines and the services they offer – not the internal operation of the software itself.

We're not interested in the UI of Outlook or the layout of a webpage. We're interested in what network service it needs to do its job – and that service lives at layer 7.

2.8 Encapsulation and decapsulation

When data moves through a network, it doesn't simply travel as is from one device to another. It goes on a journey – not just physically across cables and airwaves, but logically through layers of preparation, labelling, wrapping and unwrapping. This process is known as encapsulation and decapsulation, and it's the very heart of the how the OSI and TCP/IP models work.

To encapsulate is to wrap data with the necessary headers and, where applicable, trailers, that allow it to be carried, directed, verified, and understood. To decapsulate is to unwrap that same data on the receiving end – peeling back each label and processing it in reverse order, so the original message can be reconstructed.

Encapsulation vs decapsulation

Enc happens at the sender's side. As the data moves down through the layers (from application to physical), it is repeatedly packaged into new containers – with each layer adding its own header (and sometimes a trailer) that serves its specific purpose. This isn't just metaphorical wrapping: these headers literally become part of the data.

Decapsulation is the inverse process – it happens at the receiver's side, as data moves up through the layers. Each layer reads and removes its corresponding header (and trailer, if present) interprets what's inside, and passes it to the layer above.

The data unit at each layer has a name – collectively, these are known as protocol data units (PDUs). The name of the Protocol data unit depends on the layer it belongs to:

- Segment at the transport layer (TCP or UDP)
- Packet at the network layer (IP)
- Frame at the data link layer (Ethernet)
- Bits at the physical layer

Each Protocol data unit has a specific structure and purpose – and each layer trusts the one above it to have done its job. It's like a relay race where each runner passes a baton, but adds a note or sticker before handing it off.

The encapsulation process per layer

Application → presentation → session

The user generated data begins at the top. At this point, its just raw content, compression or encryption. For example, if TLS is in play, its encrypted here. These transformations don't usually add a header, but they shape the data before it hits the transport layer. The Protocol data unit is still just data here – sometimes referred to as the payload

Transport layer

Here's where the data first gets formally encapsulated. The transport layer breaks the payload into segments (if using TCP) or datagrams (if using UDP).

In the case of TCP, each segment receives a TCP header. This includes:

- Source and destination ports: to identify the sending and receiving applications
- Sequence and acknowledgment numbers: for ordering and reliability
- Control flags; like SYN (start connection), ACK (acknowledge), FIN (finish), RST (reset), PSH (push), URG (urgent)

UDP, by contrast, is simpler and faster. Its UDP header includes:

- Source and destination ports
- Length
- Checksum

The header here enables tracking of individual conversations and ensures the data is delivered correctly – or at least, in the case of TCP, that errors are caught and retransmitted

Network layer

The segment is now handed off to the network layer, which wraps it in an IP header, turning it into a packet. The IP header contains:

- Source IP address
- Destination IP address
- TTL (Time to Live): prevents packets from circulating forever
- Protocol field: identifies whether the payload is TCP, UDP, ICMP etc

This is where routing comes into play. While layer 4 is all about which application should receive the data, layer 3 decides how to get there.

Data link layer

The packet now gets encapsulated inside a frame, the PROTOCOL DATA UNIT of layer 22. This layer prepares the data for transmission on the local link – such as over ethernet o Wi-Fi. The ethernet frame consists of:

- MAC header

    - Source MAC address
    - Destination mac address
    - Type field
- Payload: the layer 3 packet
- Trailer: usually contains the frame check sequence (FCS) – a checksum used for error detection

At this point, one more critical idea comes in: the maximum transmission size, or maximum transmission unit (MTU). Ethernet usually has an MTU of 1500 bytes. If the IP packet is too large to fit in a frame, it may be fragmented.

Physical layer

Finally, the frame is converted to bits – electrical signals or pulses of light – and transmitted across the medium - this is the only layer where we're dealing with the physical, tangible transmission of information.

Decapsulation in reverse

As the bits arrive at the destination device, the process happens in reverse:

1. The physical layer reads the signal and converts it into bits
2. The data link layer reads the frame, checks the FCS for integrity, and removes the ethernet header/trailer
3. The network layer reads the packet, checks the IP addresses, strips the IP header, and passes the payload up
4. The transport layer reads the port numbers and flags, reorders if necessary (for TCP), and delivers the segment to the appropriate application service
5. The upper layers decrypt, decompress, or reformat the payload if needed, and finally the data is presented or stored as intended.

Section 3: ports and protocols

3.1 Introduction, TCP, UDP and ICMP

At the heart of communication between devices on a network lies a precise language of ports and protocols. While protocols define the rules and structure of communication, ports act as numbered doorways, guiding traffic to the correct application on a device. Together, they ensure that everything from a web request to a video stream ends up where it's meant to go.

Understanding ports

When a device receives data over the network, that data doesn't just get dumped into the system – it needs to be directed to the correct service. That's what ports are for. A port is a 16-bit number (from 0 to 65,535) used to identify a specific process or service listening for network traffic on a device. For example, web servers typically listen on port 80 (HTTP) or 443 (HTTPS), while email might use ports like 25 or 587.

Ports are categorised based on their number ranges and management:

- Well known ports (0-1023): these are assigned by IANA (Internet Assigned Numbers Authority) and are reserved for standard services. Examples include 22 for SSH, 80 for HTTP, and 443 for HTTPS.
- Registered ports (1024 – 49151): these are assigned to user applications or services that are widely used but not necessarily standard, like certain games or proprietary applications
- Ephemeral ports (49152 – 65535): these are temporary ports selected dynamically by a client when initiating a connection. They're short lived and used only for the duration of a communication session

Behind the scenes, when you visit a website, your system might open a connection from an ephemeral port to port 443 on the server. The server replies to that port, and your device knows which session the reply belongs to. This whole process happens constantly and invisibly – every second, thousands of these tiny handshakes are keeping your applications in sync.

TCP: reliable delivery with heavyweight precision

The transmission control protocol (TCP) is one of the most fundamental protocols at the transport layer of the TCP/IP model. What sets TCP apart is its ability to guarantee delivery, even across unreliable networks.

It achieves this reliability through a few key mechanisms:

- Error checking, using checksums to detect corrupted data
- Data sequencing, ensuring data arrives in the right order even if packets arrive out of sequence
- Acknowledgement, where the receiver confirms receipt of each piece of data

TCP begins its process with a three-way handshake. The client first sends a SYN message to initiate the connection. The server replies with a SYN-ACK, acknowledging the request and offering its own. Finally, the client replies with an ACK, completing the handshake. Only after this process can data begin flowing.

TCP also uses flow control to manage how much data is sent at a time, adapting to the network's capacity. Through a technique called windowing, it ensures the sender doesn't overwhelm the receiver, gradually increasing or decreasing the amount of in flight data depending on how acknowledgements are received.

A TCP segment contains the source and destination ports, as well as the IP addresses from the network layer, tightly tying together transport and routing information. TCP is considered connection oriented, and its overhead – while higher than simpler protocols – is what makes it so trustworthy. It's the backbone of services where accuracy and reliability matter, such as web browsing, file transfers, and emails.

Summary

- SYN/ACK; messages sent between devices as part of the three way handshake
- Flow control; manages how much data is sent at a time
    - Windowing is the technique within this that ensures the sender doesn't overwhelm the receiver

UDP: Lightweight and fast

The user datagram protocol (UDP) also lives at the transport layer, but it operates very differently. It is a connectionless protocol, meaning it does not establish a session or verify delivery. Instead, it simply sends data and hopes it arrives.

This makes UDP stateless (doesn't keep track of the communication state between entities like TCP does) and incredibly fast. It has very low latency and requires much less processing overhead, making it ideal for use cases where speed is more important than reliability. Online gaming, live voice or video calls, and DNS queries all benefit from UDP's efficiency.

A UDP datagram includes far fewer fields than a TCP segment: only the source port, destination port, length and checksum. It doesn't try to manage sequencing, acknowledgments, or flow control – it simply throws the data out into the network.

This approach isn't reckless; its deliberate. In many cases it's better for data to arrive quickly than to arrive perfectly. Missing a frame in a video call is preferable to freezing while waiting for it to be retransmitted.

ICMP: the internet control messenger

The internet control message protocol (ICMP) is not a transport layer protocol like TCP or UDP – instead, it operates at the network layer (Layer 3), alongside IP. It's not used to send user data, but rather to report errors and test connectivity.

ICMP's most well-known use is the ping utility. When you ping a device, you're sending ICMP echo request messages, and hoping for echo replies in return. This helps determine whether a device is reachable and how long it takes to communicate with it.

An ICMP message includes fields such as type, code and checksum, but does not include mechanisms for reliability or retransmission like TCP does. That's not its role.

ICMP is meant to be fast, simple and efficient, designed for network diagnostics and control, not end to end communication.

Because ping deals directly with IP, it's critical for operations like:

- Reporting unreachable destinations
- Indicating time exceeded (TTL expired)
- Supporting tools like traceroute

However, because it reveals information about the network, ICMP has been abused in the past for denial of service (DoS) attacks, such as the infamous ping od death, where malformed or oversized ICMP packets were used to crash systems.

Bringing it all together

Ports and protocols work in harmony to guide data to its rightful place and ensure it gets there in the right way. TCP provides the structure and patience needed for reliable communication. UDP provides the speed and simplicity needed for real time performance. ICMP provides the awareness and diagnostics that allow administrators to monitor and maintain the health of a network.

Together, they form the skeleton and nervous system of every communication you've ever made online.

3.2 Web, email and file transfer ports and protocols

When devices communicate across a network, they don't just fling data around randomly – they use protocols that define how that communication should take place, and they send this data to specific port numbers that identify services running on a device. These ports, standardized by the IANA, ensure that both sender and receiver know what kind of data is being transferred and how to handle it. This chapter introduces the common protocols used for web browsing, email, and file transfer, along with the port numbers associated with them.

Web ports and protocols

Web browsing relies on two primary protocols: HTTP and HTTPS. http, or hypertext transfer protocol, is the foundation of communication on the web. It operates on port 80 and is used for retrieving unencrypted web pages. However, in modern networks, http is largely considered insecure because data is sent in plaintext.

That's where https comes in – the secure version of http. It used TLS (transport layer security) to encrypt the connection the client and the web server, protecting sensitive data from interception. https communicates over port 443, and today it is the standard for nearly all legitimate websites. When a device accesses a secure website, it typically opens a random (ephemeral) port on the client side and connects to port 443 on the server, ensuring encrypted transmission from end to end.

Summary

HTTP

- Port 80
- Insecure

HTTPS

- Port 443
- Secured via TLS

Email ports and protocols

Email communications are a little more varied, involving protocols for both sending and receiving messages. The simple mail transfer protocol (SMTP) handles the sending of emails from a client to a mail server or between servers. For receiving messages, we rely on POP3 (post office protocol version 3) and IMAP (internet message access protocol).

SMTP is traditionally associated with port 25, although modern secure variants use port 465 or 587. Pop3, which downloads messages from the server and deletes them afterward, uses port 110 for standard and 995 for secure connections. IMAP, which allows messages to be stored on the server and accessed from multiple devices, uses port 143, or 993 for secure connections.

These protocols can use either TCP or UDP, but in most real-world scenarios – especially involving secure communication – TCP is used due to its reliability and connection-oriented nature. The key takeaway here is that while email feels like a single, seamless action from the user's perspective, it's really multiple protocols working together under the hood, each assigned a specific port.

Summary

SMTP

- Port 25 (insecure)
- Port 465/587 (secure)
- Only sends messages, doesn't receive

POP3

- Port 110 (insecure)
- Port 995 (secure)
- Retrieves messages from server and then deletes them

IMAP

- Port 143 (insecure)
- Port 993 (secure)
- Retrieves messages but allows them to stay stored on the server

File transfer ports and protocols

File transfer protocols provide another key example of how ports and services work together. The most basic and familiar is ftp (file transfer protocol), which operates on ports 20 and 21. FTP is an older protocol that lacks built in encryption, which makes it unsuitable for sensitive data unless combined with additional security features.

A more secure alternative is SFTP (SSH file transfer protocol). Despite the similar name, sftp is entirely different from ftp – it runs over the ssh protocol and uses port 22, providing both authentication and encryption. Another lightweight option is TFTP (trivial file transfer protocol), which used port 60 and forgoes many of the features of ftp for the sake of simplicity. TFTP uses UDP instead of TCP, making it faster but less reliable, and its often used for tasks like booting devices or pushing configuration file sin a local network.

Lastly, SMB (server message block) is commonly used for sharing files, printers, and even application access within local area networks (LANs). SMB typically runs over port 445 and is not intended for use over open internet due to its vulnerability to attack. SMB operates primarily in internal windows environments, enabling users to browse shared folders or access files from a network drive.

Summary

FTP

- Basic file transfer
- Port 20 for
- Port 21 for
- Insecure

SFTP

- Secure version of FTP
- Port 22
- Secured via SSH

TFTP

- Lightweight version of FTP
- Port 60 (UDP)

SMB

- For sharing locally within windows environments
- Port 445

Summary

Each of these protocols is carefully mapped to specific ports, enabling devices to send and receive data without confusion. Whether you're browsing a website over https, downloading an email through IMAP, or transferring files via sftp, it all comes down to using the right protocol on the correct port number, with the appropriate transport layer behaviour underneath. TCP and UDP both have roles to play, but the protocol chosen depends on whether reliability or speed is more important.

This understanding is critical for configuring firewalls, troubleshooting connectivity, and diagnosing service issues on a network – because if the ports aren't open or if the protocols aren't behaving as expected, the services simply wont work.

3.3 Remote access and network service ports and protocols

One of the foundational skills in network administration is the ability to access devices remotely. Whether configuring a router, checking logs on a server, or managing a remote desktop, these protocols make it possible.

SSH and Telnet

SSH (secure shell) is the modern standard for secure remote administration. It runs on port 22 and encrypts the entire session, making it ideal for managing servers and network equipment over an insecure network like the internet. SSH allows command line access to remote systems and also serves as the foundation for protocols like SFTP, which we saw in the file transfer chapter.

In contrast, telnet is an older protocol that also provides command line access to remote systems, but with a critical difference – it sends all data, including usernames and passwords, in plaintext. It operates on port 23 and is rarely used today outside of legacy systems or secure internal networks, due to its lack of encryption.

Remote Desktop Protocol (RDP)

RDP is another key remote access protocol, but unlike ssh and telnet, it provides a graphical interface rather than a command-line shell. It's used primarily on windows systems and communicates over port 3389. RDP allows a user to fully control a remote system's desktop as if sitting in front of it. Security concerns with RDP typically involve brute force attacks and vulnerability exploitation, which is why its often protected by firewalls or VPN access.

Summary

SSH

- Port 22
- Allows for secure access to foreign device command lines (CLI only)
- Secure; used by other secure version of protocols

Telnet

- Port 23
- Insecure; doesn't encrypt data. Not used anymore

RDP

- Port 3389
- Allows user full remote control of a foreign device (full GUI)

3.4 Network services: supporting the infrastructure

Behind the scenes of any functioning network lies a collection of services that help manage, organise and monitor communication. These network services run quietly in the background but are essential for everything from assigning IP addresses to authentication users.

Domain Name System (DNS)

DNS is one of the most foundational protocols on the internet. It resolves human readable domain names (like www.example.com) into IP addresses. DNS operates on port 53 and can use both TCP and UDP, depending on the size and type of the query. Most lookups use UDP due to its low overhead, but larger zone transfers use TCP for reliability.

Dynamic host configuration protocol (DHCP)

DHCP handles automatic IP address assignment. When a device joins a network, it sends a broadcast request, and a dhcp server responds with an IP address, subnet mask, gateway, and DNS settings. DHCP uses UDP ports 67 and 68 and is critical for keeping networks scalable without requiring manual configuration for every device.

SQL services

SQL services like Microsoft SQL server and MySQL manage access to structured databases across networks. Microsoft SQL commonly runs on port 1433, while MySQL uses port 3306. These services are accessed by applications that need to read or write data – such as inventory systems, websites, or CRMs. Because these ports are well known, they're frequent targets for attackers, so firewalls and authentication controls are essential.

Simple Network Management Protocol (SNMP)

SNMP Simple Network Management Protocol) is the protocol used to monitor and manage network devices like routers, switches, and printers. It runs on UDP port 161 for standard queries and UDP port 162 for trap messages (unsolicited alerts). SNMP is essential for building a centralised monitoring system and is often used by tools like SolarWinds or PRTG.

Syslog

Syslog is a standardised way for devices to send log messages to a centralised logging server. These logs help administrators identify errors, performance issues, or security threat. Syslog typically used UDP port 514 and is used by routers, firewalls, Linux servers, and more.

Network time protocol (NTP)

NTP ensures that all devices on a network have synchronised clocks, which is vital for accurate logging, certification validation, and time sensitive protocols. NTP runs on UDP port 123 and pulls time for higher level servers all the way up to atomic clocks or GPS sources.

SIP (session initiation protocol)

SIP is the backbone of most modern VoIP communication. It sets up and tear down calls over the internet and runs on ports 5060 and 5061, depending on whether encryption is used. SIP does not transmit voice data itself – it simply manages the session; actual audio is typically handled by protocols like RTP.

Lightweight directory access protocol (LDAP)

LDAP is used for directory services – centralised lists of users, devices, and resources. LDAP operates on port 389, while its secure counterpart, LDAPS, uses port 636. These protocols are central to centralised authentication systems like Microsoft active directory and allow users to log in to multiple services using the same credentials.

Summary

DNS

- Resolves domain names to IP addresses to facilitate web access
- Port 53
- UDP or TCP based on query size

DHCP

- Allows for request and assignment of IPS and relevant information
- Uses UDP 67 and 68

SQL services

- SQL protocols provide database management
- Microsoft SQL uses 1433
- MySQL uses 3306

SNMP

- For managing and monitoring network devices such as routers and switches
- Uses port 161 UDP for standard queries
- Uses port 162 UDP for trap alerts

Syslog

- Standardised way to message centralised logging server
- Port UDP 514

NTP

- Ensures devices on a network have synched clocks
- Port 123

SIP

- Sets up and brings down VoIP calls
- Uses 5060 or 5061 depending on if encryption is used

LDAP

- Used for directory services
- Uses port 389 (insecure)
- Uses port 636 (secure)

Section 4: Media types

4.1 Copper network media

Copper cabling forms the foundation of most wired local area networks (LANs), and despite the growing popularity of fibre optics and wireless connectivity, copper remains widely used due to its cost effectiveness, ease of installation, and sufficient performance for many environments. This chapter explores the main types of copper-based media, the standards that define them, and important physical considerations like shielding and fire safety.

IEEE 802.3 and ethernet standards

At the core of copper networking lies the IEEE 802.3 standard – more commonly known as ethernet. This standard defines how data is formatted and transmitted across a physical medium. Whether it's a simple office LAN or a data centre backbone, the structure and operation of copper ethernet networks are rooted in 802.3 this standard supports a wide range of copper cable types and speeds, from 10 mbps to multiple gigabits per second.

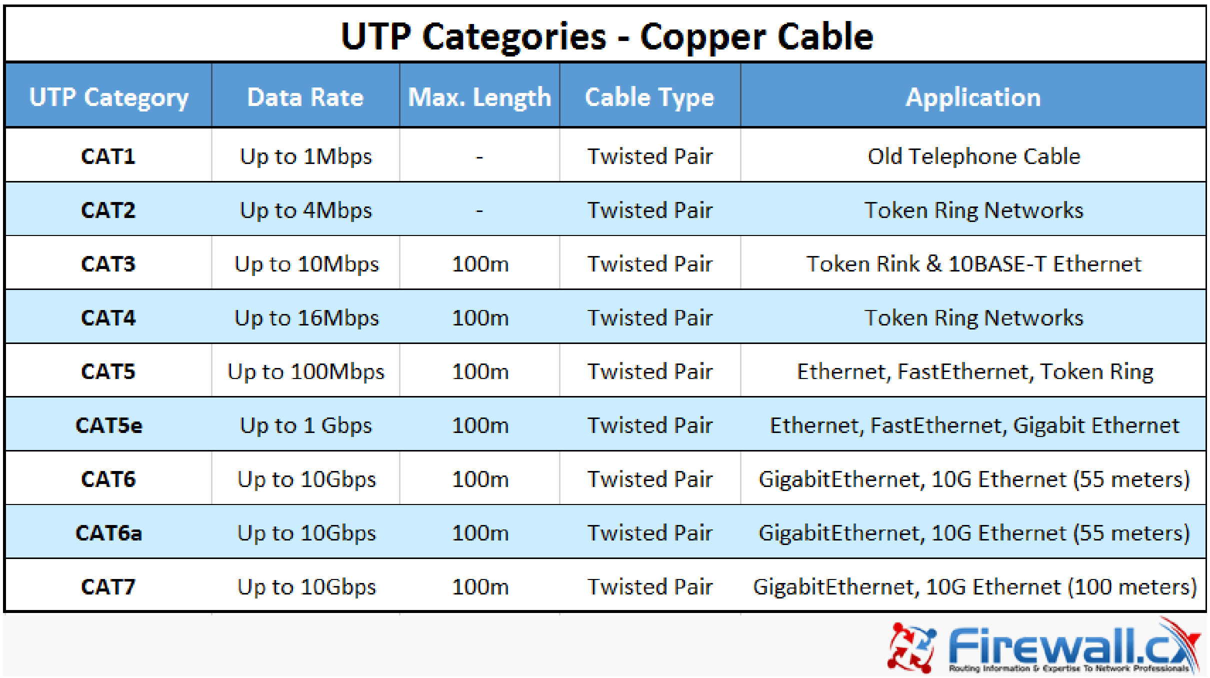
Twisted pair cables: UTP and STP

The most common form of copper cabling is twisted pair cable, found in nearly all modern ethernet deployments. This cable contains pairs of insulated copper wires twisted together to reduce electromagnetic interference.

There are two key varieties:

- **Unshielded twisted pair (UTP):** by far the most common, UTP relies on the twisting alone to resist crosstalk and electromagnetic interference. Its lightweight, flexible and cost effective, making it ideal for office environments.
- **Shielded twisted pair (STP):** this type adds a layer of foil or braided shielding around the pairs or the entire bundle to further reduce interference. STP is used in environments with high levels of electrical noise, such as factories or areas with a lot of machinery.

Twisted pair cables are categorised based on performance. These categories (Cat5 etc) define the maximum supported speed, bandwidth, and distance.



UTP Categories - Copper Cable

UTP Category	Data Rate	Max. Length	Cable Type	Application
CAT1	Up to 1Mbps	-	Twisted Pair	Old Telephone Cable
CAT2	Up to 4Mbps	-	Twisted Pair	Token Ring Networks
CAT3	Up to 10Mbps	100m	Twisted Pair	Token Rink & 10BASE-T Ethernet
CAT4	Up to 16Mbps	100m	Twisted Pair	Token Ring Networks
CAT5	Up to 100Mbps	100m	Twisted Pair	Ethernet, FastEthernet, Token Ring
CAT5e	Up to 1 Gbps	100m	Twisted Pair	Ethernet, FastEthernet, Gigabit Ethernet
CAT6	Up to 10Gbps	100m	Twisted Pair	GigabitEthernet, 10G Ethernet (55 meters)
CAT6a	Up to 10Gbps	100m	Twisted Pair	GigabitEthernet, 10G Ethernet (55 meters)
CAT7	Up to 10Gbps	100m	Twisted Pair	GigabitEthernet, 10G Ethernet (100 meters)

Firewall.cx Routing Information & Expertise To Network Professionals

Coaxial cable

Though largely phased out of mainstream ethernet use, coaxial cables still have a role in specific application like cable broadband or certain CCTV systems. Coax cables have a central conductor surrounded by insulation, a metallic shield, and an outer jacket. This structure gives them strong resistance to signal loss and interference.

The two main types you should be aware of are:

- RG-6: used primarily for internet and TV cabling
- RG-59: thinner and used for short runs, like video surveillance systems. Not used so much anymore
- DAC cable (direct attached copper): while not a traditional coaxial cable, DAC falls under copper media. It is a short-range, high-speed cable used to directly connect network equipment, such as switches and routers in data centres.

Twinaxial cable

Twinaxial cable is a high-performance alternative to standard coax or twisted pair, featuring two conductors twisted together inside a shielded cable. Its often used in high speed data centre interconnects, especially with storage networks or short-range switch to switch connections. Twinax is valued for its ability to deliver high speeds with low latency over short distances.

Plenum vs non-plenum cabling

Beyond performance, cable installations must also consider fire safety. Plenum rated cables have a fire-retardant jacket made from special low smoke material. These are required in building spaces where air circulates – like HVAC air ducts – because in the event of a fire, they emit less toxic smoke. Non plenum cables are cheaper but cannot be used in those same air handling spaces. Selecting the right type depends not on performance, but on building codes and safety regulations.

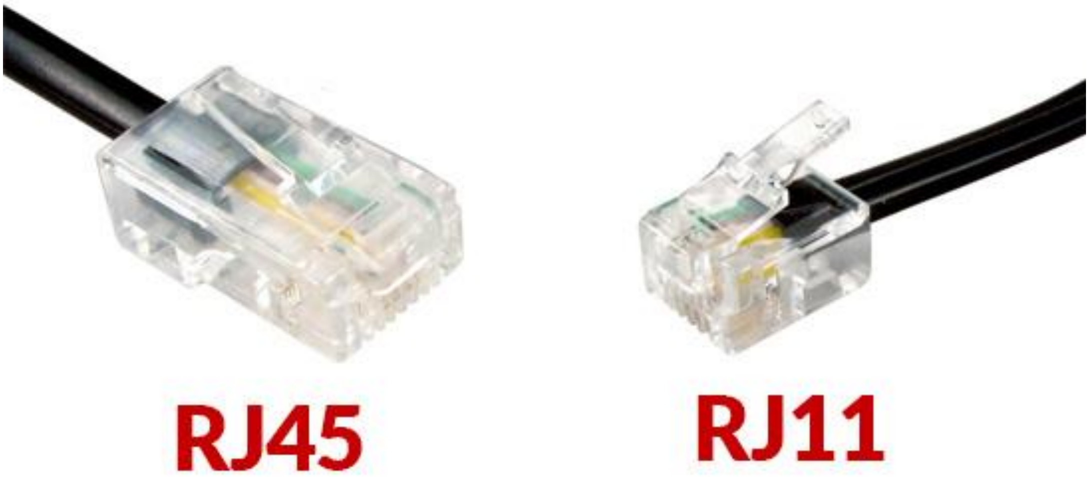
4.2 Copper network connections

While the cables themselves carry the signal, it's the connectors on each end that make copper media usable in real work networking. A connector must match the cable type, maintain electrical integrity, and be physically durable enough for repeated use. In this chapter, we explore the most common types of connectors used with copper cabling, their design characteristics, and the scenarios in which they are typically deployed.

RJ connectors (registered jack series)

The most widely used connectors in copper networking are part of the RJ family – specifically RJ-11 and rj-45. Despite looking similar, they are functionally different and used for separate tasks.

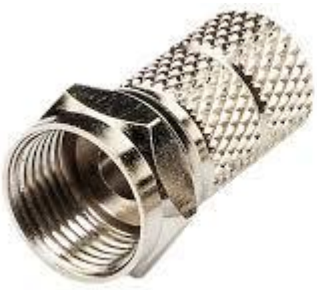
- **Rj11**: this has a 6 position, 2 pin layout (6P2C) and is typically used for telephone lines and legacy modem connections. Its much slower than ethernet, and supports analog connections. It's ideal for applications where speed is not a priority, due to its compact size which gives it narrow bandwidth.
- **Rj45**, on the other hand, is the standard connector for ethernet networking. It uses an 8 position, 8 contact (8P8C) interface, allowing for higher data throughput compared to rj11. Rj45 supports fast and gigabit ethernet standards and is found on the ends of Cat5e, Cat6, and Cat6a cables. It clicks into place securely and is used for both end user devices and patch panel or switch ports.



Coaxial connectors

For coaxial cabling, the type of connector depends on the cable and the function its serving.

- **F Type**



- **BNC (bayonet Neill-Concellman)** connectors are another common coax option. These connectors offer a quick connect and quick disconnect mechanism using a simple quarter turn bayonet mount. They are widely used in professional video, early ethernet (10BASE2), and radio frequency (RF) applications. BNC connectors are known for being secure, stable, and reliable, especially in environments where the cable might be move or unplugged frequently.

Summary and practical insight

choosing the right connector I as important as selecting the right cable. Connectors need to align with the cable type, the network speed requirements, and the environment in which they're deployed. Whether crimping RJ-45s not Cat6 cables or attaching BNCs to RG-59 for a surveillance system, understanding each connector's role ensures both physical durability and consistent signal performance.

4.3 Fiber media

Fiber optic cabling represents a leap forward in data transmission, offering speed, distance, and reliability far beyond the limits of non-copper based solutions. At the core of fiber technology is the use of light – not electrical signals – to transmit data. This fundamental shift gives fiber a unique set of strengths that make it indispensable in high performance and enterprise networks.

Core benefits of fiber

The transmission of light through glass or plastic fibres provides electromagnetic interference (EMI) immunity, making fiber especially useful in environments where interference from power lines, machinery, or other devices might compromise copper signals. Unlike copper, fiber optics ca span very long distances without experiencing significant signal degradation. This makes them ideal for connecting buildings, data centers, or distant switches across campuses and cities.

Fiber also delivers much higher data transfer speeds, often reaching 10-100GBPS . In fact, in many networks today, fiber isn't the bottleneck – other infrastructure components (like switches or routers) tend to limit performance before the fiber does.

Its superior bandwidth and low latency make it ideal for real time services, high volume data transfers, and modern applications like cloud computing and 4k/8k video streaming.

Limitations and drawbacks

Fiber is not without its challenges. It is more expensive, both in terms of the cables themselves and the equipment required to support them. Specialised transceivers, modules, and infrastructure upgrades often accompany fiber deployments. It’s also harder to work with than ethernet. Splicing, terminating, and testing fiber requires specialised tools and training, which can increase deployment time and costs.

Types of fiber media

There are two primary types of fiber optic cables: ingle mode and multi-mode. Both transmit light, but they differ in how the light travels, their core diameters, and their ideal use cases.

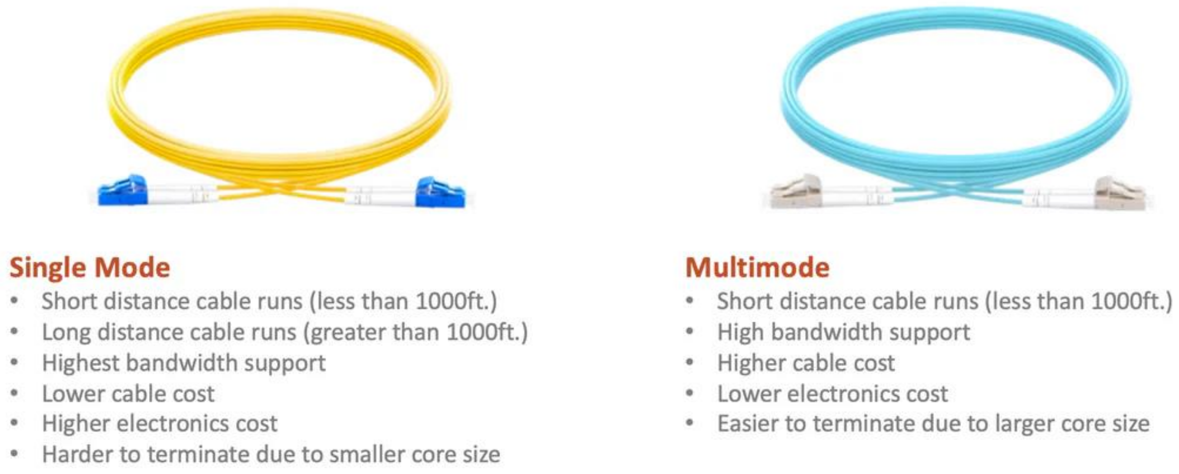
Single mode fiber (SMF)

Single mode fiber has an extremely narrow core – 9 microns in diameter – which allows for only one path of light to travel. Because the light moves in a straight line down the fiber, it can travel much longer distances without dispersion. This makes single mode fiber the go to choice for long range communications, such as metro area networks or backbone links between data centres. Its commonly sheathed in a yellow coat.

Multi- mode fiber (MMF)

Multi-mode fiber has a larger core, typically 50 to 62.5 microns, which allows multiple light paths to travel simultaneously. This results in more modal dispersion, limiting its effective range, but enabling cheaper and simpler equipment for short range communication. Because of its lower cost and ease of use, multi-mode fiber is often found in LAN environments, data closets, and short haul interconnections. Its commonly sheathed in an orange or blue coat.

In summary, fiber media forms the backbone of modern high-speed networking. Whether deployed in the form of long haul single-mode connections or dense short range multi mode links, its advantages in speed, immunity, and bandwidth make it an essential technology in any serious network infrastructure.



Single Mode

- Short distance cable runs (less than 1000ft.)
- Long distance cable runs (greater than 1000ft.)
- Highest bandwidth support
- Lower cable cost
- Higher electronics cost
- Harder to terminate due to smaller core size

Multimode

- Short distance cable runs (less than 1000ft.)
- High bandwidth support
- Higher cable cost
- Lower electronics cost
- Easier to terminate due to larger core size

4.4 Fiber connectors

Fiber network connectors and transmission

Fiber optic cabling transmits data as pulses of light rather than electrical signals. This allows for extremely high bandwidth and long-distance communication, with much lower signal degradation than copper.

Most fiber links use two strands of fiber:

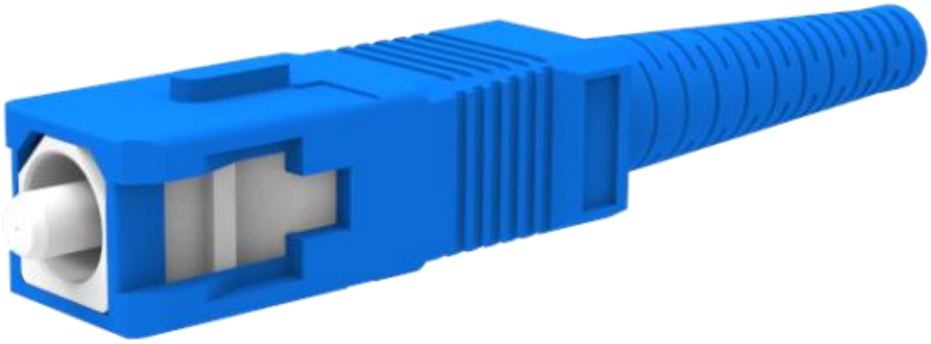
- One for transmitting data (TX)
- One for receiving it (RX)

Each end of the connection has its own connector type, and understanding these connectors is essential for working with fiber systems in enterprise networks.

SC connector

The SC, or subscriber connector, is one of the most established fiber connectors in use. It features a square shape and employs a simple push-pull snap in locking mechanism. SC connectors gained widespread popularity due to their low cost, ease of use, and durability. They are commonly found in both single mode and multimode fiber installations, particularly in older enterprise environments. Although newer connectors have since replaced SC in many modern deployments, it remains a recognisable and reliable option.

- Uses a simple push/pull snap in mechanism



LC connector

The LC, or lucent connector, is a smaller, more compact alternative to the SC. It features a latching mechanism similar in style to an RJ-45 clip, making it secure yet easy to release. What makes LC especially valuable is its small form factor – it occupies roughly half the space of an SC connector. This allows for much higher port density, which is critical in modern environments such as data centers and telecommunications rooms. LC connectors are now the standard in most single mode fiber networks due to their efficient use of space and consistent performance.

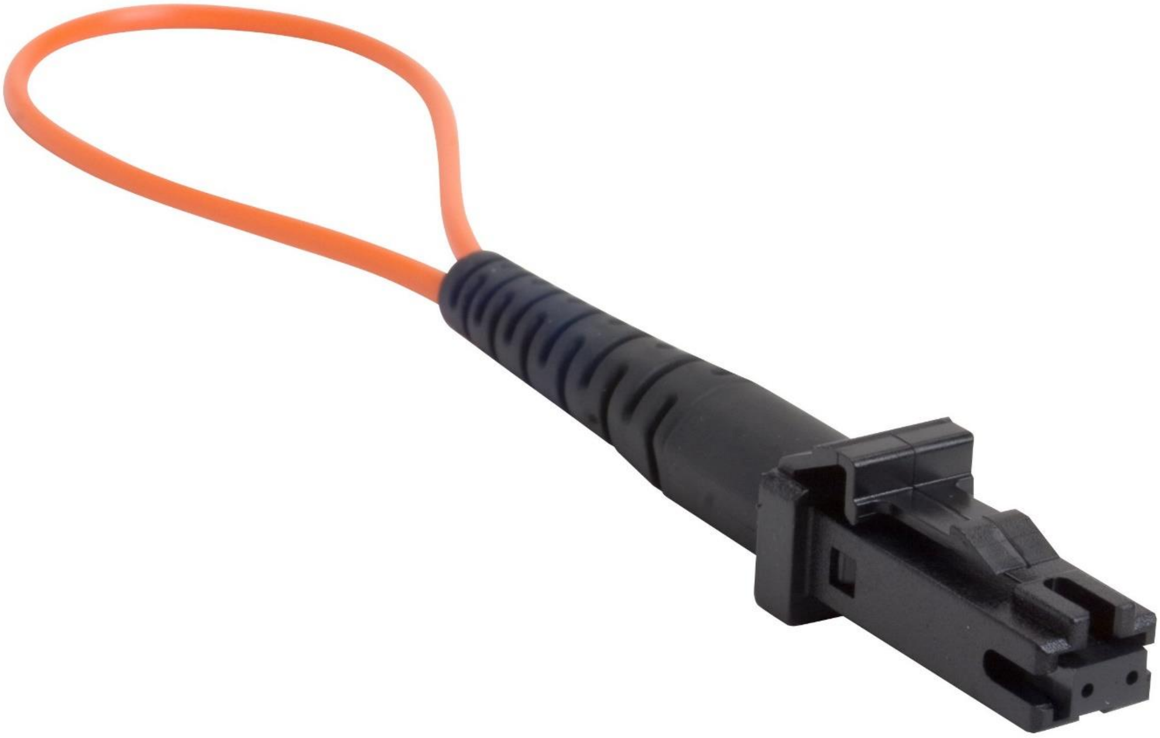
ST connector

The ST, or straight tip connector, is a round connector that uses a twist lock bayonet style mechanism. To secure the connection, the user inserts the connector and twists it until it locks into place. ST connectors were once widely used in multimode fiber networks, especially in campus and industrial environments. However, their bulkier design and more cumbersome installation process have led to a decline in popularity, especially where high density or easily managed connections are preferred.

MTRJ connector

The MTRJ, or mechanical transfer registered jack, is a compact fiber connector designed to carry duplex fiber within a single connector body. It visually resembles an ethernet port and uses a similar snap in locking mechanism. MTRJ connectors were intended to simplify installation and save space by combining both transmit and receive fibres into a single connector. Though not ass widely adopted as LC, MTFJ is still encountered in legacy systems and specialised network hardware.

- Combines transmit and receive fibres for a compact design (TX and RX)
- Resembles and ethernet jack



MPO connector

The MPO, or multi fiber pish on connector, is designed for high-capacity applications where multiple fiber strands need to be terminated in a single connection. It is rectangular and relatively wide, often carrying 12 or 24 fibers simultaneously. The MPO uses a push pull latch and includes precision alignment pins to ensure that each fiber is properly mated. This connector is commonly used in data centre backbone links and high-speed environments such as 40Gbps and 100Gbps s deployments, where parallel optics are required.

Back reflection and fiber polish types

In fiber optics, back reflection – also known as return loss- refers to the reflection of light back toward the source due to imperfections or gaps at connection points. This can degrade signal quality, particularly in single mode systems where precision is critical. To reduce back reflection, fiber connectors are polished using one of several techniques.

The standard physical contact (PC) polish uses a domed end face that brings the fiber cores into close physical contact. This reduces the air gap between connectors and minimises reflection. An enhancement of this is the ultra-physical contact (UPC) polish, which applies a finer finish to further lower the return loss. UPC connectors are commonly used in modern single mode systems for their reliable performance.

For applications that demand even stricter control over back reflection, such ass in telecommunications and high precision optical systems, angled physical contact (APC) polishing is used. APC connectors are polished at an 8-degree angle, causing any reflected light to be deflected into the fiber cladding rather than back into the core. This results in extremely low return loss, though it requires APC to APC mating and is not interchangeable with flat polished connectors.

4.5 Transceivers

A transceiver, short for transmitter-receiver, is a compact, modular device used to send and receive data over a network medium. Transceivers are used in both fiber optic and copper-based networks, and play a critical role in modern networking hardware, especially in switches, routers, and media converters.

Transceivers operate at layer 1 of the OSI modern, meaning they do not interpret or process the data itself 0 they simply handle the electrical or optical transmission of bits between devices. Their primary role is to bridge different physical media types or extend signal reach, all while maintaining protocol compatibility.

Transceiver function and use cases

Transceivers are often used to connect ethernet or fiber channel devices over various media types. Even when the underlying protocol remains the same – such as ethernet – a transceiver may be required to support transmission across different media. For instance, a switch with a transceiver port can use copper cabling for a short connection, a fiber optic cable for a longer distance connection, simply by swapping out the transceiver module

This flexibility is what makes transceivers so valuable: they convert network connections between different physical layers (e.g.) from electrical to optical), while keeping the logical data form intact. This allows a network administrator to build local networks over copper while extending backbone connections over fiber, all using the same switch or router.

In data centre environments, transceivers are also used to interconnect switches across long distances, particularly where high speed links are required and signal integrity must be maintained over hundreds of meters or more. For example, transceivers allow for the use of fiber channel for storage area networks (SANs), or ethernet for uplinks between switches.

Transceivers and media converters

Though they serve similar roles, it's important to distinguish between trasceivers and media converters. A media converter is typically a standalone device that bridges tow media types (such as copper to fiber), while a transceiver is a modular insert that plugs directly into a network device such as a switchport or router interface.

ts can be swapped in and out of ports depending on the desired connection type or distance. This modularity makes them extremely practica for scalable, flexible network design.

Transceiver form factors

over time, a range of transceiver form factors have been developed to support different speeds, physical sizes, and use cases. These form factors define the size of the module, the connector type, and the data rate range.

- SFP (small form factor pluggable4) supports speeds up to 1 Gbps. It is widely used for both fiber and copper connections in enterprise networks
- SFP+ is an enhanced version of SFP and supports speeds up to 10 Gbps. It is backward compatible with SFP modules but only at lower speeds
- WSFP (quad small form factor pluggable) supports 4 lanes of 1 Gbps, making it suitable for 4 Gbps applications
- QSFP28 supports 4 lanes of 25 Gbps, for a total of 100 Gbps. It is commonly used in high performance data centre networks
- QSFP56 builds upon the QSFP28 form factor and supports 4 lanes of 50 Gbps, totalling 200 Gbps of throughput. This form factor is used in extremely high bandwidth network environments, such as backbone cloud infrastructure.

Each of these transceiver types is hot swappable and physically distinct, though many are electrically backward-compatible within the same socket type. for example, an SFP module can typically be inserted into an SFP+ port, though the port will operate at the lower speed.

Section 5

5.1: Cable distribution systems

A cable distribution system refers to the organised cable infrastructure that carries network cabling from the central switching equipment out to end user devices throughout a building or campus. Rather than running individual cables directly from switches to wall jacks, distribution systems are designed to create scalable, serviceable, and logically segments networks using intermediate hardware and structured cabling.

The main purpose of a cable distribution system is to establish a centralised point of control and access for network connections, ensuring that cable runs are manageable, serviceable, and compliant with performance and safety standards. This system defines how devices like patch panels, punchdown blocks, and switches are connected to support structured cabling from a central location to distributed access points.

Main distribution frame (MDF)

At the heart of most structured cabling systems is the main distribution frame. This is the primary wiring hub where the core switches, routers, and internet service entry point are typically located. The MDF acts as a central demarcation point between the internal network and the outside world – often housing the buildings backbone switch or router.

Within the MDF, cabling originates and is routed either directly to patch panels or out to other parts of the building via backbone cables. In larger networks, the MDF distributes connectivity to one or more intermediate distribution frames (IDFs) throughout the facility.

Intermediate distribution frame (IDF)

An IDF serves as a satellite wiring hub, positions away from the MDF but connected to it via horizontal or vertical backbone cabling. This allows network signals to be distributed across multiple floors or wings of a building while maintaining manageable cable lengths.

Each IDF contains its own patch panels, switches, and cable management hardware. From there, connections are extended to individual rooms and workstations. This hierarchical structure reduces cable congestion in the MDF and simplifies troubleshooting and upgrades in localised areas.

Patch panels and punchdown blocks

At both the MDF and IDF levels, patch panels are used to organise and terminate cable runs. Cables from the wall jacks or keystone jacks terminate at the rear of the patch panel using punchdown tools. The front facing ports are then connected via patch cables to the network switches or routers. This setup allows technicians to easily reroute, relabel, or disconnect endpoints without disturbing the actual cable runs.

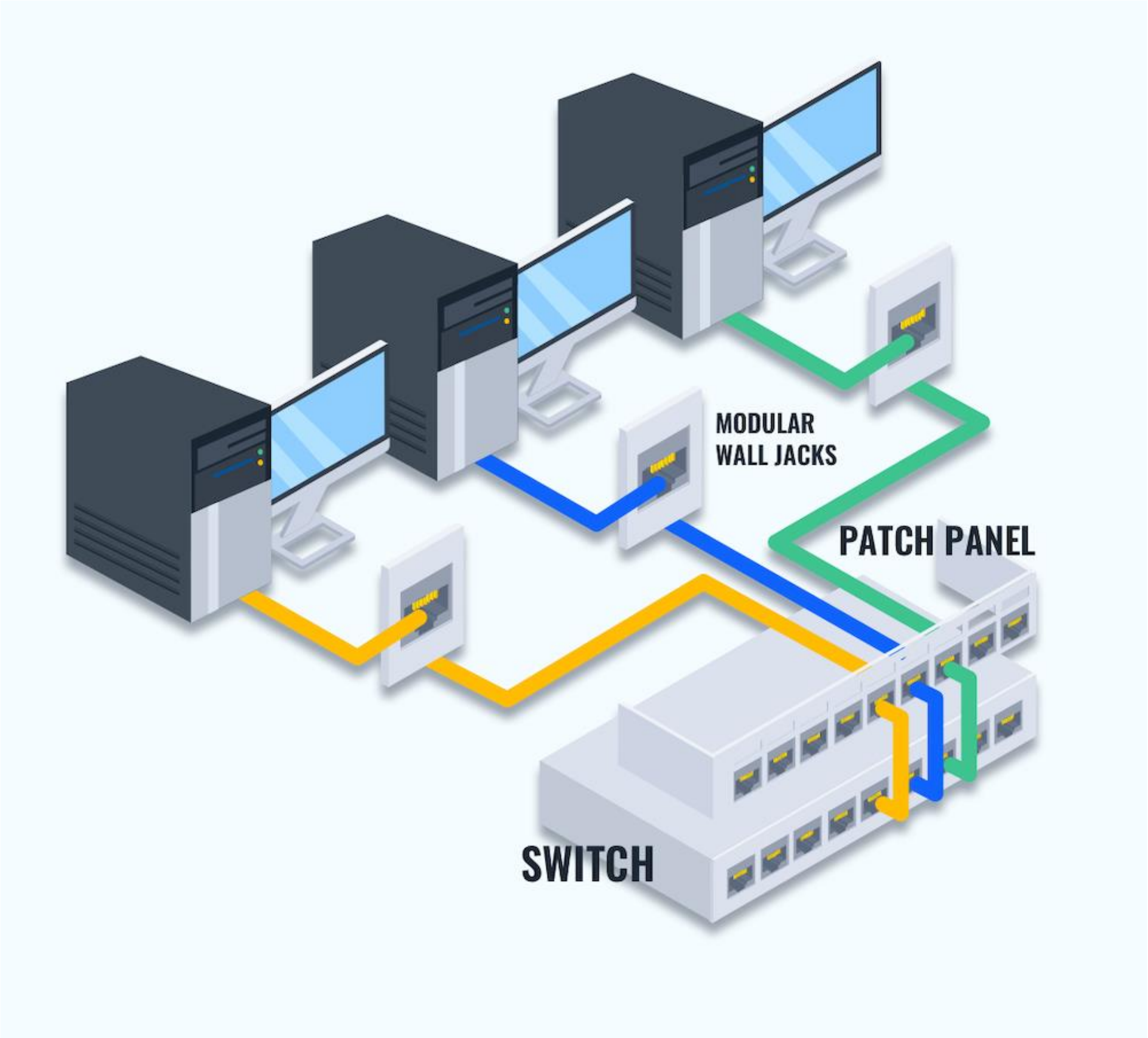
In some cases, especially for voice cabling or older installations, punchdown blocks may be used as part of the physical distribution layer. These blocks also serve as termination points but are more commonly associated with telephone systems or legacy networks.

Path of connectivity

In a typical structured cabling system, the flow of connectivity might look something like this:

A network switch inside the MDF or IDF connects to a patch panel via short patch cables. From the rear of that patch panel, ethernet cables run through cable trays or conduits toward the building’s walls or ceilings. These cables terminate at punchdown blocks or directly into a fibre distribution panel, depending on the media type. From there they are routed to keystone jacks mounted in wall plates, providing an access point for end user devices like computers, VoIP phone, or wireless access points.

The structured nature of this system means that upgrades, repairs and changes can be made with minimal disruption. It also allows for easier documentation and clearer physical labelling of ports and endpoints.



5.2 Power distribution systems

In any enterprise IT environment, reliable and regulated power distribution is just as important a network connectivity. A power distribution system is the infrastructure that ensures electricity is safely and effectively delivered to all computing equipment, networking gear, and support systems within a facility. These systems are essential for maintaining uptime, protecting equipment, and responding to power disruptions with minimal impact.

At its core, a power distribution system is concerned with how power is routed, regulated, backed up, and protected. It includes a range of components such as uninterruptible power supplies (UPS), power distribution units (PDUs), and, in larger facilities, backup generators. Each component serves a specific role in ensuring stable and continuous power delivery.

UPS (uninterrupted power supply)

The UPS is a central feature of most power distribution systems. It serves as an intermediary between the wall power source and the equipment, providing temporary battery power in the event of an outage. UPS units not only maintain operation during short term power loss but also offer surge protection and voltage regulation. This prevents damage to sensitive components during brownouts or voltage spikes.

UPS systems come in various forms, from small desktop units to large, rack mounted or even room sized solutions. In data centres, UPS arrays are used to sustain critical systems ling enough for backup generators to come online or for controlled shutdowns to be performed.

PDU (power distribution unit)

Once power enters the rack or enclosure, it is managed and distributed by a PDU. A PDU is responsible for delivering power from the UPS or wall outlet to each individual server, switch, or appliance in the rack. Basic PDUs are essentially high-capacity power strips, but more advanced models may include features like current monitoring, remote power cycling, load balancing, and environmental monitoring.

Some PDUs are metered, allowing admins to track real time power usage per outlet or phase. Others are switched, meaning individual ports can be toggled on or off remotely – useful for rebooting devices without physical access. In high density environments, intelligent PDUs help manage power more efficiently and reduce the risk of overloads.

Managing power loads

Proper power load management ensures that no component of the distribution system is overburdened. Each pdu and UPS has a maximum power rating, and equipment connected to them must be balanced to stay within those limits. Overloading a UPS can result in failure during outages, while overloading a PDU may trip circuit breakers or cause overheating.

Network admins must account for the wattage and amperage of all connected devices and distribute them across circuits accordingly. Power planning also includes allowing for future expansion – critical when provisioning racks in growing environments.

Voltage considerations: Europe VS USA

Different regions of the world supply electrical power at different voltages and frequencies. In the United States, standard power is delivered at 120 volts and 60 Hz. In much of Europe and other parts of the world, the standard is 230 volts at 50 Hz. This affects the selection and configuration of UPS units and PDUs, especially in global data centres or when importing equipment.

Many enterprise grade power devices are designed to be dual voltage or support automatic voltage switching. Nonetheless, its’s essential to verify voltage compatibility and connector types when installing equipment internationally. Failure to do so can lead to hardware damage, safety hazards, or unnecessary downtime.

5.3 HVAC systems and fire suppression

In data centers and other critical computing environments, both environmental control and fire safety are essential to ensure continuous, safe operation. HVAC (heating, ventilation and air conditioning) systems and fire suppression systems work together to maintain optimal conditions for servers, networking equipment, and storage devices.

Temperature control

Servers and networking devices generate significant heat during operation. If not managed, this heat can lead to equipment failures or reduced lifespan. HVAC systems must maintain a steady temperature, typically between 20-25 C, to ensure safe and stable performance. Sudden temperature spikes or drops can cause condensation or component instability.

Humidity control

Humidity plays a crucial role in protecting equipment from two major risks:

- Condensation: if humidity is too high, moisture can condense on internal components, potentially short-circuiting delicate electronics
- Static electricity: if humidity is too low, there is an increased risk of static discharge, which can damage sensitive circuits

To mitigate these risks, humidity should be maintained between 40 and 60%, which is considered ideal for most data centers.

Airflow management

Proper airflow is vital to the cooling process. Systems must be designed with both exhaust and intake paths to allow hot air to be efficiently removed and cool air to be evenly distributed. Techniques like hot aisle/cold aisle configurations and blanking panels can improve airflow efficiency and prevent hot air recirculation.

The correct orientation of server racks is port side exhaust, because it allows hot air from the server to be expelled efficiently, leading to better cooling. This orientation optimizes airflow and prevents the equipment from drawing in warmer air, which enhances overall performance and lifespan of the servers.

Fire suppression systems

Fires in data centers can be catastrophic. While fire prevention is a priority, suppression systems are necessary as a last line of defence. However, traditional water-based systems can destroy electrical equipment, so alternative methods are used.

- Pre action systems: these use a dry pipe approach where water is only released into the pipes once a fire is detected and confirmed. This reduces the risk of accidental leaks or damage from false alarms
- Suppression systems: specialised suppression systems are used in most data centres to avoid the use of water. Common methods include gas-based agents, though some gases can be dangerous to people and must be carefully deployed with appropriate safety measures

Section 6: Wireless networking

6.1 Wireless network types

Like wired, wireless networks come in different types, with each reflecting the need of the hosts on the network when it's made, and the limitations/abilities of its infrastructure.

Ad hoc wireless network

There are times when wireless internet is required where there is no underlying infrastructure to facilitate this. This includes temporary networks, such as networks involved in disaster relief, the sharing and presentation of data during conferences, or just a spontaneous network setup away from the home or office, or maybe in the field.

In this situation an ad hoc wireless network can be set up. This is where there is no centralised routing device and each node is linked to other nodes directly. They may all be linked to each other, or some nodes may only be linked to some other nodes; the connections are set up to facilitate whatever is required from the network.

This is a very dynamic network in nature, where devices can join or leave at any time. Each node can behave as a router too, by forwarding the data for other nodes. Its flexible and cost effective as no existing infrastructure is needed, but this topology doesn't offer strong security, network efficiency, redundancy and management, or the many other characteristics of organised network setups.

This type of connection can generally be set up through the device's network settings, with the option for ad hoc mode specified. These types of connections are typically meant for local sharing and collaboration, and not web access.

Infrastructure wireless network

Infrastructure mode is the alternative to ad hoc, and the standard for most permanent wireless networks. unlike decentralised ad hoc mode, it relies on centralised routing of data around the network via routers and access points.

The access point is the core feature of infrastructure networks. it is these that allow devices to connect to the network, and they handle the security authorisation, traffic management, and the general facilitation of data flow from network to host.

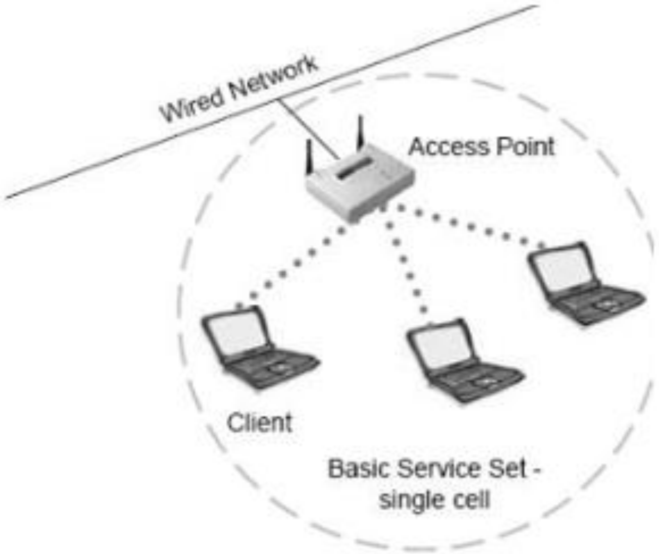
A BSS (basic service set) is one AP plus all the clients connected to it, so in other words a BSS is a mini network. Each BSS is identified by a BSSID, which is the MAC address of the AP's wireless interface. Two APs may broadcast the same SSID, but they have different BSSs.

Due to its organised, centralised nature, infrastructure mode provides:

- Strong security via authorisation at the WAP and more
- Efficiency of data flow via network management and organisation
- Redundancy in communications via multiple routes set up
- Smoother performance as all data is going through an AP, which is ensuring each device gets optimal bandwidth
- Easy management via network device interfaces

This is the standard wireless network type and can be found in schools, homes, cafes, and more.

To join an infrastructure wireless network, you'll have to select the SSID from your device's network settings, or search for it if it's not available. You'll then have to enter the password. Access points use BSSIDs for their own mini networks, distinguished from the SSID of the parent network.



When multiple APs are configured with the same security settings and SSID, they form an ESS – Extended Service Set. They are generally connected through a wired backbone or distribution system so they form a larger network, allowing for seamless roaming connection for users in the area covered by the APs.

Point to point wireless (P2P)

This is where a direct link between is set up between two points known as end points through the use of radio waves. They're often used for connecting buildings, CCTV cameras, sensors, and can be used for either short or long range connections. Antenna are installed and used to point the wireless signal at the corresponding antenna.

Typically, they're used where its inconvenient or overly costly to install cables to bridge the distance between the devices that need to be connected. It may be that at either end are wired networks, and the P2P connection is simply a wireless addition to an otherwise wired network. In this case, the data would be converted from ethernet to radio waves by hardware installed along with the antenna.

The distance these wireless connections can extend is dependent on the data throughput, with 200Mbps of data being sent across 70 Km, and then 10 Gbps of data being sent 5 Km. The throughput also determines whether they can penetrate trees, with throughputs of over 100 Gbps needing clear lines of sight. Wireless bridges can be used to extend the range just as in ethernet networks.

In this wireless network type the communication is highly efficient due to the bandwidth being dedicated to this singular connection.

Wireless mesh network

This is a wireless network made up of nodes (access points) organised in a mesh topology.

This means they aren't bound to the layout of a specific topology, whether it's having to deal with the next node point to point, or the hub in a spoke. Every node can connect to every other, providing fast and effective Wi-Fi over a large area, minimizing dead-zones. The nodes communicate too, updating one another on configurations and settings, providing a dynamic and non-hierarchical system where the best option for data flow is selected on the go by the nodes.

This different from having multiple Wi-Fi repeaters to extend coverage; in a mesh, the nodes behave as part of the same network, whereas additional repeaters simply extend the existing signal and are seen as another device on the network. Meshed APs share information not only on data flow, but configurations and settings, meaning they don't all have to be individually configured as in the case of Wi-Fi repeaters.

This is a decentralised network, without the hierarchies of gateways and switches and the like that may be found in more rigidly structured networks.

Due to these characteristics, a mesh network is resilient to failure as it offers multiple routes, so singular points of failure don't result in the entire network going down (good redundancy). However, due to each client acting as both a node and client, a large attack surface is available. To mitigate these risks, secure configuration, robust security practices and regular updates are crucial.

While ESS and mesh networks are similar in their use of multiple connected APs creating a seamless roaming area, they are different; an ESS uses multiple APs broadcasting the same SSID, all connected via some sort of backbone, whereas in a mesh there is no need for a backbone as the APs are all communicating with one another wirelessly to form the network. In ESS, the APs are subsidiaries of the network – in mesh, they *are* the network.

Autonomous vs lightweight access points

Access points can be either autonomous or lightweight. This section will discuss the difference between these and their use cases.

The autonomous AP is the most common type of AP, being the first that was developed. This stands alone as a network node, creating its own wireless network. These can broadcast and support multiple logical wireless networks with their own SSIDs, which can then be mapped to their own VLAN.

Advantages:

- Simple set up
- Independence: each operates on its own, so the failure of one AP doesn't impact others
- Cost effective: works well for small networks as no need for expensive wireless controllers and subscriptions

Disadvantages

- No centralised management; as each operates independently, they must be configured one by one; no central way for control
- Scalability challenges; more manual config and maintenance than if there was a central controller

Lightweight APs are the opposite to autonomous APs, and were developed to address their limitations. They rely on a central controller, which handles configurations and security policies of all the lightweight APs in a network, leaving the AP just to handle wireless data transmission.

Advantages

- **Centralised management**: unlike with autonomous APs, security policies and configurations don't need to be done for each AP, the controller handles it all
- **Scalability**: new lightweight APs automatically connect to the controller without manual set up needed, making scaling the network far more practical

- **Improved security**: the controller enforces security policies, ensuring consistent security across the network

When it comes to mesh and ESS networks, lightweight APs are more desirable due to the above mentioned advantages being beneficial to the network.

6.2 Wireless Antennas

Wi-Fi antennas convert radio frequency (RF) waves, which contain packets of information, into electrical signals, or electrical signals into RF. This conversion method permits wireless devices such as routers, smartphones, laptops, and tablets to communicate wirelessly. The receiving antennas of the devices pick up RF frequencies and transform them into electrical signals that the devices process. Transmitting antennas, on the other hand, convert electrical signals back into RF for transmission.

Transmission Process

- **Data Encoding**: A specific encoding scheme converts digital data from a device into electrical signals.
- **Signal Modulation**: The electrical signals are modulated onto a carrier wave. A carrier wave is a high-frequency radio wave.
- **Antenna Radiation**: The modulated carrier wave is sent into the Wi-Fi antenna. The antenna acts like a transducer by converting the electrical signal into RF frequencies that are subsequently radiated into the neighbouring environment.

Reception Process

- **RF Reception**: RF frequencies are transmitted by other devices or access points to the Wi-Fi antenna.
- **Signal Demodulation**: The RF is demodulated to obtain the original electrical signals. This process includes separating the modulation carrying the data from the carrier wave.
- **Data Decoding**: Applying the same encoding scheme that was used for transmission, the electrical signals are decoded back into digital data.

Factors That Affect Wi-Fi Antenna Performance

- **Antenna Gain**: This measures an antenna’s capability to amplify a signal. Higher gain antennas give stronger signals and longer range.
- **Antenna Placement**: Proper placement reduces obstruction and interference, enabling a strong signal reception and transmission.

- **Environmental Factors**: Obstructions like walls and furniture can weaken Wi-Fi signal transmissions. By using strategic antenna placement and high-gain antennas, such obstacles can be mitigated.

Omnidirectional antennas

This type of antenna broadcasts their signal uniformly in all directions, so is used when a device needs to provide a wide area of coverage. These means these are used in access points (including in routers with wireless access points inside) and mobile phones.

They're ideal for day-to-day use like with user devices and home/public Wi-Fi.

Unidirectional antennas

These types of antennas focus their signal in one precise spot in one direction, and so are used for point to point wireless connections, such as between two buildings in a campus or metropolitan area network. They may even be used across many miles to avoid the need for extensive cabling.

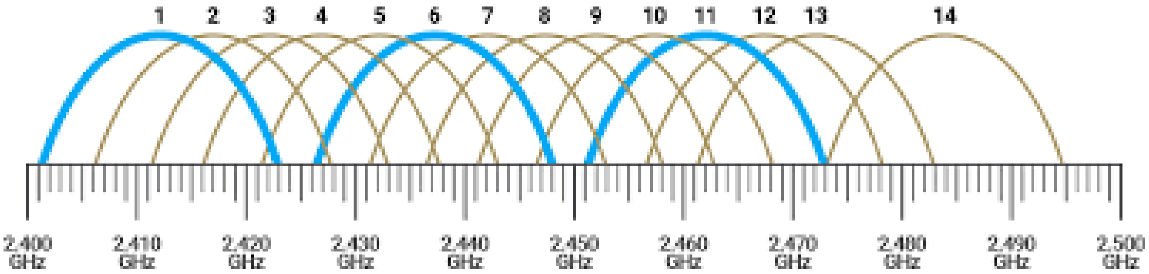
A type of unidirectional antenna is Yagi. These are excellent for long distance connections as the design of the antenna makes them good at picking up signals from the direction they're pointed. These are the go-to choice when creating long distance links to remote places where the signal needs to go to a specific point.

6.3 Wireless Frequencies

Like how wired communication relies on cables for sending data, wireless communication relies on radio frequencies. These are the physical medium on which the data travels, with different portion of the frequency band acting as different cables. Different bands offer different speeds and coverage, with the available portions being controlled and allocated by government agencies.

2.4 GHz range

This is an image of the 2.4 GHz range.



Here it can be seen how the 2.4 range is broken into 14 bands, with each band consisting of a portion of the range. Most of the bands overlap, with only four – 1, 6, 11 and 14 – not overlapping with one another at all; overlapping of channels causes the data on one band to interfere with the data on another, resulting in poorer network performance. This means that channels 1, 6, 11 and 14 are the safest choices when configuring a network.

2.4 GHz range This is the most widely used band as it offers good range and penetration through solid objects. It spans from 2.400 to 2.495, with each band (1,2,3 etc) using 22 of the gigahertz within the 2.4 megahertz range.

The regulations on available bands vary by location, with north America only have 1 through 11 available, Japan 1 through 14, and most other places 1 through 13.

5 GHz range

The 5 GHz range was introduced to address the limitations caused by the overlapping nature of the 2.4 GHz range, with 5 GHz being a larger frequency and offering 24 non-overlapping channels of 20 MHz each. This allows for less interference and therefor better performance, meaning this band is better for applications requiring higher performance. The drawback is this band has shorter range and poorer penetration so it's less suited to applications involving longer distances but better suited to those requiring high transfer speeds.

This range allows channel bonding, where two or more of the 20 MHz channels are combined into a wider channel, with the maximum being 8 channels combined to form a band of 160 Mhz. This is the equivalent of making a cable wider so more data can travel on it at once, so channel bonding may be helpful when lots of data needs to be transferred.

6 GHz range

This channel was opened up for use by newer Wi-Fi protocols, further adding more channels and available bandwidth. It allows for less congestion and higher transfer speeds so is ideal for applications with lots of data transfer that needs to be fast.

6.4 Wireless Standards

The standard at which data is sent across the above discussed channels has been an ongoing development over the years.

The first development was 802.11, which had a maximum data rate of 2Mbps, which was too slow for most applications so was quickly replaced, however it laid the groundwork for the subsequent standards. Each built upon this standard in various ways.

Wi-Fi 1, 802.11b, and Wi-Fi 2, 802.11a, were developed around similar times and addressed various needs while both improving on the too slow 802.11. b offered better range by using the 2.4GHz band but slower speeds, while a offered higher speeds at the trade off of shorter range by utilising the 5GHz band.

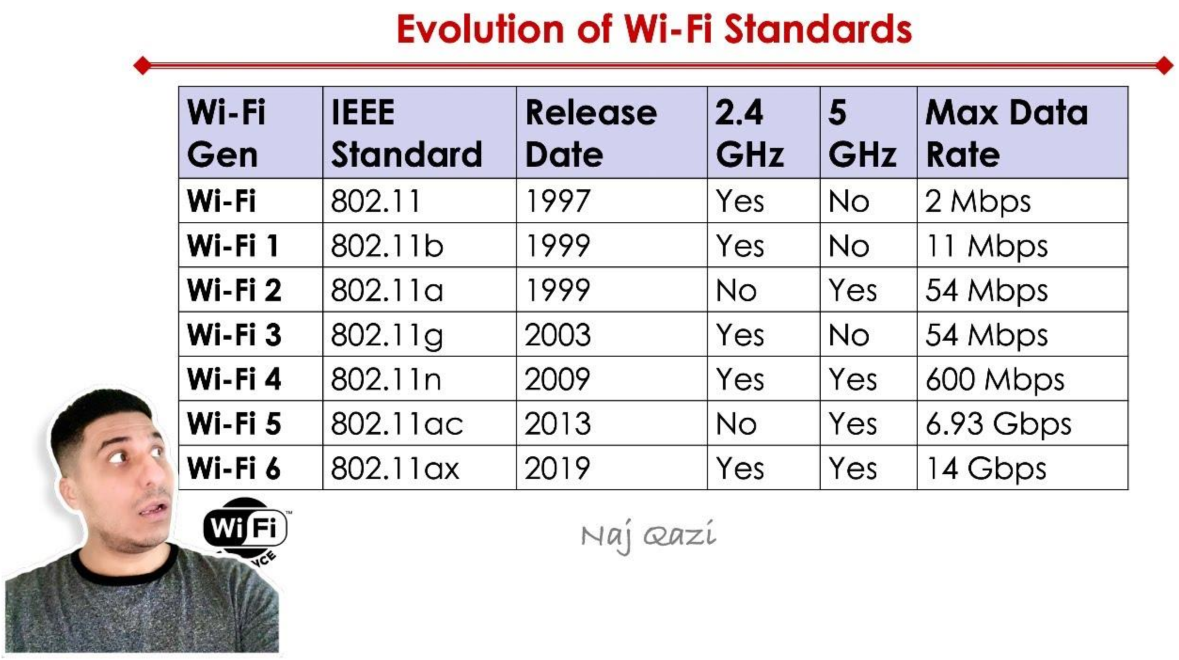
802.11g (Wi-Fi 3) improved on these standards by combining their traits; it used the 2.4GHz band like b, but had a higher speed (54Mbps) like a. This was important because it facilitated backwards compatibility between devices while ensuring the higher data transfer speeds necessary for evolving technology.

802.11n (Wi-Fi 4) made a huge improvement by introducing the technology MIMO (multi-input, multi-output). This used multiple antennas to allow for multiple data streams at once, allowing for a huge jump to transfer speeds of 600Mbps. It also provided improved range, as well as backwards compatibility with all previous standards by using both the 2.4 and 5GHz bands.

802.11ac (Wi-Fi 5) introduced MU-MIMO (multi-user multi-input multi-output), which amplified the benefits of MIMO by allowing for multiple users to utilise MIMO at once. This again drastically improved the data speeds allowed by Wi-Fi 5. This standard operates only in the 5GHz band and so does not offer backwards compatibility.

Wi-Fi 6 (802.11ax) introduced another technology, OFDMA, which increased speeds further. This standard also utilised both 2.4 and 5Ghz, as well as the newly in use 6GHz band, meaning it offers backwards compatibility and access to a whole new band for communication.

Specifics on each standard can be seen in the chart below.



Evolution of Wi-Fi Standards

Wi-Fi Gen	IEEE Standard	Release Date	2.4 GHz	5 GHz	Max Data Rate
Wi-Fi	802.11	1997	Yes	No	2 Mbps
Wi-Fi 1	802.11b	1999	Yes	No	11 Mbps
Wi-Fi 2	802.11a	1999	No	Yes	54 Mbps
Wi-Fi 3	802.11g	2003	Yes	No	54 Mbps
Wi-Fi 4	802.11n	2009	Yes	Yes	600 Mbps
Wi-Fi 5	802.11ac	2013	No	Yes	6.93 Gbps
Wi-Fi 6	802.11ax	2019	Yes	Yes	14 Gbps

Naj Qazi

6.5 Wireless Security

The nature of wireless networks means there are various attack surfaces that are vulnerable. Wireless networks are protected by security standards which have developed over the years and provide varying levels of protection.

The main aspects of wireless security are client authentication on logging into a network, and security of traffic by the access points via encryption.

The two main mechanisms by which these are ensured are:

- Pre-Shared key (PSK); this is a shared password used by both the APs and clients
- Enterprise Authentication System: where individual credentials are managed by a centralised security server

Pre shared keys are appropriate in home setups where one; there are only a handful of users of the Wi-Fi and two; there aren't masses of sensitive data that is likely to be the target of cyber-attacks. Were a PSK to be used in an enterprise environment, the PSK would have to be changed, and therefor all users and their devices updated, when an employee leaves the company. There is also a lack of accountability when it comes to a PSK, as it's unclear who logged into the network at a given time; such tracking is allowed for by using a centralised security server. It's simply not practical for companies with many employees, in terms of logistics or security considerations.

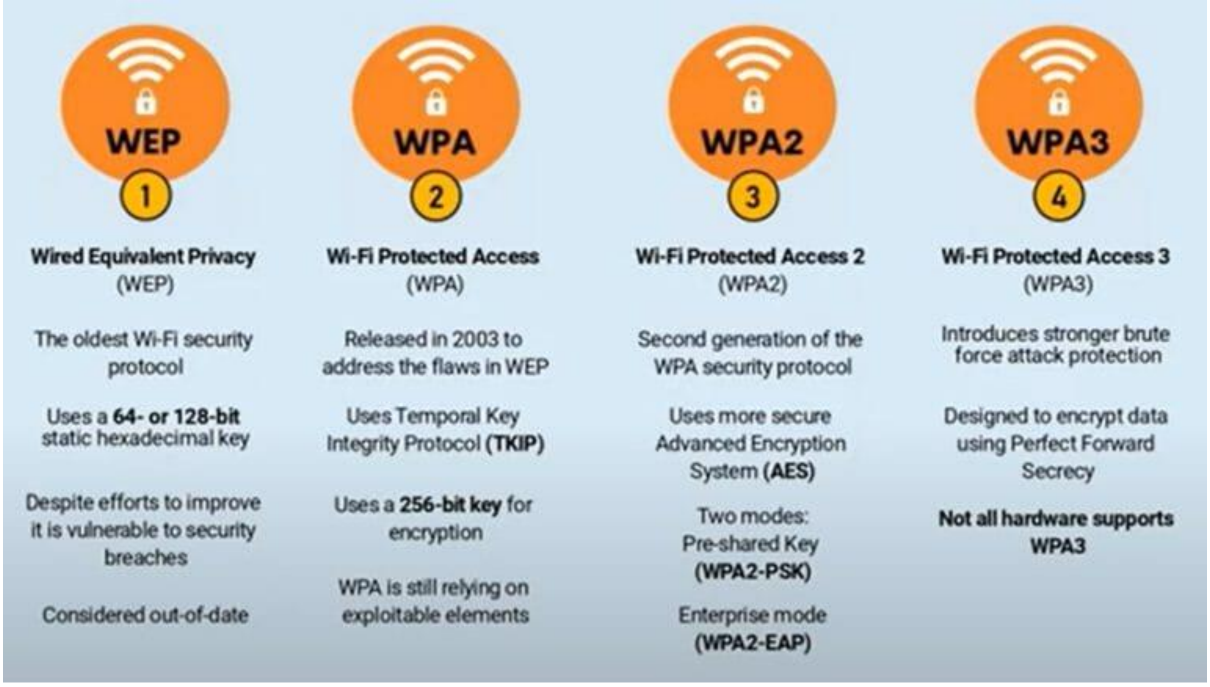
PSK limitations:

- Not scalable

- No user specific accountability
- Security risks (employee leaving with password knowledge)

The 802.1x standard is a security standard that addresses these limitations and relies on an authentication server such as RADIUS, which provides user specific credentials. This means user specific passwords can be deactivated when an employee leaves and user log ons are trackable.

Wireless security and encryption protocols



Wired Equivalent Privacy (WEP)

The oldest Wi-Fi security protocol

Uses a **64- or 128-bit** static hexadecimal key

Despite efforts to improve it is vulnerable to security breaches

Considered out-of-date

Wi-Fi Protected Access (WPA)

Released in 2003 to address the flaws in WEP

Uses Temporal Key Integrity Protocol **(TKIP)**

Uses a **256-bit key** for encryption

WPA is still relying on exploitable elements

Wi-Fi Protected Access 2 (WPA2)

Second generation of the WPA security protocol

Uses more secure Advanced Encryption System **(AES)**

Two modes: Pre-shared Key **(WPA2-PSK)**

Enterprise mode **(WPA2-EAP)**

Wi-Fi Protected Access 3 (WPA3)

Introduces stronger brute force attack protection

Designed to encrypt data using Perfect Forward Secrecy

Not all hardware supports WPA3

Wi-Fi Protected Setup (WPS)

WPS is not an encryption or integrity protocol but a network security standard designed to simplify the setup of secure Wi-Fi connections. It allows users to connect using a PIN, push button, or Near Field Communication (NFC).

Typically, WPS is implemented via a push button on the wireless access point, enabling users to connect without manually entering complex passwords.

However, WPS has a significant vulnerability in its eight-digit PIN, which is effectively reduced to two four-digit halves, drastically lowering the number of possible combinations from 100 million to about 20,000. This makes it susceptible to quick brute force attacks.

Due to this vulnerability, it is recommended to disable WPS in networks requiring high security.

In summary, WPA3 is the gold standard for wireless security. WPA2 is usable, but none of the others are secure, and any network using any other should be changed.