Computer Networks

Sai Krishna

Contents

1 History of Internet

1.1 Dial-up Connections

Long before the internet as we know it today came into existence. **Dial-up connection** was used to exchange messages with each other. A dial-up connection uses public switched telephone networks (PSTN) or also called as Plain Old Telephone Service (POTS) for data transfer. Transferring data across a dial-up connection is done through devices called **modems** (modulator/demodulator)

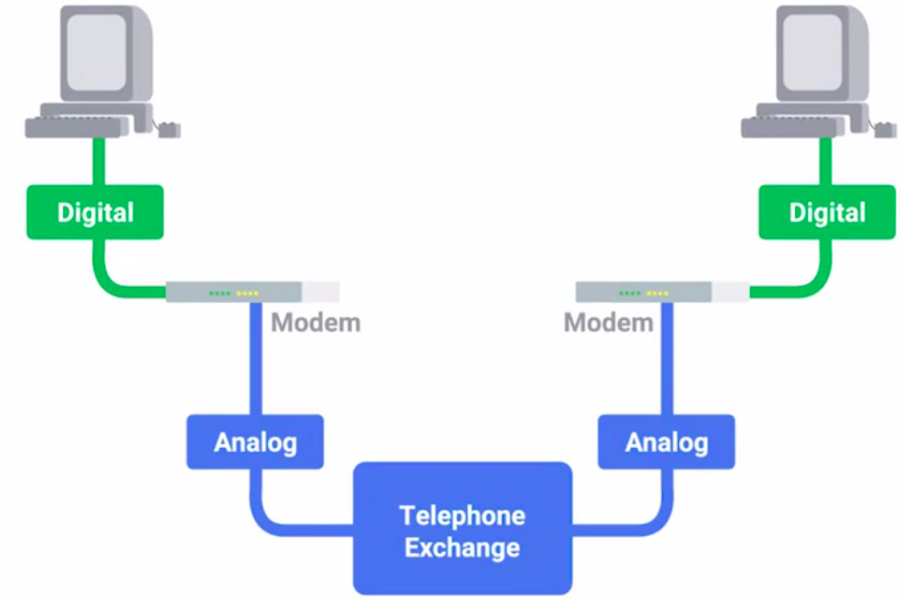


Figure 1: modem

1.2 Broadband

Any connectivity technology that isn't dial-up internet is called broadband. Broadband is much faster than dial-up internet and it refers to connections that are always on. To improve the data transfer rates, T-carrier technologies[1] were used. T-carrier technology requires dedicated lines which make them more expensive. The other commonly used broadband solutions are

- Digital Subscriber Line (DSL)
- Cable broadband
- Fiber connections

[1] T-carrier technology transmit multiple phone calls over a single link

1.2.1 T-carrier technology

First invented by AT&T in order provision a system that allowed multiple phone calls to travel across a single cable. Initially only one telephone call was being transmitted over one piece of twisted copper wire but with **Transmission System 1 (T1)** the first T-carrier specification, 24 simultaneous calls could be transmitted across a single piece of twisted copper wire. Later the same was also used for data transfer instead of transferring voice signals, each of the 24 phone channels was capable of transmitting data at rates of 64kbps. Hence a single T1 line was capable of transmitting data at rates of 1.544Mbps.

Further T3 lines were introduced. 28 T1 lines were multiplexed to form a single T3 line, with this data rates upto 44.736Mbps was achieved. But now all these have been replaced with other broadband solutions.

1.2.2 Digital Subscriber Lines (DSL)

With the ever increasing need for higher data rates, research showed that twisted copper wires are capable of transmitting data at much higher rates than what was needed for voice-to-voice calls by operating at a frequency range that did not interfere with normal phone calls, this technology is called **DSL**. DSL allows data transfer and voice phone calls to occur at the same time on the same line.

Instead of modems DSL uses DSLAMs (Digital Subscriber Line Access Multiplexer). Unlike dial-up connections DSL connections are long-running i.e. the connection is established when DSLAM is powered on and isn't turned down until the DSLAM is powered off.
The commonly used types of DSL are

1. **ADSL (Asymmetric Digital Subscriber Line)**
   They support faster downloading speeds and slower uploading speeds.
2. **SDSL (Symmetric Digital Subscriber Line)**
   Supports the same upload and download speed.
3. **HDSL (High bit-rate Digital Subscriber Line)**
   Supports data rates greater then 1.544 Mbps

1.2.3 Cable Broadband

Coaxial cables used for televisions, were capable of transferring data at much higher rates than what they were being used for. By using frequencies that dont interfere with television broadcast. This technology is called **Cable broadband**. The connections are usually managed by cable modem (at consumer end) and connects it to **cable modem termination system (CMTS)**. CMTS connects lot of different cable connections to an ISPs core network

1.2.4 Fiber communication

The maximum distance over which copper cables can be used for data transfer is limited, as the signal gets attenuated. So for long distance communication optic fibers are used which use light to carry data instead of electrical signals. As light is used to carry data, attenuation is very less.

Instead of modems the demarcation point for fiber communications is called **Optical Network Terminator (ONT)**. ONT converts data from protocols the fiber network can understand to those that twisted-pair copper networks can understand.

The commonly used protocols for broadband connections are **Point-to-Point Protocol**(IETF RFC1661) and **Point-to-Point Protocol over Ethernet**(IETF Informational RFC2516)

1.3 Connecting Computers to form a network

Initially computers were connected using a RG58U coaxial cable and a T connector. The main drawback with this was, if one of the terminators for cable could not be detected during communication, then the entire network would come down. To overcome this hubs were used, and the network topology shifted from BUS to a Star topology, where all the nodes would be connected to the HUB through a RJ45 cable. With this setup if one connection is down, it wont affect the entire network.

Computer networking is the name given to the full scope of how computers communicate with each other.

2 Jargons

- **Protocol**: A defined set of standards that computers must follow in order to communicate properly.
- **Network model**: A model is a way to organize a system's functions and features to define its structural design. A design can help us understand how a communication system accomplishes tasks to form a protocol suite. TCP/IP is the prevalent architecture today
- **Internetwork**: A collection of networks connected together through routers, the most famous of these is the **internet**.
- **Crosstalk:** When an electrical pulse on one wire is accidentally detected on another wire
- **Payload**: This is the actual data being transported.
- **Internet Assigned Numbers Authority (IANA):** A non profit organization that helps manage things like IP address allocation, Autonomous System Number (ASN) allocation
- **Internet Engineering Task Force (IETF):** An open community charged with developing and maintaining the standards required for the internet to continue to operate.
- **Wireless Access Point** is a device that bridges the wireless and wired portions of a network.
- **Network Fabric** is an industry term that describes a network topology in which components pass data to each other through interconnecting switches.
- **Network orchestration** refers to automating interactions across multiple types of devices, domains, and even potentially other related systems in the network.

2.1 Enterprise Campus

The main office in case of MAN which holds most of the corporate resources is called Enterprise Campus. Enterprise Campuses also typically include a separate Data Center which is home to the computational power, storage, and applications necessary to support an enterprise business.

Many corporate environments require deployment of wireless networks on a large scale and they use **Wireless LAN Controllers (WLC)** for centralizing the management of wireless deployments.

2.2 Web and Internet

Internet is a global network that interconnects various networks and hence provides a world-wide communication infrastructure.

The World Wide Web decribes one way to provide and access information over the internet using the web browser. It is a service that relies on the connections provided by the internet.

2.3 RFC

Since internet is a global connection, for various networks to communicate with each other, they should follow some set of rules common to all networks that constitute the internet. These rules (also called **protocols**) are defined by a set of documents called **Request for Comments (RFCs)**.

Part I
Basic Concepts of Networking

3 Network Models

3.1 TCP/IP model

#	Layer Name	Protocol	Protocol Data Unit	Addressing
5	Application	HTTP, SMTP, etc..	Messages	n/a
4	Transport	TCP/UDP	Segment	Port #'s
3	Network	IP	Datagram	IP address
2	Data Link	Ethernet, Wi-Fi	Frames	MAC Address
1	Physical	10 Base T, 802.11	Bits	n/a

Figure 2: TCP/IP model

- **Physical Layer**: Represents the physical devices that interconnect computers.
- **Data link layer**: Responsible for defining a common way of interpreting the signals used in the physical layer, so network devices can communicate.
  - A well known protocol of the data link layer is **Ethernet**, ethernet standards specify physical layer attributes and also define a protocol responsible for getting data to nodes on the same network or link.
- **Network layer**: Allows different networks to communicate with each other through devices known as routers.
  - Data link layer is responsible for transmitting data across a link, network layer is responsible for transmitting data across multiple networks
  - The most common protocol used in this layer is **Internet Protocol (IP)**. IP is the heart of the internet.

- **Transport layer**: Sorts out which client and server programs are supposed to get that data. i.e. We may request for a webpage and simultaneously send an email on a different tab, we receive data packets corresponding to these applications simultaneously, the transport layer determines which packets to be sent which programs.
  - The most common protocol used in this layer are **Transmission Control Protocol (TCP)**, **User Datagram Protocol (UDP)**
  - The network layer is responsible for transmitting data from one node to another. But it is the task of Transport layer to ensure the data packets reach the desired program on the node.
- **Application layer**: This is an abstraction layer with which the user interacts.

Figure 3: Analogy for the different layers of the network model

3.2 OSI Model

The **Open Systems Interconnection (OSI)** reference model was a widely used network model, and the original objective is was to

Provide a set of design standards for equipment manufacturers so they could communicate with each other.

The OSI model has 7 layers, each of which has a different level abstraction and performs a well defined function. The prinicples that were applied to arrive at the 7 layers are:

- A layer should be created where a different level of abstraction is needed
- Each layer should perform a well defined function
- The function performed by each layer should be compatible with the internationally standardized protocols
- The layer boundaries should be chosen to minimize the information flow across the interfaces.

The 7 layers of OSI model are

1. Application
2. Presentation
3. Session
4. Transport
5. Network
6. Data link
7. Physical

Session and Presentation are the two additional layers in the OSI model as compared to the TCP model. The functionalities of these two layers are incorporated within the Application layer of the TCP model.

Session Layer facilitates the communication between actual applications and transport layer. It is part of the operating system that takes the application layer data that is unencapsulated from all the layers below it and hands it to presentation layer.

Presentation Layer is responsible for making sure that the unencapsulated application layer data is able to be understood by the application. This layer handles compression/encryption of data.

3.3 De encapsulating the received packet

As the data goes down from the Apllication layer to the physical layer various headers are added at each layer to ensure the contents are not corrupted and also contain some instructions on how to read the data at the receiving end.

At the receiver the data in the form of electrical signals/ light is received at the physical layer. Which is then sent to the data link layer where the Ethernet header is read, and device identifies its MAC address and further sends the data to the network layer. But to identify which protocol of the network layer was used at the transmitting end, the ethernet header contains **ethertype (0x0800 for IP** which indicates which protocol of the network layer was used. The network layer header consists of a **protocol number (6 for TCP, 17 for UDP)** that indicates which transport layer protocol was used. Further in the transport layer header there is a field called **port number (80 for HTTP)** which indicates the protocol used in the application layer.

4 Basics of Networking Devices

4.1 Cables

Cables connect different devices to one another, allowing data to be transmitted over them. Two main categories of network cables are

- Copper cables: Different voltage levels are used to represent the binary data, and they are transmitted through these copper cables. The most common forms of Copper twisted-pair cables used in networking are: **Cat5, Cat5e, Cat6** cables
  - The way in which the twisted-pairs inside the cable are arranged has a significant impact on the data transfer rates.
  - Cat5 (Category 5) cables introduce a lot of crosstalk and hence they are mostly replaced by Cat5e or Cat6
  - Cat6 cables offer a more reliable and faster data transfer, but one limitation with Cat6 is that they have shorter maximum distance when used at higher speeds.
- Fiber: Contain individual optical fibers, which are tiny tubes made out of glass about the width of a human hair. Pulses of light are used to represent the binary data.
  - In environments where there is a lot of Electromagnetic interference, optic fiber cables are preferred as these EM interferences can affect the data being transferred via copper cables.
  - Fiber cables are more expensive than copper cables, at the same time they transfer data at a faster rate as compared to copper cables.

4.2 Hubs and Switches

Cables allow you to form point to point networking connections.
Hub is a physical layer device that allows for connections from many computers at once. Every computer connected to the hub will be able to communicate with every other computer at the same time. Its up to each system connected to hub to determine if the incoming data was meant for them or to ignore it. This causes a lot of noise on the network and creates a collision domain[2] Hence hubs are not that preferable.

[2]Collision domain is a network segment where only one device can communicate at a time. If multiple systems try sending data at the same time, the electrical pulses sent across the cable can interfere with each other.

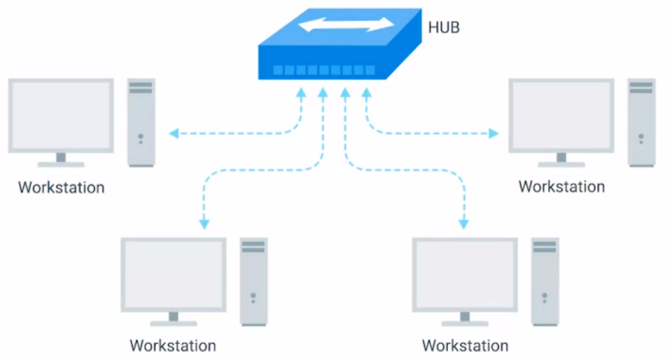


Figure 4: Hub

Network switch is very similar to a hub, many devices can be connected to it. Hub is a physical layer device but a switch is a data link layer device. Hence the switch is much smarter than a hub , it can determine to which system the data is to be sent (by inspecting the ethernet data) and sends the data to only that particular system. This reduces the size of collision domains on a network. A switch remembers which devices are connected on each interface, while a hub does not.

Hubs and switches are the primary devices used to connect computers on a single network, usually referred to as LAN (Local Area Network)

4.3 Routers

A router is a device that knows how to forward data between independent networks. The router is a network layer device, it can inspect the IP data to determine where to send things. Routers store internal tables containing information about how to route traffic between lots of different networks all over the world.

The main goal of the routers we use at are houses is to transfer the data from within the home to the Internet Service Provider (ISP). In the ISP a more sophisticated router takes over. The routers used at our houses will not have a very huge routing table, but the routers used in the ISP will have giant routing tables.

A **core router** used in the ISP will have to deal with a lot of traffic and will have to take complex decisions regarding where to send the traffic. This router will have many connections to various other routers spread across the globe.

Routers share data with each other via **Border Gateway Protocol**, which lets them learn about the most optimal paths to forward traffic.

5 Physical Layer

Its main concern is about transferring bits across the cable, it has nothing to do with what information is contained in the bits. Varying voltages to represent 1's and 0's is called **Line coding**. This allows devices on either end of the link to understand that an electric charge in a certain state is a zero and another state is 1.

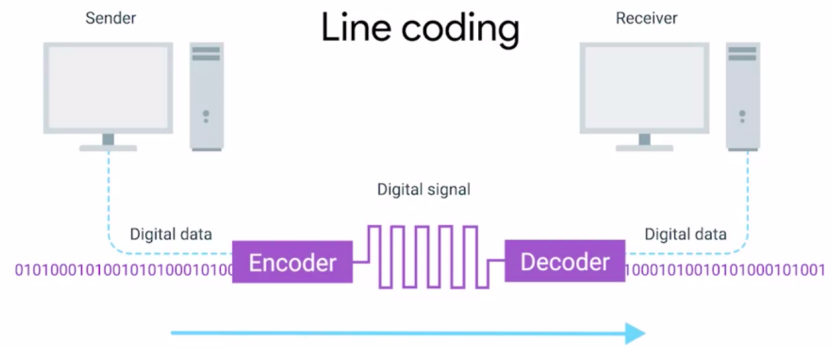


Figure 5: Line coding

The copper wires used to transmit these bits (in the form of voltages) are twisted copper wires. These copper wires are twisted to protect the data from Electromagnetic interference and crosstalk. These cables allow **Duplex communication**. One pair of wires are used for communicating in one direction, and the remaining pair is used for communicating in another direction.

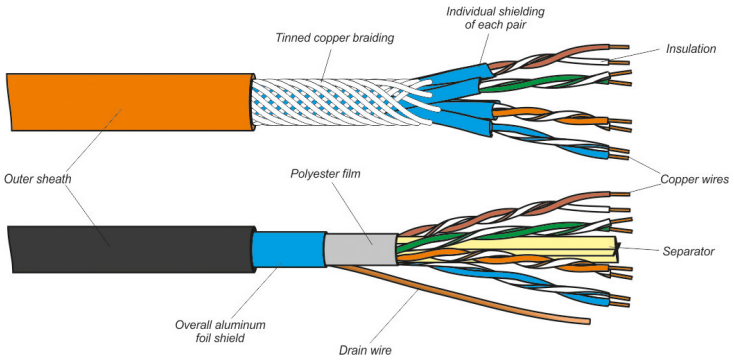


Figure 6: Twisted copper wire

The twisted pair cables terminate in a RJ45port or RJ45 plug.[3]

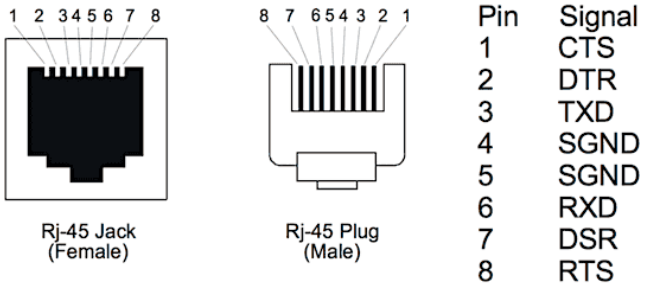


Figure 7: RJ45 connectors

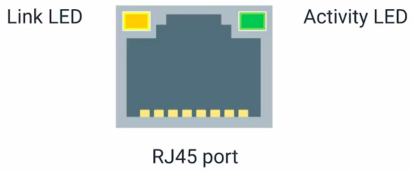


Figure 8: RJ45 port

- The link LED will glow when a cable is properly connected and both the devices are powered ON
- The activity LED will flash when data is actively transferred across the cable. [4]

In case of a few switches a single LED is used to indicate both link and activity.

Usually in offices and other large institutions these cables are not directly connected to the networking devices, instead they are connected to patch panel (usually wall mounted)

[3]**Network ports** are directly attached to the devices that make up a computer network

[4]Previously the led flashing corresponded to 1's and 0's that are being transmitted on the cable. But nowadays the data transfer rate is soo to high and the LEDs cannot blink at that speed.

Figure 9: patch panel

These are just the end points to many cables. These patch panels are further connected to the networking devices.

6 Data link layer

The main functionality of the Data link layer is to abstract away the need for any other layer to care about the physical layer and what hardware is in use. i.e. The web browser does not care if you are using a wired connection or a wireless connection. It just needs the underlying layers to send and receive data.

The protocol most widely used to send data across individual links is **Ethernet**.

6.1 Ethernet

Ethernet was introduced in 1980s and was standardized in 1983. Around that period switches were not invented yet, and hence all devices on a network shared a single collision domain. Ethernet as a protocol solved this problem using a technique called **Carrier sense Multiple Access with collision detection (CSMA/CD)**[5]

6.1.1 CSMA/CD

The working of CSMA/CD is pretty simple. If there is no data currently being transmitted on the network segment, a node will feel free to send data. If multiple nodes try to send data at the same instance, the computers detect this collision and wait for a random interval of time before they try to send data again. Waiting for random intervals ensures that these computers do not end up in a collision again.

6.2 MAC address

Due to the absence of switches, all computers on a network would receive the data shared by one of the computers on the network. So there was a requirement for a mechanism to uniquely identify each computer on the network, so that the shared data can be directly sent, only to that computer. For this purpose the **Media Access Control (MAC) Address** was used.

MAC Address is a globally unique identifier attached to an individual network interface. It is a **48-bit** number, normally represented by 6 groupings of two hexadecimal numbers.

MAC address is split into two sections:

- The first 3 octets of a MAC address are known as **Organizationally Unique Identifier (OUI)**[6]

[5] CSMA/CD is used to determine when the communication channels are clear, and when a device is free to transmit data

[6] Using MAC address (specifically OUI) it is possible to get manufacturer details for any networking device.

- The last three octets can be assigned in any way according to the manufacturer's wish, but it should be ensured that no two devices manufactured should have the same MAC address.

Ethernet uses the MAC address to ensure that the data it sends has both an address for the machine that sent the transmission, as well as the one the transmission was intended for.

6.3 Ethernet Frame

Data packet: An all-encompassing term that represents any single set of binary data being sent across a network link. Data packets at the ethernet level are called **Ethernet frames**.

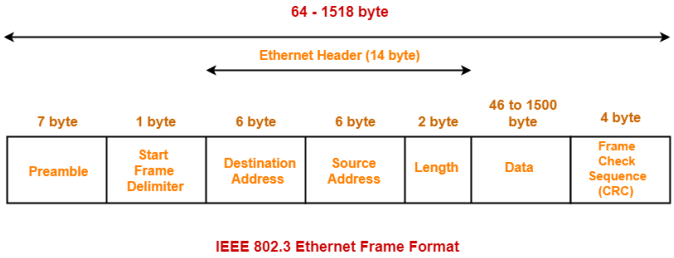


Figure 10: Ethernet frame

- The first 7 bytes consists of alternating 1's and 0's. These act as partially as a buffer between frames and can also be used by the network interfaces to synchronize the internal clocks they use to regulate the speed at whihc they send data.
- The last byte of the preamble is called **Start Frame Delimiter (SFD)**. This signals to the receiver that the preamble is over and that the actual frame content will now follow
- The next 6 bytes are the destination MAC address
- The next 6 bytes are the source MAC address
- Optionally the next 4 bytes constitute the VLAN (Virtual LAN)[7] header. Any frame with a VLAN tag will only be delivered out of a switch interface configured to relay that specific tag.
  - Usually VLANs are used to segregate network traffic
- The next two bytes are called **EtherType field**, it is used to describe the protocol of the contents of the frame.

[7]VLAN is a technique that lets you have multiple logical LANs operating on the same physical equipment

- Following this is the actual data being transported. It can be from 46 - 1500 bytes long
- The last 4 bytes constitutes the **frame check sequence (FCS)**. This represents the checksum value for the entire frame. The checksum value is calculated by performing **Cyclic redundancy check** against the frame.
- CRC is done at the receiving end and the obtained checksum value is compared against the data in FCS field. If they both match then the contents are not corrupted. If they do not match, then it is up to the protocols in the higher layers to decide what is to be done with the frame.

Ethernet only comments about data integrity and does not perform data recovery

6.3.1 Cyclic Redundancy Check

CRC is an error-detecting code. It uses a generator polynomial that is available at both the sender and the receiver's side.

Assuming **n** bits of data is to be sent, and the key (generator polynomial) is of **k** bits.
At the sender's side:

- k-1 zeros are padded at the end of the data
- modulo-2 binary division is used to divide binary data by the key
- Append the remainder at the end of the data to form the encoded data and send the same

At the receiver's side:

- Perform modulo-2 division again and if the remainder is 0, then there are no errors.

7 Network Layer

In a local area network, nodes can communicate with each other using their MAC addresses. But this scheme does not scale well.

Consider a case of an employee and his office laptop. During the weekdays the laptop will be connected to the office network. In case the employees decides to work from home on some day, then he will be connected to his home network. In this case it is not possible to figure out how to reach a particular computer with that MAC address. As at different times that computer will be part of different networks.

To overcome this issue we have **IP address**.

7.1 IP Address

IP Addresses are 32 bit long numbers made up of 4 octets. IP addresses are distributed in large sections to various organizations and companies, instead of being determined by hardware vendors.

IP addresses belong to networks, not to the devices attached to those networks.

The office laptop as considered in the previous case will always have the same MAC Address, but its IP Address will change based on the network it is connected to.

When a computer connects to a network, it is assigned an IP Address through a technology known as **Dynamic Host Configuration Protocol (DHCP)**. The IP Address assigned this way is called **Dynamic IP Address**. Alternatively, there is **Static IP Address** which must be configured on a node manually.

In most cases Static IP Addresses are reserved for servers and network devices, while dynamic IP Addresses are reserved for clients.

7.2 IP Datagram

The data packet in the network layer under the IP protocol is called **IP Datagram**

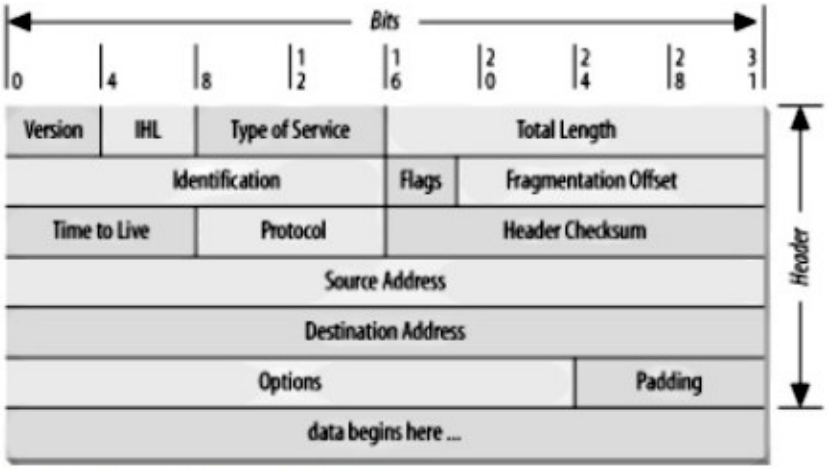


Figure 11: IP datagram

- The first 4 bits indicates the version of the IP being used. The most common version is version 4 (IPv4)
- The **header length** field is also a 4 bit field that indicates how long the header is. (The header length is not fixed as there are some optional fields in the IP datagram). The minimum size of the IP datagram header is 20 bytes.
- **Type of Service** is used to specify details about QoS. (Routers can use this field to decide which IP datagram is more important, and then prioritise packet transfer accordingly)
- **Total Length** field indicates the total length of the IP datagram. The maximum size of IP datagram is: $2^{16} = 65{,}535$ bytes.

  If the total amount of data that needs to be transmitted is larger than what can fit in a single datagram, the IP layer needs to split this data into many individual packets.
- **Identification** field is a 16-bit number that's used to group messages together.

  When data is split into multiple datagrams, all those datagrams will have the same value for the **identification** field. At the receiving end this indicates that the data in these packets needs to combined.

- **Flag field** is used to indicate if a datagram is allowed to be fragmented[8], or to indicate that the datagram is already fragmented.
- At the receiving end when the fragmented IP datagrams are received, the order in which they must be combined to get back the original data is specified by the **Fragmentation offset field**.
- **Time to live (TTL)** field is an 8-bit number that indicates how many router hops a datagram can traverse before it's thrown away.

  Every time a datagram reaches a router, it decrements the TTL field value by 1. This field is used to make sure that, if there is any misconfiguration in routing that causes an endless loop, datagrams wont spend all eternity trying to reach their destination.

  A possible endless loop is when router A thinks router B is the next hop and router B thinks router A is the next hop.
- **Protocol** field indicates which transport layer protocol is being used.
- **Header checksum** is a checksum of the contents of the entire IP datagram header. It is used to verify if the contents of the header is corrupted or not.

  Since the TTL field changes at every router along the path, the Header checksum needs to be recomputed at every router.
- Source and Destination IP address fields are self explanatory
- **IP options** field is used to set special characteristics for datagrams primarily used for testing purposes.
- Since the options field is optional, **Padding** field is a series of zeros used to ensure the header is the correct total size.

IP datagram constitutes the data payload section of the Ethernet frame.
This process is called **Encapsulation**

Food for thought:

Is it possible to attack a network by setting a very high value for the **Type of Service** field and a very small value for the **TTL** field.

[8]Fragmentation is the process of taking a single IP datagram and splitting it up inyo several smaller datagrams

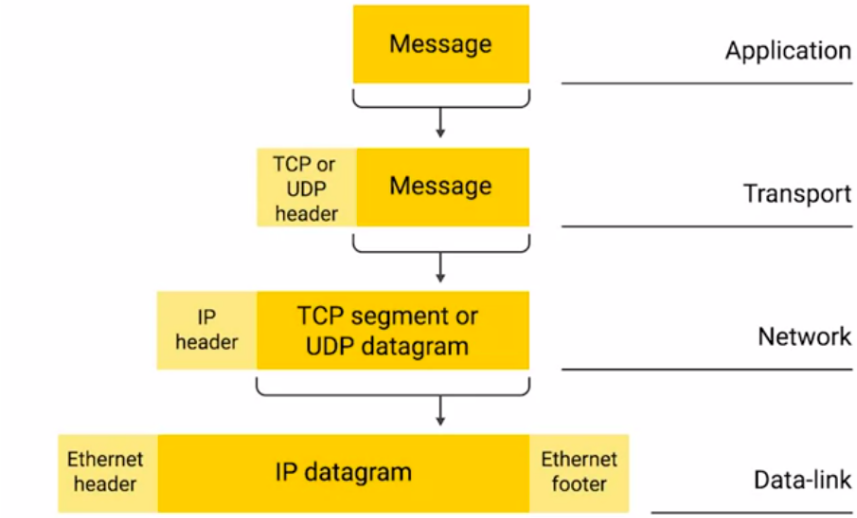


Figure 12: Packet formation at different layers

7.3 IP Address Class

IP Addresses can be split into two sections:

- network ID:
- host ID:

Address class system is a way of defining how the global IP address space is split up. The three types of classes are:

- Class A
  - first octet: network ID
  - remaining octets are used for host ID
  - first bit of network ID must be 0, hence there are 127 Class A networks and 16,777,214 hosts on each network (2^{24} - 2)
- Class B
  - first two octets: network ID
  - remaining octets are used for host ID
  - first two bits of network ID must be 10, hence there are 16,384 Class B networks and 65,534 hosts on each network (2^{16} - 2)
- Class C
  - first three octets: network ID

- remaining octet is used for host ID
- first three bits of network ID must be 110, hence there are 2,097,152 Class C networks and 254 hosts on each network (2^8 - 2)

The above three classes are used for UNICAST/ BROADCAST purposes. CLASS D is used only for MULTICAST purposes.

Class	Leading bits	Size of *network number* bit field	Size of *rest* bit field	Number of networks	Addresses per network	Total addresses in class	Start address	End address
Class A	0	8	24	128 (2^7)	16,777,216 (2^{24})	2,147,483,648 (2^{31})	0.0.0.0	127.255.255.255
Class B	10	16	16	16,384 (2^{14})	65,536 (2^{16})	1,073,741,824 (2^{30})	128.0.0.0	191.255.255.255
Class C	110	24	8	2,097,152 (2^{21})	256 (2^8)	536,870,912 (2^{29})	192.0.0.0	223.255.255.255
Class D (multicast)	1110	not defined	not defined	not defined	not defined	268,435,456 (2^{28})	224.0.0.0	239.255.255.255
Class E (reserved)	1111	not defined	not defined	not defined	not defined	268,435,456 (2^{28})	240.0.0.0	255.255.255.255

Figure 13: IP address classes

NOTE: In Class A the network ID cannot be 0 (used by a new host in the connecting to the network, who does not have an IP address assigned to it yet) or 127 (loopback address), as they are used for other special purpose. The class system is now being replaced with **Classless inter-domain routing (CIDR)**

7.4 Address Resolution Protocol (ARP)

ARP is a protocol used to discover the hardware address of a node with a certain IP address.

We know that the IP datagram is encapsulated in the ethernet frame, and the ethernet frame requires the destination MAC address to complete the ethernet frame header.

Almost all networking devices maintain a local **ARP table**. An ARP table is a list of IP addresses and the MAC addresses associated with them.

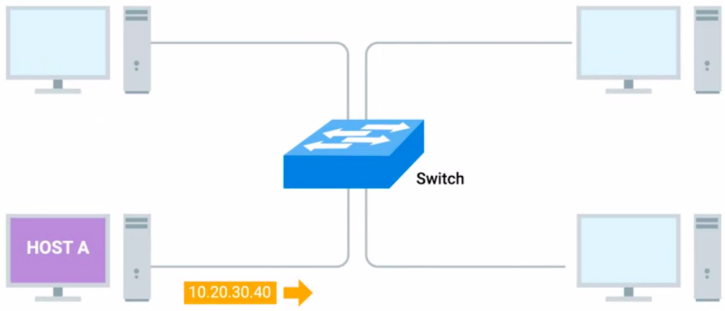


Figure 14: Computer to send data to another system with IP address 10.20.30.40

When a computer wants to send some data to another computer, it will look at its local ARP table to get the MAC address of the desired destination. And uses it to complete the ethernet header. In case the computer does not know the MAC address of the destination system

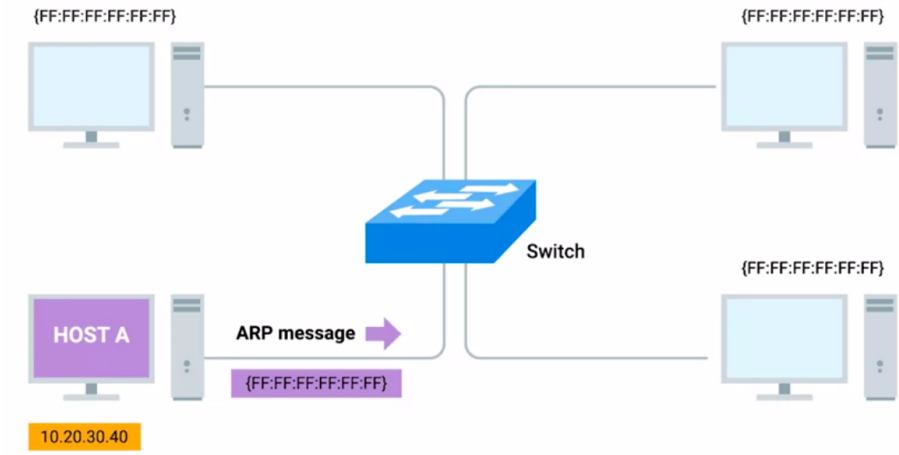


Figure 15: Computer sending ARP broadcast to get MAC address of destination system

The sending node transmits a broadcast ARP message using FF:FF:FF:FF:FF:FF as the destination MAC address. Hence all the nodes connected to the network receive this ARP message. When a network interface whose IP address is 10.20.30.40 receives this ARP broadcast, it sends back an **ARP response** that contains its MAC address.

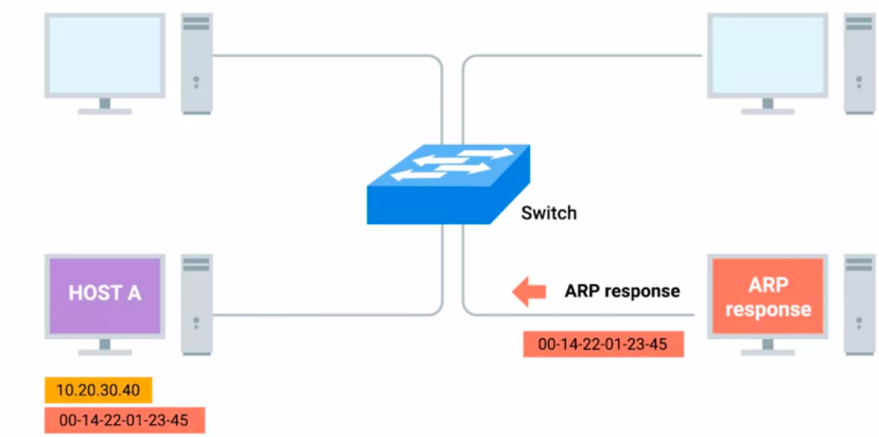


Figure 16: ARP Response

The sending node uses this MAC address to complete the ethernet header and sends the packet for delivery. The sending node also stores this MAC address in its local ARP table.

ARP table entries generally expire after a short amount of time to ensure changes in the network are accounted for.

7.5 Subnetting

The process of taking a large network and splitting it up into many individual and smaller subnetworks is called subnetting. Few of the bits assigned for HOST ID can be used for subnet ID. The core routers are only concerned about the network ID. As this is sufficient to reach the **gateway router** of a particular network. Reaching the particular host within the network is taken care by the gateway router.

7.5.1 Subnet Masks

To reach the particular host, first the subnet to which the host belongs should be reached. To calculate the subnet ID the **subnet mask** is used. Subnet masks are 32-bit numbers that are normally written as 4 octets in decimal.

A subnet mask is used by computers to determine if an IP address exists on the same network.

- The part of the subnet mask that contains only 1's tells the router what part of the IP address is the subnet ID.
- The part of the subnet mask that contains only 0's corresponds to the HOST ID.

255.255.255.0 is a commonly used subnet mask.

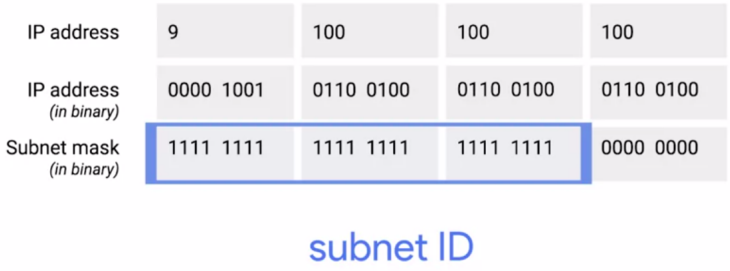


Figure 17: Subnet mask

Looking at the leading bits we can determine to which class an IP address belongs and from that we can get to know the Network ID. This can be used to reach a particular network. Further using the subnet mask, we can obtain the subnet ID. This is used to reach a particular subnet of the network to which the host belongs. Further the remaining bits indicate the Host ID.

The size of a subnet is given by the subnet mask. For example consider the subnet mask 255.255.255.0
We know that the part that consists 0's correspond to the HOST ID. And in the above subnet mask there are 8 zeros and hence the number of possible hosts in the subnet are $2^8 = 256$.
NOTE: A subnet can contain only 2 less than the total number of Host IDs available. This because **0** is generally not used and **255** is used for broadcast purposes within the subnet. Hence only the numbers from 1 to 254 are available to be assigned to a host.

Alternate representation for the subnet mask is: /<number>
It is called the **CIDR notation** For example consider a subnet mask 255.255.255.224. The binary representation is as follows

11111111.11111111.11111111.11100000

The IP address 9.100.100.100 with the above subnet mask can be represented as

9.100.100.100/27

This is because the subnet mask has 27 1's and followed by 5 0's. This means that the subnet can contain 32 host IDs.

7.6 Classless Inter-domain routing

With IP address classes, we can possibly have only 254 Class A networks, and similarly some fixed number of Class B and Class C networks.

In a Class C network, number of possible hosts are 254. This may be very small for many practical cases, where as having 65,534 hosts (Class B network) may be too large for an organization. This would result in a wastage of a lot of IP addresses. Hence IP address classes is not an efficient way of sorting the IP addresses. To overcome these inefficient techniques CIDR was used.

CIDR uses the subnet mask to demarcate[9] networks. With CIDR, the network ID and subnet ID are treated as a single entity. And this is why we get the **slash/CIDR notation**. CIDR completely abandons the usage of IP address classes.

9.100.100.100
subnet mask: 255.255.255.0

Using the CIDR notation the above can be represented as

9.100.100.100/24
Network ID: 9.100.100
Host ID:100

With IP address classes, the size of the network was fixed, and subnets within it could be of different sizes. But with CIDR the entire network can be of any size. There are no restrictions on the size of the network. This leads to more efficient utilization of the IP addresses.

Another advantage with CIDR notation:
Previously when a company using Class C network, wanted more IP addresses it could use another Class C network. With this the number of hosts that could be handled is:

$$254 + 254 = 508$$

But with CIDR, the company must reduce the subnet mask by 1. So that 9 bits are allocated for Host ID. With this the number of hosts that can handled is:

$$2^9 - 2 = 510$$

Hence we get 2 more host IDs as compared to the case of IP Address classes. This process of increasing the number of host bits by reducing the number of network bits is called **supernetting** or **route aggregation**.

NOTE: For supernetting the smaller networks that will be combined need to be contiguous.

Using the supernetting the number of entries required in the routing table can also be minimized.

[9]Demarcation point is used to describe where one network ends and another network begins

7.6.1 Numerical Example for Subnetting

Consider the IP address given to you is 200.1.1.0/24
You have to divide this network into 15 smaller subnetworks.

The procedure to be followed is

1. Find the closest power of two, to the required number of networks

   In this case it is 16. Lets call this as **n**

2. Once the number of networks to be formed is obtained, calculate the additional number of network bits needed which is $\log_2(\mathbf{n})$

   In this case it is 4. Hence the new subnet mask will be /28

3. Divide 256 by 2^n to get the increment value, which is necessary to obtain the network addresses for each subnetwork.

   In this case it is $256/16 = 16$.

4. The network addresses for the various subnets are:

   - 200.1.1.0/28
   - 200.1.1.16/28
   - 200.1.1.32/28
   - 200.1.1.48/28

     .
     .
     .
   - 200.1.1.224/28
   - 200.1.1.240/28

7.7 Types of communication

Communication may not always be between one sender and one receiver also known as **Unicast communication**.[10] The other types of communication are:

1. Broadcast

   - Data is sent to every single device on the LAN
   - to accomplish this, a special destination address called **Broadcast address** is used.
   - Ethernet broadcast address is: **FF:FF:FF:FF:FF:FF**

[10] In case of unicast communication the least significant bit in the first octet of a destination address is set to zero, which means that ethernet frame is intended for only the destination address

- Further to send message to all nodes connected to a particular network on the LAN, and not to all networks of the LAN **direct broadcast** is used. The IP address in this case will be the network ID followed by all 1's (This is the broadcast address for the devices on the network only).

2. Multicast:

   - Multicast frame is sent to all computers on a local network. But all the computers wont accept it. This is because the network interface can be configured to accept lists of configured multicast addresses for these sort of communications
   - The least significant bit in the first octet of a destination address is set to one, to indicate you are dealing with a multicast frame
   - Usually IP addresses in the range 224.0.0.0 through 239.255.255.255 are designated as multicast addresses.

8 Basic Concepts of Routing

Router is a network device that forwards traffic depending on the destination address of that traffic. The basic working of a router is as follows

- Receive a data packet
- Examine the destination IP address of the packet
- Looks up for the destination IP address in its routing table
- Forwards the data packet to the next router which is the closest to the destination address.

A router is the connector between two independent networks, each interface of the router is connected to a different network, and hence has a different IP address.

In the following example when a computer on network A wants to send data to a computer on Network B. The sending node knows that the destination node belongs to a different network, hence it uses the destination MAC address of the Gateway router. The router's interface on Network A receives the packet. And it sees that the destination MAC address belongs to it. The router then strips the data-link layer encapsulation to get the IP datagram. Now the router can inspect the IP datagram to obtain the destination IP address (10.0.0.10). The router then looks at its routing table, and understands that the packet is to be transferred to Network B. And since Network B is directly connected to one of the interfaces of the router, it has the MAC address for the destination IP in its ARP table.

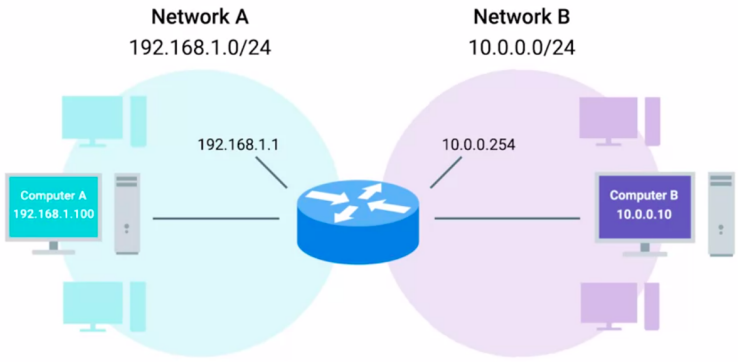


Figure 18: Basic setup for routing

The router then decrements the TTL field value by 1, and calculates a new checksum. It then encapsulates this new IP datagram inside another Ethernet frame, this time it sets its own MAC address of the interface on Network B as the source MAC address. Since the router also has the MAC address of 10.0.0.10 in its ARP table, it uses that as the destination MAC address.

Consider another example where the router is not directly connected to the Destination network. The computer with an IP Address of 192.168.1.100 wants to send data to 172.16.1.100

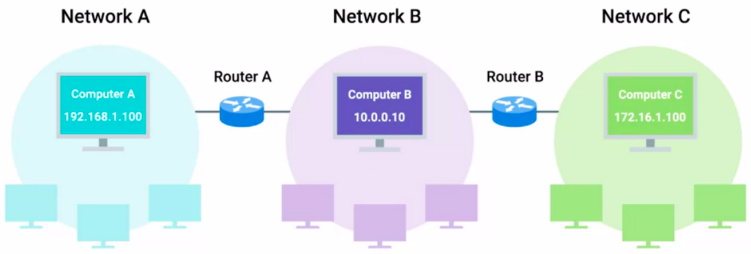


Figure 19: A more complicated setup for routing

The source node knows that the destination node is not a part of its local network, hence sends the packet to its gateway router (Router A). Upon receiving the packet and getting to know the IP address, Router A searches its ARP table for the MAC address corresponding to the destination IP. It gets to know that the shortest path to reach the IP is via another Router. Router A decrements the TTL and calculates the checksum and sends it to Router B. Router B then goes through its ARP table and gets to know destination MAC address, it decrements the TTL and then sends the packet to 172.16.1.100

8.1 Routing Table

The most basic routing table will have 4 columns.

- **Destination network:** Consists of IP and subnet mask/ CIDR of the destination network
- **Next hop:** IP address of the next router that should receive data intended for the destination network.
- **Total hops:** Routers usually route the packet through the shortest path (which may keep changing dynamically possible causes - failure of a router/link). So whenever a packet reaches a router, the router should keep track of how far the packet is from the destination and the shortest path to reach that destination.
- **Interface:** Router should know to which of its interfaces it should forward the traffic matching the destination network out of.

NOTE: A routing table will generally have a **catch-all entry** that matches any IP address that it does not have a explicit network listing for, this is called the **default route**.

The other types of routes are

1. Static route
2. Dynamic route

When there are multiple routes between the source and destination, then the route suggested by the protocol with the least **administrative distance** is considered.

8.2 Gateway Protocols

The routing tables should always be updated with the new information about the shortest path to destination networks. **Routing protocols** are used to keep constantly updating the routing tables.

Routing protocols are used by routers to speak to each other in order to share the information they have. This is how routers get to know the best/shortest path to destination network. The two main categories of routing protocols:

1. Interior Gateway protocol
2. Exterior Gateway protocol

The interior gateway protocols are further split into two categories

- Link state routing protocols (Share complete information)
- distance-vector routing protocols (Share info only about next hop)

8.2.1 Interior Gateway protocols

These are used by routers to share information within a single autonomous system.[11]

Distance-vector protocol: A router simply sends its routing table (which contains the distance of various destination networks from it, in terms of number of hops required to reach them) to the other routers which are directly connected to it.

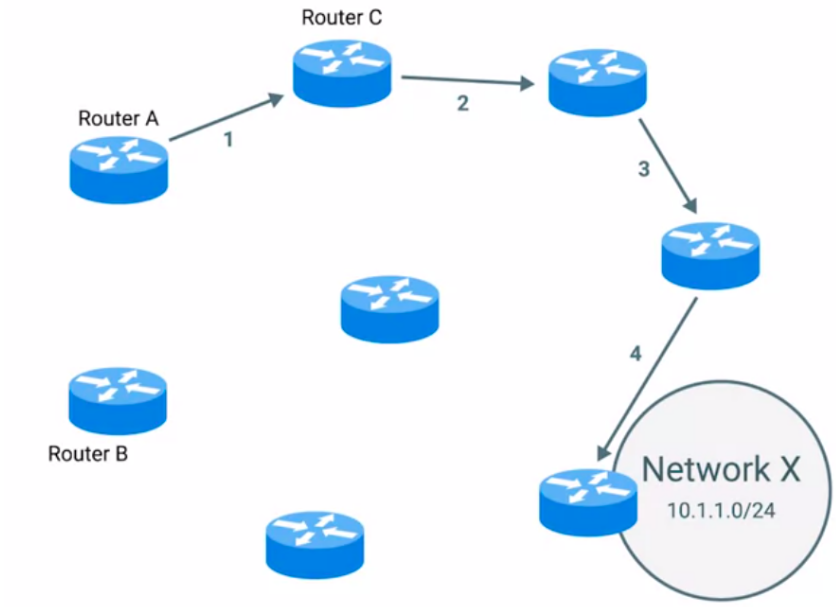


Figure 20: Basic router setup to explain IGP

Consider the above shown set up for a bunch of routers. Router A knows that to reach to Network X requires 4 hops next hop being router C.

[11] Autonomous system is a collection of networks that all fall under the control of a single network operator.

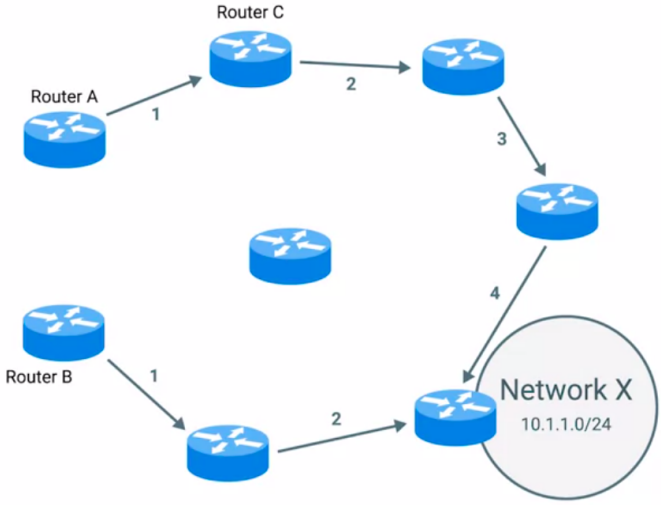


Figure 21: Basic router setup to explain IGP

Router B needs only two hops to reach network X. So when router B shares its routing table with router A. Router A will observe that even with the extra hop needed to reach router B, the total number of hops required to reach network X would be 3. Hence it updates its routing table with the new information that network X is at a distance of 3 hops and the next hop is router B.

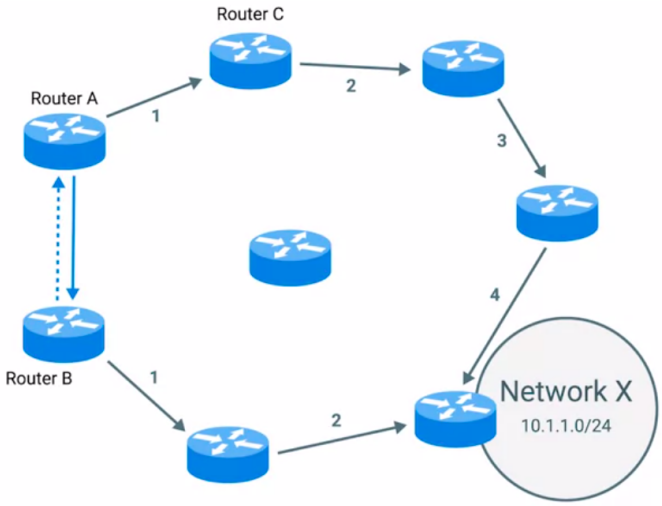


Figure 22: Basic router setup to explain IGP

Common distance-vector protocols are:

1. Routing Information Protocol (RIP) [**IETF RFC2453**]
2. Enhanced Interior Gateway Protocol (EIGRP)

- So basically every router sends a list/vector of distances to other routers. Hence the name distance-vector protocol.
- Routers do not have much info about the total state of the autonomous system. They have info about their immediate neighbours. Because of which a router may be slow to react to any changes that has happened in the network far from it.

This drawback with Distance-vector protocol led to Link-state protocol.

Link state protocol: Routers use a more sophisticated approach to determine the best path as compared to that used with distance-vector protocol. In a link state protocol each of the router advertises the state of the link of each of its interfaces. Due to this every router in the autonomous system has info about every other router in the autonomous system.

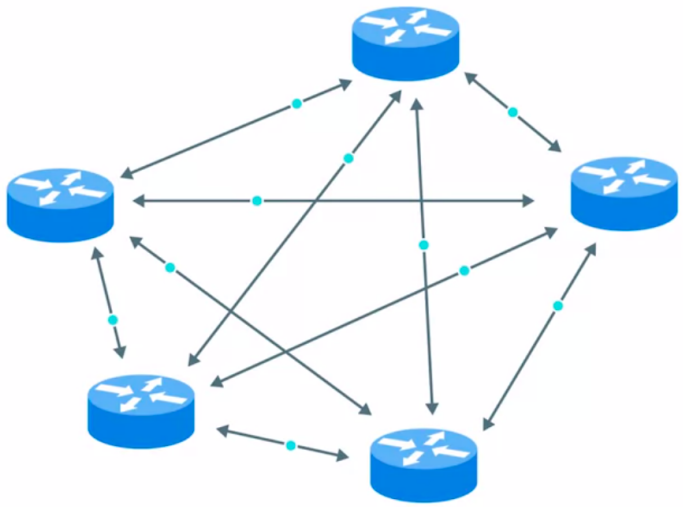


Figure 23: Basic router setup to explain IGP

- Link state protocols use complex algorithms with the available data from every router, to determine the shortest path
- To implement link state protocols, routers need more memory and processing power.
- Common link state protocols are: Open Shortest Path First (OSPF) [**IETF RFC2328**]

Link state protocols have made distance-vector protocols outdated

8.3 Exterior Gateway Protocol

These are used to communicate data between routers representing the edges of an autonomous system. Exterior Gateway protocols are used to share info between multiple autonomous systems.

Every autonomous system is assigned an unique ASN(Autonomous System Number). ASN are 32-bit numbers that are represented by a single decimal number.

The only exterior Gateway protocol in use today is Border Gateway protocol (BGP) [**IETF RFC4271**]

8.4 Non routable address space

These are ranges of IP set aside for use by anyone that cannot be routed to. Computers with IP address within this range on a local network will be able to communicate with each other, but no gateway router will attempt to forward traffic to this type of network. **RFC1918** defined three range of IP address that are non routable[12]

1. 10.0.0.0/8 (**Class A private address range**)
2. 172.16.0.0/12 (**Class B private address range**)
   16 contiguous networks
3. 192.168.0.0/16 (**Class C private address range**)
   256 contiguous networks

The main reason to have such non routable address space is that there are around 7.5 billion people on earth but with IPv4 the total number of addresses possible is only 4 billion. So using one IP address for the Gateway router which inturn communicates on behalf of systems on a network which use these non routable IP addresses, allows a means to use same IP address from this range repeatedly in different autonomous systems. The gateway router of these autonomous systems, is representative for all systems part of the autonomous system.

8.5 Network Address Translation

NAT takes one IP address and translates it into another IP address. This provides a solution to the limited number of IPv4 addresses that are available.

NAT is a technology that allows a gateway router to rewrite the source IP of an outgoing IP datagram while retaining the original IP in order to rewrite it into the response.

[12]These IP address belong to no one and everyone is free to use them

Using NAT, the gateway router ensures that the computer outside the network does not get to know the actual IP address of the computer it is communicating with this is called **IP masquerading**.

Though the router changes the source IP during transmission, but handling the return traffic i.e. deciding which data goes to which system is a tedious job. The simplest way to do this is through **Port preservation**. Port preservation is a technique where the source port chosen by a client is the same port used by the router. So the gateway router keeps track of the source port number and uses this to redirect the return traffic.

If more than one node uses the same source port number at the same time, then the router normally selects an unused port at random to use instead.

8.5.1 Port forwarding

This is a technique where specific destination ports can be configured to always be delivered to specific nodes.
If the clients want to communicate with a server at 10.1.1.5:80 using Port forwarding, the client don't even have to know the destination IP, if they know the IP of the gateway router and use port number 80 as destination port, then these will directly be forwarded to 10.1.1.5:80

This technique allows for IP masquerading and also simplifies how external users interact with lots of services all run by the same organization.

8.6 Limitations of IPv4

IANA is responsible for assigning address blocks to the five **Regional Internet Registries (RIR)**.

1. AFRINIC (continent of Africa)
2. ARIN (US, Canada and parts of Caribbean)
3. APNIC (Asia, Australia, New Zealand)
4. LACNIC (Central and South America)
5. RIPE (Europe, Russia, Middle East)

These 5 RIRs are responsible for assigning IP address blocks to organizations within their geographic areas and most have already run out of IPs to be assigned. The solution to this problem is to shift to IPv6, but shifting to IPv6 will take a lot of time. Another simpler workaround can be achieved using NAT and Non-routable address space.

Non-routable address space (RFC1918) can be used by unlimited number of networks, this is because the gateway router changes the source IP while forwarding the packets out of the network (NAT). Hence the only necessity is that the gateway router should have an unique IP address (to avoid global collision of IP addresses), the other computers within the network can have dynamic IP address from non-routable address space, which changes after every session.

Hence with NAT, a network with a 1000 nodes or more than that just needs a single IP address. This addresses the issue of exhaustion of IPv4 addresses. So this can be used as a solution until IPv6 is globally available.

8.7 IPv6 Addressing

Since IANA ran out of IPv4 address, IPv6 addresses were developed. These addresses are 128 bits long(8 groups of 16 bits each). There are two rules to shorten the representation of an IPv6 address.

1. Leading zeros from a group can be removed.
2. Any number of consecutive groups composed of just zeros can be replaced with two colons. (Can be used only once for an address)

2001:0db8:0000:0000:0000:ff00:0012:3456
2001:db8:0:0:0:ff00:12:3456
2001:db8::ff00:12:3456

Loopback address: 0000:0000:0000:0000:0000:0000:0000:0001 = ::1

- Address beginning with **ff00::** are used for multicast
- Addresses beginning with **fe80::** are used for link-local[13] unicast.
- Addresses beginning with **001** are global unicast addresses.
- Addresses starting with **fc00** refers to Private Addresses.

The first 64 bits constitute the network ID and the next 64 bits constitute the host ID. IPv6 uses the same CIDR notation for subnetting like IPv4 addresses.

[13]link-local address is used by an IPv6 host to receive their network configuration

8.7.1 IPv6 header

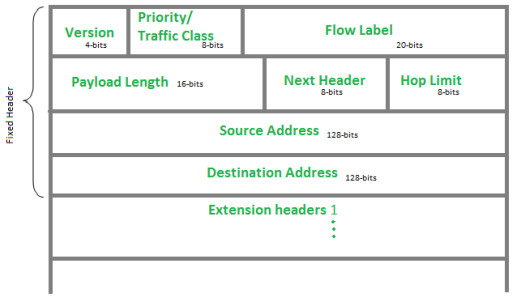


Figure 24: IPv6 header

- **version** field: 4-bit field that defines which version of IP is in use.
- **traffic class** field: 8-bit field that defines the type of traffic contained within the IP (useful for routers to prioritize packets while routing)
- **flow label** field: 20-bit field used in conjunction with the traffic class field for routers to make decisions about the quality of service.
- **payload length**: 16-bit field that defines how long the data payload section is
- **next header** field: Defines what kind of header follows after this datagram
- **hop limit** field: 8-bit field similar to TTL field in IPv4
- **source address** and **destination address** is followed by a next header if one is specified in the next header field. If not data payload follows the destination address field.

8.7.2 Ipv4 mapped IPv6 address space

Any IPv6 address space that begins with 80 zeros, followed by 16 ones is understood to be part of IPv4 address space. The remaining 32 bits of the IPv6 address is the same as the IPv4 address. **IPv6 tunnels** are used to ensure that IPv6 datagrams can be transported across IPv4 networks. IPv6 tunnels servers are available at either end of a connection, they encapsulate IPv6 datagram within a IPv4 datagram. At the receiving end the IPv6 tunnels server will perform de-encapsulation. An **IPv6 tunnel broker** is a company that provides IPv6 tunneling endpoints, so that additional equipments need not be used in the network. Different protocols used for this purpose are **6in4**, **Tunnel Setup Protocol** and **Anything in Anything (AYIYA)**

9 Transport Layer

Allows traffic to be directed to specific network applications. The transport layer does the task of multiplexing and demultiplexing at the nodes. By multiplexing the various requests made by a single node are combined and delivered to different nodes on the network. By demultiplexing all the requests/data aimed at a particular node are delivered to it.

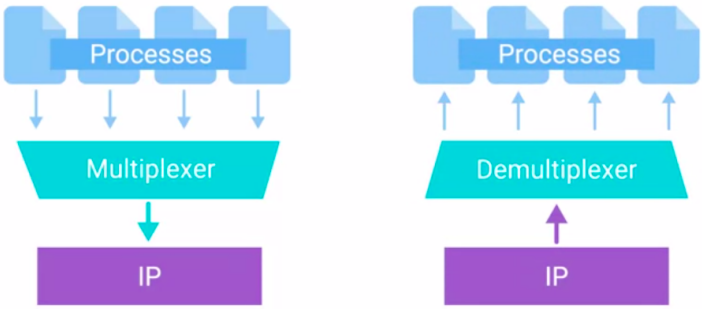


Figure 25: Multiplexing and demultiplexing action of Transport Layer

But once the various data is multiplexed, to identify which data corresponds to which process, transport layer makes use of ports[14].

A **server** is a program running on a computer waiting to be requested for data. A **client** is a program that is requesting for the data. Various servers run while listening for data requests on specific ports.

http uses port 80
ftp uses port 21

So in order to request a webpage from a web server (10.1.1.100), the request made by the client will be directed to port 80 of that server i.e. the address to which the traffic will be directed to will be 10.1.1.100:80. This address is called **socket address**.

If another client wants to request a file from the server (10.1.1.100) the traffic is directed to 10.1.1.100:21. As port 21 is used by the server to listen to requests made for files.

So a single server is capable of answering to different type of queries made by clients. Note that all the queries are directed to the IP address 10.1.1.100 . But the server uses only specific ports for particular type of requests, all these requests directed to the server are demultiplexed by the transport layer and directed to specific ports of the server.

[14]Port is a 16 bit number that is used to direct traffic to specific services running on a networked computer

9.1 TCP segment

Payload section of IP Datagram consists of the TCP segment. TCP segment is made up of TCP header and a data section. The application layer places its data in the data section.

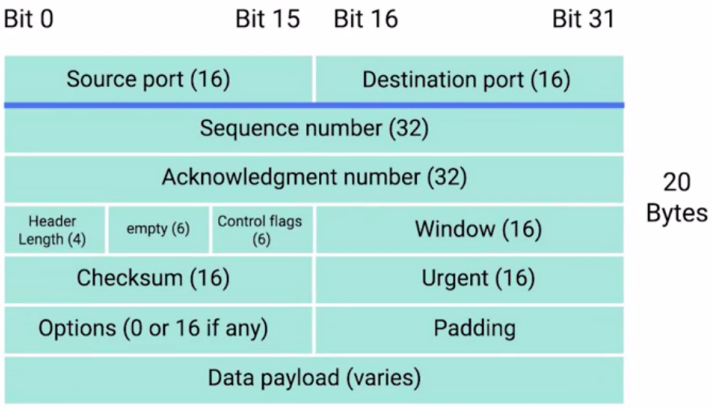


Figure 26: TCP Segment

- **Destination Port** is the port of the server/other node the traffic is intended for
- **Source port** is a high numbered port chosen from a special section of ports known as ephemeral ports.
  Source port is necessary to demultiplex the server response at the node and redirect the data to the corresponding processes.
- **Sequence number** is a 32-bit number that's used to keep track of the position of a particular TCP segment in a sequence of TCP segments.
- **Acknowledgement number** is the number of the expected segment. i.e. In a particular TCP segment is sequence number is 1, if the next segment has a sequence number 2. Then acknowledgement number in the first TCP segment will be 2.
- **Header length/ Data offset field** is a 4-bit number that communicates how long the TCP header for the particular segment is.
- **TCP Window**: Data processing capabilities of the server and client are different. So a client uses a buffer to store the data received from the server if it is not able to process it in realtime. Window size indicates the amount free space in the buffer at the client end. If the buffer gets filled up, the server will wait for sometime before sending the next set of data.
- **Checksum** is similar to the checksum field at IP and ethernet levels.

- **Urgent pointer field** is used in conjunction with one of the TCP control flags to point out that a particular segment might be more important than others.
- **Options field** is used for more complicated flow control protocols
- **Padding** adds some zeros to ensure that data payload section begins at the expected location.
- **Control flags**
  - **URG (urgent)**
    A value of 1 indicates that the segment is considered urgent and the urgent pointer field has more data about this.
  - **ACK (acknowledgement)**
    A value of 1 in this field means that the acknowledgement number field should be examined
  - **PSH (push)**
    A value of 1 in this field means that the transmitting device wants the receiving device to push currently buffered data to the application as soon as possible.
  - **RST (reset)**
    A value of 1 in this field means that one of the participants in the TCP connection is unable to make sense of the packets it received. So that particular participant asks for resetting the communication and begin the data transfer from the beginning again.

    It is also used to reject a connection establishment request
  - **SYN (synchronize)**
    Used while establishing a TCP connection and makes sure the receiving end knows to examine the sequence number field
  - **FIN (finish)**
    When this flag is set, it means the transmitting computer does not have any more data to send and the connection can be closed.

9.1.1 TCP handshake

A handshake is a way for two devices to ensure that they are speaking the same protocol and will be able to understand each other. to establish a TCP connection requires an handshake that usually involves 3 steps and hence it is called a **three way handshake**

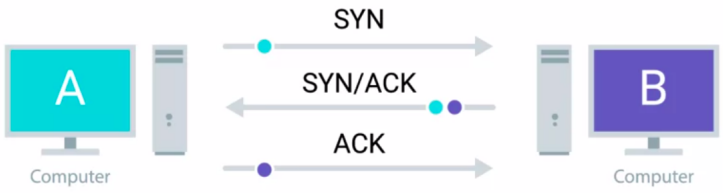


Figure 27: Three way handshake

If computer A wants to establish a communication with Computer B, it first sends a TCP segment with the SYN flag set. If computer B wants to establish the communication it will respond with another TCP segment where the SYN, ACK flags are sent. By doing this computer B responds to A that it has noted the sequence number of A, and asks A to make a note of the sequence number it is using and also indicates that it is ready to establish the communication. And sends an acknowledgement (i.e. the next sequence number it is expecting). Computer A will now send the next set of packets that include the data request.

A 4-way handshake is used while terminating the connection.

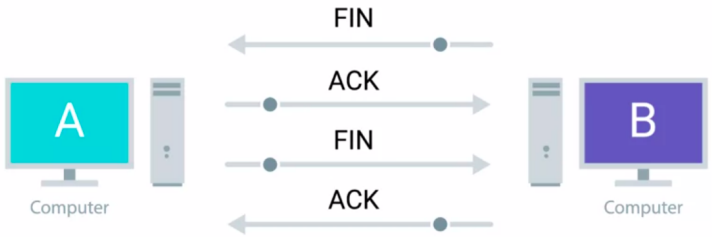


Figure 28: Four way handshake

When Computer B is ready to close the connection it sends a segment where FIN is set. Computer A responds with a segment where the ACK flag is set, saying it has got to know that B wants to end the connection. If A is also ready to close the connection it further sends another segment where FIN flag is set. B now responds to this with an ACK.

9.1.2 TCP Socket

Socket is an instantiation of an end-point in a TCP connection.
NOTE: Traffic can be sent to any port but to get a response a socket must be opened on that port.

The different states of a TCP socket are

1. **LISTEN**
   TCP socket is ready and listening for incoming connections
2. **SYN_SENT**
   A synchronization request has been sent, but connection has not been established yet.
3. **SYN_RECEIVED**
   A socket previously in a LISTEN state has received a synchronization request and sent a SYN/ACK back
4. **ESTABLISHED**
   TCP connection is in working order and both sides are free to send each other data.
5. **FIN_WAIT**
   A FIN has been sent, but the corresponding ACK from the other end has not been received yet.
6. **CLOSE_WAIT**
   The connection has been closed at the TCP layer, but the application that opened the socket has not released its hold on the socket yet.
7. **CLOSED**
   The connection has been fully terminated and no further communication is possible.

Socket states and names can change from one operating system to another.

If checksum fails in the ethernet and IP data is discarded. It is upto TCP to determine how/when to resend the data. Whenever the checksum fails, the receiving node does not send the acknowledgement for that segment. So if the sending node does not receive the ACK, it will retransmit that particular data.

So sequence numbers are also used to arrange the data, when the packets arrive out of order at the receiving node.

9.2 UDP

TCP is a connection oriented protocol[15]. Hence data transmission takes a lot of time. There are certain applications where ensuring that all the data has been successfully sent to the client is not necessary, but requires fast transmission of data to the client. In situations such as there **User Datagram Protocol (UDP)** is used. UDP is a connectionless protocol i.e. it does not require the client and server to have a socket which is dedicated for the communication. The client and server do not have to talk to each other and can directly jump to data transfer. It does not even use Acknowledgement. The best example where UDP is preferred over TCP is video streaming.

[15] A connection oriented protocol first establishes a connection, and uses it to ensure that all data has been properly transmitted

9.3 Ports

Ports are representee by 16-bit numbers. Hence they have a range of 0-65535. IANA has split this range into independent sections.

- Port 0 is not used for network traffic. But it is sometimes used in communications taking place between different programs on the same computer
- Ports 1 - 1023 are referred to as **system ports**. These ports represent the official ports for well-known network services.
- Ports 1024 - 49151 are known as **registered ports**. These ports are used for network services that are not quite common.
- Ports 49152 - 65535 are known as **private or ephemeral ports**. These ports are used for establishing connections and cant be registered with IANA.

NOTE: The range of ephemeral ports varies depending on the platform being used.

9.4 Firewall

Firewall is a device that blocks traffic that meets a certain criteria. Firewalls are critical in keeping a network secure. Firewalls can operate at different layers. But they are commonly used at the transportation layer.

Firewalls operating at the transportation layer will have a configuration that enables them to block traffic to certain ports, while allowing traffic to other ports.

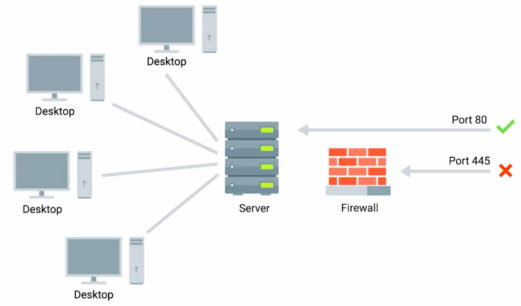


Figure 29: Firewall

From the above diagram it can be seen that firewall allows a computer outside the network to send data to port 80, but blocks all the requests that are sent to port 445.

For many home users the functionality of the firewall is performed by the router.

10 Application Layer

The data section of the TCP segment is filled with whatever data the applications want to send. It can be the url of some webpage that the browser is requesting for, or it can be the email we are sending. Or it can be some webpage/images and other stuff when we are trying to upload something. At the application layer web browser act as the client for the user. Few of the well known web browsers are:
Chorme, Safari, Microsoft Edge etc. Similarly few of the well known web servers are Microsoft IIS, Apache, nginx etc.
Even though there are many different web browsers and web servers all of them need to talk the same language (i.e. protocol). And one such common protocol used is **Hyper Text Transfer Protocol (HTTP)**

11 What happens when you request for a webpage?

Consider a simple case with 3 networks.

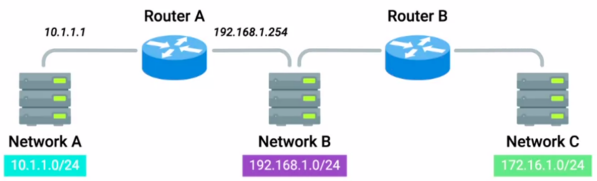
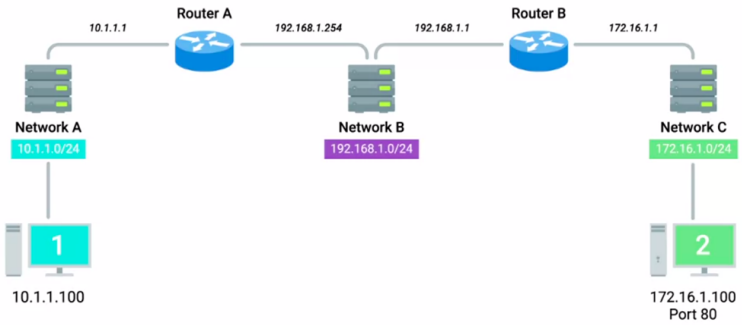
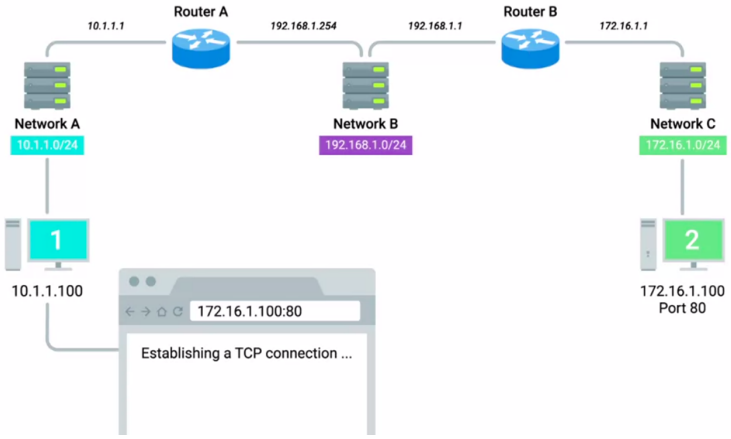


Figure 30: Simple network set up to understand working of all layers

Let computer 1 connected to Network A act as a client and computer 2 connected to network C be the server

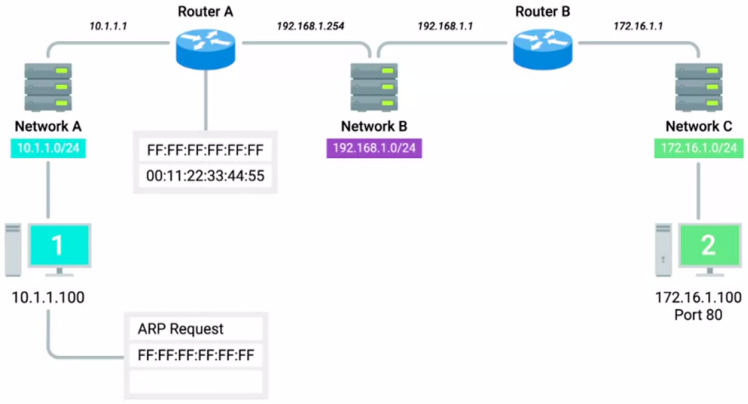


The web server is listening on port 80.(http)
An end user sitting at computer 1 opens up a web browser and enters 172.16.1.100 into the address bar.

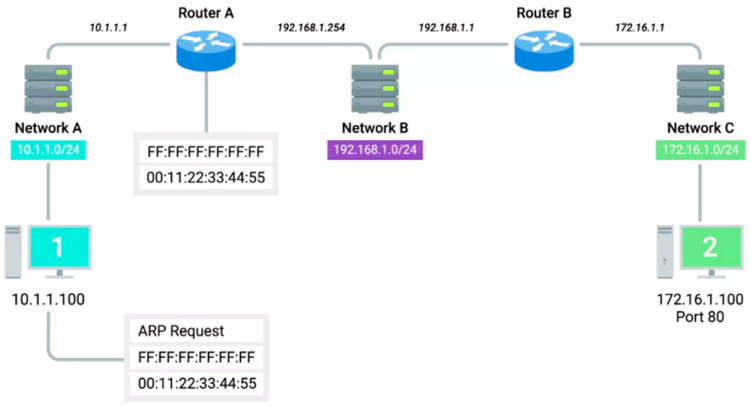


The web browser running on computer 1 knows it's been ordered to retrieve a web page from 172.16.1.100. The web browser communicates with the local networking stack, which is the part of the operating system responsible for handling networking functions. The web browser explains that it's going to want to establish a TCP connection to 172.16.1.100, port 80. The networking stack will now examine its own subnet. It sees that it lives on the network 10.1.1.0/24, which means that the destination 172.16.1.100 is on another network. At this point, computer 1 knows that it'll have to send any data to its gateway for routing to a remote network. And it's been configured with a gateway of 10.1.1.1.

Next, computer 1 looks at its ARP table to determine what MAC address of 10.1.1.1 is, if it doesn't find any corresponding entry. Computer A crafts an ARP request for an IP address of 10.1.1.1, and sends it to the hardware broadcast address (FF:FF:FF:FF:FF:FF)

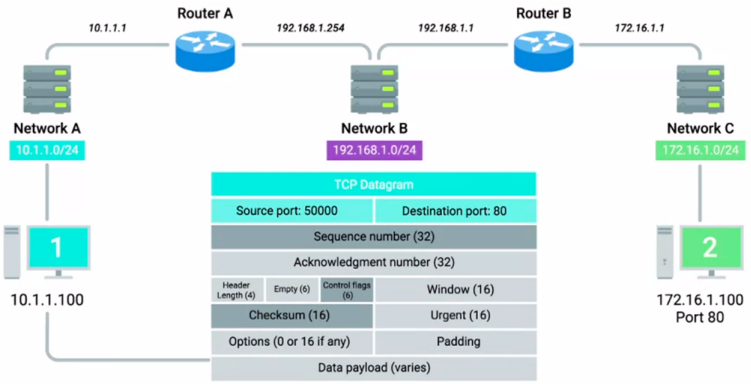


When router A receives this ARP message, it sees that it's the computer currently assigned the IP address of 10.1.1.1. So it responds to computer 1 to let it know about its own MAC address of 00:11:22:33:44:55. Computer 1 receives this response and now knows the hardware address of its gateway.



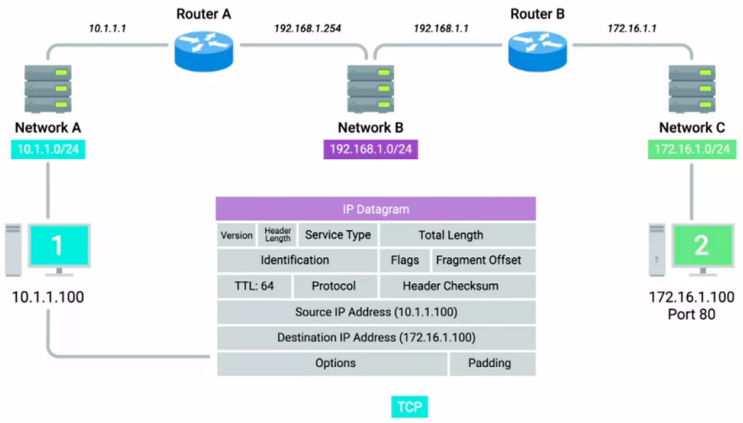
Computer 1 knows that it's being asked by the web browser to form an outbound TCP connection, which means it'll need an outbound TCP port. The operating system identifies the ephemeral port of 50000 as being available, and opens a socket connecting the web browser to this port.

Since this is a TCP connection, the networking stack knows that before it can actually transmit any of the data the web browser wants it to, it'll need to establish a connection. The networking stack starts to build a TCP segment.

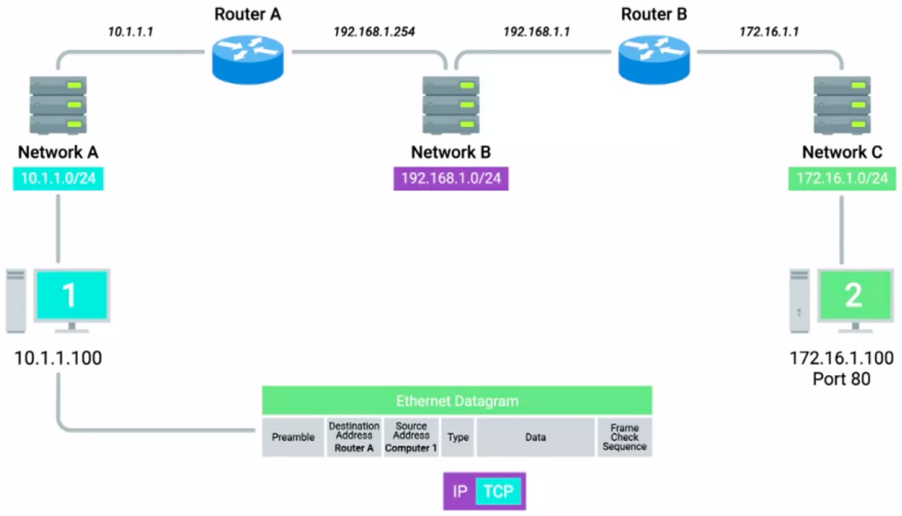


In the TCP segment the source port and destination port fields are filled, an appropriate 32-bit sequence number is chosen, the SYN flag is set, and a checksum for the segment is calculated and written to the checksum field.

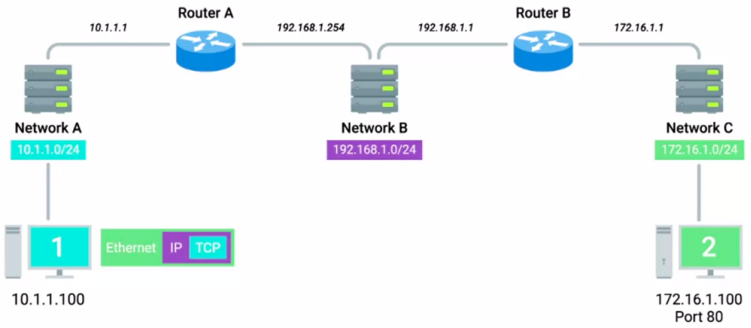
The newly constructed TCP segment is now passed along to the IP layer of the networking stack. This layer constructs an IP header. This header is filled in with the source IP, the destination IP, and a TTL of 64, which is a pretty standard value for this field.



The TCP segment is inserted as the data payload for the IP datagram. And a checksum is calculated for the whole thing. Now that the IP datagram has been constructed, computer 1 needs to get this to its gateway, which it now knows has a MAC address of 00:11:22:33:44:55, so an Ethernet Datagram is constructed.

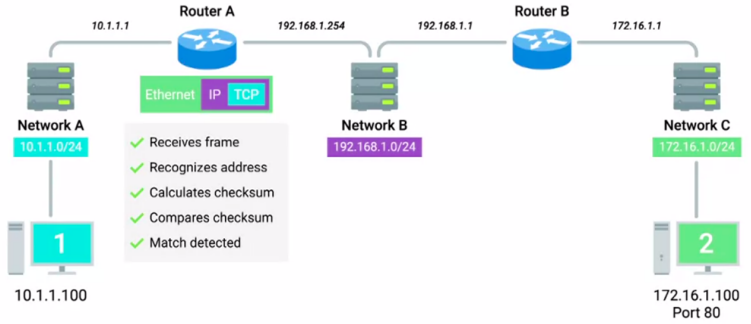


IP datagram is inserted as the data payload of the Ethernet frame, and another checksum is calculated. The network interface connected to computer 1 sends this binary data as modulations of the voltage of an electrical current running across a CAT6 cable that's connected between it and a network switch.

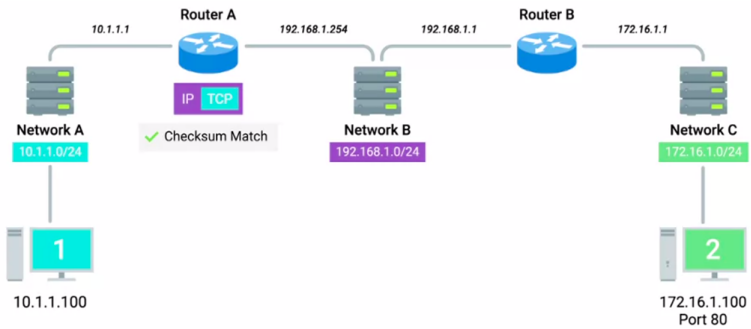


The switch receives the frame and inspects the destination MAC address. The switch knows which of its interfaces this MAC address is attached to, and forwards the frame to router A.

Router A knows that this frame is intended for itself. It calculates a checksum against it. Router A compares this checksum with the one in the Ethernet frame header and sees that they match. Meaning that all of the data has made it in one piece.

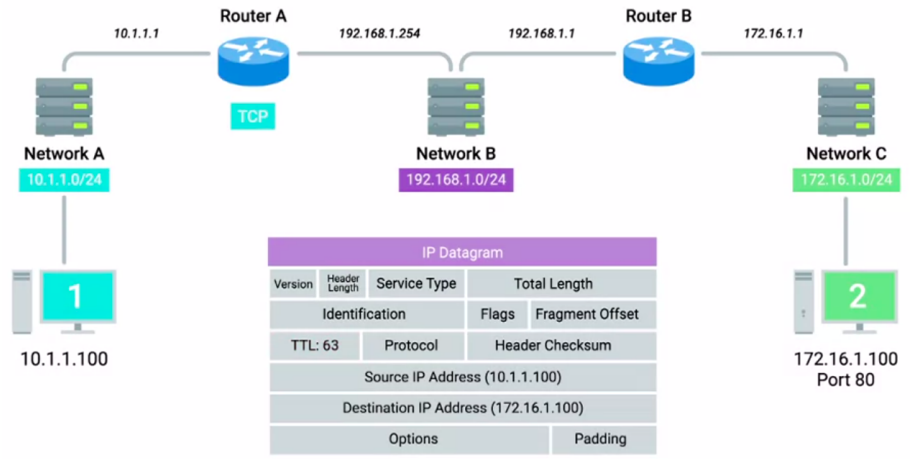


Next, Router A strips away the Ethernet frame, leaving it with just the IP datagram. Again, it performs a checksum calculation against the entire datagram. And again, it finds that it matches, meaning all the data is correct.

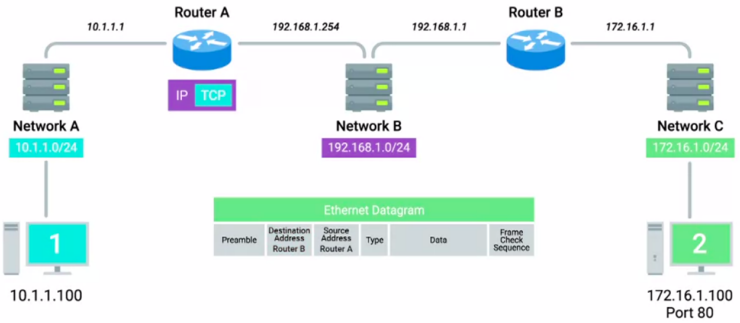


It inspects the destination IP address and performs a lookup of this destination in its routing table. Router A sees that in order to get data to the 172.16.1.0/24

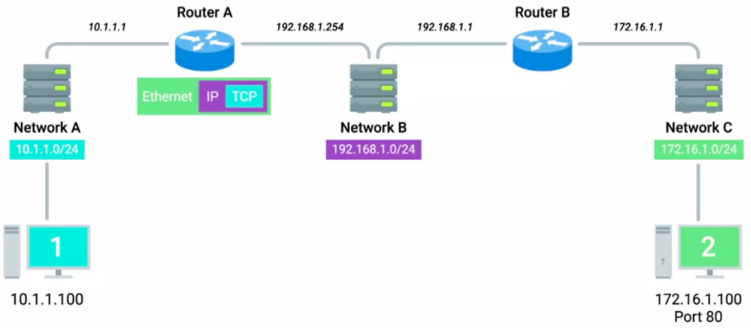
network, the quickest path is one hop away via Router B, which has an IP of 192.168.1.1. Router A looks at all the data in the IP datagram, decrements the TTL by 1, calculates a new checksum reflecting that new TTL value, and makes a new IP datagram with this data.



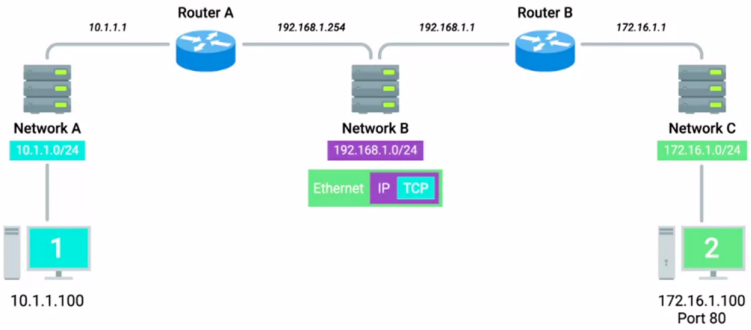
Router A knows that it needs to get this datagram to router B, which has an IP address of 192.168.1.1. It looks at its ARP table, and sees that it has an entry for 192.168.1.1. Now router A can begin to construct an Ethernet frame with the MAC address of its interface on network B as the source. And the MAC address on router B's interface on network B as the destination.



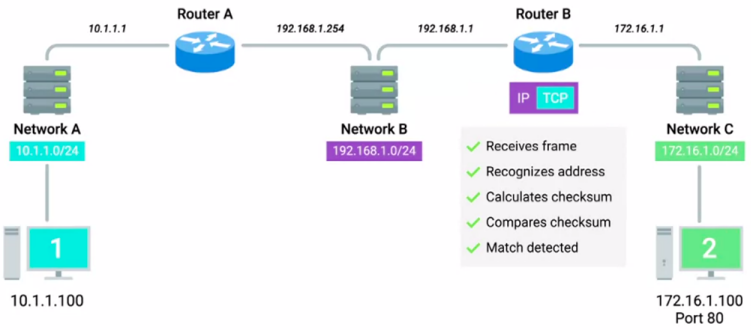
Once the values for all fields in this frame have been filled out, router A places the newly constructed IP datagram into the data payload field. Calculates a checksum, and places this checksum into place



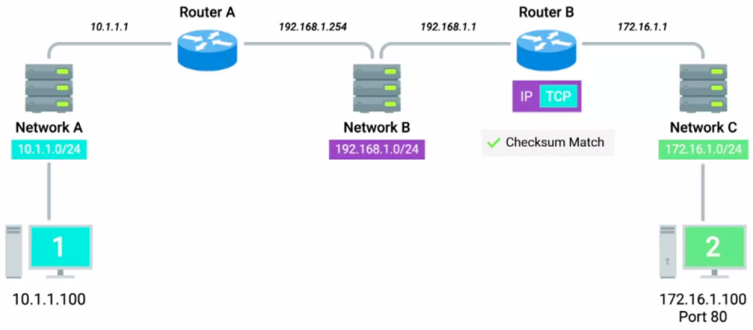
The ethernet datagram is now ready to be sent to network B.



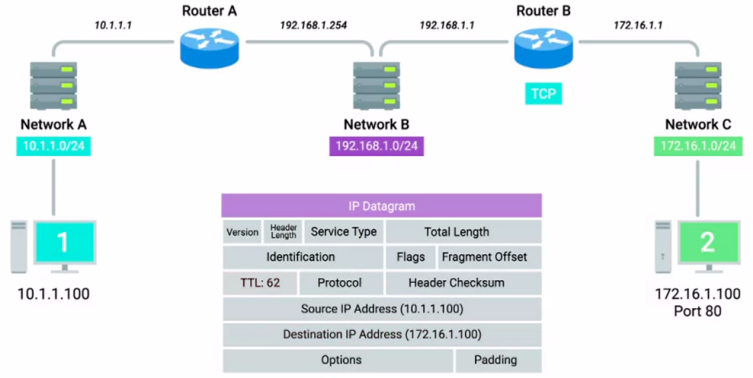
This frame makes it across network B, and is received by router B. Router B performs all the same checks



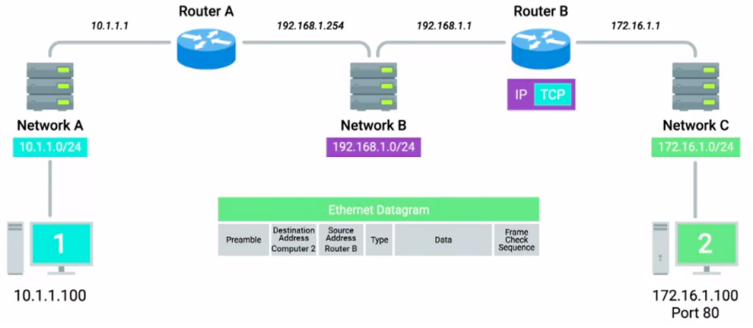
Next it removes the the Ethernet frame encapsulation, and performs a checksum against the IP datagram.



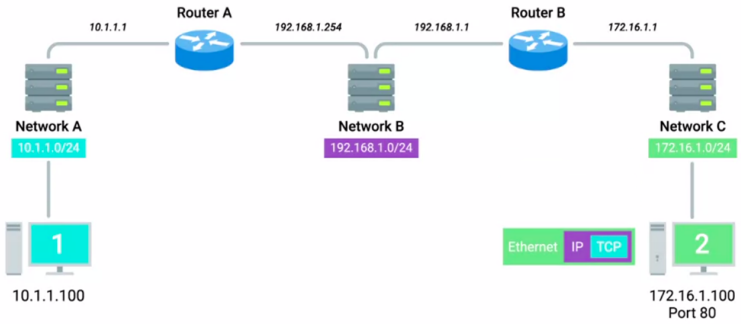
It then examines the destination IP address. Looking at its routing table, router B sees that the destination address of computer 2, or 172.16.1.100, is on a locally connected network. So it decrements the TTL by 1 again, calculates a new checksum, and creates a new IP datagram.



This new IP datagram is again encapsulated by a new Ethernet frame. This one with the source and destination MAC address of router B and computer 2.

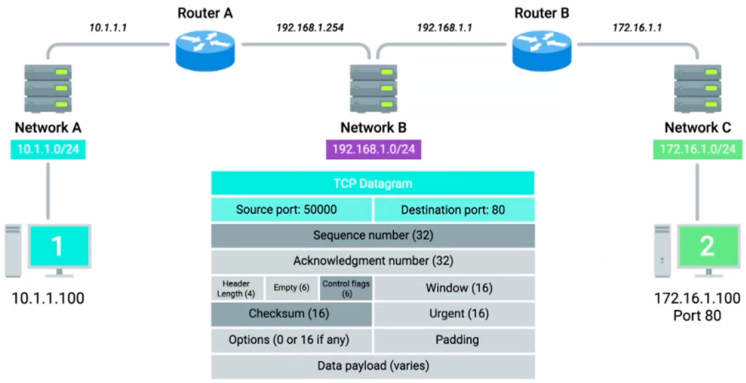


The whole process is repeated one last time. The frame is sent out onto network C, a switch ensures it gets sent out of the interface that computer 2 is connected to. Computer 2 receives the frame, identifies its own MAC address as the destination, and knows that it's intended for itself.



Computer 2 then strips away the Ethernet frame, leaving it with the IP datagram. It performs a CRC and recognizes that the data has been delivered intact. It then examines the destination IP address and recognizes that as its own.

Computer 2 strips away the IP datagram, leaving it with just the TCP segment. Again, the checksum for this layer is examined, and everything checks out. Next, computer 2 examines the destination port, which is 80.



The networking stack on computer 2 checks to ensure that there's an open socket on port 80, which there is. It's in the listen state, and held open by a running Apache web server. Computer 2 then sees that this packet has the SYN flag set. So it examines the sequence number and stores that, since it'll need to put that sequence number in the acknowledgement field once it crafts the response. Everything would have to happen all over again for computer 2 to send a SYN-ACK response to computer 1. Then everything would have to happen all over again for computer 1 to send an ACK back to computer 2. Once this is done a connection is established between computer 1 and computer 2. Then there can be multiple request-response cycles. And when either of them have nothing else to transfer, they can close the connection.

12 Domain Name System

Computers use binary language to communicate with each other, but for humans it is very hard to communicate in 1's and 0's. Humans prefer words as compared to a long string of 1's and 0's. So it is hard for a human to remember the IP address of every website he uses, as he is not good at remembering numbers, but however if each of these IP addresses is given an unique name then it is possible for the human to remember the different websites. This is where **Domain Name System** comes into picture.

DNS or Domain Name System is a global and highly distributed network service that resolves strings of letters into IP addresses. The IP address for a domain name[16]. For example

www.weather.com
is the domain name it may resolve into multiple IP address based on various factors.

For example if the server hosting the website changes, then the IP address for the website also changes. In such a scenario
Assume the DNS service was not available, then each and every end user must be informed about the change in the IP address to access the website. But if DNS service is available all that the company has to do is,

Change the IP address to which the domain name resolves into, the rest is taken care by DNS. The end users still use the same domain name, but this time they will be redirected to the new IP address. Hence DNS solves this issue very easily.

The farther we have to route data the slower is the transmission rate, so to improve the performance data transfer should take place between two nodes that are geographically close to each other. So to enhance user experience, one possible thing to do is to route the requests to the nearest (geographically) server, to ensure that the user does not have to wait for long. So the same domain name must resolve to different IP address based on the geographical location of the user. This is also possible with DNS.

The task that DNS performs of resolving domain names into IP address is called **Name resolution**.
The 4 essential things that must be configured for a host to operate on a network are

1. IP address
2. Subnet mask
3. Gateway for a host
4. DNS server

[16]Domain name is the term used for something that can be resolved by DNS

12.1 DNS Servers

A DNS server is a computer server that contains a database of public IP addresses and their associated hostnames, and in most cases serves to resolve, or translate, those names to IP addresses as requested.
The give primary types of DNS servers are:

1. Caching name servers
2. Recursive name servers
3. Root name servers
4. TLD name servers
5. Authoritative name servers

Note that the same server can be used for all these different purposes.

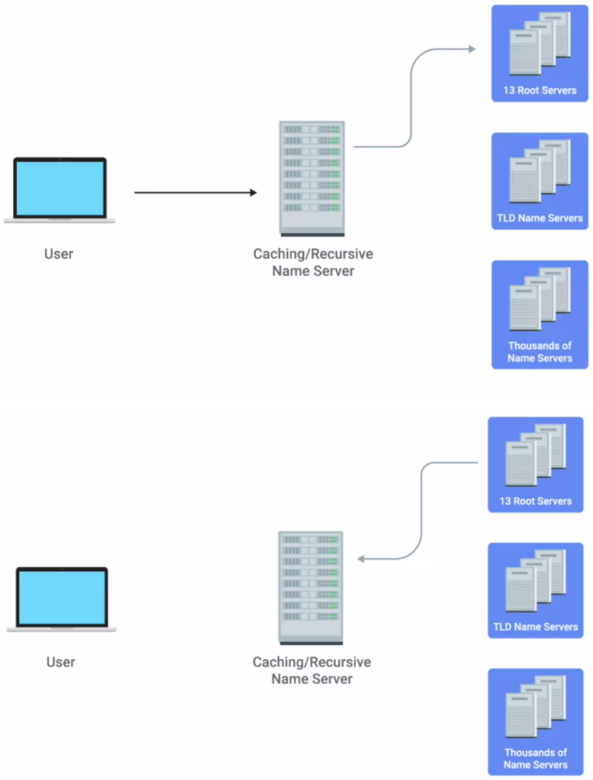
Caching and recursive name servers are generally provided by ISP. The main purpose of these is to store known domain name lookups for a certain amount of time.

Name resolution is a time consuming process, so to avoid this everytime a TCP connection is established to a frequently used website. Caching and Recursive name servers are used to store these frequently visited domain name lookups to save time.
Most Caching name servers are also recursive name servers. Recursive name servers are the ones that perform full DNS resolution requests.

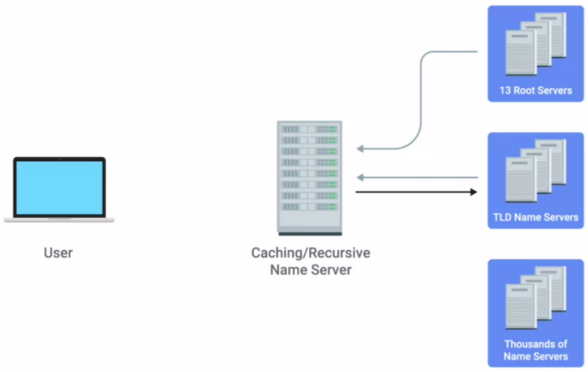
NOTE: All domain names in the global DNS system have a TTL. This is a value (in seconds) that can be configured by the owner of a domain name for how long a name server is allowed to cache an entry before it should discard it and perform a full resolution again.

12.2 Recursive Name resolution

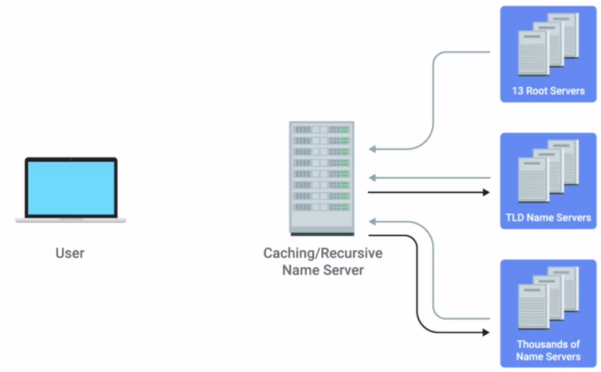
The first step is always to contact a root named server, there are 13 total root name servers and they are responsible for directing queries toward the appropriate TLD name server. **Anycast** is a technique that's used to route traffic to different destinations depending on factors like location, congestion, or link health. Using anycast, a computer can send a data gram to a specific IP but could see it routed to one of many different actual destinations depending on a few factors. The 13 root servers can be thought of as 13 authorities that provide route name lookups as a service. The root servers will respond to a DNS lookup with the TLD name server that should be queried.



Top level domain represents the top of the hierarchical DNS name resolution system. A TLD is the last part of any domain name. The TLD name servers will respond again with a redirect, this time informing the computer performing the name lookup with what authoritative name server to contact.



Authoritative name servers are responsible for the last two parts of any domain name which is the resolution at which a single organization may be responsible for DNS lookups.



This strict hierarchy is very important to the stability of the internet, making sure that all full DNS resolutions go through a strictly regulated and controlled series of lookups to get the correct responses, is the best way to protect against malicious parties redirecting traffic.

12.2.1 DNS and UDP

DNS queries uses UDP protocol in the transport layer. A single DNS request and its response can fit inside a single UDP datagram. DNS generates a lot of traffic, though DNS entries are cached on local systems and caching name servers, after refresh whenever full resolution needs to be processed it takes a lot of time. Hence TCP is not suitable for this purpose, because every time a full resolution needs to take place the number of packets to be transferred between various DNS servers will be very high, thus increasing the network traffic and also delay.

If TCP is used for DNS a minimum of 44 data packets is to be transferred to complete the DNS resolution process.

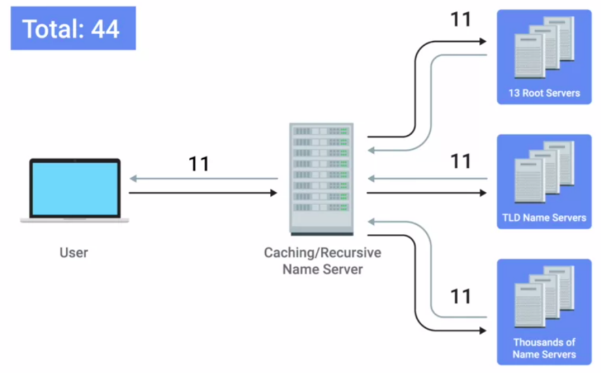


Figure 31: DNS with TCP

Using UDP for DNS resolution requires a total of 8 data packet transfers.

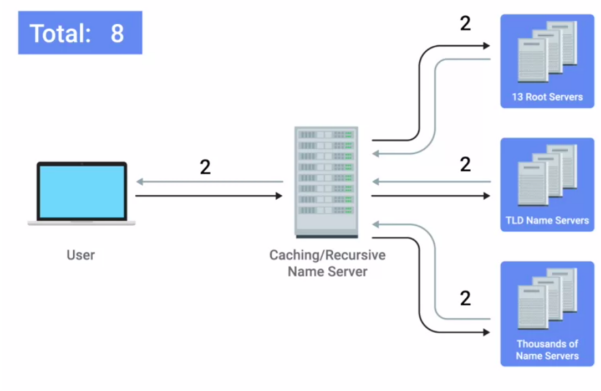


Figure 32: DNS with UDP

Since UDP does not take care of reliable data transfer, if DNS request is lost, the computer again makes a request for DNS resolution. Basically the functionality provided by TCP in transport layer is provided by DNS in application layer.

If DNS lookup response does not fit in a single UDP packet, the DNS name server would respond with a packet explaining that the response is too large, the DNS client would then establish a TCP connection in order to perform the lookup.

12.3 Resource record types

DNS operates with a set of defined resource record types. These allow different types of DNS resolution to take place. The most common resource record is called **A record**. An A record is used to point a certain domain name at a certain IPv4 IP address. A single **A record** is configured for a single domain name. But a single domain name can have multiple Arecords.

The main website of some big company will have many users every day. Hence multiple Arecords are used for this company. Assume 4 Arecords are used the corresponding IP addresses are 10.1.1.1, 10.1.1.2, 10.1.1.3, 10.1.1.4 Whenever an user makes a request for the DNS resolution of this company's website, all the 4 Arecords are returned. The DNS name server sends the first IP address (10.1.1.1) to the user and places it at the end. If another user requests for the same domain name resolution, this time the DNS server returns the second IP address (10.1.1.2) and places this at the end. This is done to equally divide the network traffic. This technique is called **DNS round robin**.

AAAA - Quad A record is similar to A record, the only difference being it returns an IPv6 address instead of IPv4.

CNAME record is used to redirect traffic from one domain name to another. CNAME is short for Canonical Name

MX (Mail exchange) record is used to deliver the email to the correct server. A large companies will have multiple servers, MX resource record ensures that emails are sent to the email server, where as the other network traffic is redirected to other web servers.

SRV (service) record is used to define the location of various specific services. MX is used only for mail services, SRV record can be defined to return the specifics of many different service types.

TXT (text) record is used to communicate extra data with the user. Like configuration preferences about network services.

12.4 Anatomy of a Domain name

Every domain name consists of majorly three parts

1. Sub Domain
2. Domain
3. Top level Domain

Combining all the three parts together we obtain something called as **Fully qualified domain name (FQDN)**. DNS can support up to 127 levels of domain in total for a single FQDN. Each individual section can be only 63 characters long. A complete FQDN is limited to 255 characters.

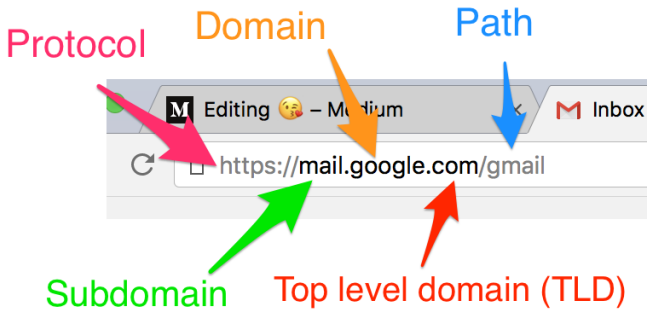


Figure 33: Different components of domain name

The number of different TLDs are ever increasing. Few of the very frequently used ones are *.com*, *.edu*, *.net*. There are some TLD specific to a county. The administration and definition of a TLD is handled by **ICANN (Internet Corporation for Assigned Names and Numbers**

Domain are used to demarcate where control moves from a TLD name server to an authoritative name server. Any individual can purchase a certain domain name for a certain period of time from a registrar[17]. However subdomains can be freely chosen and assigned by anyone who controls such a registered domain.

12.4.1 Hosts files

Before DNS, the original way that numbered network addresses were correlated with words was through **hosts files**. Hosts files contains a network address followed by the host name it can be referred to as, on each line.
All modern operating systems still have hosts files, one of the main reason for this is the **loopback address**[18]. For IPv4 this address is 127.0.0.1 Most operating systems contain a hosts file with atleast one entry that reads

127.0.0.1 localhost

In case of IPv6 the loopback address is ::1

NOTE: Hosts files are a popular way for computer viruses to disrupt and redirect user's traffic to malicious sites.

So even before a name resolution request is sent to the local DNS server, the hosts files are examined by the operating system.

12.5 DNS Zones

An authoritative name server is responsible for a specific DNS zone. DNS zones are a hierarchical concept, root name servers are responsible for root zone, each TLD name server is responsible for the zone covering its specific TLD. Authoritative name servers are responsible for some even finer-grained zones.

NOTE: Zones do not overlap.

DNS Zones allow for easier control over multiple levels of a domain. As the number of resource records for a domain increase, it can be split into multiple zones to make the handling easier. Zones are configured through **Zone files**. Zone files are simple configuration files that declare all resource records for a particular zone. A zone file contains

[17] A registrar is a company that has an agreement with ICANN to sell unregistered domain names

[18] Loop back address is a way of sending network traffic to yourself

- Start of Authority (SOA) records: Declares the zone and the name of the name server that is authoritative for it
- NS records: Indicate other name servers that might also be responsible for this zone.
- Other resource records can also be found

Reverse lookup zone files allow DNS resolvers ask for an IP and get the FQDN associated with it returned. These are similar to zone files, but instead of resolving domain names to IPs, they contain pointer resource record[19] declaration.

[19]Pointer resource record (PTR) resolves an IP to a name

13 Dynamic Host Configuration Protocol (DHCP)

Every time you connect to a network, your system will first be assigned an IP address. Every system on a network should have 4 things specifically configured.

1. IP address
2. Subnet mask for the local network
3. Primary gateway
4. Name server

3 out of the 4 things mentioned above are common to every system on the network. The remaining one IP address needs to be different for every single node on the network. This involves a lot of work, which is handled by **DHCP**.

DHCP is an application layer protocol that automates the configuration process of hosts on a network.
With DHCP any new machine connecting to the network can query the DHCP server and receive all the network configuration details.

For devices like the DNS server or the gateway router, a static IP address that is known to all nodes is required. As they are very essential for the operation of the network and different nodes keep sending data to these. However for other devices like a laptop, mobile etc they just need an IP address which need not be static as they wont be permanently connected to the network. The IP address assigned to these devices will be valid for that session. They can be assigned a different IP address in the next session.

13.0.1 Dynamic Allocation

Using DHCP it is possible to configure a range of IP addresses that is set aside for these client devices. And one of these IPs is issued to the device when they request for one. Hence IP given to a device may change in every session. This is called **dynamic allocation**.

13.0.2 Automatic Allocation

The other mode of working is **Automatic Allocation** in which a range of IP addresses is set aside for assignment purposes. But the drawback here is that the DHCP server must keep track of which IP address was assigned to which device so that the same IP is assigned to the machine each time it connects to the network if possible.

13.0.3 Fixed Allocation

This mode of operation requires a manually specified list of MAC address and their corresponding IPs. When a computer requests for an IP, the DHCP server looks for the MAC address of the requesting device in the table and then assigns

the corresponding IP address. If MAC address is not found then the DHCP server may use automatic/dynamic allocation or it might even refuse to assign IP address.

NOTE: DHCP discovery can also be used to assign NTP servers[20]

13.1 Working

DHCP is an application layer protocol that is used to set up configurations in the network layer.

13.1.1 DHCP Discovery

This is the process by which a client configured to use DHCP attempts to get network configuration information. This is a 4 step process

1. **DHCP Discover**
   The DHCP client sends a DHCP discover message to the network. Since the client does not have an IP address (0.0.0.0 is used), and it does not know the address of DHCP server (255.255.255.255 is used), it sends a broadcast message to the network. The DHCP server makes note of the MAC address of the new client.

   DHCP server listens on UDP port 67, DHCP discover messages are sent from UDP port 68.

Figure 34: DHCP Discover

2. **DHCP Offer**
   Upon receiving the request DHCP server responds with another broadcast message which includes the IP address it can offer

   It is possible that the client can reject the offer, as multiple DHCP servers will be operating in the network, the client may be configured to only respond to an offer of an IP within a certain range.

[20]Network Time Protocol (NTP) servers are used to keep all computers on a network synchronized in time

Figure 35: DHCP Offer

3. **DHCP Request**
   The client responds to the DHCP Offer message with a DHCP Request message (broadcast message). In which the acceptance of the offer is mentioned.

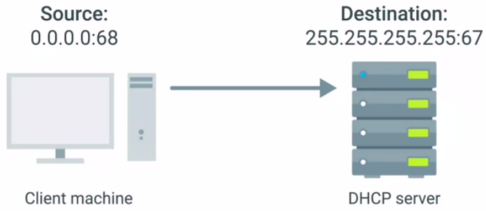


Figure 36: DHCP Request

4. **DHCP Acknowledgement**
   The DHCP server upon receiving the REquest message responds with the DHCP Acknowledgement message. This is again a broadcast message. the new client uses the Destination MAC address to understand that the broadcast message is intended for it.

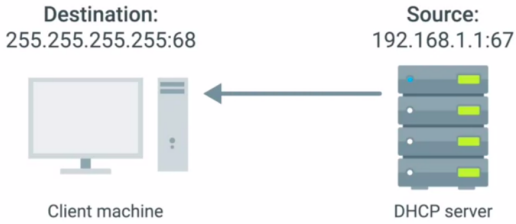


Figure 37: DHCP Acknowledgement

Upon receiving the acknowledgement message, the new client has all the required configuration set up, to be fully operational. The above operation is called **DHCP Lease**. Once the lease has expired, the client must again request for

an IP address. The above process repeats again. The lease period can extend from few minutes to a few days.
A client can also release its lease by disconnecting to the network, the DHCP server then adds this IP address freed by the client, to the available pool of IP address which can be used for some other client.

14 Wireless networking

Since many of the devices that connect to the internet are portable, it does not make sense to have wired technology to connect the user to the internet. Wireless networking connects the user to the internet without any wires.

The most common specifications for how wireless networking devices should communicate, are defined by **IEEE 802.11** standards. Radio waves are used by different devices to communicate with each other.

802.11 standards define how operations take place at physical and data link layers. The most commonly used specifications are 802.11b, 802.11a, 802.11g, 802.11n, 802.11ac. Each newer version of the 802.11 standards has seen some improvement, interms of higher access speeds or the ability for more devices to use the network simultaneously. All these different standards operate with the same datalink protocol, but how they operate at physical layer varies. Each of these specifications can have different ranges, can use different modulation techniques, can have different transmission bit rates, operate at different frequency bands.

Switches are used to connect the computers to the network in case of a wired network, this role is placed by Access points in case of wireless networks.

Figure 38: 802.11 frame

- **frame control** field contains a number of sub fields used to describe how the frame should be processed
- **duration** field specifies how long the total frame is.
- **Address 1** field represents source MAC address
- **Address 2** field represents destination MAC address
- **Address 3** field represents the MAC address of the access point that should receive the frame
- **Address 4** field represents the MAC address of whatever that transmitted the frame.
  NOTE: In some cases source and transmitter address can be same and receiver and destination address can be same.
- **sequence control** field contains a sequence number used to keep track of ordering the frames

- **data payload** contains all the data of the protocols further up the stack
- **Frame check sequence (FCS)** field contains a checksum used for Cyclic redundancy check.

WiFi operating at 5GHz frequency band is generally faster and has a lesser range as compared to that operating in 2.4GHz frequency band.

14.1 Wireless network configurations

Different ways in which a wireless network can be configured are:

1. **Ad-hoc networks**: All nodes directly speak to each other
2. **Wireless LANs (WLANs)**: One or more access points act as a bridge between a wireless and a wired network
3. **Mesh networks**: Hybrid of the above two.

14.2 Wireless Channels

Channels are individual smaller sections of the overall frequency band used by a wireless network. With wired networks, switches can be used to reduce collision domains. But when it comes to wireless networks, a similar functionality is achieved by channels.

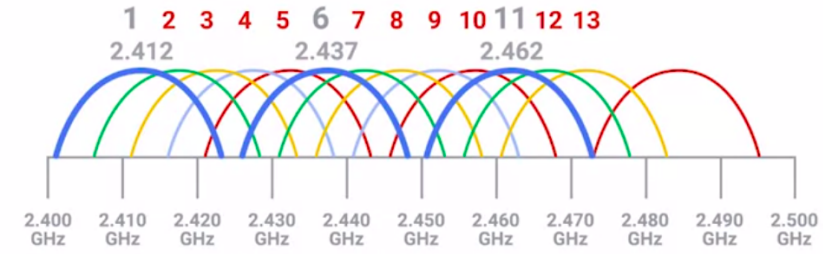


Figure 39: Wireless network channels

The bandwidth of every channel is 22MHz. From the above diagram it can be seen that in a 802.11b network channels 1,6,11 do not overlap.

Most networking equipment can automatically sense what channels are most congested. Some access points only perform this analysis when they start up. So it is essential to switch channels dynamically when more congestion is experienced on a given channel.

14.3 Wireless Security

Because of wireless transmission of data, anyone within the range can intercept it. To provide more security to the data **Wired Equivalent Privacy (WEP)** was invented. WEP is an encryption technology that provides a very low level of privacy. WEP uses 40 bits for its encryption keys. Due to the low security offered by WEP it was replaced by **WiFi Protected Access (WPA)**. WPA uses a 128 bit key, hence making it harder to crack than WEP. Today the most commonly used encryption algorithm for wireless networks is **WPA2**. WPA2 uses 256-bit key and hence makes it even harder to crack it.

14.3.1 MAC filtering

An alternative to WPA is MAC filtering. With MAC filtering the access points are configured to only allow for connections from a specific set of MAC addresses belonging to devices that are part of the trusted network. This provides an additional barrier preventing unauthorized devices from connecting to the wireless network itself.

14.4 CLOUD

Cloud is just a concept. **Cloud computing** is a technological approach where computing resources are provisioned in a shareable way, so that lots of users get what they need, when they need it.

Hardware virtualization[21] is the core concept of how cloud computing technologies work. It allows the concept of a physical machine and a logical machine to be abstracted away from each other. Hardware virtualization platforms employ **hypervisor**[22]

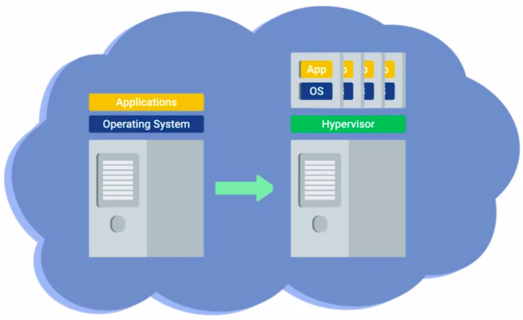


Figure 40: Cloud computing

[21] Virtualization is a technique where a single physical machine, called a host could run many individual virtual instances, called guests.

[22] Hypervisor is a piece of software that runs and manages virtual machines, and also offers the guests a virtual operating platform that's indistinguishable from actual hardware.

With a huge cluster of interconnected machines that can all function as hosts for lots of guests, it is possible to share resources amongst all of those instances.

14.4.1 Infrastructure as a Service (IaaS

Companies dont have to worry about building their own network, they can pay some one else to provide those infrastructure as a service. The other service types include

- **Platform as a Service (PaaS)**:
  A platform is provided to customers to run their services.
- **Software as a Service (SaaS)**:
  A way of licensing the use of a software to others while keeping that software centrally hosted and managed. Ex: Gmail, Outlook

15 Multicast

This is a communication technique where packets are transferred from a single source to multiple interested receivers. This is different from broadcast as the packets are not sent blindly to all nodes on the network and it is only sent to a subset of users who are interested to receive the packets. Similarly it is also different from replicated unicast as in case of multicast the source transmits the packet only once. But in case of replicated unicast the source has to transmit the packet as multiple (= number of users) times.

The advantages of using Multicast over broadcast/ replicated unicast

- Better resource utilization
  - Source has to transmit a single copy
  - Receiver does not receive unwanted packets
- Better bandwidth utilization
  - No single link will have multiple copies of the same packet
  - Packets are only sent to the users who have requested for it

15.1 L3 Multicast

The source IP address in case of multicast/unicast remains the same. However the destination IP address is different for a unicast/multicast packet. The destination IPs represent the mutlicast group. The destination multicast IP addresses belong to class D.

224.0.0.0 - 239.255.255.255

Also note that Class D IP space is considered to be flat and there is no concept of subnetting here. Every single IP represents a multicast group.

There are sub-ranges within ClassD IP addresses

- 224.0.0.0/24 - Local Network Control Block
  - Operate within a link-local scope (ex: Within a VLAN)
  - TTL of 1 or 2
- 232.0.0.0/8 - Source specific Multicast block
- 239.0.0.0/8 - Organization-Local Scope (Private IPs)

Since the destination IP is multicast, a suitable MAC address must be used for the destination in the L2 frame. So the sender and receiver must agree upon a single MAC address, and the interested receivers must accept frames with this destination MAC address in the L2 frame. The source IP is unicast hence the traditional ARP/NAT concepts hold with respect to source MAC and IP address.

15.2 L2 Multicast

Incase of unicast communication, the MAC address used for communication will be that of the device (Burnt-In Address (BIA)). And hence packets with that destination address will be received. But in case of mluticast a different destination IP address is used and hence a different MAC address is used. The multicast IP address to be used by the device is generated by the application being run at the application layer. NIC will be tuned to the multicast address. But in order to receive the L3 packet NIC must be aware of the L2 address as well.
For multicast the address must be dynamic and well-known

- Dynamic because the L3 multicast groups are highly arbitrary
- Well-known because a unique L3 multicast address must have a consistent L2 address.

To provide the above a mapping is created from the Multicast IP address to the MAC address to generate a well-known multicast MAC address.

- The first 24 bits will be 01-00-5E (This identifies IP multicast)
- The next bit will be 0
- The last 23 bits of the MAC address will be the same as the last 23 bits of the Multicast IP address

NOTE: If the last bit of the first octet in the MAC address is

- 0 $\rightarrow$ individual address (unicast)
- 1 $\rightarrow$ group address (multicast)

Note that 32 bit multicast IP address is mapped to the 23 bits of the MAC address. Out of the 32 bits, 4 bits are reserved for ClassD IP address space. So the remaining 28 bits are dynamic and are mapped to the 23 bits of the MAC address. Hence there is a possiblity that 32 different IP addresses can have the same MAC address for the multicast communication. But given a subnet mask of /8 only 2 nodes will have the same MAC address, hence the same mapping is still being used.

15.3 IP Multicast Routers

The same routers can be used to route unicast as well as multicast packets. The mechanisms and logic used to forward unicast packets is different from that used for multicast packets.

- In unicast packet forwarding the packets are guided towards a destination. In multicast packet forwarding the packets are guided away from a source. Hence it is termed as **Reverse Path Forwarding (RPF)**

- Unicast routing table will have destination IP, next hop, exit interface. In case of multicast the routing table consists of two major types of interfaces
  - Upstream interface:
    Metrically closest to the source.
  - Downstream interfaces:
    Interfaces with interested receivers

  In case of multicast the source transmits the packet only once. This packet is replicated and sent to multiple interfaces of the router from where they reach different networks. While doing so the interface closest to the source is used as an upstream interface (the interface from which the router receives the multicast packet) because in case of packet drop, it will be easier to retransmit the packet through upstream interface (as it is closest to the source) as compared to any other interface.

- Multicast routing tables stores and organizes information by multicast groups (G) which are also called as forwarding states

- Each multicast group has a source
  1. can be a known source (IP address is known) indicated by "S"
  2. can be any/unknown source indicated by "*"

  In case of (1) the node is ready to accept packets directed towards a particular muticast group from a particular source. In case of (2) the node is ready to receive packets from any source directed towards the particular multicast group.
  The table is organized by a combination of sources for the groups.
  - (S,G) is the forwarding state for known source
  - (*,G) is the forwarding state for unknown sources

  .

- Each forwarding state (S,G) or (*,G) has an associated set of upstream and downstream interfaces.
  - (upstream) Single incoming interface is called IIF
  - (downstream) Outgoing interface list are called OIL

15.3.1 Responsibilities of Multicast Routing protocol

- Using the unicast forwarding table, the interface on the shortest path to the source is chosen as the upstream interface (IIF)

- The routing protocol will provide procedures for the receivers to signal interest. As the request comes in to the router, OIL must be updated

- Add and remoce both IIFs and OILs dynamically as the sources and receivers come online and go offline.

15.4 Data flow in multicast

- Multicast packets arrive at a router on IIF for a group
- Multicast packets are forwarded by the router on the OIL for the group
- The correct execution of the above two functionalities is very critical. As breaking of anyone of those may lead to the formation of loops.
- Multicast loops are very dangerous as multicast forwarding includes packet replication at each router. And after every iteration of the loop the number of multicast packets keep increasing and this results in **multicast storms**.

To prevent these multicast storms routers perform **Reverse path forwarding (RPF)** check. This is done as follows

- Upon receiving a multicast packet the router inspects the source IP of the packet.
- The upstream interface (IIF) for the source is identified
- If the packet was not received on a valid IIF, it is dropped.

16 Types of Networks

- Personal Area Network (PAN)
- Local Area Network (LAN)
- Metropolitan Area Network (MAN)
- Wide Area Network (WAN) - private
- Wide Area Network (WAN) - public

16.1 Personal Area Network

This is the smallest and the most basic type of network. Typically found in homes. Example: IOT, Smart homes etc

16.2 Local Area Network

LANs are networks that are within a single institution (set of buildings). Routers are used to connect different networks with each other. Within a LAN switches are used to connect the different computers to the network.

In case of a LAN the internal wiring that connects different networks is done and owned by the company itself.

16.3 Metropolitan Area Network

MAN covers a larger geographical area than a LAN, hence the external wiring done to connect the networks in a MAN is owned by some other company. MAN requires a service provider to connect the geographically separated(within the same metropolitan area) networks.

16.4 Wide Area Network

WAN is a network that covers a very large geographical area, it may be spread across countries. The services of a Service Provider (SP) are required to connect the sites to each other.

Connecting LAN and MANs of the same company results in a **private WAN**. Connecting WAN, MAN and LANs of different institutions results in a **public WAN**. Example: Internet

17 Virtual Private Networks (VPN)

Organizations have proprietary information that needs to remain secure, network services that are only intended for employees to access and other things that are only accessible when physically connected to LAN. But employees are not always in office they may be working from home as well, to still be able to access all the network services requires the usage of Virtual Private Network (VPN).

VPN is a technology that allows for the extension of a private or local network to hosts that might not be on that local network. VPNs are a tunneling protocol i.e. they provide access to something not locally available.

Figure 41: VPN setup

17.1 Working

VPN provides the remote client a virtual interface with an IP that matches the address space of the network that they established a VPN connection to. By sending data through this virtual interface, the remote client can access internal resources just like if it was physically connected to the private network.

VPNs work by using the payload section of the transport layer to carry an encrypted payload that actually contains an entire second set of packets. The network, transport and application layers of a packet intended to traverse the remote network. This payload is carried to VPNs end point, where all the other layers are stripped and the payload is unencrypted leaving the VPN server with the top3 layers of a new packet. This is encapsulated with a proper datalink layer information and sent out across the network. The same process is completed in the inverse for traffic in opposite direction

17.2 Security

VPNs usually require strict authentication procedures to ensure that they can only be connected to by employees and users authorized to do so. VPNs were one of the first technology to use **two-factor authentication**[23].
VPNs are a general technology concept not a strictly defined protocol

[23]Two-factor authentication is a technique where more than just a username and password are required to authenticate

17.3 Proxy Services

Proxy service is a server that acts on behalf of a client in order to access another service. Proxies sit between clients and other servers providing additional benefits like

- Anonymity
- Security
- Content filtering
- Increased performance

A gateway router is an example that acts as a proxy for the nodes within the network.

17.3.1 Web Proxy

Web proxies are used to improve the performance. All the network traffic is directed to these web proxies, where the data received as a response from the server is cached, so that if any other user requests for the same data, it is directly sent to the user without actually making another request to the server.

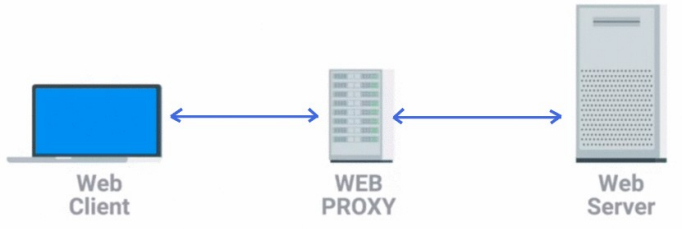


Figure 42: Web proxy

Web proxies are used for a different purpose now. Since all the web traffic is directed to web proxy, it can be used to filter out the web requests. For example, a company may restrict access to social media sites during work hours for the employees.

17.3.2 Reverse Proxy

This is a service that might appear to be a single server to external clients, but actually represents many servers living behind it. Websites like google, twitter receive so much traffic that there's no way a single web server could possibly handle all of it. So many servers are required in order to keep up with processing all incoming requests. Similar to DNS round robin this is also a technique to uniformly divide the traffic to multiple servers to ensure optimal performance for the users.

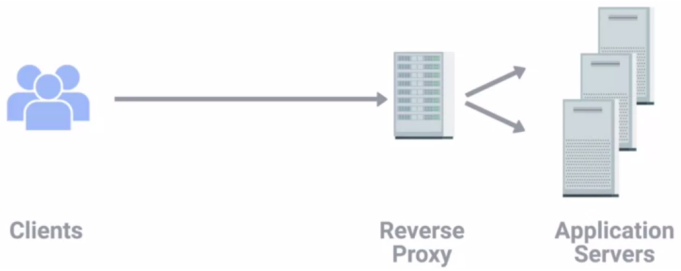


Figure 43: Reverse proxy

Another common application of reverse proxies is to deal with decryption. The traffic on internet is encrypted, in order to respond to these requests the servers must first decrypt the data. Encrypting and decrypting takes a lot of processing power. So reverse proxies are now implemented to perform encryption and decryption work, so that the servers have to just serve the content, without having to worry about encryption and decryption.

18 Virtual LAN (VLAN)

LAN is a layer2 broadcast domain. Many a times it is preferable to have larger network break down into smaller ones, as this increases security and makes maintenance easier. The nodes containing more sensitive information can be separated, and a firewall can be used to monitor the network traffic through these nodes.

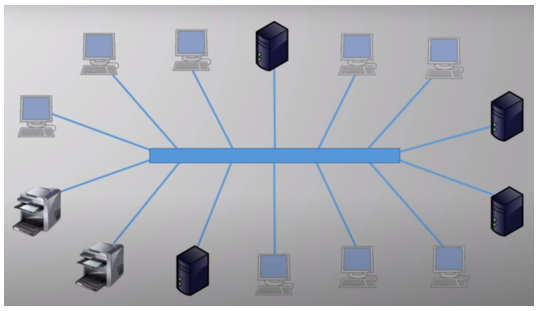


Figure 44: LAN

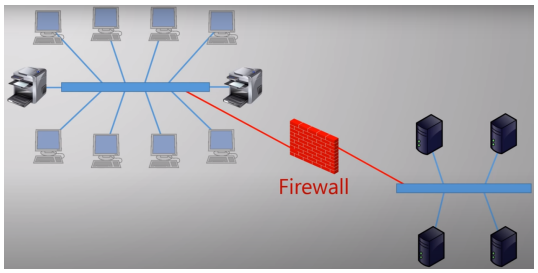


Figure 45: Splitting a LAN

The simplest approach to divide a LAN is to have multiple switches and connect them using a router. This option however is not scalable. The other alternative is to use a **VLAN**.

With VLAN it is possible to break up one Physical switch into multiple Virtual switches. This is done using VLAN Ids. Traffic on one VLAN ID does not interfere with the traffic on a different VLAN ID.

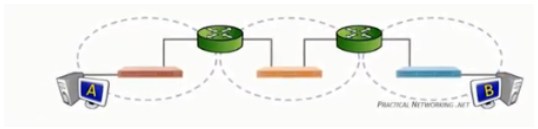


Figure 46: Three different networks communicating via 3 switches and two routers

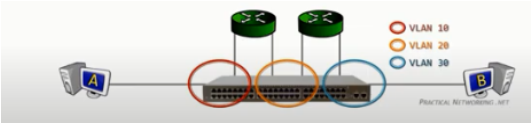


Figure 47: The same network setup using a VLAN (2 routers and a single switch(3 VLAN IDs))

As the network size grows the number of broadcast messages also increase, which increases congestion on the network. But by VLAN the size of the network is reduced and hence it also avoids congestion.

It is also possible to further extend the virtual switches to other physical switches. But in order to extend, requires more ports

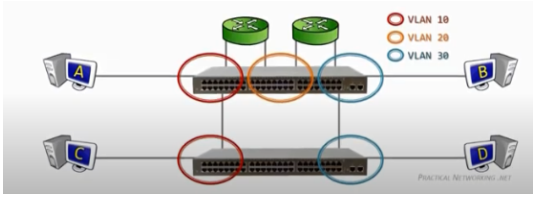


Figure 48: Extending virtual ports to physical ports

It can be observed that each extension requires another port on the switch. This again causes scalability issues. To avoid this a technology was developed where multiple VLANs traverse through the same link, this is called **trunk ports**.

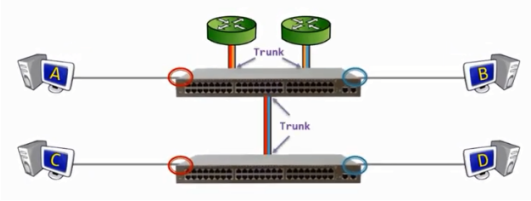


Figure 49: Trunk ports

Ports that carry a single VLAN are called **Access ports**. VLAN ID (VLAN# in the below image) is used to segregate the data on the trunk port.

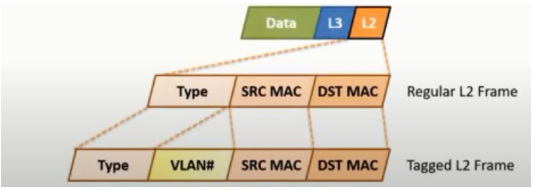


Figure 50: VLAN frame vs LAN frame

VLAN tag field in the frame is used to make sure that the frame is isolated from ports belonging to other VLANs when it is within the switch fabric.

When a switch receives a packet from an end-station it tags it with a VLAN ID. The tag is removed when the frame moves out of the switch fabric. VLAN ID is a 12 bit number and hence the maximum number of VLANs possible in a given network is 4096. Even if VLANs are not configured on a VLAN capable Cisco switch all the ports on the switch by default belong to VLAN 1

18.1 Configuring a VLAN

To configure (for example eth 0/0) as an access port belonging to VLAN 10

```
R1# conf t
R1(config)# int e0/0
R1(config-if)# switchport mode access
R1(config-if)# switchport access vlan 10
```

It is possible to use comma separated list of ports and configure them as access ports for a particular VLAN using **interface range** command.

```
R1# conf t
R1(config)# interface range e0/0, e0/1, e0/2, e0/3
R1(config-if)# switchport mode access
R1(config-if)# switchport access vlan 10
```

When the VLAN spans across multiple physical switches, trunk ports are used to connect these switches. But in this case when the frame leaves the switch fabric the tag should not be removed as the next switch would not know to which VLAN this frame belongs to. Hence tagging and untagging should not be performed when the frame enters/exits the switch via trunk ports.

18.1.1 Configuring trunk ports

Since trunk ports connect different switches, the encapsulation used by one should be understood by the other. The standard encapsulation techniques used are

- Cisco Switch ISL
- dot1q (Preferred option)

To configure a trunk port with dot1q encapsulation

```
R1# conf t
R1(config)# int e0/0
R1(config-if)# switchport trunk encapsulation dot1q
R1(config-if)# switchport mode trunk
```

By default all VLANs are allowed to travel accross the trunk. This can be restricted using

```
R1(config-if)# switchport trunk allowed vlan 10-20
```

In the above example the trunk port only allows traffic from VLANs 10,11,...,20. To add more VLANs to this list use keyword **add** in the command else it will overwrite the previous command

```
R1(config-if)# switchport trunk allowed vlan add 21, 22
```

This will allow traffic from VLANs with ID 21 and 22 to pass through the trunk port.

18.2 VLAN Trunking Protocol

VTP is a Cisco layer 2 messaging protocol that manages addition, deletion and renaming of VLANs on a network wide basis. When a new VLAN is configured on one VTP server, the VLAN is distributed through all switches in the domain. This reduces the need to configure the same VLAN everywhere. But configuring the ports to belong to a particular VLAN must be done on each switch separately.

A switch participating in VTP can have 3 different modes

- Server mode:
  VLAN creation. deletion is done here
- Client mode:
  Mimic the VLAN creation and deletion that has happened in the server
- Transparent mode:
  A switch not participating in VTP but involved in transferring VTP messages is said to be in transparent mode.

18.3 Inter VLAN Routing

Inter-VLAN routing can be accomplished either by using a **Router** or a **Layer 3 Switch**[24]

[24]Layer 3 switch also has the ability to perform routing functions

18.3.1 Using Router

Assume there are two VLANs on a switch, then if a router has two interfaces connected to this switch it can be used to perform inter vlan routing.

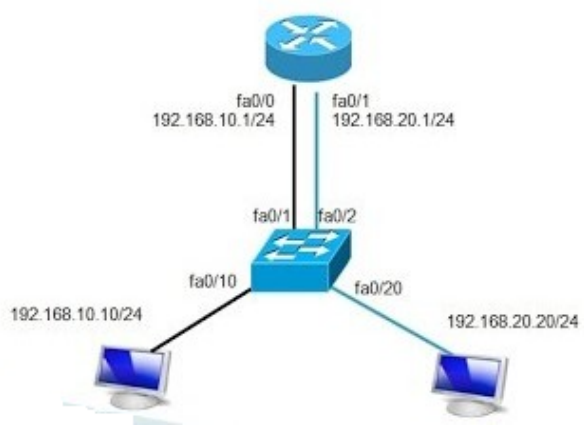


Figure 51: inter VLAN routing using a router

In the above topology there are two VLANs 192.168.10.0 and 192.168.20.0 hence two interfaces of the router is connected to this switch.

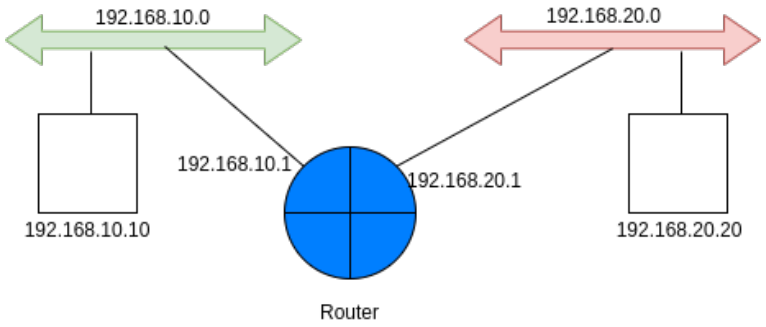


Figure 52: Logical diagram of the previous network

Red arrow represents one VLAN which is logically isolated from the Green arrow that represents the second VLAN. Though both the VLANs are physically existing on the same switch they are logically separated.

Similarly if there are N VLANs on a switch, N interfaces of the router would be needed. To address this a different technique is used which is called **Router on a stick**. Where only one of the interfaces of the router is connected to the switch and all the messages from different VLANs reach the router through this connection only.

For this reason **sub interfaces** are used. Sub interfaces are logical interfaces and one sub interface is used per VLAN. And the port of the switch that is connected to the router should be a trunk port. This is because the VLAN tag should not be removed by the switch when it is forwarding the packet to the router. As the router will use this tag to further send the packet to the appropriate sub interface. To configure the above network

```
Router(config)# int e0/0
Router(config-if)# no shut
Router(config-if)# int e0/0.1
Router(config-if)# encap dot1q 10
Router(config-if)# ip add 192.168.1.1 255.255.255.0
Router(config-if)# int e0/0.2
Router(config-if)# encap dot1q 20
Router(config-if)# ip add 192.168.2.1 255.255.255.0
```

"encap dot1q 10" ensures that frames tagged with VLAN ID 10 are allowed in the logical interface e0/0.1. Similarly "encap dot1q 20" ensures that frames tagged with VLAN ID 20 are allowed in the logical interface e0/0.2

18.3.2 Using L3 Switches

Switch Virtual Interfaces are created L3 switches to handle inter vlan routing. So it is like there is a router within the switch. The IP address of the SVI becomes the default gateway for the devices within the VLAN. The most commonly used technique to achieve inter VLAN routing is via L3 switch.