I) Introduction to Networks

1) Network models

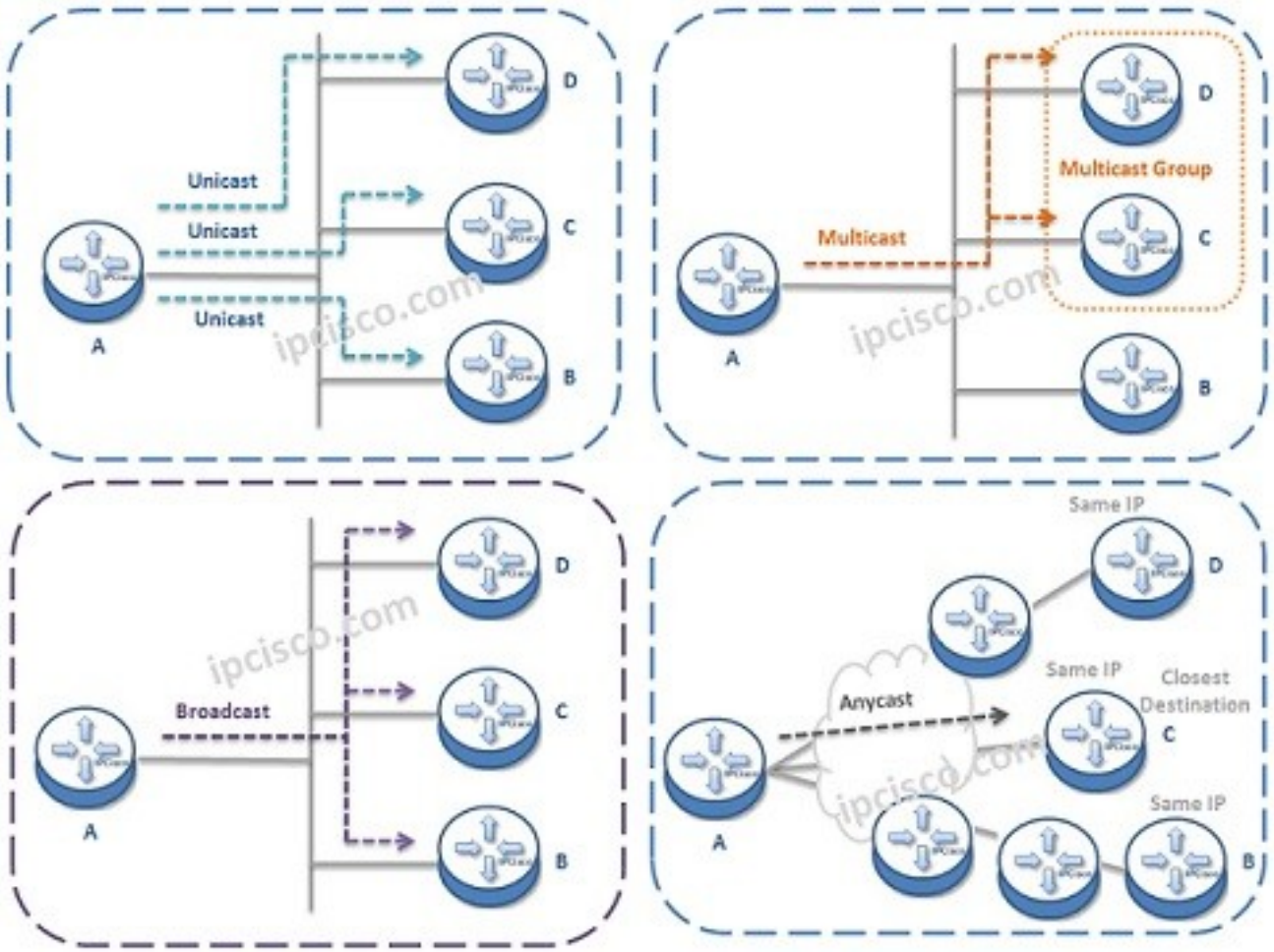
Unicast, Multicast, Broacast, Anycast

Unicast is the communication that there is only one receiver. This is one-to-one communication.

Multicast is the communication that there is one more receiver. Only the members of the multicast group receive the multicast traffic.

Broadcast is also the communication that there is one more receiver but this time, all the receivers receive broadcast traffic.

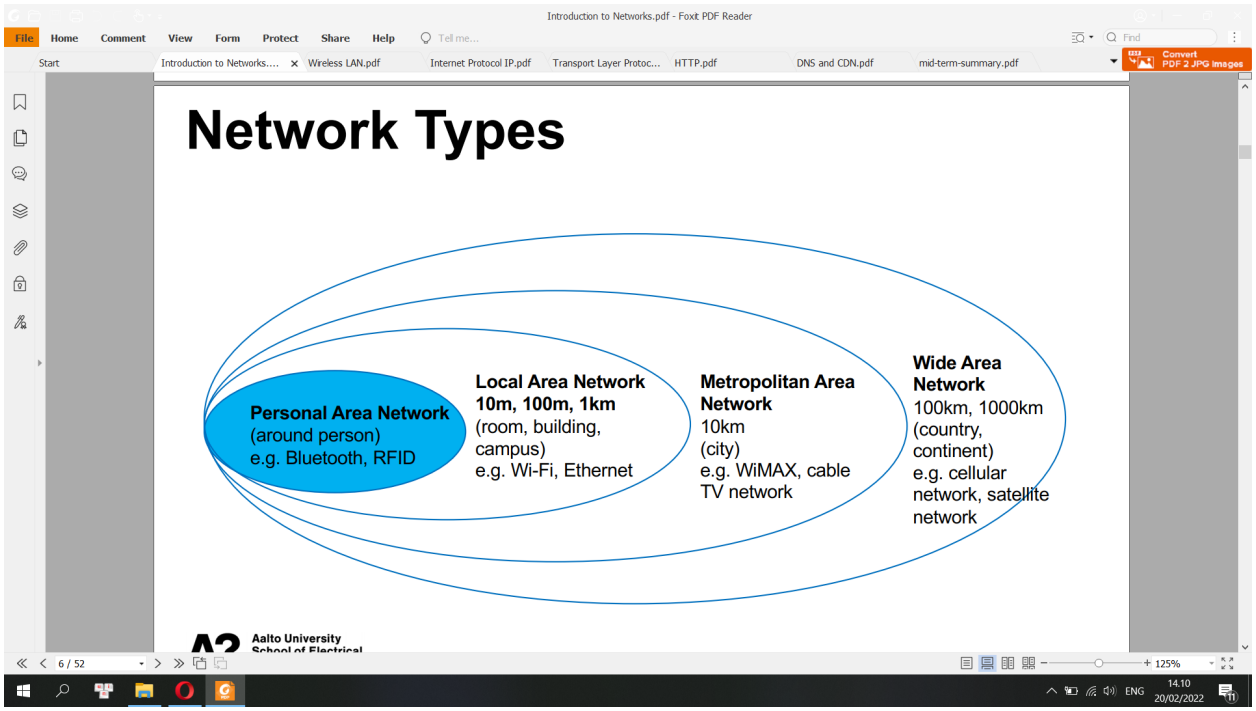
Anycast is the communication that is developed with IPv6. With anycast, the traffic is received by the nearest receiver in a group of the receivers that has the same IP.



Network Types

- RFID: Radio-frequency identification uses electromagnetic fields to automatically identify and track tags attached to objects. An RFID system consists of a tiny radio transponder, a radio receiver and transmitter.

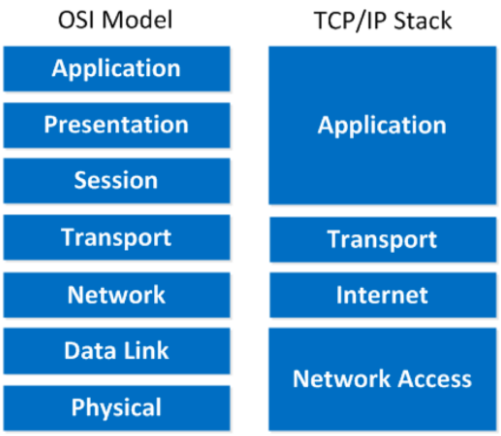
- Wi-Fi is the wireless technology used to connect computers, tablets, smartphones and other devices to the internet
- The internet connects users from all over the world in a single massive network. Devices on the internet can talk to one another using the global infrastructure. Ethernet connects devices in a local area network (LAN), which is a much smaller collection of interconnected devices.
- LAN: local area networks: room, building, campus
- MAN: metropolitan area networks: 10km, city
- WAN: wide area networks: 100km, 1000km, country, continent
- LAN < MAN < WAN in terms of coverage



- ISP: Internet service provider: Elisa, Telia, DNA
- IXP: Internet exchange Point
- A protocol is an agreement between the communicating parties on how communication is to proceed
- The interface defines which primitive operations and services the lower layer makes available to the upper one.

- OSI: The Open Systems Interconnection model is a conceptual model that characterises and standardises the communication functions of a telecommunication or computing system without regard to its underlying internal structure and technology.
- The Internet protocol suite, commonly known as TCP/IP, is the set of communications protocols used in the Internet and similar computer networks. The current foundational protocols in the suite are the Transmission Control Protocol and the Internet Protocol

OSI Model and TCP/IP Stack



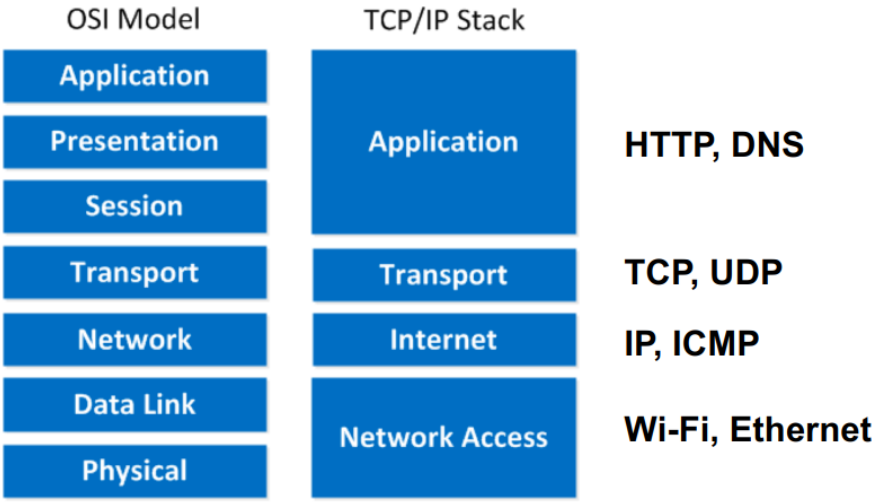
Aalto University
School of Electrical
Engineering

Open Systems Interconnection (OSI)

OSI Model

7 Application	Serves as the window for users and application processes to access the network services.	Data
6 Presentation	Formats the data to be presented to the Application layer. It can be viewed as the “Translator” for the network.	Data
5 Session	Allows session establishment between processes running on different hosts	Data
4 Transport	Ensures that messages are delivered error-free, in sequence, and with no losses or duplications	Segments
3 Network	Determine how packets are routed from source to destination	Packets
2 Data Link	Provides error-free transfer of data frames from one node to another over the Physical Layer.	Frames
1 Physical	Transmission and reception of the unstructured raw bit stream over the physical medium.	Bits

Example Protocols



DNS: The Domain Name System is the hierarchical and decentralized naming system used to identify computers, services, and other resources reachable through the Internet or other Internet Protocol networks. The resource records contained in the DNS associate domain names with other forms of information

HTTP: The Hypertext Transfer Protocol is an application layer protocol in the Internet protocol suite model for distributed, collaborative, hypermedia information systems

UDP:In computer networking, the User Datagram Protocol is one of the core members of the Internet protocol suite. With UDP, computer applications can send messages, in this case referred to as datagrams, to other hosts on an Internet Protocol network.

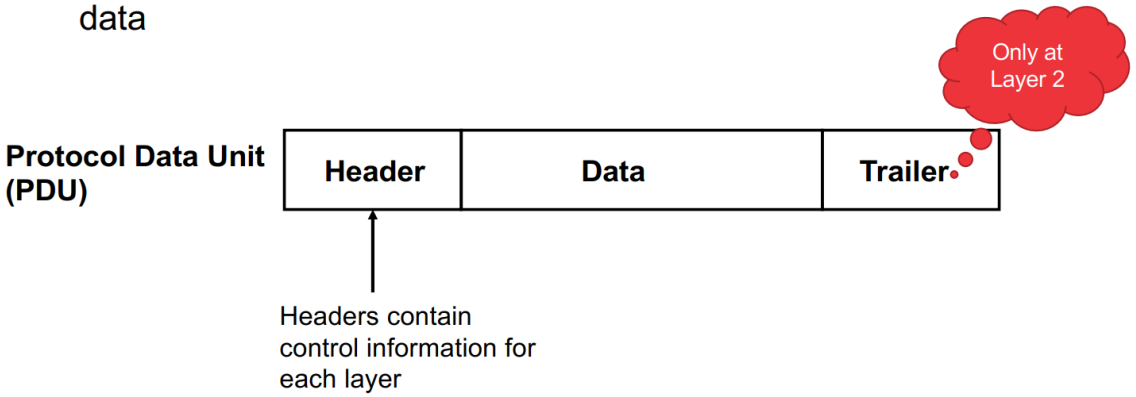
ICMP: The Internet Control Message Protocol is a supporting protocol in the Internet protocol suite. It is used by network devices, including routers, to send error messages and operational information

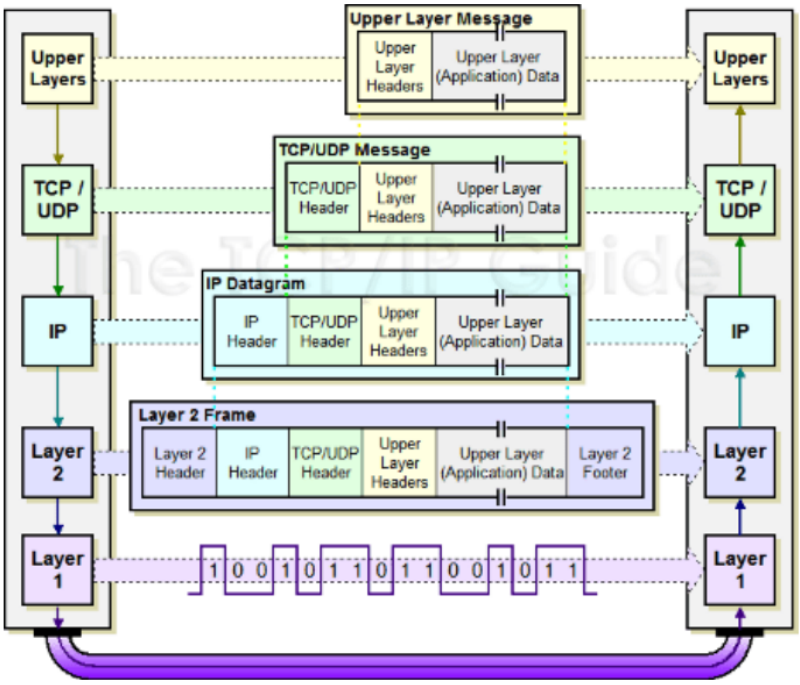
PDU: In telecommunications, a protocol data unit is a single unit of information transmitted among peer entities of a computer network. A PDU is composed of protocol-specific control information and user data

Data Encapsulation: The process of adding headers and trailers to data

Data Encapsulation

- **Encapsulation:** The process of adding headers and trailers to data





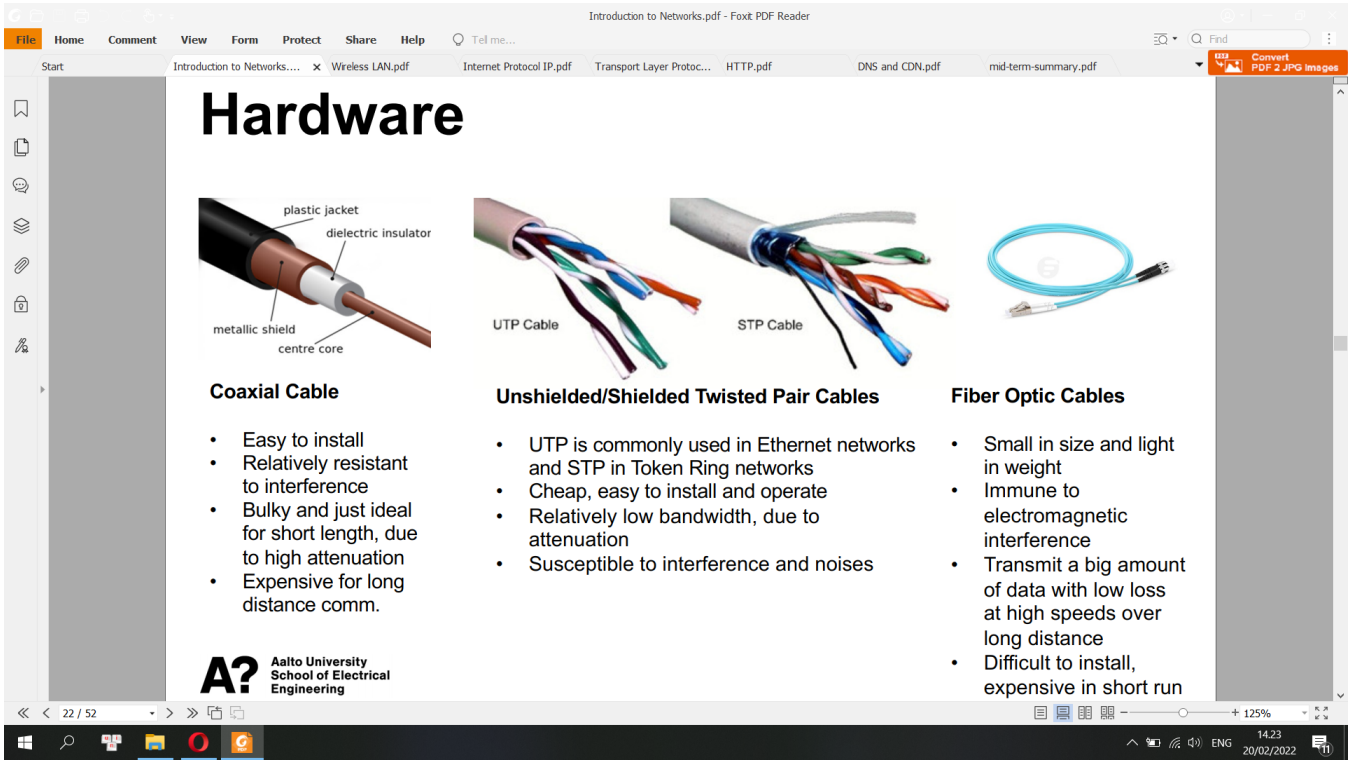
2) Ethernet

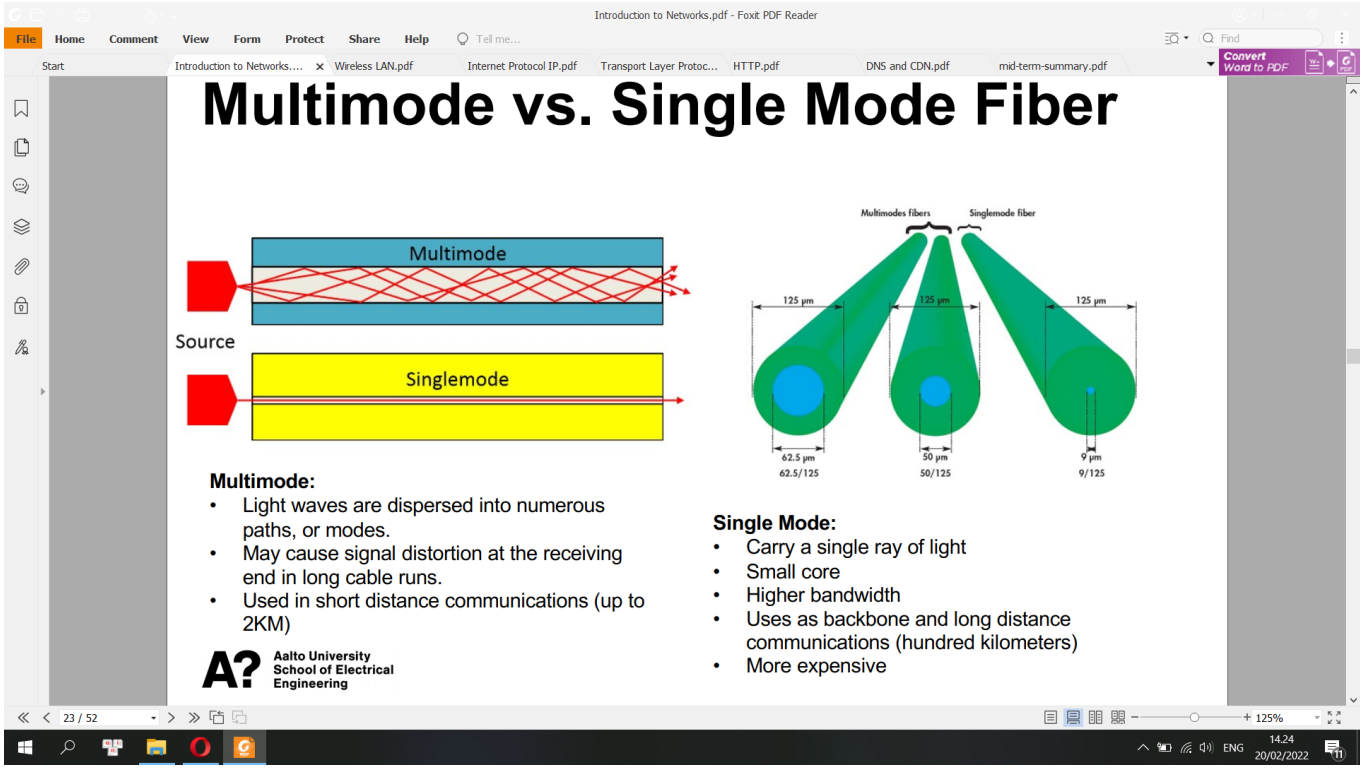
- Ethernet (/ˈiːθərnɛt/) is a family of wired computer networking technologies commonly used in local area networks (LAN), metropolitan area networks (MAN) and wide area networks (WAN)

Network Access Layer (Layer 1& 2 in OSI Model) e.g. 802.3 Ethernet, 802.11 Wi-Fi, 802.15.4 ZigBee

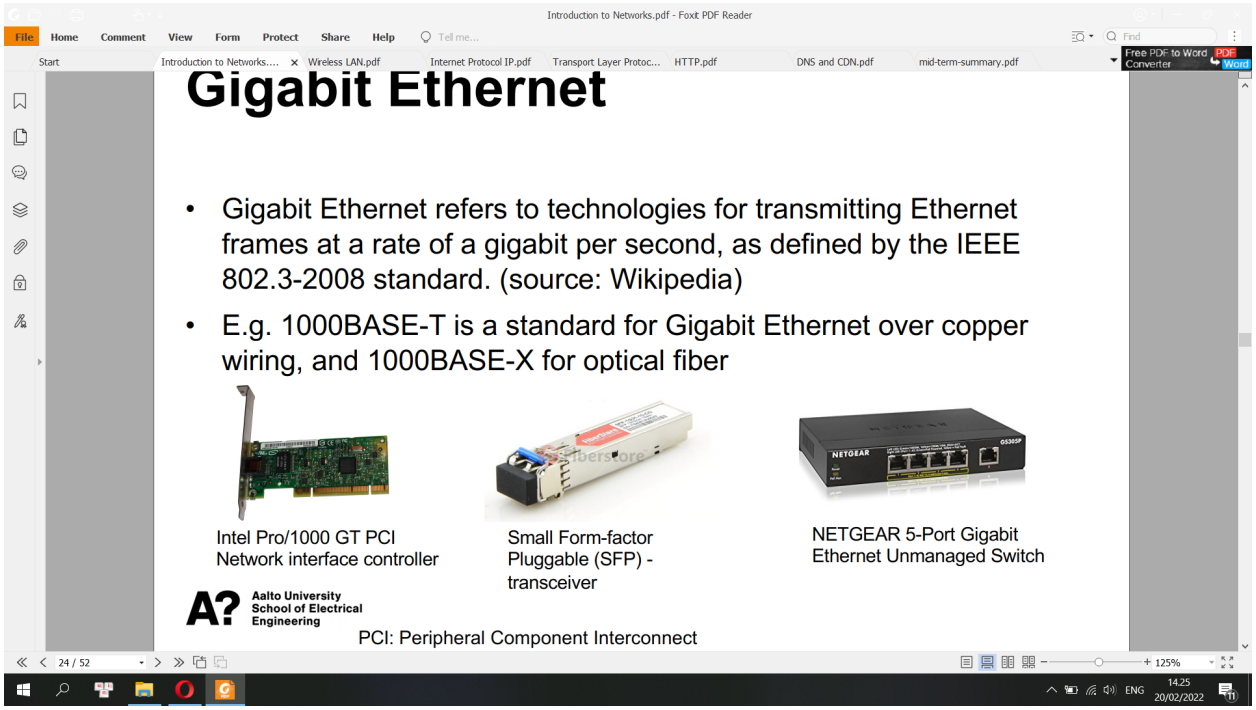
IEEE: The Institute of Electrical and Electronics Engineers is a professional association for electronic engineering and electrical engineering with its corporate office in New York City

- Hardware of Ethernet





Gigabit Ethernet



Data Link Layer (Layer 2) has two sub-layers.

• IEEE 802.2 Logical Link Control (LLC)

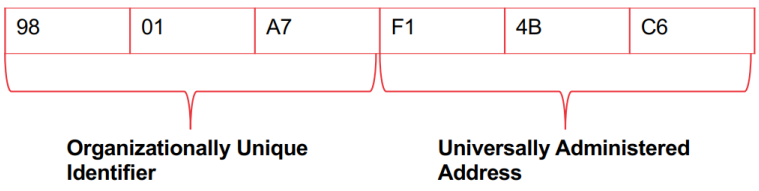
LLC: In the IEEE 802 reference model of computer networking, the logical link control data communication protocol layer is the upper sublayer of the data link layer of the seven-layer OSI model. The LLC sublayer acts as an interface between the media access control sublayer and the network layer.

• IEEE 802.3 Media Access Control (MAC)

MAC: In IEEE 802 LAN/MAN standards, the medium access control sublayer is the layer that controls the hardware responsible for interaction with the wired, optical or wireless transmission medium. The MAC sublayer and the logical link control sublayer together make up the data link layer.

A MAC address is a unique identifier assigned to a network interface controller (NIC) by manufacturer

- A MAC address is a **unique identifier** assigned to a network interface controller (NIC) by manufacturer
- Used for communications at the data link layer
- Ethernet addresses are 6 bytes long



```
Wireless LAN adapter Wi-Fi:

   Connection-specific DNS Suffix  . :
   Description . . . . . . . . . . . : Intel(R) Wireless-AC 9260 160MHz
   Physical Address. . . . . . . . . : 58-A0-23-F8-DD-AB
   DHCP Enabled. . . . . . . . . . . : Yes
   Autoconfiguration Enabled . . . . : Yes
   IPv6 Address. . . . . . . . . . . : 2001:14ba:a0bd:dd00:c10c:de5e:2cbf:132c(Preferred)
   Temporary IPv6 Address. . . . . . : 2001:14ba:a0bd:dd00:39c7:84f5:ce41:67f6(Preferred)
   Link-local IPv6 Address . . . . . : fe80::c10c:de5e:2cbf:132c%9(Preferred)
   IPv4 Address. . . . . . . . . . . : 192.168.1.208(Preferred)
   Subnet Mask . . . . . . . . . . . : 255.255.255.0
   Lease Obtained. . . . . . . . . . : Sunday, 20 February 2022 12.36.22
   Lease Expires . . . . . . . . . . : Monday, 21 February 2022 12.36.22
   Default Gateway . . . . . . . . . : fe80::1eb7:2cff:fe7c:c420%9
                                       192.168.1.1
   DHCP Server . . . . . . . . . . . : 192.168.1.1
   DHCPv6 IAID . . . . . . . . . . . : 72917027
   DHCPv6 Client DUID. . . . . . . . : 00-01-00-01-28-C4-50-5E-58-A0-23-F8-DD-AB
   DNS Servers . . . . . . . . . . . : 2001:14ba:a0bd:dd00::1
                                       192.168.1.1
   NetBIOS over Tcpip. . . . . . . . : Enabled
```

Physical address 58-A0… is the MAC address of this laptop.

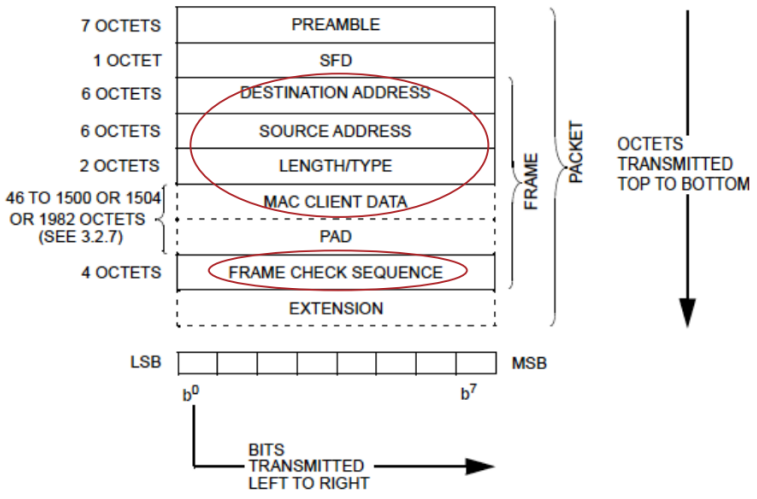
Packet-based Data Transmission

Ethernet Frame

802.3 Ethernet packet and frame structure

Layer	Preamble	Start frame delimiter	MAC destination	MAC source	802.1Q tag (optional)	Ethertype (Ethernet II) or length (IEEE 802.3)	Payload	Frame check sequence (32-bit CRC)	Interpacket gap
	7 octets	1 octet	6 octets	6 octets	(4 octets)	2 octets	46-1500 octets	4 octets	12 octets
Layer 2 Ethernet frame			← 64–1522 octets →						
Layer 1 Ethernet packet & IPG	← 72–1530 octets →								← 12 octets →

- Ethernet is commonly described as being a packet delivery system



Aalto University
School of Electrical
Engineering

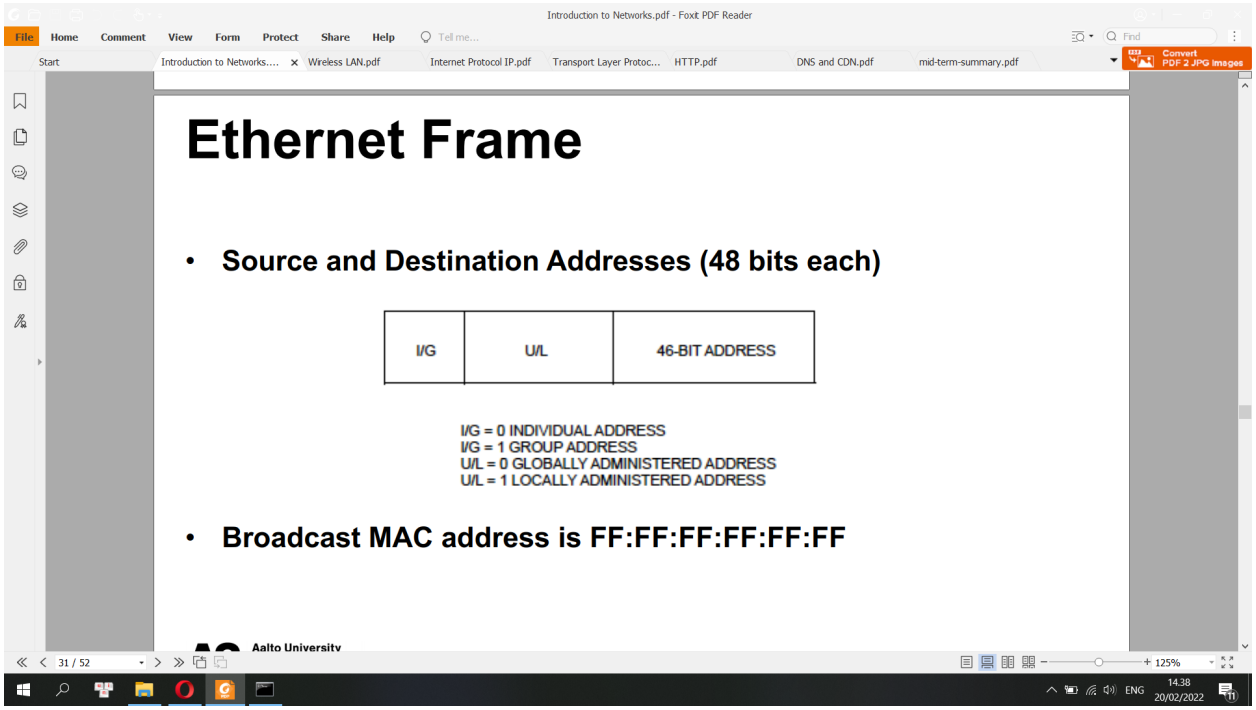
Source: IEEE 802.3-2018

A preamble is a signal used in network communications to synchronize transmission timing between two or more systems. In general, preamble is a synonym for "introduction."
Preamble is a bit sequence used for physical medium stabilization and synchronization.

10101010 10101010 10101010 10101010 10101010 10101010 10101010

SFD: The Start Frame Delimiter (SFD) field is the sequence 10101011. It immediately follows the preamble pattern. A MAC frame starts immediately after the SFD. Note: 100 and 1000 Mb/s Ethernet systems signal constantly and do not need preamble or start frame delimiter fields.

- **Ethernet Frame**



Length/Type Field: the number of MAC client data octets contained in the following MAC Client Data field, if the field value < or = 1500 decimal. Otherwise, the Ethertype of the MAC client protocol (e.g. 1500 basic frames, 1504 Q-tagged frames, 1982 envelop frames)
• PAD field: A minimum MAC frame size is required for correct CSMA/CD protocol operation
The length of PAD is max [0, minFrameSize – (clientDatasize + 2 x addressSize + 48)] bits.
minFrameSize is typically 64 octets
• Frame Check Sequence (FCS) contains a 4-octet CRC (cyclic redundancy check) value.
The CRC value is computed as a function of the contents of the protected fields of the MAC frame (from destination address to Pad)

All frames and the bits, bytes, and fields contained within them, are susceptible to errors from a variety of sources. The FCS field contains a number that is calculated by the source node based on the data in the frame. This number is added to the end of a frame that is sent. When the destination node receives the frame the FCS number is recalculated and compared with the FCS number included in the frame. If the two numbers are different, an error is assumed and the frame is discarded.

The FCS provides error detection only. Error recovery must be performed through separate means. Ethernet, for example, specifies that a damaged frame should be discarded and does not specify any action to cause the frame to be retransmitted. Other protocols, notably the Transmission Control Protocol (TCP), can notice the data loss and initiate retransmission and error recovery.

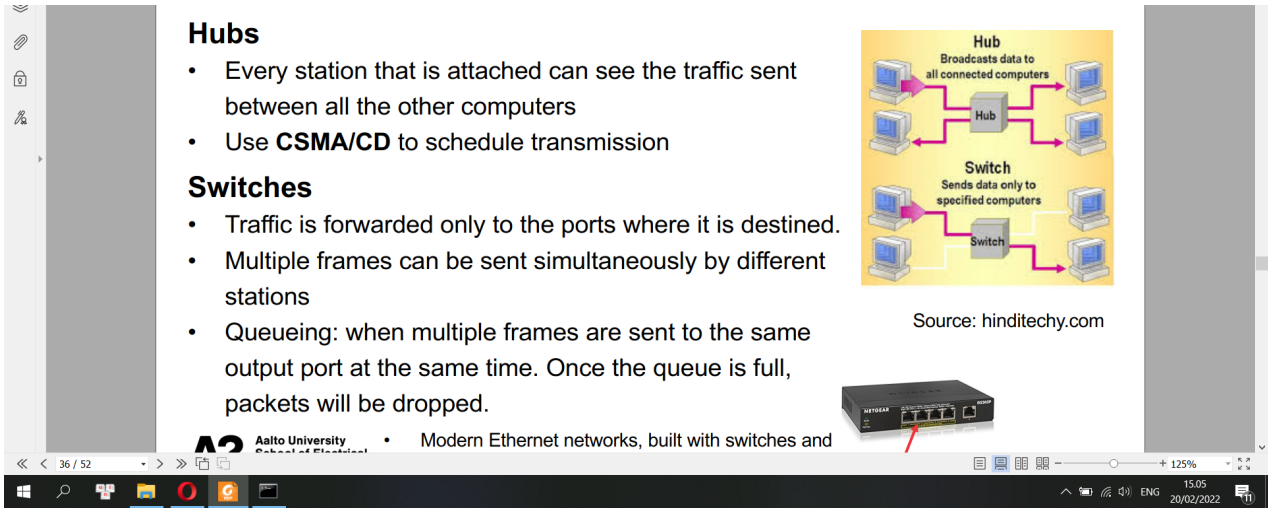
• An Extension Field is added, if required (for 1000Mb/s half duplex operation only)

In computing and telecommunications, the payload is the part of transmitted data that is the actual intended message. Headers and metadata are sent only to enable payload delivery. In the context of a computer virus or worm, the payload is the portion of the malware which performs malicious action.

Invalid MAC Frame
A MAC frame is invalid when it meets at least one of the following conditions:
• The frame length is inconsistent with a length value specified in the length/type field
• It is not an integral number of octets in length
• The bits of the incoming frame (exclusive of the FCS field itself) do not generate a CRC value identical to the one received
Invalid frames should not be passed to the LLC or MAC sublayers.

Two Operating Modes of MAC SubLayer
• Half Duplex: A host can only send or receive at one time: CSMA/CD
• Full Duplex: A host can send and receive simultaneously. No collision.
• Duplex configuration: either manually set or auto negotiated by connected devices
• Duplex mismatch -> poor performance

Hubs and Switches
Hub and Switch are both network connecting devices. Hub works at physical layer and is responsible to transmit the signal to port to respond where the signal was received whereas Switch enable connection setting and terminating based on need. ... Hub works in Physical Layer. Switch works in Data Link Layer.



CSMA/CD

CSMA: Carrier-sense multiple access is a media access control protocol in which a node verifies the absence of other traffic before transmitting on a shared transmission medium, such as an electrical bus or a band of the electromagnetic spectrum

Carrier Sense Multiple Access with Collision Detection (CSMA/CD) defines how Ethernet frames get onto an Ethernet network

• CSMA/CD is designed to allow fair access by all transmission devices to shared network channels.

• The Physical Layer performs the task of generating the signals on the medium that represent the bits of the frame. Only one signal at a time can be transmitted on an Ethernet network

• Simultaneously, the Physical Layer monitors the medium and generates the collision detect signal, which in the contention-free case, remains off for the duration of the frame.

• Whenever two stations transmitted at the same time, the signals would collide; frames that collide must be retransmitted

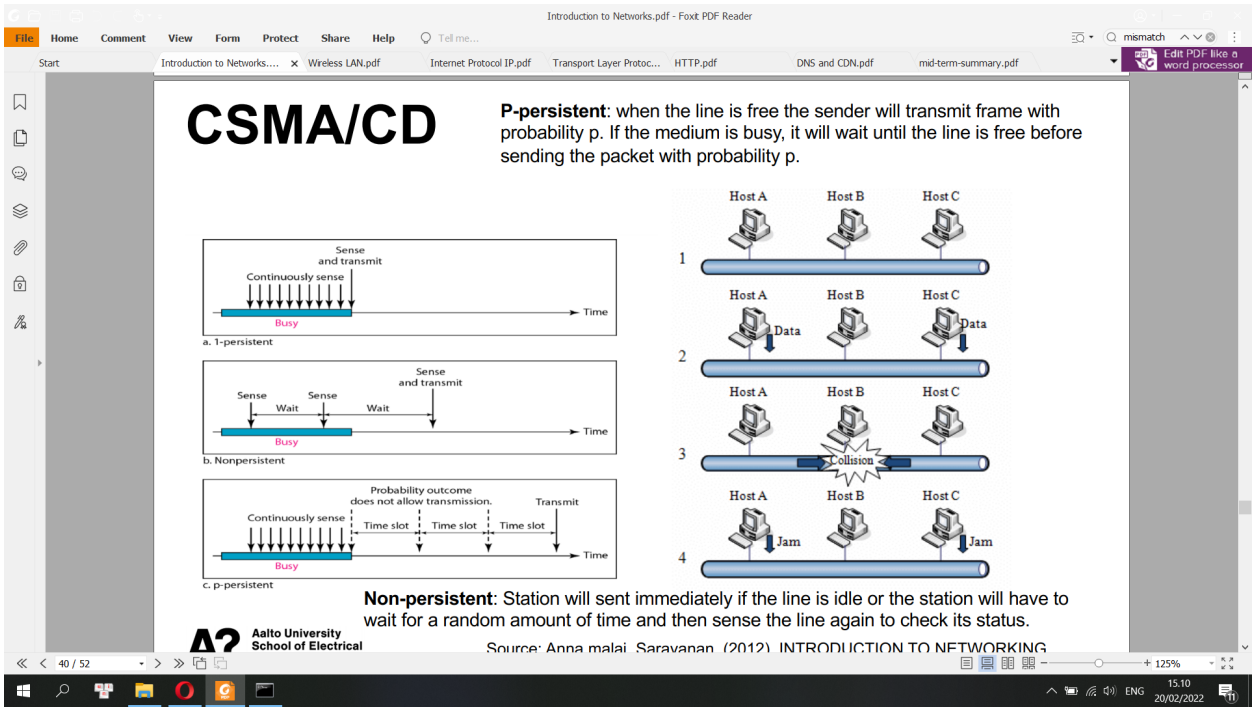
In order to minimize collision loss, each station implemented the following:

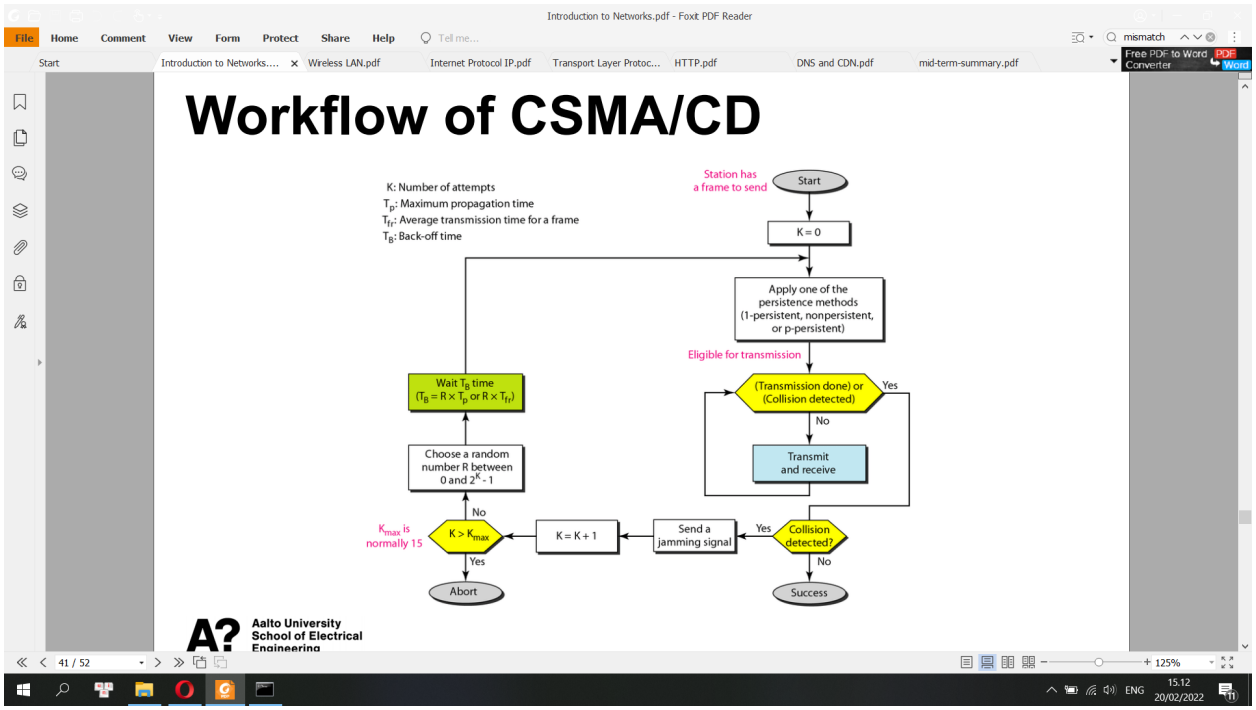
• Every Ethernet device listens to hear if another device is already transmitting. When the medium is clear, frame transmission is initiated (after a brief interframe delay).

• While transmitting, continually monitor the carrier sense signal provided by the physical layer to detect collisions; if a collision is detected, cease transmitting

• If a collision occurs, use a backoff-and-retransmit strategy

1-persistent, p-persistent, non-persistent





Backoff-and-retransmit

• Transmit Media Access Management component of the MAC sublayer enforces the collision by transmitting a bit sequence called jam.

• Terminate the transmission and schedule another transmission attempt after a randomly selected time interval.

• In case of repeated collision, adjust the medium load by backing off (voluntarily delaying its own retransmission to reduce its load on the medium)

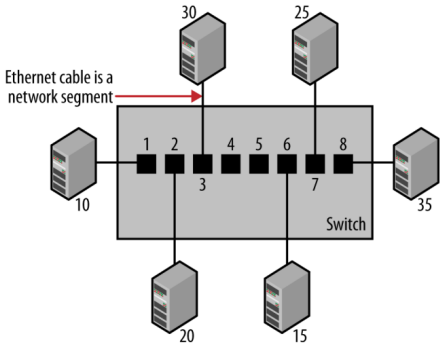
Datagram Forwarding

Header: destination address

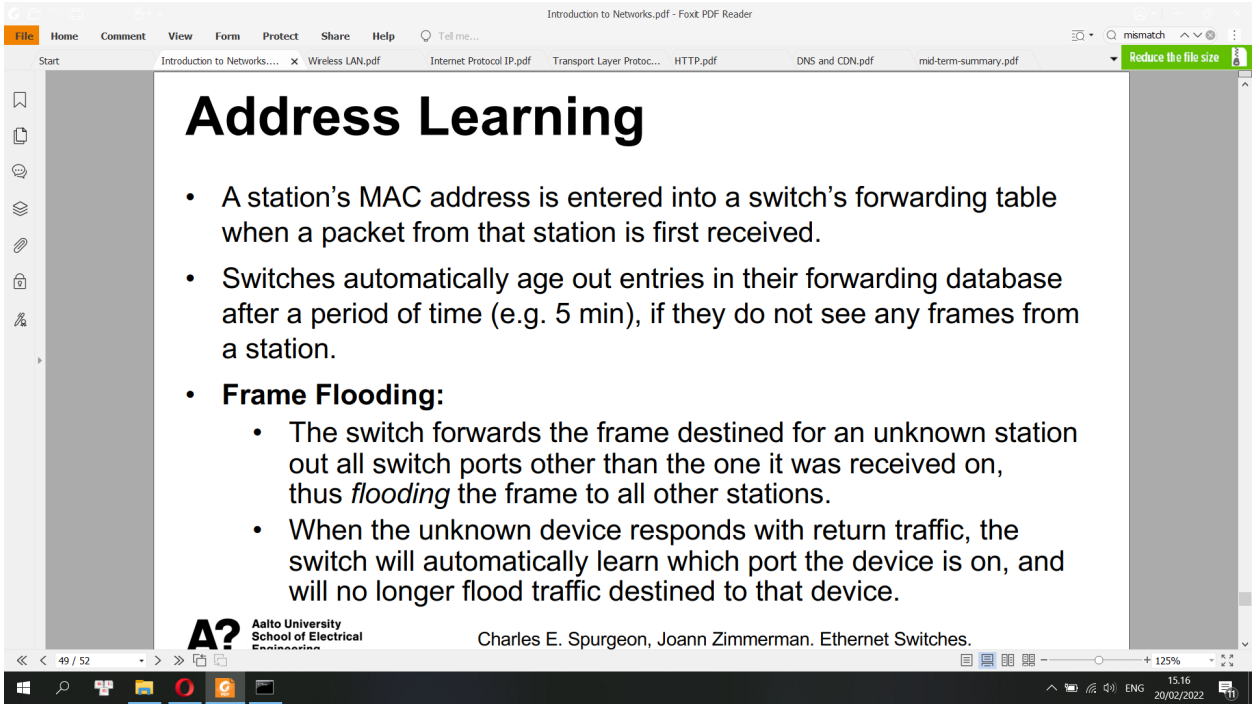
Forwarding Table

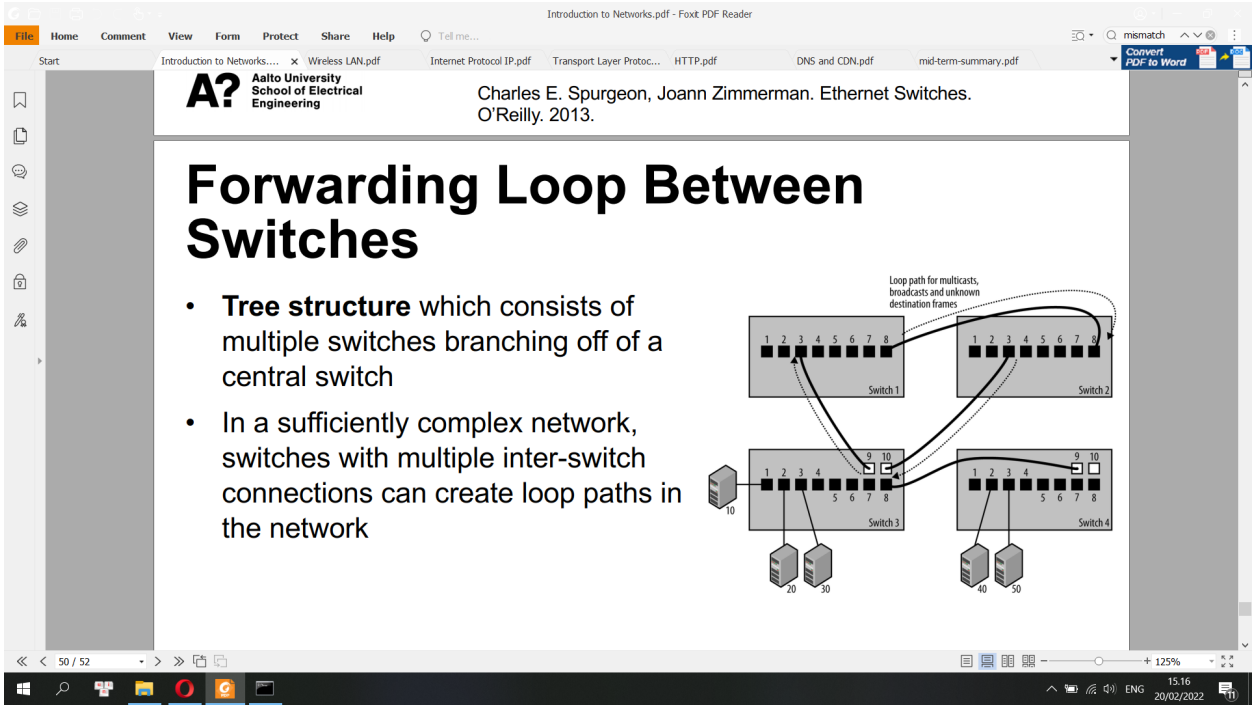
<destination, next_hop>

Address Learning



Port	Station
1	10
2	20
3	30
4	No station
5	No station
6	15
7	25
8	35





3) Self-test: Network models

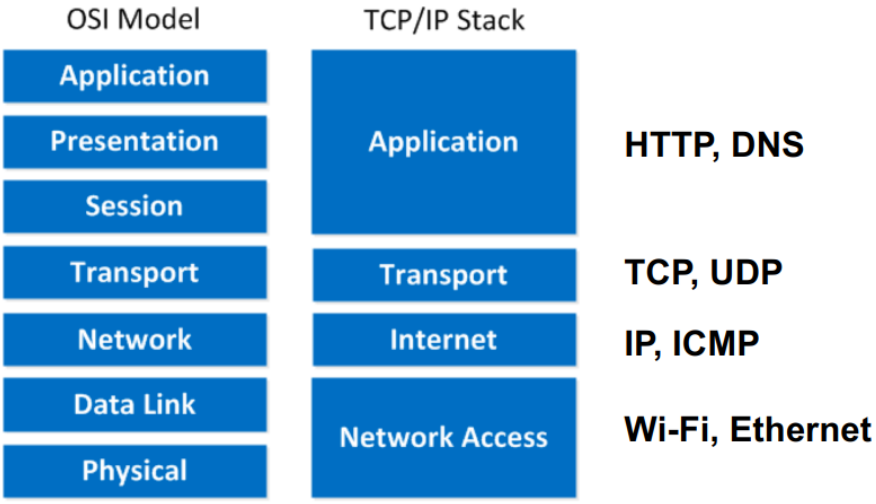
• Unicast vs. Multicast vs. Broadcast: check above

• Can you remember the name of each layer in the OSI model?: check above. Abbreviation:

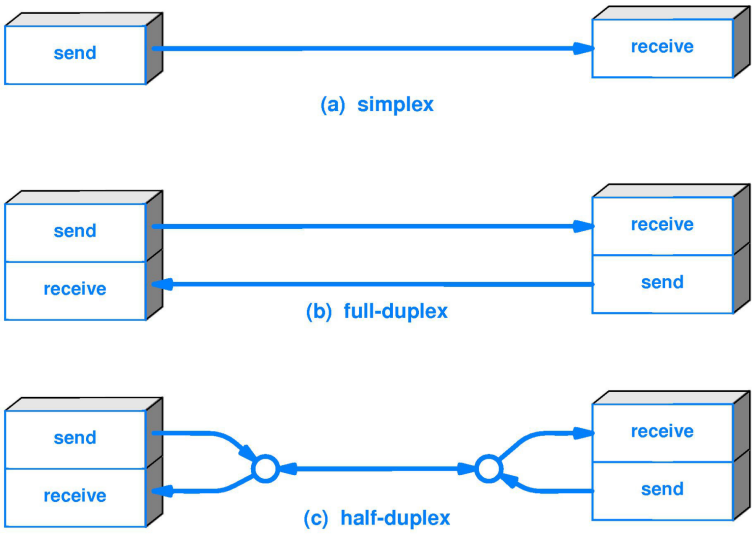
For OSI model: APSTNDP: application presentation session transport network data-link physical

For TCP-IP model: ATIN : application | transport | internet | network-access

Example Protocols



- **What is MAC address?: check above**
- **Full duplex vs. Half duplex:**



A full-duplex device is capable of bi-directional network data transmissions at the same time. Half-duplex devices can only transmit in one direction at one time. With half-duplex mode, data can move in two directions, but not at the same time.

- **Hub vs. Switches: check above**
- **How does CSMA/CD work?: check above**
- **How to create a switch's forwarding table?**

Layer 2 switches (bridges) have a MAC address table that contains a MAC address and physical port number. Switches follow this simple algorithm for forwarding frames:

When a frame is received, the switch compares the SOURCE MAC address to the MAC address table. If the SOURCE is unknown, the switch adds it to the table along with the physical port number the frame was received on. In this way, the switch learns the MAC address and physical connection port of every transmitting device.

The switch then compares the DESTINATION MAC address with the table. If there is an entry, the switch forwards the frame out the associated physical port. If there is no entry, the switch sends the frame out all its physical ports, except the physical port that the frame was received on (Flooding). If the DESTINATION is on the same port as the SOURCE (if they're both on the same segment), the switch will not forward the frame.)

Note that the switch does not learn the destination MAC until it receives a frame from that device.

• What do you think are the advantages and disadvantages of CSMA/CD?

Advantages

- CSMA/CD control software is relatively simple and produces little overhead.
- CSMA/CD network works best on a bus topology with bursty transmission. Bursty traffic is characterized by short, sporadic transmissions. Example: interactive terminal-host traffic.
- This technique is efficient for light to moderate load.

Disadvantages

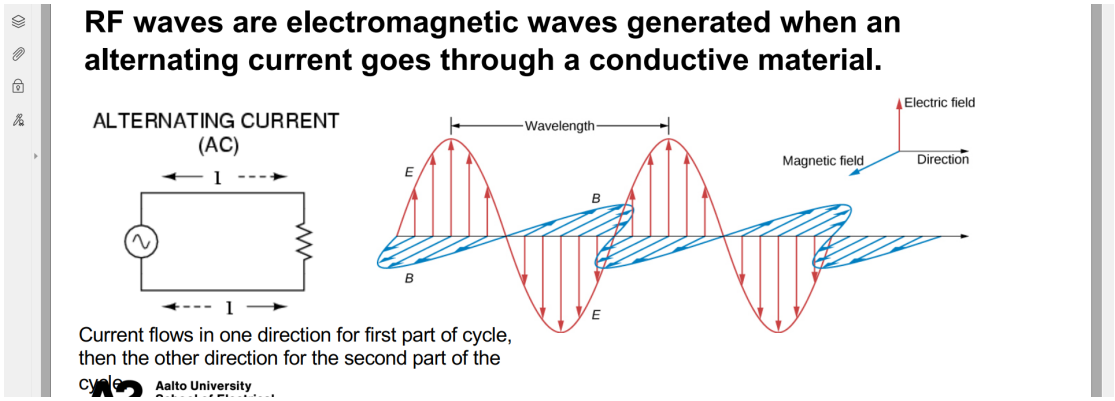
- CSMA/CD protocols are probabilistic and depends on the network (cable) loading.
- Considered unsuitable for channels controlling automated equipment that must have certain control over channel access. (This could be OK for different channel access).
- We can set priorities to give faster access to some devices (This is, probably, not an issue in some applications)

• Is a Hub half duplex or full-duplex? Can switches support full-duplex?

A hub is always Half Duplex, so only one device at a time can communicate. A switch can handle Full Duplex communication, so multiple computers can talk with each other simultaneously through a switch

II) Wireless LAN

Radio Frequency (RF): RF waves are electromagnetic waves generated when an alternating current goes through a conductive material



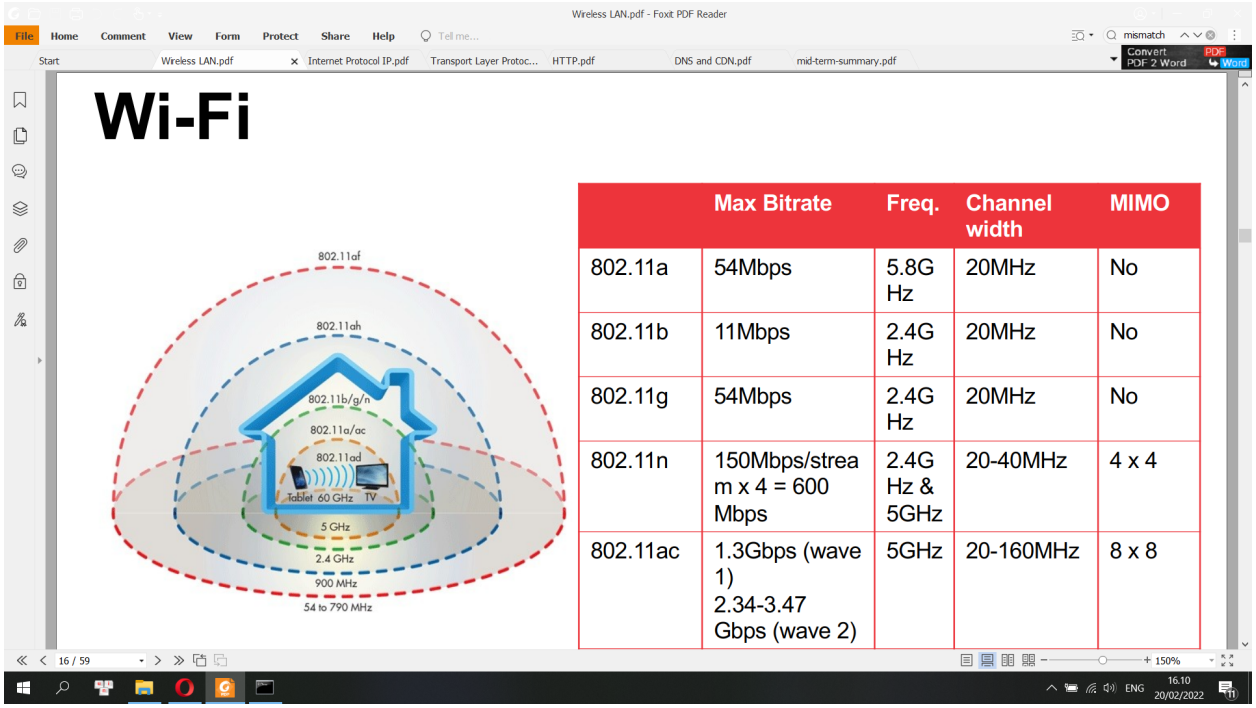
Received Signal Strength Indicator (RSSI): a measurement of how well your device can hear a signal from an access point or router
• Frequency (Hz): how many cycles per second
• Wavelength
• Amplitude
• An antenna is a device used to emit and receive RF waves.

Path Loss
• The reduction in power density of an electromagnetic wave as it propagates through space
• Causes:
• Propagation losses caused by the natural expansion of the radio wave front in free space
• Absorption losses (e.g. when radio signals pass through dense materials such as walls)
• Diffraction losses
• And etc.

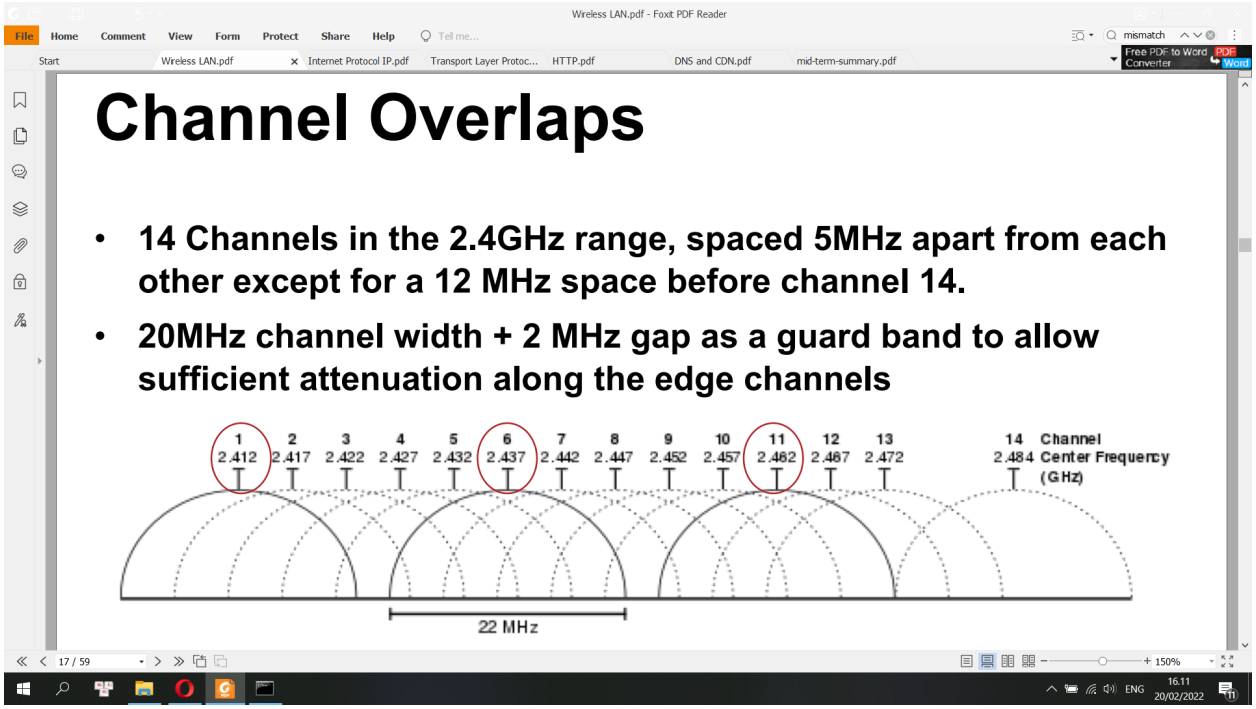
Diffraction
• Diffraction is the bending of a wave as it either passes through a barrier or passes through an opening.
• Frequency, wavelength and speed of waves do not change.
• Direction of propagation and the pattern of waves can change.

RF Waves
• RF waves travel at the speed of light in free space
  Speed of light (c) = frequency x wavelength
• When frequency increases, wavelength decreases
• Coverage area: the area in which receiving stations can successfully receive and “understand” the signal
- Depends on frequency and transmit power
• RF spectrum: the range of frequencies that are available

Wi-Fi Bands
• A band is a range of frequencies that can be used (e.g. 2.412-2.462 GHz)
• A band can be divided into smaller chunks of frequency called channels (e.g. 20MHz or 40MHz wide)
• Wi-Fi bands (2.4GHz, 5GHz) may be shared with other types of communication device
 E.g. Bluetooth (2.4GHz), Radar (5GHz)



	Max Bitrate	Freq.	Channel width	MIMO
802.11a	54Mbps	5.8G Hz	20MHz	No
802.11b	11Mbps	2.4G Hz	20MHz	No
802.11g	54Mbps	2.4G Hz	20MHz	No
802.11n	150Mbps/stream x 4 = 600 Mbps	2.4G Hz & 5GHz	20-40MHz	4 x 4
802.11ac	1.3Gbps (wave 1) 2.34-3.47 Gbps (wave 2)	5GHz	20-160MHz	8 x 8



RF Interference

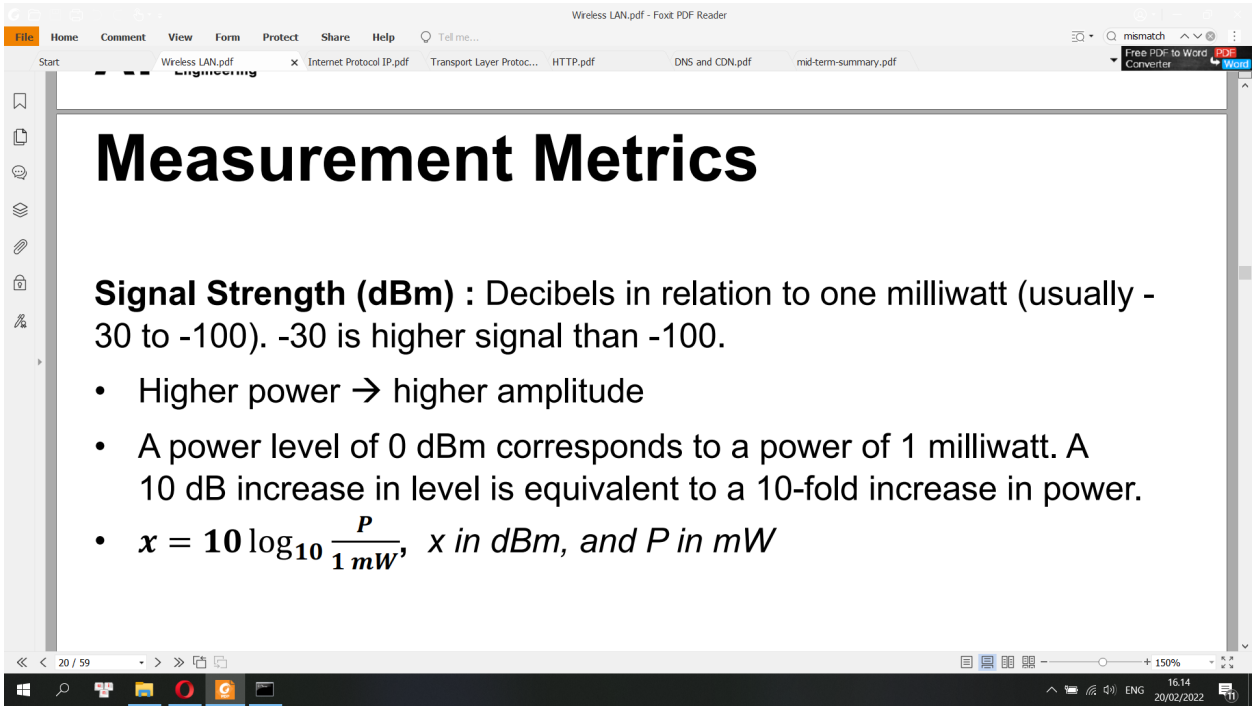
• Anything which modifies, or disrupts a signal as it travels along a channel between a source and a receiver.

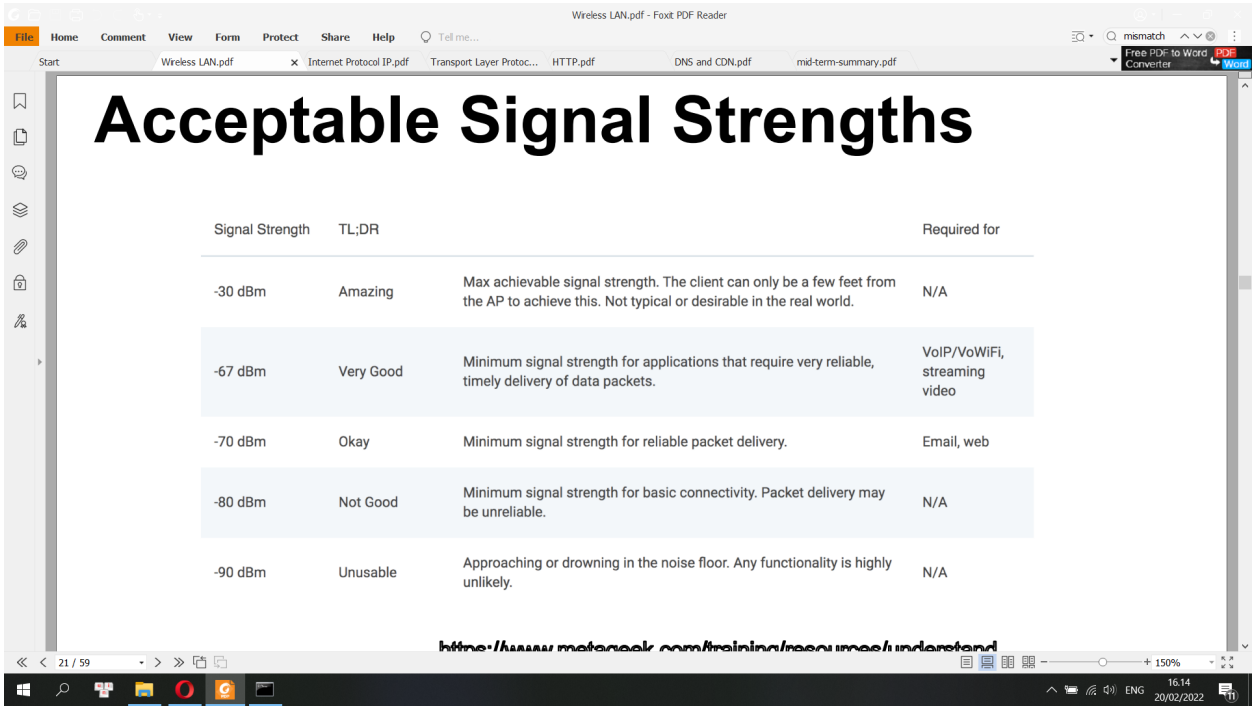
• It is an unwanted signal that occurs at the same time and frequency as a data signal.

• It causes wireless receivers to sporadically make mistakes when decoding packets, which results in retransmissions of data.

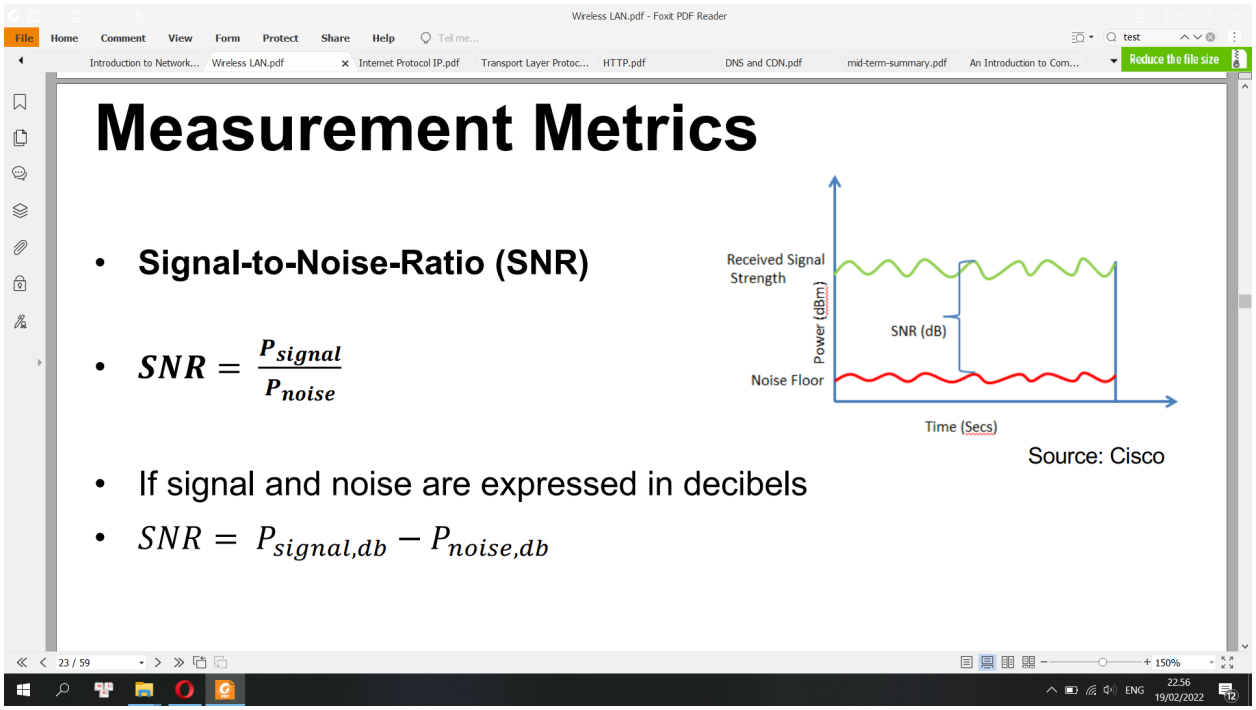
• Co-Channel interference results when there are numerous devices all competing for time to talk on the same channel.
• Adjacent-Channel interference occurs when devices from overlapping channels are trying to talk over each other.

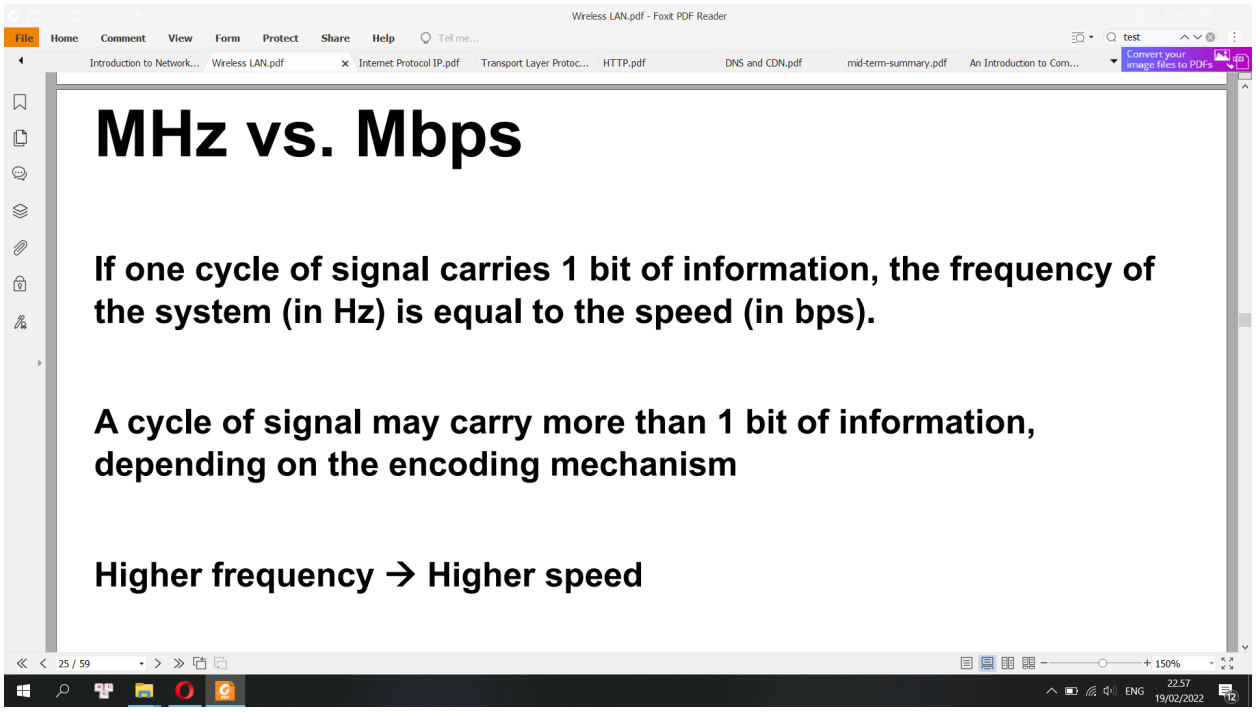
Measurement Metrics





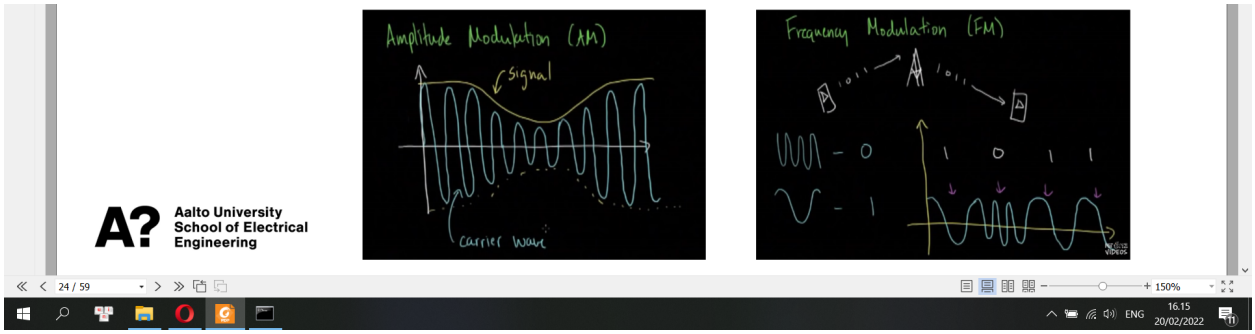
• RSSI is a relative index, while dBm is an absolute number representing power levels in mW





Modulation

- Change a characteristic of the radio signal to represent data
- Schemes:
- Amplitude modulation
- Frequency modulation (shifting between two frequencies called frequency shift keying)

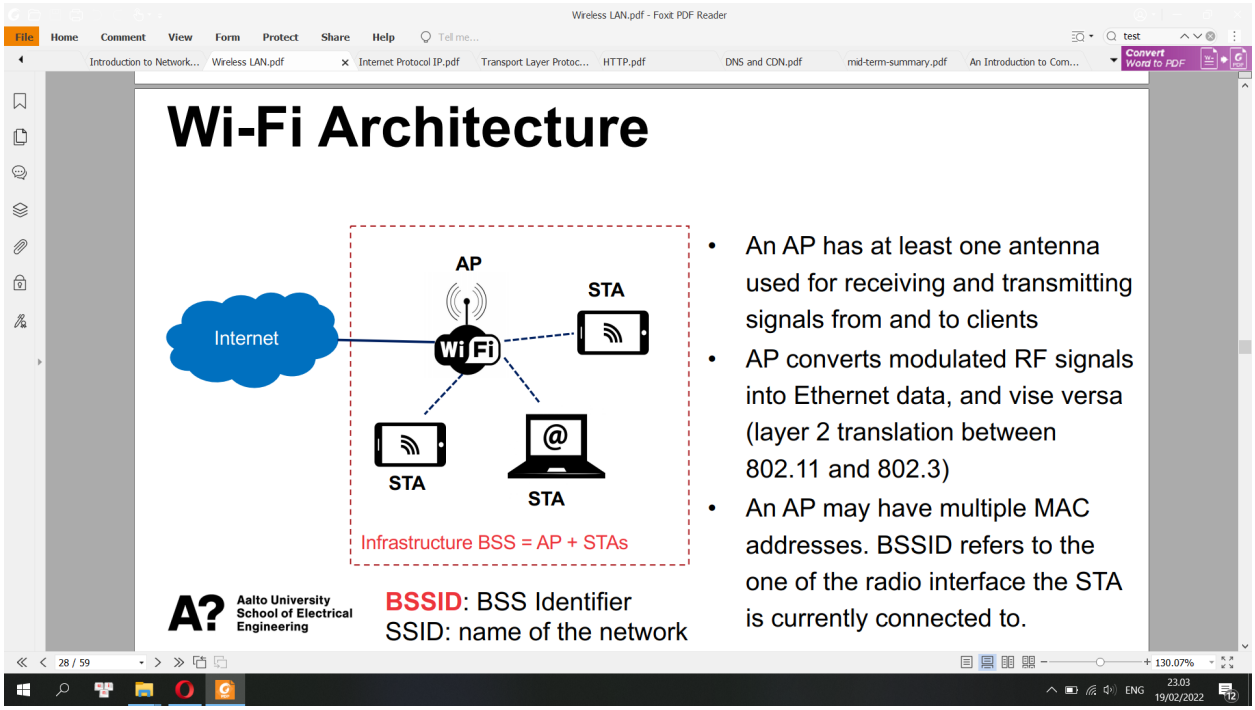


MHz vs. Mbps

If one cycle of signal carries 1 bit of information, the frequency of the system (in Hz) is equal to the speed (in bps).

A cycle of signal may carry more than 1 bit of information, depending on the encoding mechanism

Higher frequency -> Higher speed



Basic Service Set Identifier (BSS) is the basic building block of an IEEE 802.11 LAN. The members of a BSS can communication with each other directly

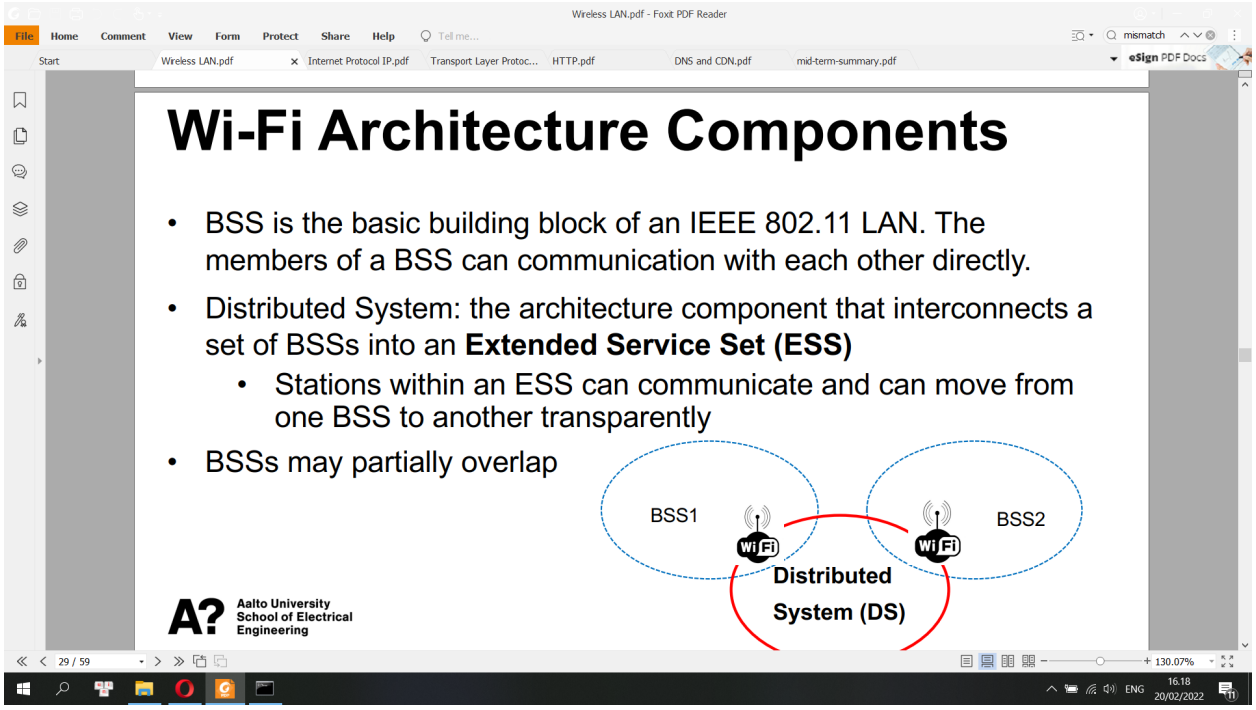
Service Set IDentifier (SSID)

In IEEE 802.11 wireless local area networking standards (including Wi-Fi), a service set is a group of wireless network devices which share a service set identifier (SSID)—typically the natural language label that users see as a network name. (For example, all of the devices that together form and use a Wi-Fi network called Foo are a service set.) A service set forms a logical network of nodes operating with shared link-layer networking parameters; they form one logical network segment.

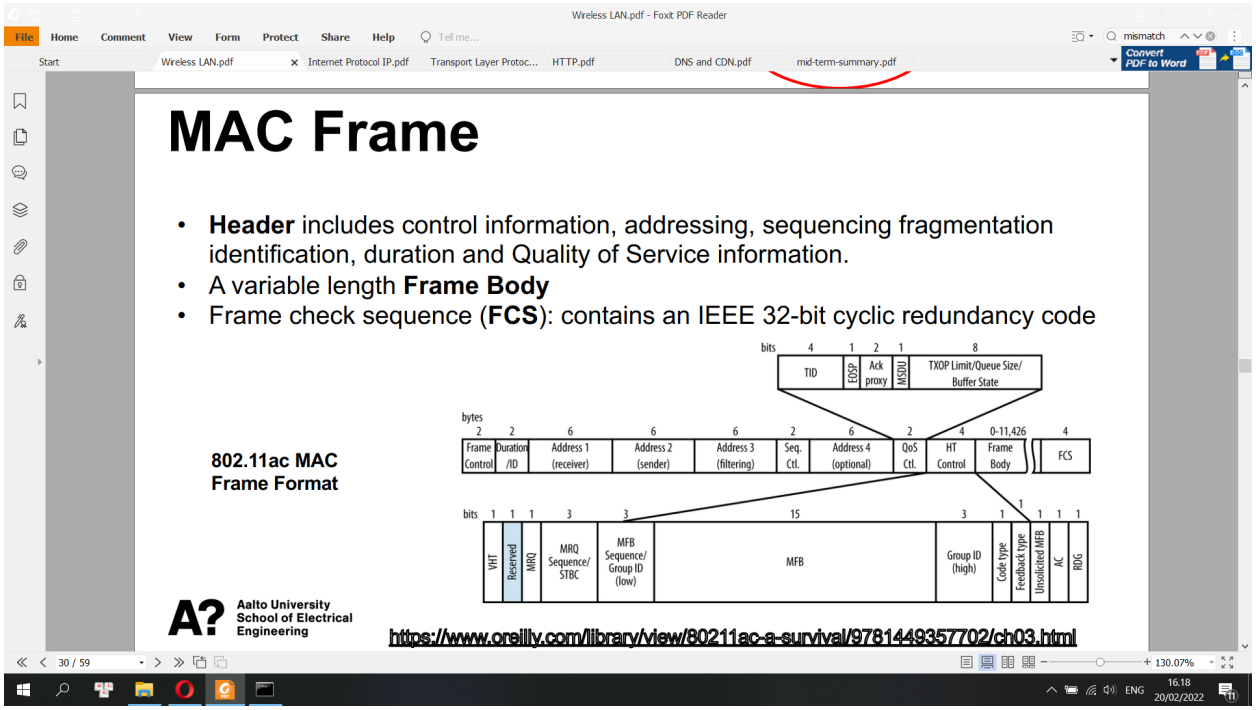
A service set is either a basic service set (BSS) or an extended service set (ESS).

A basic service set is a subgroup, within a service set, of devices that share physical-layer medium access characteristics (e.g. radio frequency, modulation scheme, security settings) such that they

are wirelessly networked. The basic service set is defined by a basic service set identifier (BSSID) shared by all devices within it. The BSSID is a 48-bit label that conform to MAC-48 conventions. While a device may have multiple BSSIDs, usually each BSSID is associated with at most one basic service set at a time.

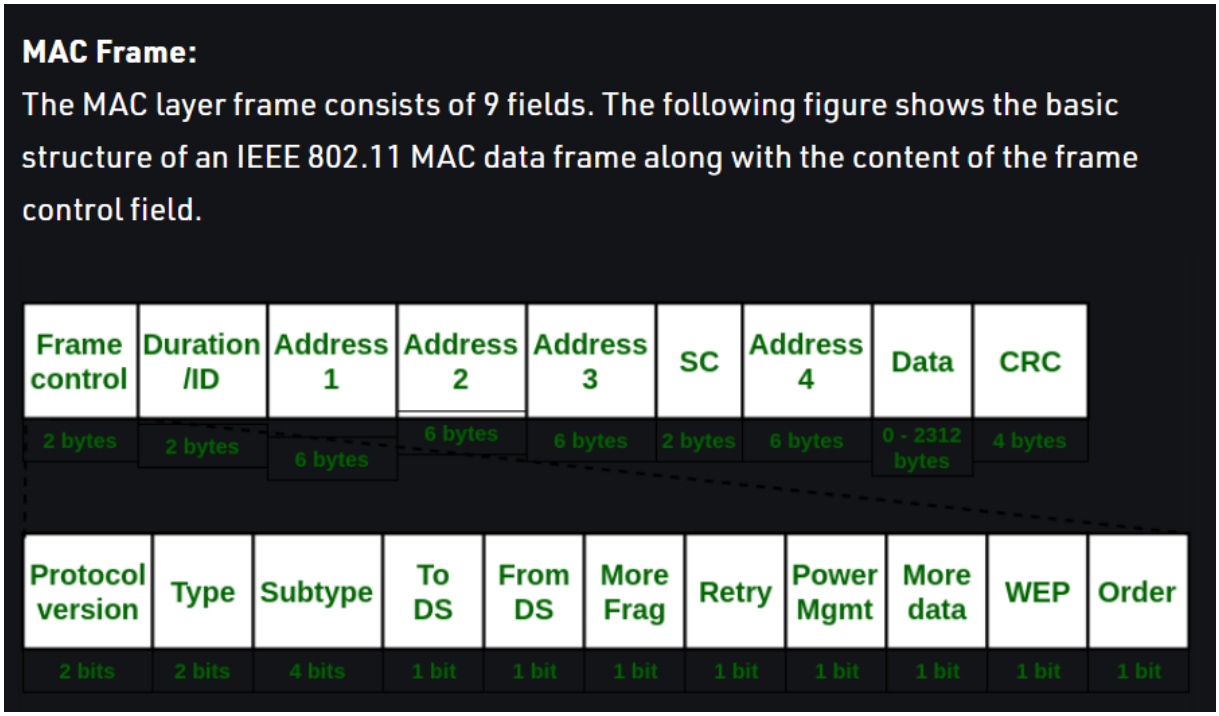


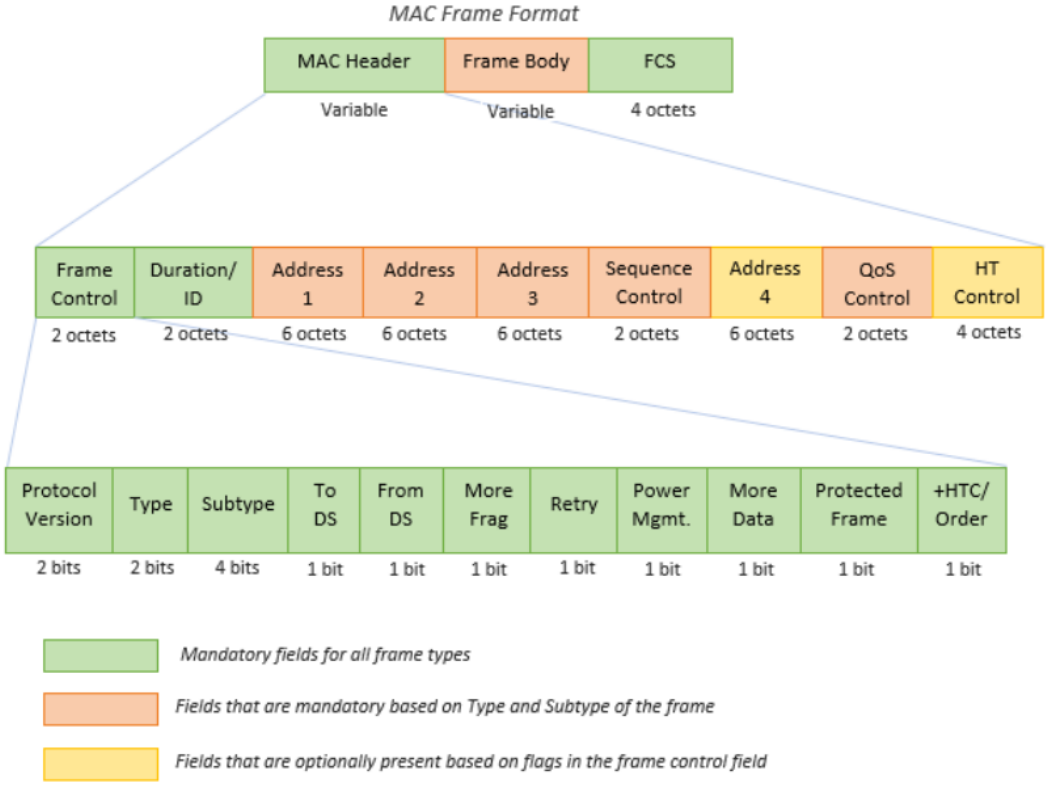
Mac Frame



MAC Frame:

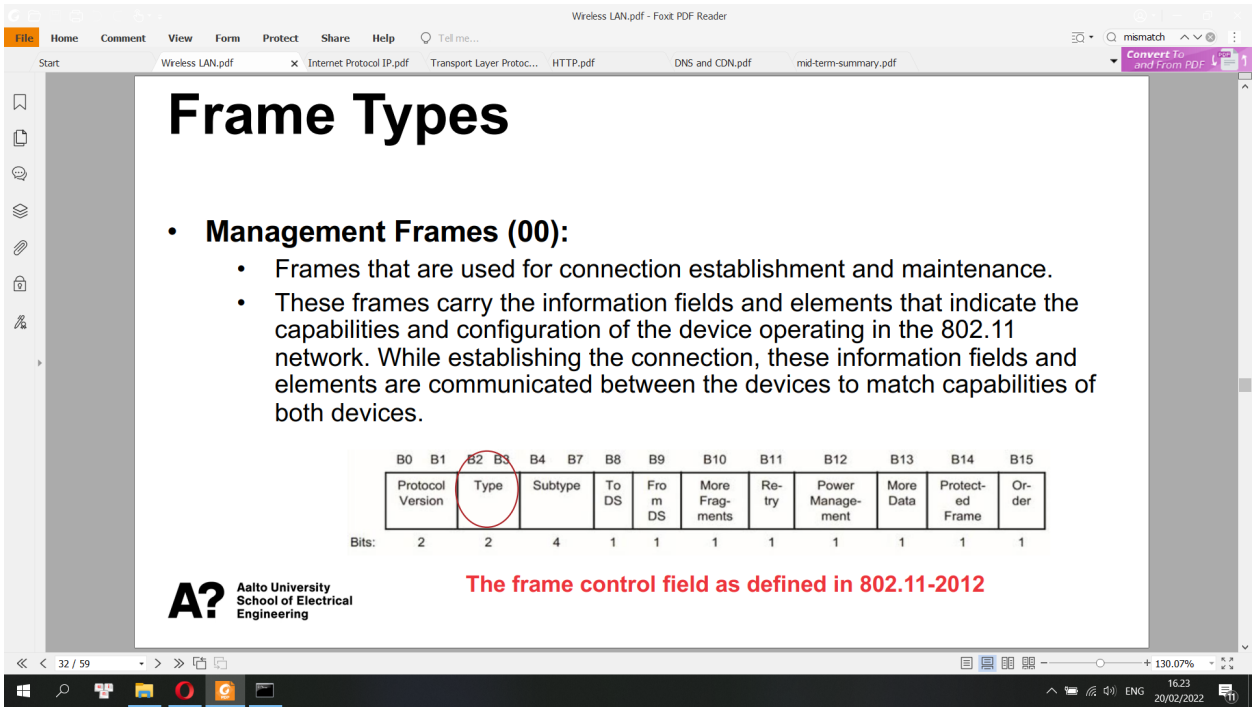
The MAC layer frame consists of 9 fields. The following figure shows the basic structure of an IEEE 802.11 MAC data frame along with the content of the frame control field.

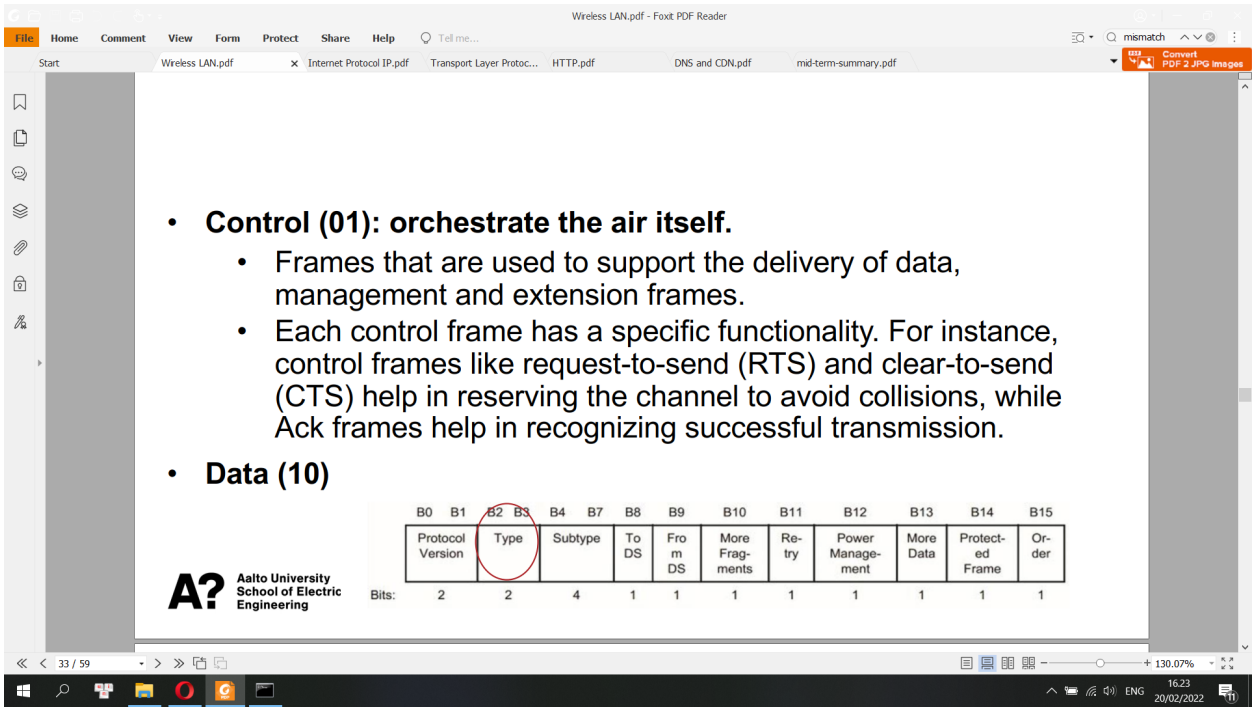




Frame Types

• **Management Frames (00), Control (01), Data (10)**





Subtype

SubTypes

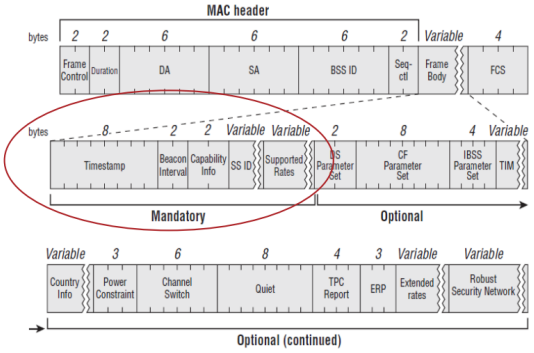
- **Management**
  - *Beacon*: used by the AP to advertise information about the BSS
  - *Probe*: used by clients so that they can find a BSS/SSID to connect to
  - Association
  - Deassociaton
  - Authentication
  - Deauthentication
  - Action

- **Control**
  - *ACK*: acknowledge the receipt of a frame
  - *RTS*: Request to Send
  - *CTS*: Clear to Send
  - BlockAckReq: request a BlockAck
  - BlockAck: acknowledge multiple frames that were sent in a row, instead of for every individual one
  - Control Wrapper
- Data (e.g. standard data frame, Null Data Frame, QoS data frame)

How to inform the presence of APs?

Beacon

- Beacon frames are transmitted **periodically** (by default every 100ms)
- Announce the presence of a wireless LAN and synchronize the members of the BSS
- Timestamp used for clock synchronization

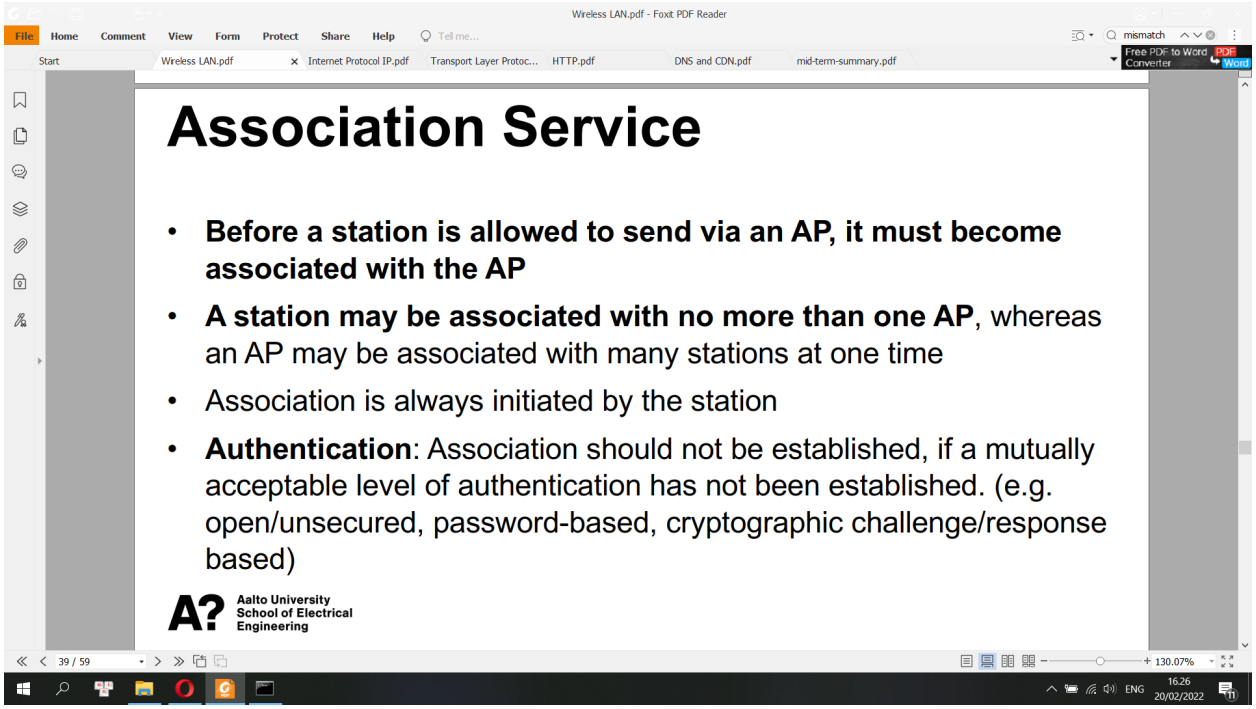


Beacon frame is one of the management frames in IEEE 802.11 based WLANs. It contains all the information about the network. Beacon frames are transmitted periodically, they serve to announce the presence of a wireless LAN and to synchronise the members of the service set. Beacon frames are transmitted by the access point (AP) in an infrastructure basic service set (BSS). In IBSS network beacon generation is distributed among the stations.

Scanning

- **Passive Scanning:** the client radio listens on each channel for beacons sent periodically by an AP
- **Active Scanning:** the client radio sends a probe request and listens for a probe response from an AP
    - Probe requests can be sent to the broadcast address (ff:ff:ff:ff:ff:ff). The client sets a Probe timer and collects answers received until the end of the timer.
    - May send a probe request to a specified SSID

Association Service

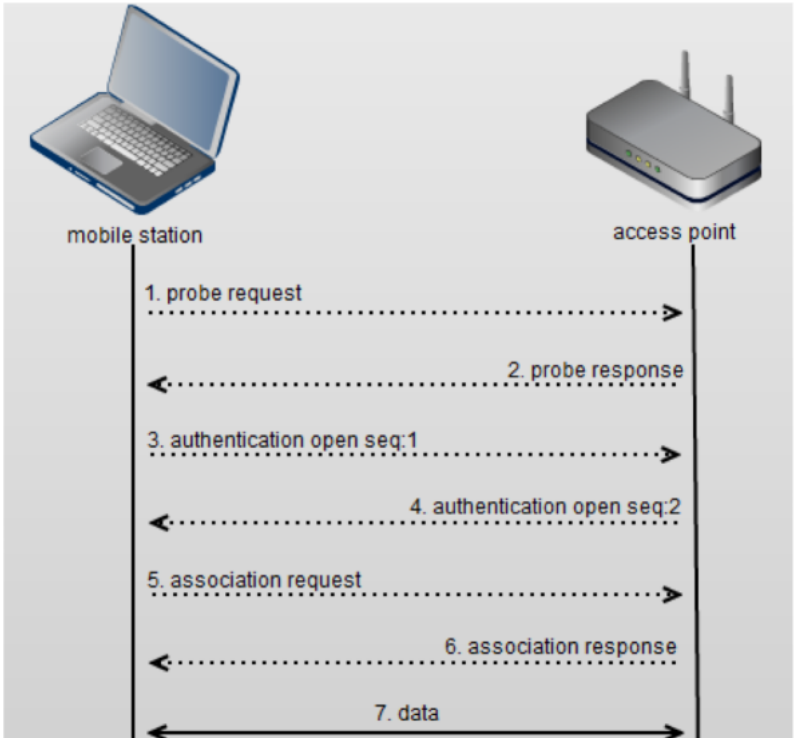


Reassociation and Deassociation

Reassociation and Deassociation

- **The association between a station and a BSS is dynamic**
- **Reassociation** (initiated by STA): enables an established Association of a STA to be transferred from one AP to another AP within an ESS
- **Deassociation** (initiated by either STA or AP): deletes an exsiting association
  - Disassociation is a notification (not a request) and can not be refused by either party to the association.

Association allows the AP/router to record each mobile device so that frames are properly delivered. Association only occurs on wireless infrastructure networks, not in peer-peer mode. A station can only associate with one AP/router at a time.
Devices that connect to WiFi network are called stations (STA). Connection to Wi-Fi is provided by an access point (AP), that acts as a hub for one or more stations. The access point on the other end is connected to a wired network.



Difference from Ethernet
• Limited coverage area
• Shared transmission medium (radio)
• Dynamic topology
• Wi-Fi APs are half-duplex
• Unprotected from outside signals
• Less reliable than wired PHY

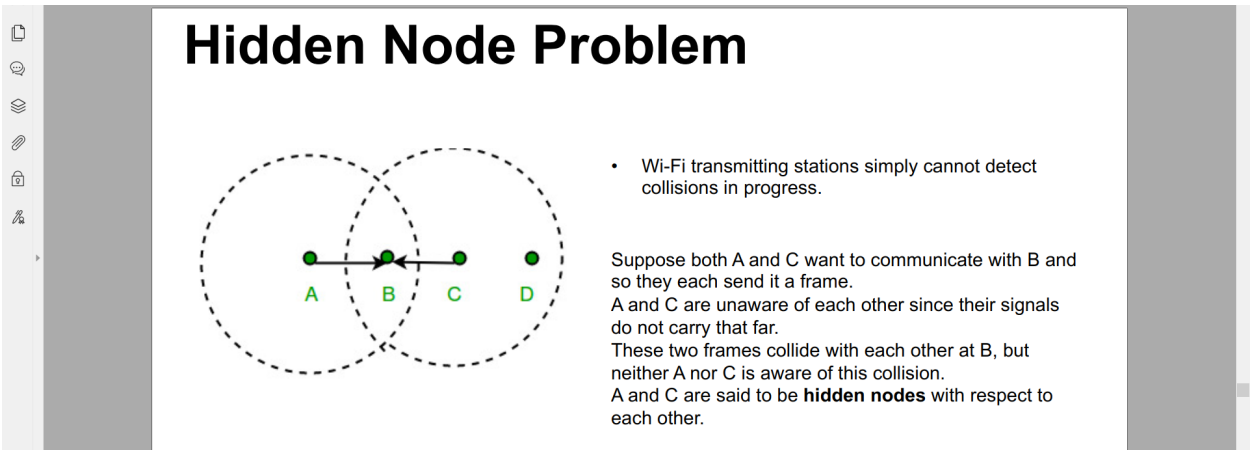
Collision

• A packet collision is defined as any case where a node is receiving more than one packet at a time, resulting in neither packet being correctly received.

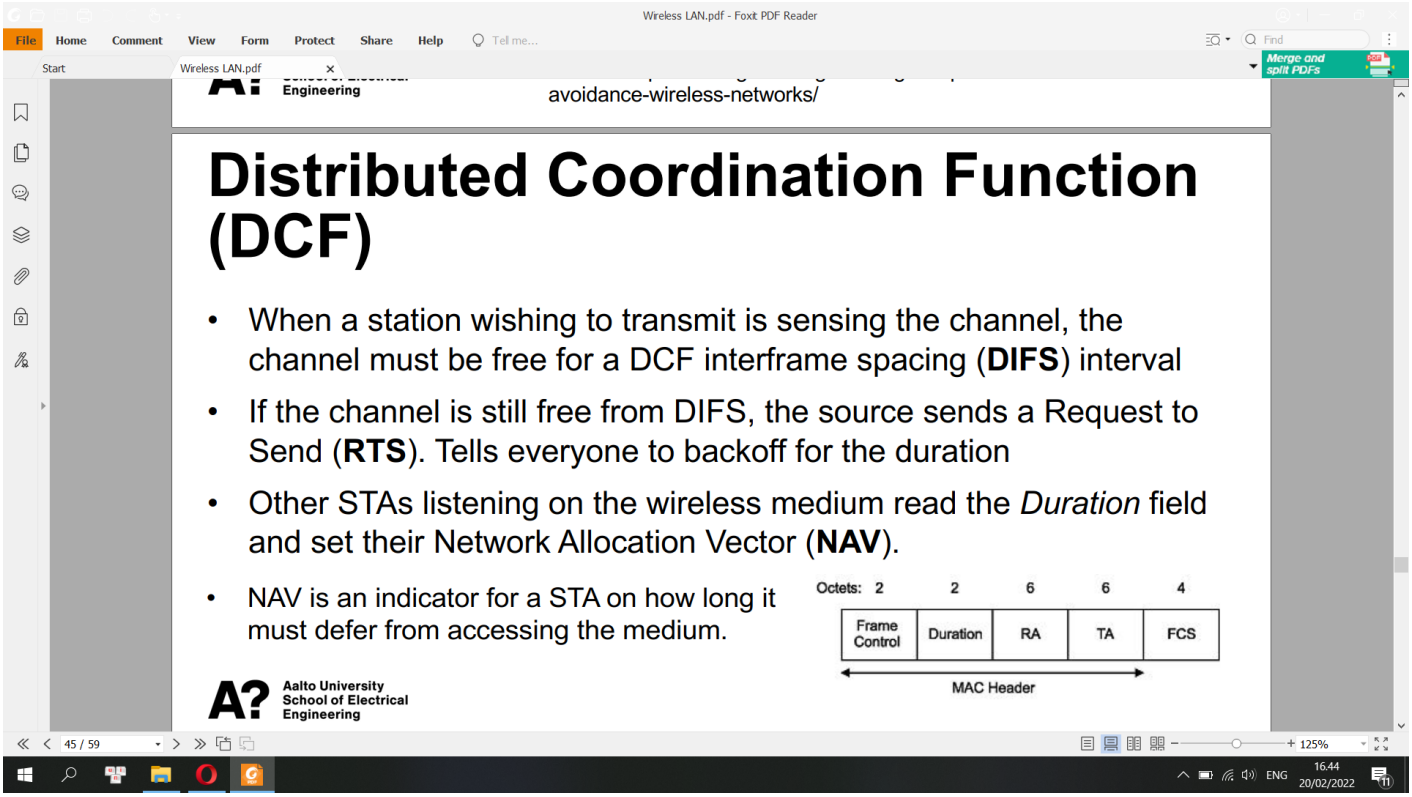
• Carrier Sense: before transmission, a node first listens to the shared medium to determine whether another node is transmitting.

• Collision Avoidance: if another node was heard, we wait for a period of time (usually random) for the node to stop transmitting before listening again for a free communications channel

Hidden Node Problem



Distributed Coordination Function (DCF)



Wireless LAN.pdf - Foxit PDF Reader

School of Electrical Engineering

DCF

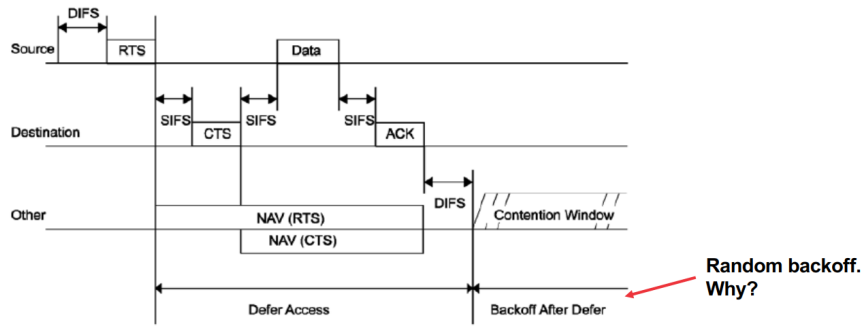
- The destination will respond with a Clear to Send (**CTS**) if it is available to receive data
- After correct reception of the data, the destination will transmit an acknowledgment (**ACK**) back to the sender.
- **Cannot detect collision → each packet is acked**
- MAC-layer retransmission if not acked
- The short interframe spacing (**SIFS**) is used as the wait time between the RTS, CTS, DATA and ACK frames.

Aalto University
School of Electrical Engineering

46 / 59

Wireless LAN.pdf - Foxt PDF Reader

DCF

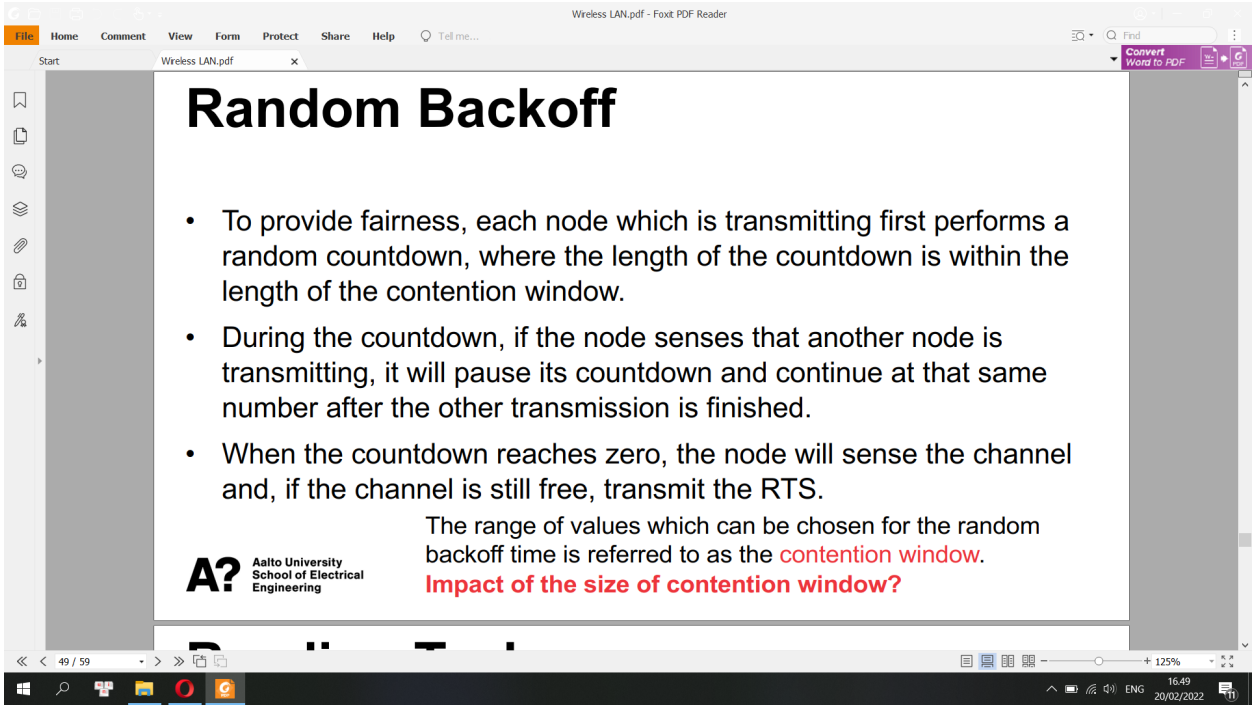


SIFS is always shorter than the DIFS. Do you know why?

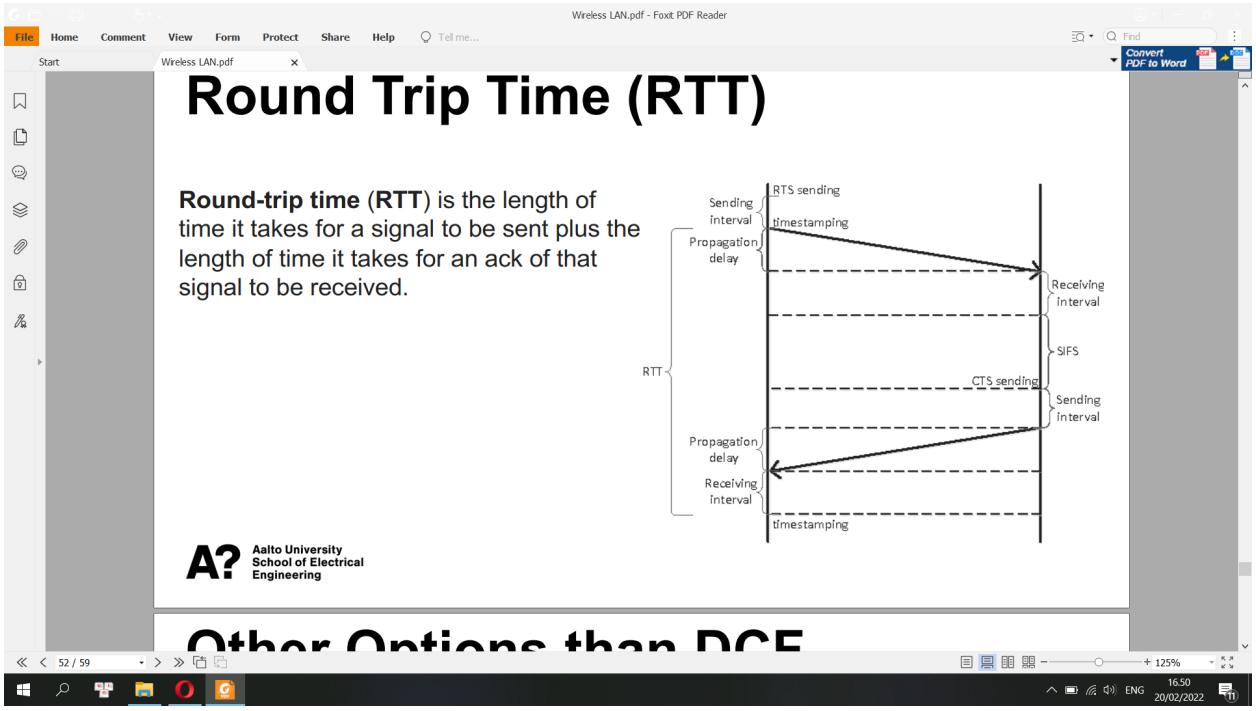
Aalto University
School of Electrical Engineering

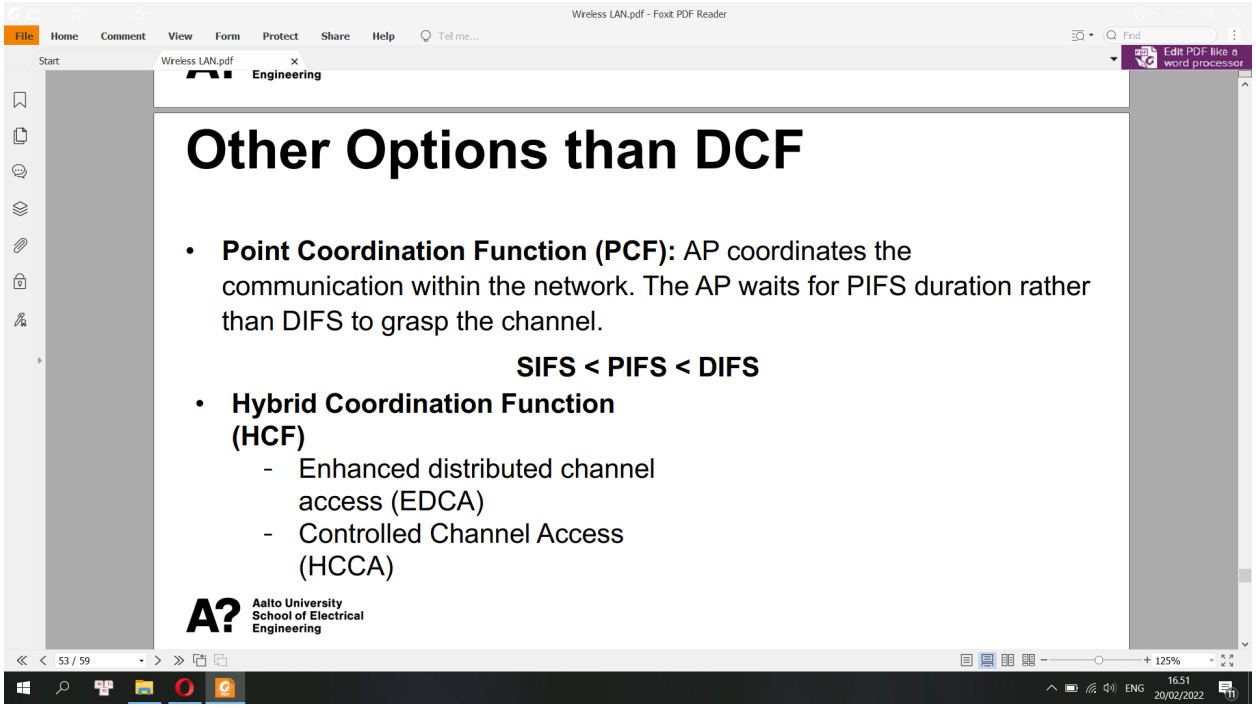
48 / 59

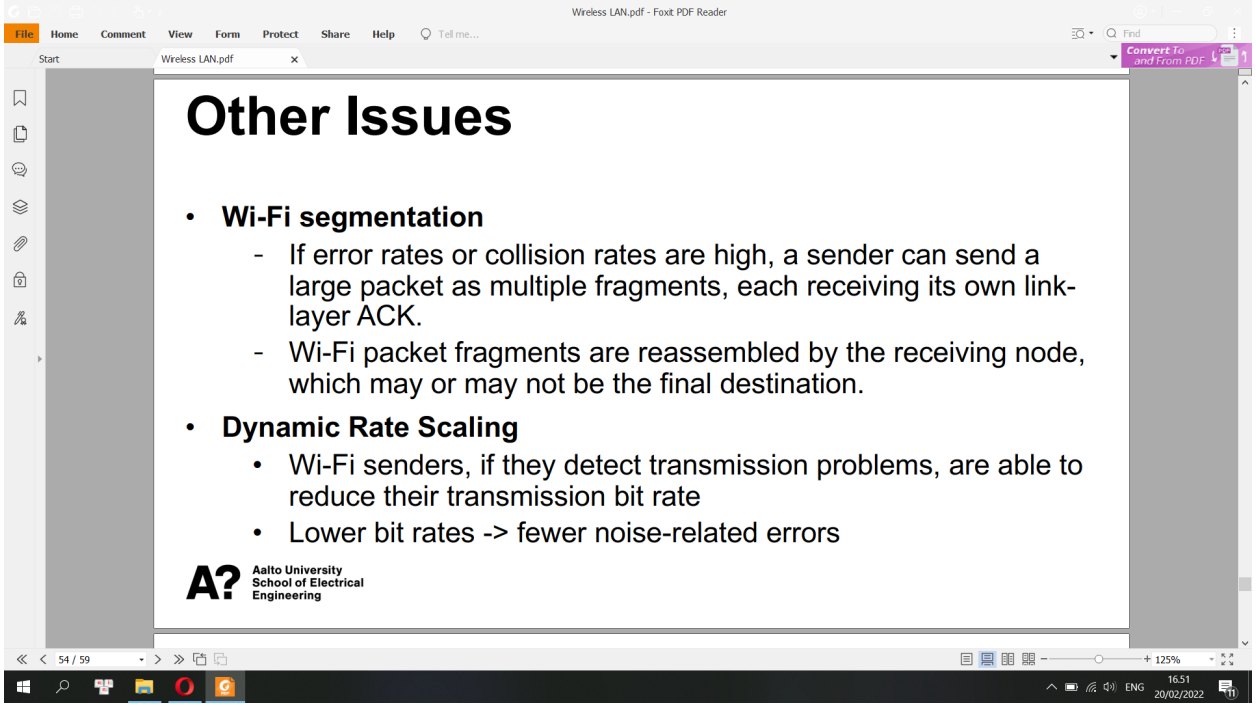
Random Backoff



Round Trip Time (RTT)



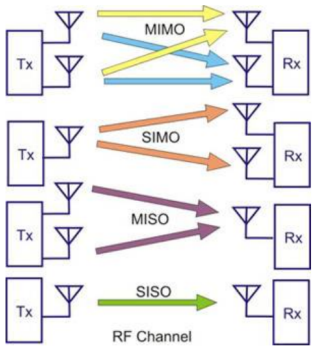

Wireless LAN.pdf - Foxit PDF Reader
File Home Comment View Form Protect Share Help
Wireless LAN.pdf
Aalto University
School of Electrical
Engineering
Other Options than DCF
• Point Coordination Function (PCF): AP coordinates the communication within the network. The AP waits for PIFS duration rather than DIFS to grasp the channel.
SIFS < PIFS < DIFS
• Hybrid Coordination Function (HCF)
- Enhanced distributed channel access (EDCA)
- Controlled Channel Access (HCCA)
Aalto University
School of Electrical
Engineering
53 / 59
125%
ENG
16.51
20/02/2022


Wireless LAN.pdf - Foxit PDF Reader
File Home Comment View Form Protect Share Help
Wireless LAN.pdf
Other Issues
• Wi-Fi segmentation
- If error rates or collision rates are high, a sender can send a large packet as multiple fragments, each receiving its own link-layer ACK.
- Wi-Fi packet fragments are reassembled by the receiving node, which may or may not be the final destination.
• Dynamic Rate Scaling
• Wi-Fi senders, if they detect transmission problems, are able to reduce their transmission bit rate
• Lower bit rates -> fewer noise-related errors
Aalto University
School of Electrical
Engineering
54 / 59
125%
ENG
16.51
20/02/2022

Wireless LAN.pdf - Foxit PDF Reader

Multiple Input Multiple Output (MIMO)

- MIMO is commonly used in Wi-Fi, WiMax and cellular networks
- To use N streams, both sender and receiver must have N antennas; all the antennas use the same frequency channels but each transmitter antenna sends a different data stream.
- More antennas → higher data rate, but also more power, space?

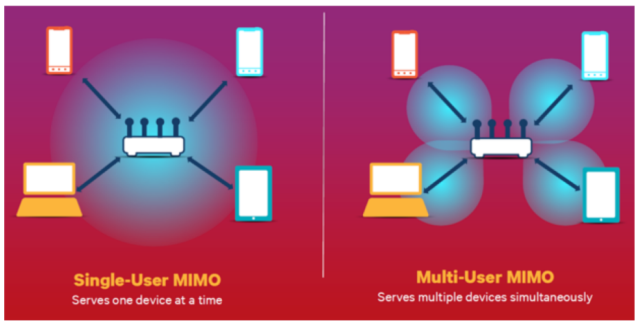


Aalto University
School of Electrical
Engineering

55 / 59

Wireless LAN.pdf - Foxit PDF Reader

SU-MIMO vs. MU-MIMO

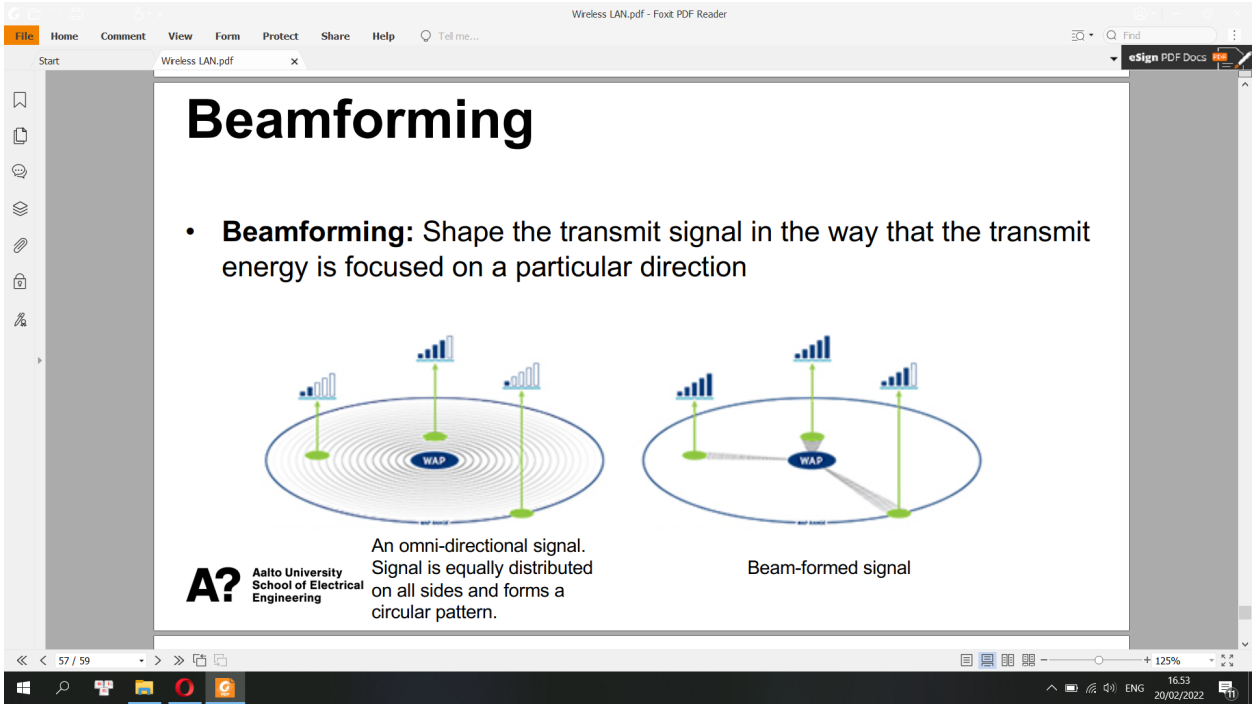


Source: Qualcomm

Aalto University
School of Electrical
Engineering

MU-MIMO becomes available in 802.11ac Wave 2 but only applies to downlink.

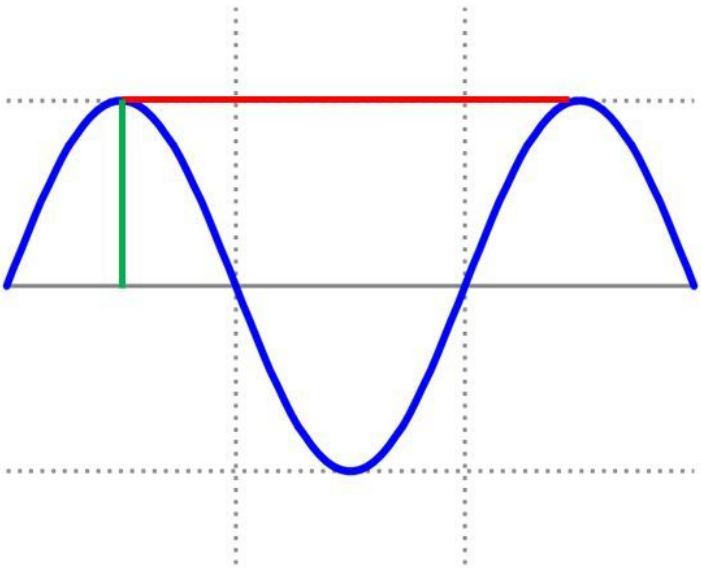
56 / 59



Beamforming uses antenna arrays to dynamically alter the transmission pattern of the AP, and the transmission pattern can be changed on a per-frame basis.

Self-test: Wireless LAN

• What is frequency, wavelength and amplitude of RF wave? What is the relationship between frequency and wavelength?



Wavelength (λ)

Distance between identical points on consecutive waves

Amplitude

Distance between origin and crest (or trough)

Frequency (ν)

Number of waves that pass a point per unit time

Speed

= wavelength x frequency

• Does higher frequency mean higher or lower data rate? Does it mean larger or smaller coverage?

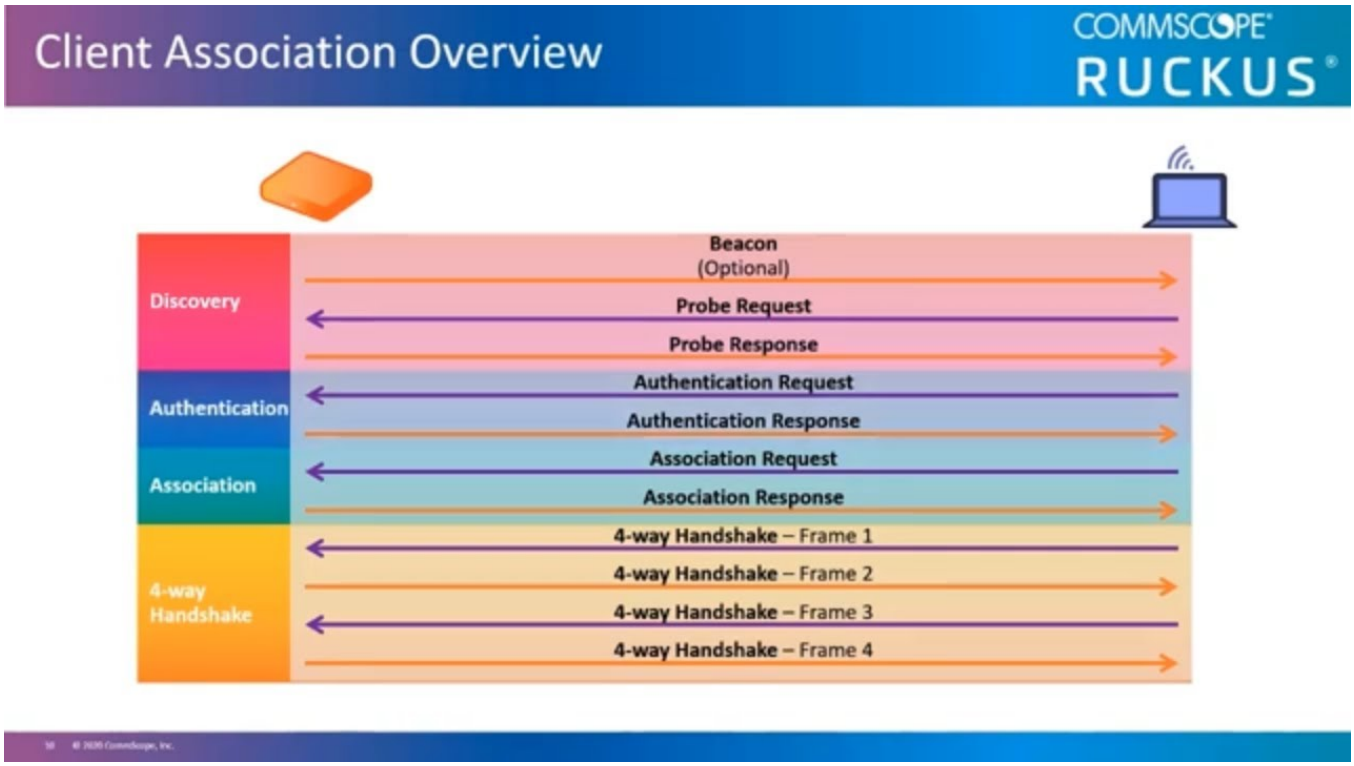
Higher frequency means higher data rate and larger bandwidth/coverage

• In case of Wi-Fi, what is BSS? What is beacon used for? How does active scanning work? How to associate with a Wi-Fi AP?

Basic Service Set (BSS), as the name suggests, is basically a network topology that allows all wireless devices to communicate with each other through a common medium i.e. AP (Access point). It also manages these wireless devices or clients. It basically provides a building block to all wireless LAN (Local Area Network). BSS basically contains only one AP that is connected to all stations i.e. all wireless devices within the network. Here, AP is a common access point that acts as a medium and creates WLAN (Wireless Local Area Network).

AP allows all wireless devices to get connected to a wired network and start communicating with each other. Therefore, AP is considered a master that controls all wireless devices or stations with BSS or WLAN. BSS contains only one AP, but it may contain one or more stations. BBS is generally considered simplest if it contains one AP and one station. Before connecting to a wireless network, wireless device or station need to send an association request to AP and then AP checks whether the client fit any of the following criteria or not

- beacon: check above
- active scanning: A client can use two scanning methods: active and passive. During an active scan, the client radio transmits a probe request and listens for a probe response from an AP. With a passive scan, the client radio listens on each channel for beacons sent periodically by an AP.
- Wifi-AP association:



• What is hidden node problem?: check above

• Is Wi-Fi half-duplex or hall-duplex? If you can explain how CSMA/CA work, that is even better

WiFi is a half duplex form of data transmission, which is to say, data packets are sent back and forth in sequence. It happens so quickly that it mimics seamless, two-way data transmission, but in fact, data cannot be both sent and received simultaneously.

Collision Avoidance (CSMA/CA) avoids collisions by listening for a transmission signal before sending data. If a signal is detected, the sender starts a counter with a random value and then waits. Once this counter runs down, the sender will try again. This process repeats until the sender can send the data.

CSMA/CA is used with Wifi and it's useful. Since "air" is generally a shared medium, no two stations must transmit simultaneously.

=> Wifi is half duplex due to CSMA/CA

• SIFS is always shorter than the DIFS. Do you know why?

DIFS = distributed inter frame spacing, a new transmission can begin only after DIFS of idle time. PIFS = spacing after which point coordinator can take over. SIFS = spacing between transmission and ACK, between polling and response.

Having PIFS < DIFS allows point coordinator to take over and separate contention-free and contention periods. Having SIFS smaller than both PIFS and DIFS prevents ACK and important control packets from getting killed.

=> SIFS is always shorter than DIFS

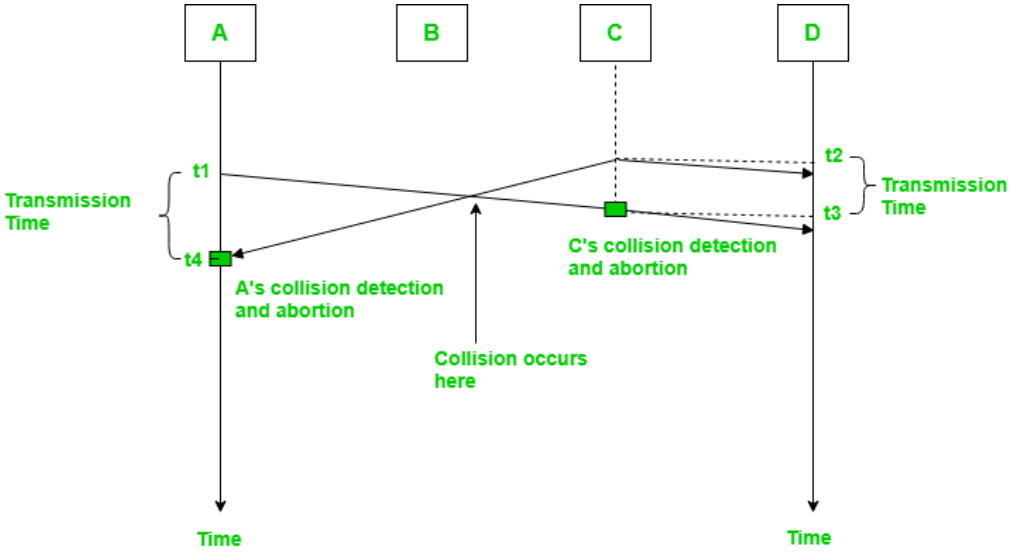
• Impact of the size of contention window?

Contention Window – It is the amount of time divided into slots.A station which is ready to send frames chooses random number of slots as wait time.

The intention with different priority classes is to use a smaller contention window to get faster access to the channel for high-priority traffic while low-priority traffic uses a larger contention window, increasing the likelihood of high-priority data being transmitted before low-priority data.

1. Carrier Sense Multiple Access with Collision Detection (CSMA/CD) –

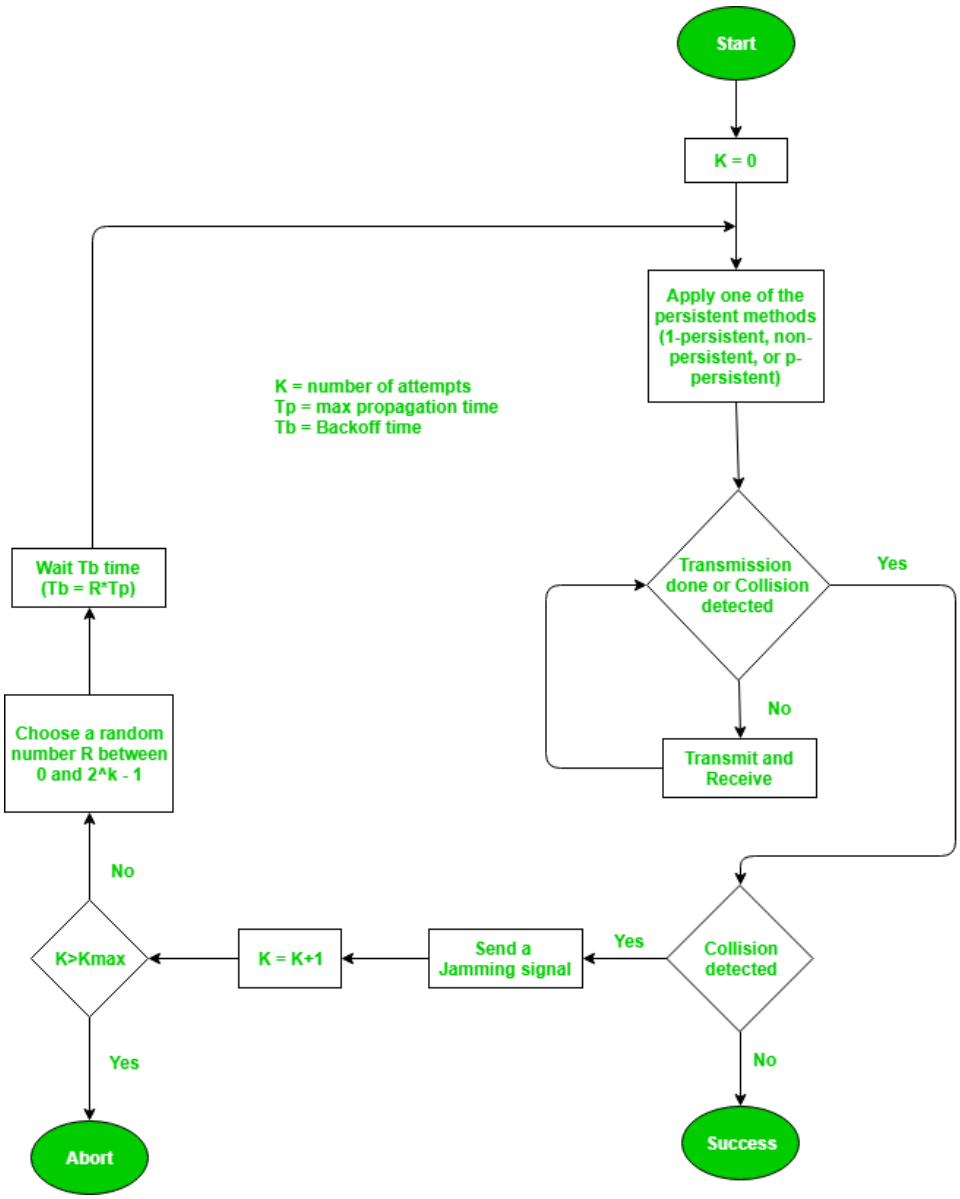
In this method, a station monitors the medium after it sends a frame to see if the transmission was successful. If successful, the station is finished, if not, the frame is sent again.



In the diagram, A starts send the first bit of its frame at t1 and since C sees the channel idle at t2, starts sending its frame at t2. C detects A's frame at t3 and aborts transmission. A detects C's frame at t4 and aborts its transmission. Transmission time for C's frame is therefore t3-t2 and for A's frame is t4-t1 .

So, the frame transmission time (Tfr) should be at least twice the maximum propagation time (Tp). This can be deduced when the two stations involved in collision are maximum distance apart.

Process –
The entire process of collision detection can be explained as follows:

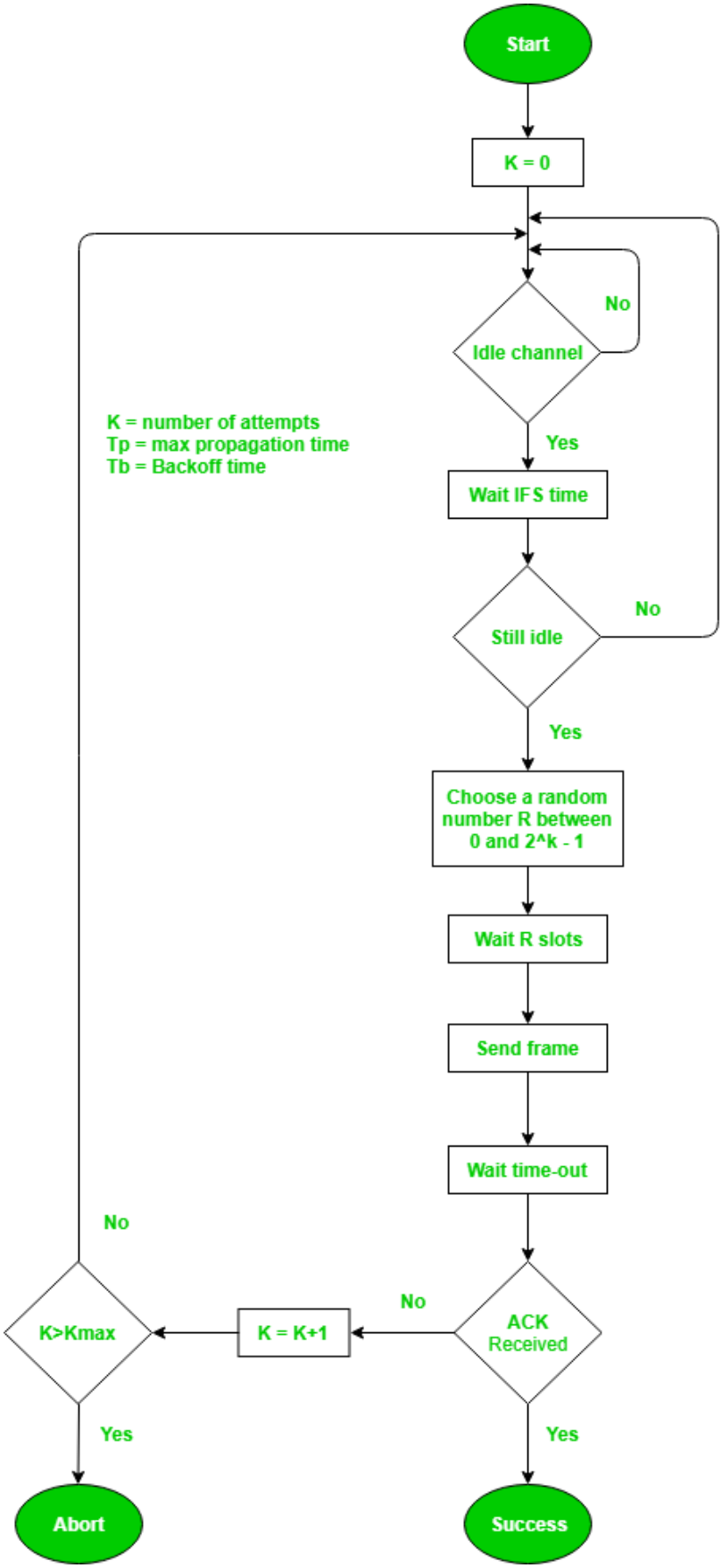


Throughput and Efficiency – The throughput of CSMA/CD is much greater than pure or slotted ALOHA.

For 1-persistent method throughput is 50% when G=1.
For non-persistent method throughput can go upto 90%.
2. Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) –
The basic idea behind CSMA/CA is that the station should be able to receive while transmitting to detect a collision from different stations. In wired networks, if a collision has occurred then the energy of received signal almost doubles and the station can sense the possibility of collision. In case of wireless networks, most of the energy is used for transmission and the energy of received signal increases by only 5-10% if a collision occurs. It can't be used by the station to sense collision. Therefore CSMA/CA has been specially designed for wireless networks.

These are three types of strategies:

InterFrame Space (IFS) – When a station finds the channel busy, it waits for a period of time called IFS time. IFS can also be used to define the priority of a station or a frame. Higher the IFS, lower is the priority.
Contention Window – It is the amount of time divided into slots.A station which is ready to send frames chooses random number of slots as wait time.
Acknowledgments – The positive acknowledgments and time-out timer can help guarantee a successful transmission of the frame.
Process –
The entire process for collision avoidance can be explained as follows:


Start
K = 0
Idle channel
No
Yes
K = number of attempts
Tp = max propagation time
Tb = Backoff time
Wait IFS time
Still idle
No
Yes
Choose a random number R between 0 and 2^k - 1
Wait R slots
Send frame
Wait time-out
No
ACK Received
No
K = K+1
K>Kmax
Yes
Yes
Success
Abort

Internet Protocol (IP)

IP Addresses

• An IP address identifies a host or network interface on a network, and provides the location of the host in the network

• Each device that wants to communicate with other devices on a TCP/IP network needs to have an IP address configured

In contrast to MAC address, an IP address is a logical address. It can be configured manually or obtained from a DHCP (Dynamic Host Configuration Protocol) server

An IPv4 address consists of 4 octets (32 bits)

[octet] . [octet] . [octet] . [octet]

1000 1000.1110 0011.1110 1101.0110 1000 à 136.227.237.104

• Each number can be 0 to 255

Number of IPv4: 4 otctets: 4 x 8 = 32 bits

=> 2^32 different IPv4 addresses = 4,294,967,296

DHCP (Dynamic Host Configuration Protocol): The Dynamic Host Configuration Protocol is a network management protocol used on Internet Protocol networks for automatically assigning IP addresses and other communication parameters to devices connected to the network using a client–server architecture

• Subnetting: dividing a network into two or more smaller networks.

• Purposes of subnetting

Reducing the size of the broadcast domain

Increasing routing efficiency

Enhancing the security of the network

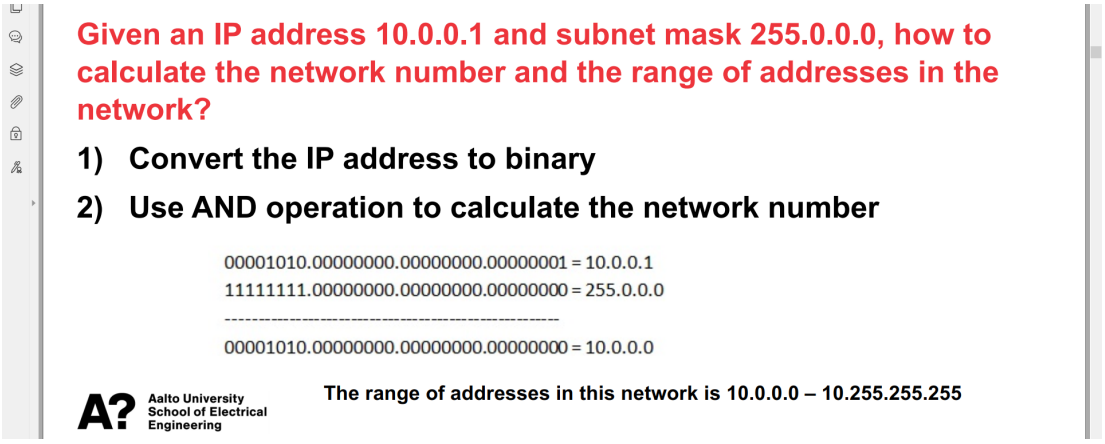
IPv4 Subnet Mask

• An IP address is divided into two parts: network and host parts.

• Subnet mask is used for determining the network part and the host part of an IP address.

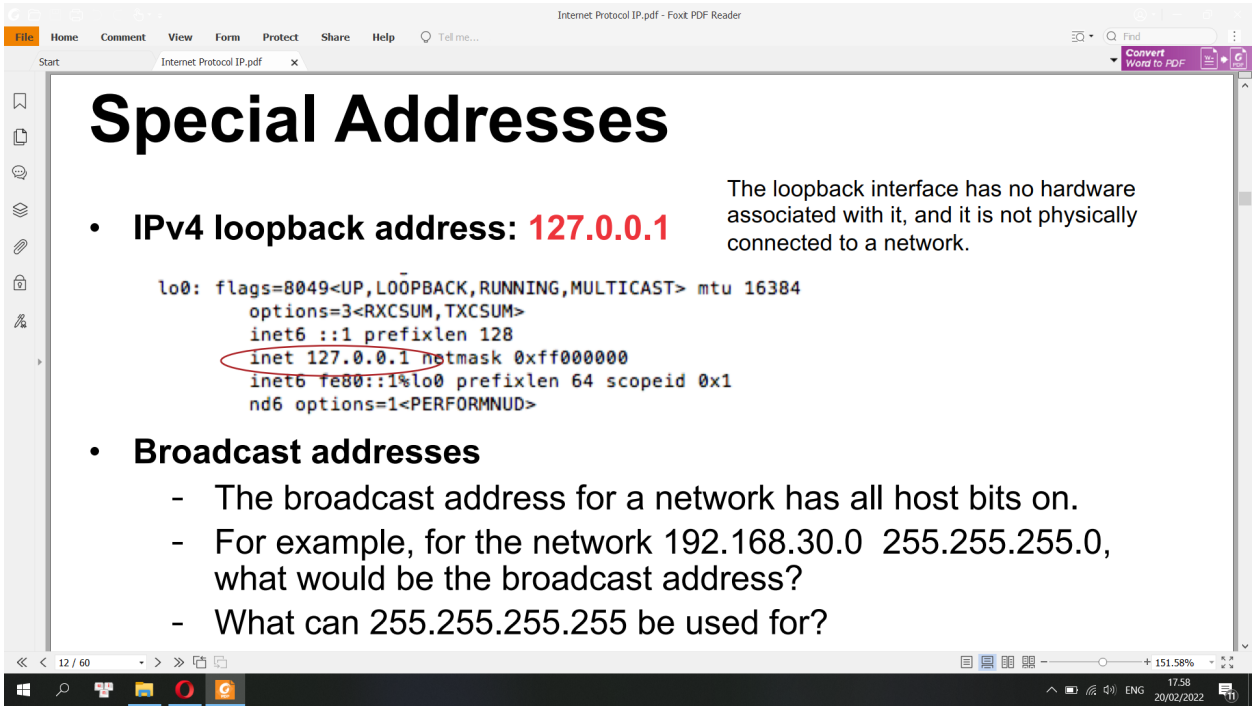
• A subnet mask consists of 32 bits. The 1s in the subnet mask represent a network part, the 0s a host part.

• A subnet mask must always be a series of 1s followed by a series of 0s. E.g. 255.255.0.0, 255.0.0.0.

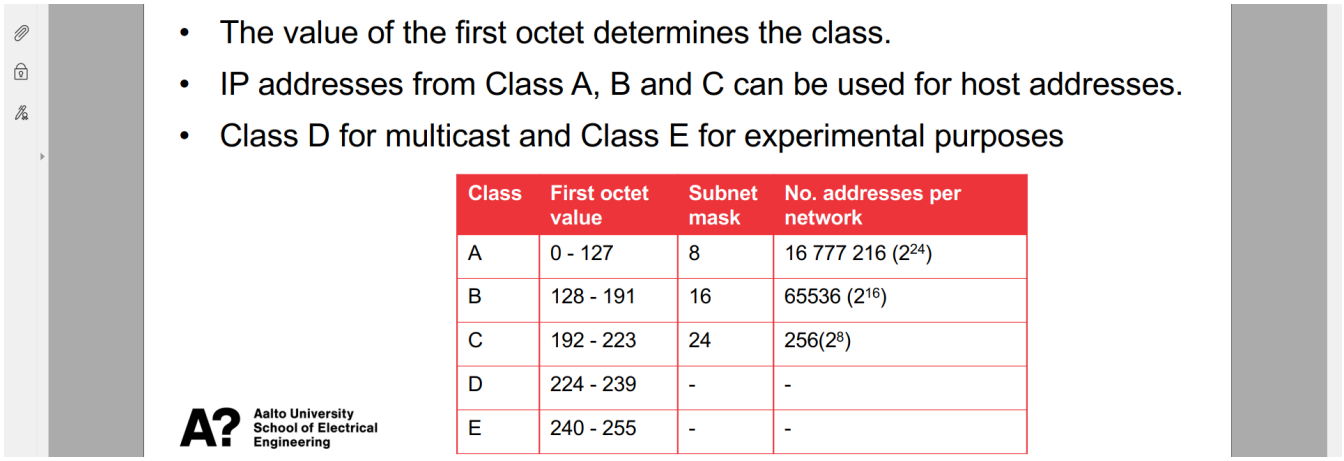


Classless Inter-Domain Routing (CIDR or supernetting) is a way to combine several class-C address ranges into a single network or route. This method of routing adds class-C Internet Protocol (IP) addresses. These addresses are given out by Internet Service Providers (ISPs) for use by their customers
VLSM stands for Variable Length Subnet Mask where the subnet design uses more than one mask in the same network which means more than one mask is used for different subnets of a single class A, B, C or a network. It is used to increase the usability of subnets as they can be of variable size
CIDR is based on variable-length subnet masking (VLSM) to allow allocation and routing based on arbitrary-length prefixes

Special Addresses



Classes of IPv4 addresses:



Class	First octet value	Subnet mask	No. addresses per network
A	0 - 127	8	16 777 216 (2^{24})
B	128 - 191	16	65536 (2^{16})
C	192 - 223	24	256(2^{8})
D	224 - 239	-	-
E	240 - 255	-	-

Prefix

Prefix

192.0.1.0/24 **prefix length is 24.**

32-24=8 bits are left for host addresses

10.0.0.0/8 **32-8=24bits are left for host addresses**

IP address: 136.227.237.104/27
What is the subnet mask?
255.255.255.224, where 224 = 11100000
Number of useable IPv4 host addresses

- **The number of usable IP addresses** can be calculated from the following formula:

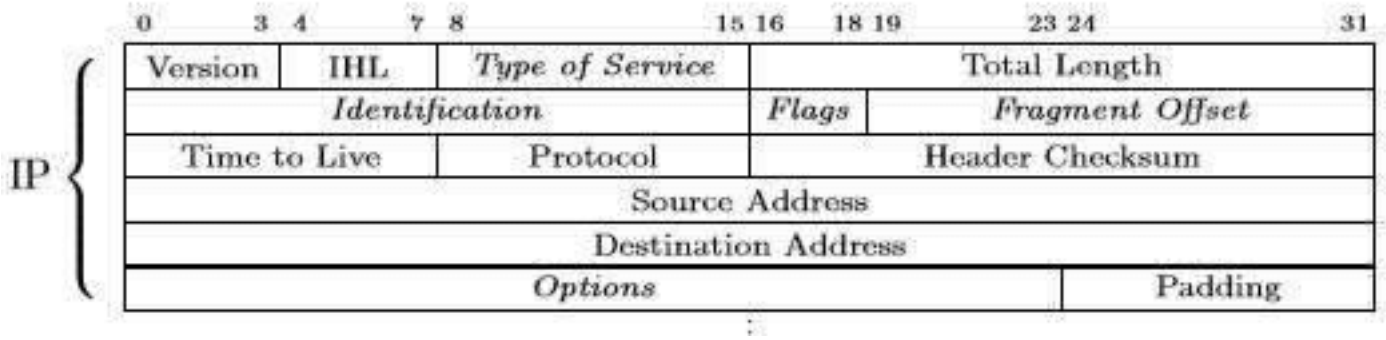
2 to the power of host bits – 2

The first and the last address are the network address and the broadcast address, respectively. All other addresses inside the range could be assigned to Internet hosts.

Private Addresses

- **Private addresses are IPv4 addresses intended only for site internal use**
- **Reserved private IPv4 network ranges**
  - 10.0.0.0/8
  - 172.16.0.0/12
  - 192.168.0.0/16

IPv4 Header



Version: IP version field, set to 4 for IPv4.

Internet Header Length
IHL is the number of 32-bit words making up the header field. The minimum available number is 5, as the first 20 bytes are mandatory.

Differentiated Services Code Point (DSCP). It specifies the type of service for differentiated services, such as voice over IP.

Explicit Congestion Notification (ECN)
When supported, ECN carries a network congestion notification without dropping packets or wasting bandwidth.

Total Length
It's the total size of the datagram, including both header and data parts.

Identification
It's mainly used to indicate a group of fragments if the datagram is fragmented.

Flags
Different combinations of the flags control the fragmentation and indicate fragmented datagrams.

We assume the don't fragment flag is set when the destination can't assembly the fragmented packet. Therefore, the packet is dropped if the flags mark the datagram as not to be fragmented, but it must be.

Fragment Offset
It contains the offset of a fragmented packet.

Time to Live (TTL)
The number of hops a packet lives. Each router on the way decrements the TTL field. If it reaches 0, the packet is discarded. This way, looping packets are eliminated.

Protocol
It's the protocol for the data part like TCP or UDP.
Protocol: identifies the content of the packet body
1: an ICMP packet
4: an encapsulated IPv4 packet
6: a TCP packet
17: a UDP packet
41: an encapsulated IPv6 packet
50: an encapsulating security payload

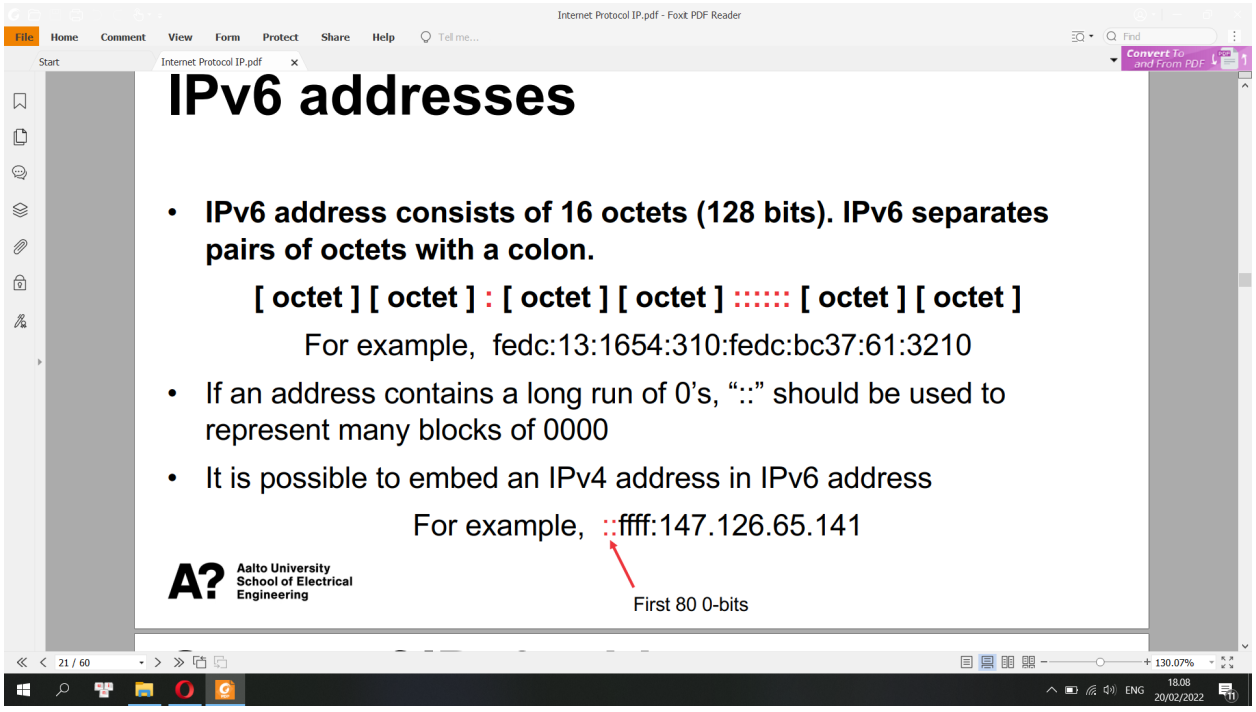
Header Checksum
It's a checksum for error detection, covering only the header field. Each router checks this value and discards the packet if an error is detected.
The data part's checksum is not calculated.

Source Address
Source host IPv4 address.

Destination Address
Destination host IPv4 address.

Options
Additional options are stored in this field when present.

IPv6 addresses



Number of IPv6: 16 octet: 4 x 16 = 128 bits
=> 2^128 different IPv6 addresses = 340,282,366,920,938,463,463,374,607,431,768,211,456

Scope of IPv6 addresses

- The scope of a unicast address is either **global**, meaning it is intended to be globally routable, or **link-local**, meaning that it will only work with directly connected neighbors
- E.g. the loopback address **::1** **(127 0-bits followed by 1 1-bit)** is considered to have link-local scope
- **Link-local** addresses begin with the 64-bit link-local prefix consisting of the ten bits 1111 1110 10 followed by 54 more zero bits; that is, **fe80::/64**.

```
Wireless LAN adapter Wi-Fi:

   Connection-specific DNS Suffix  . :
   IPv6 Address. . . . . . . . . . . : 2001:14ba:a0bd:dd00:c10c:de5e:2cbf:132c
   Temporary IPv6 Address. . . . . . : 2001:14ba:a0bd:dd00:d840:627c:3545:4dd5
   Link-local IPv6 Address . . . . . : fe80::c10c:de5e:2cbf:132c%9
   IPv4 Address. . . . . . . . . . . : 192.168.1.208
   Subnet Mask . . . . . . . . . . . : 255.255.255.0
   Default Gateway . . . . . . . . . : fe80::1eb7:2cff:fe7c:c420%9
                                       192.168.1.1
```

where % sign is the scope id
Predefined and Reserved Scopes

Predefined and Reserved Scopes

Scope values

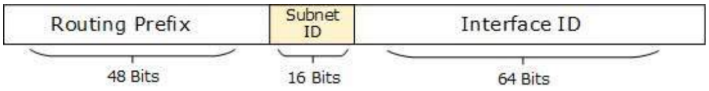
Value	Scope name	Notes
0x0	*reserved*	
0x1	interface-local	Interface-local scope spans only a single interface on a node, and is useful only for loopback transmission of multicast.
0x2	link-local	Link-local scope spans the same topological region as the corresponding unicast scope.
0x3	realm-local	Realm-local scope is defined as larger than link-local, automatically determined by network topology and must not be larger than the following scopes.[13]
0x4	admin-local	Admin-local scope is the smallest scope that must be administratively configured, i.e., not automatically derived from physical connectivity or other, non-multicast-related configuration.
0x5	site-local	Site-local scope is intended to span a single site belonging to an organisation.
0x8	organization-local	Organization-local scope is intended to span all sites belonging to a single organization.
0xe	global	Global scope spans all reachable nodes on the internet - it is unbounded.
0xf	*reserved*	

```
en0: flags=8863<UP,BROADCAST,SMART,RUNNING,SIMPLEX,MULTICAST> mtu 1500
        ether 98:01:a7:90:7d:db
        inet6 fe80::9a01:a7ff:fe90:7ddb%en0 prefixlen 64 scopeid 0x4
        inet 192.168.0.154 netmask 0xffffff00 broadcast 192.168.0.255
        nd6 options=1<PERFORMNUD>
        media: autoselect
        status: active
```

Source: wikipedia

IPv6 Interface Identifier

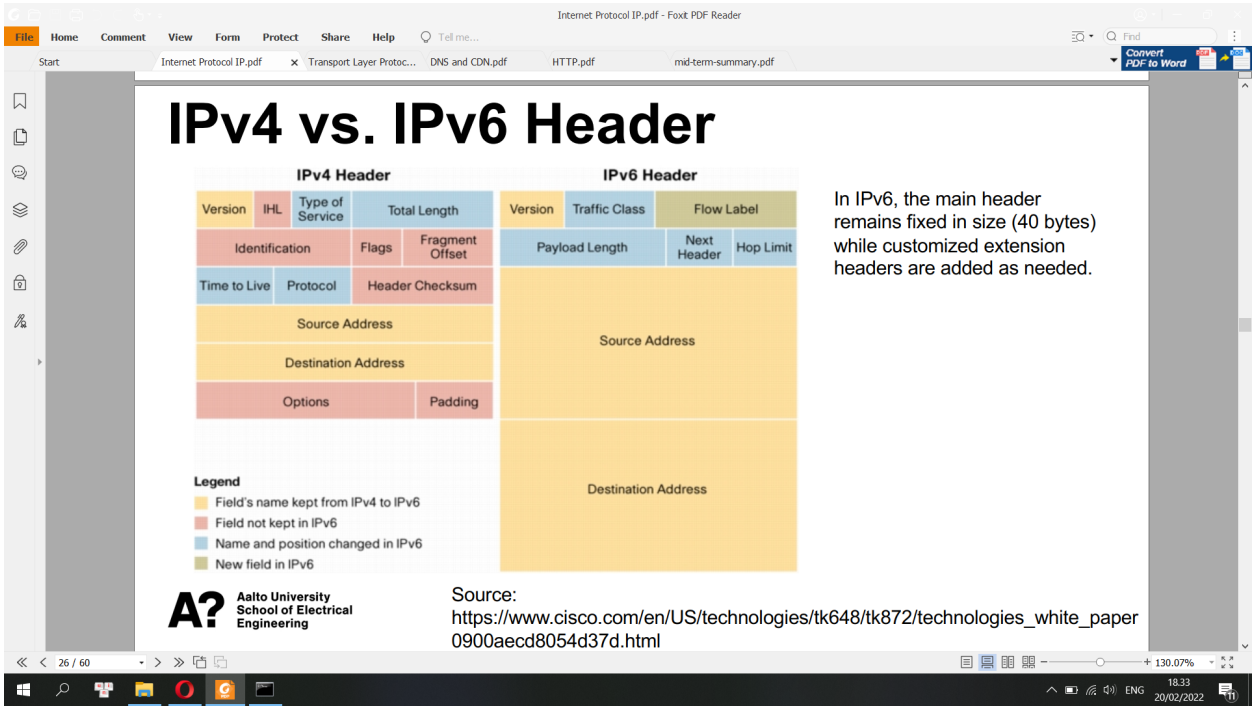
• Most IPv6 addresses can be divided into a 64-bit network prefix and a 64-bit “host” portion. These host-portion bits are known officially as the interface identifier

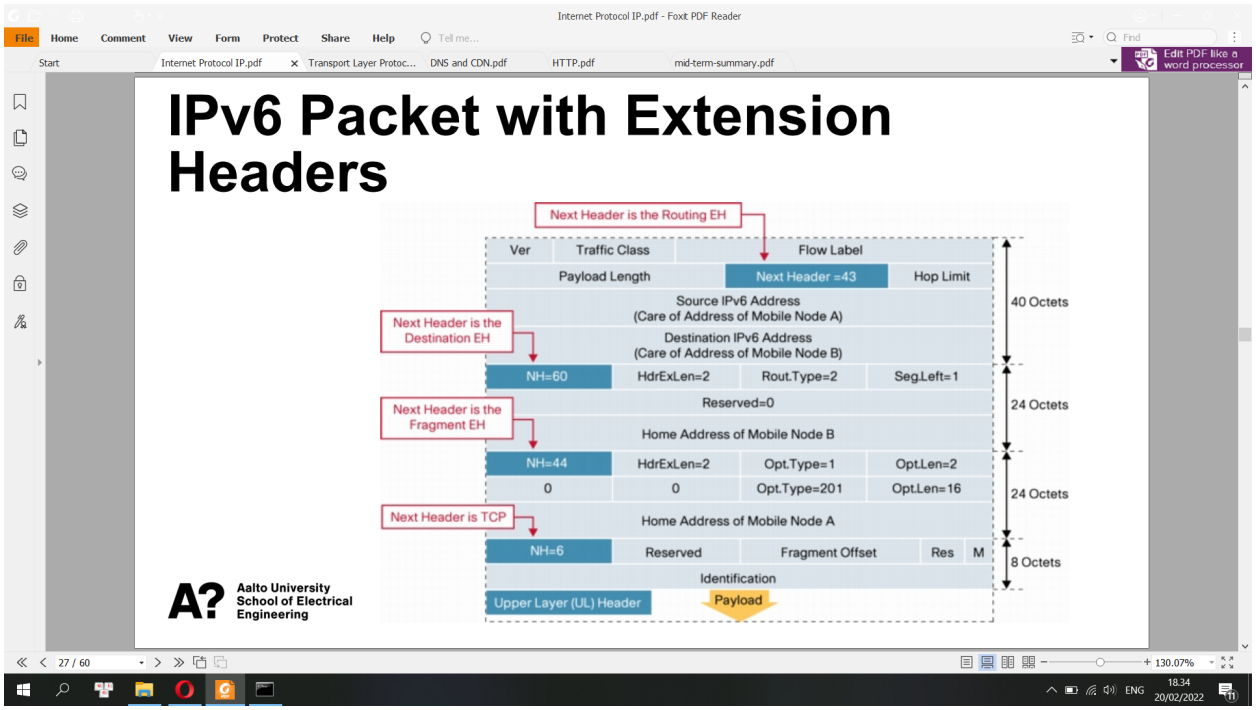


AnyCast

IPv6 has introduced Anycast mode of packet routing. In this mode, multiple interfaces over the Internet are assigned same Anycast IP address. Routers, while routing, send the packet to the nearest destination.

IPv6 does not support broadcast

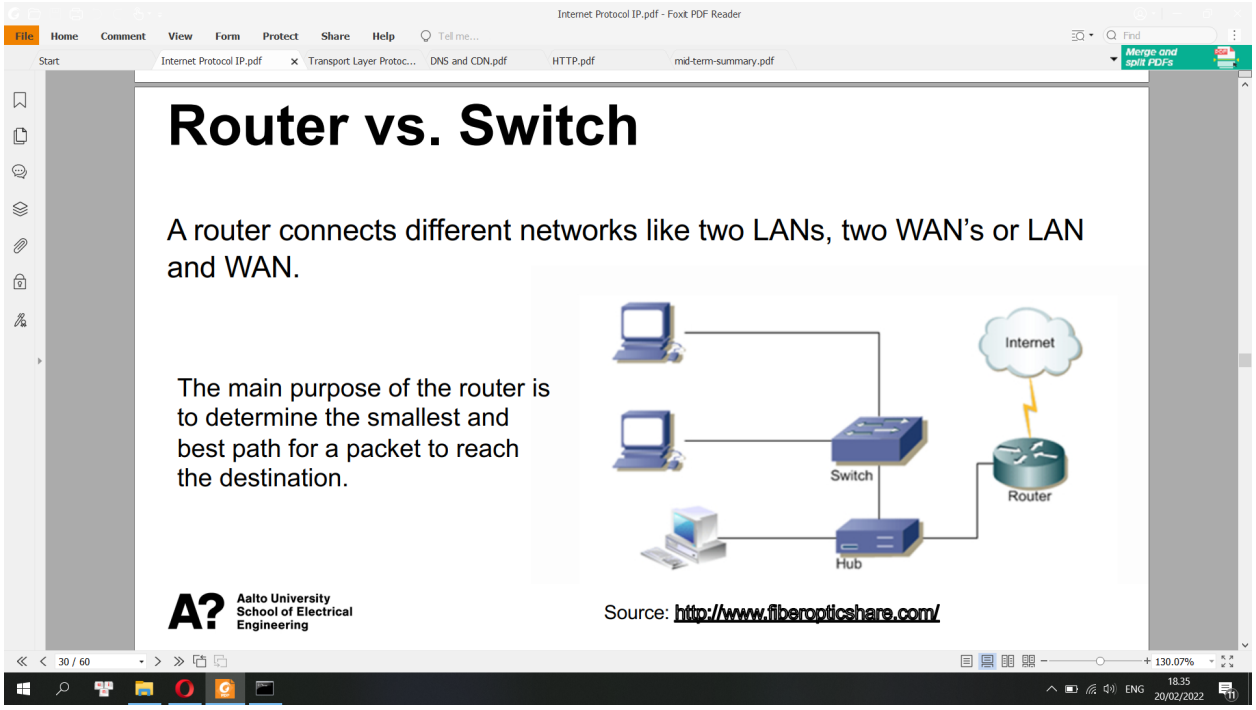




IP Routing

IP routing is the process of sending packets from a host on one network to another host on a different remote network. This process is usually done by routers

Router vs Switch vs Hub

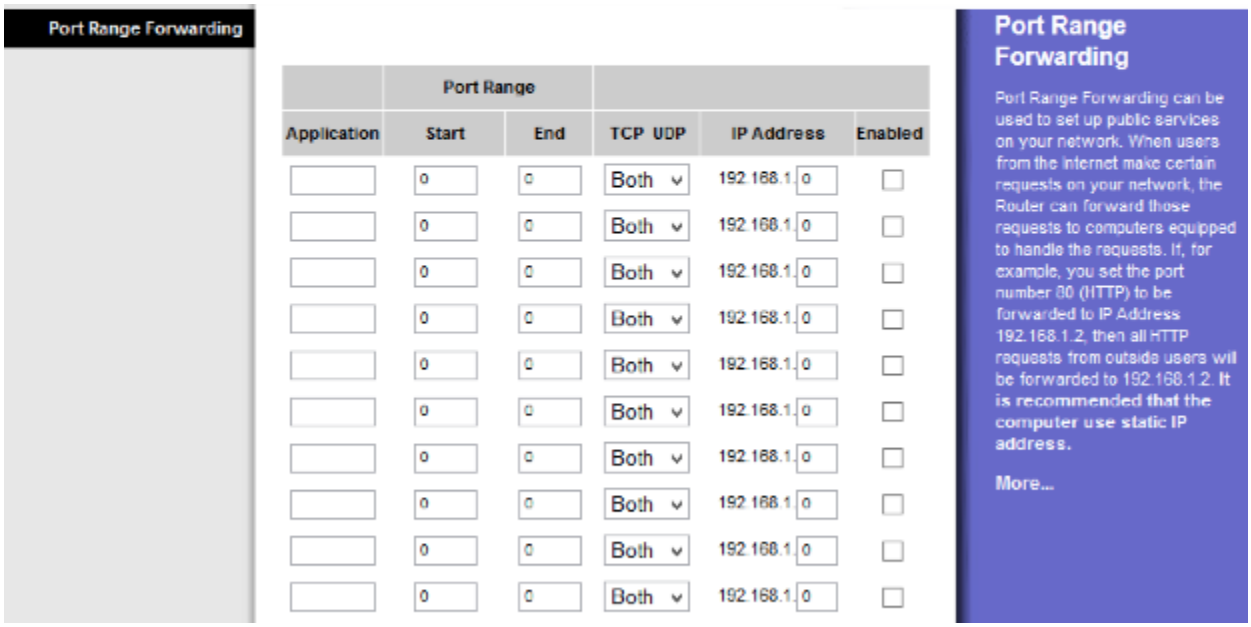


IP Routing

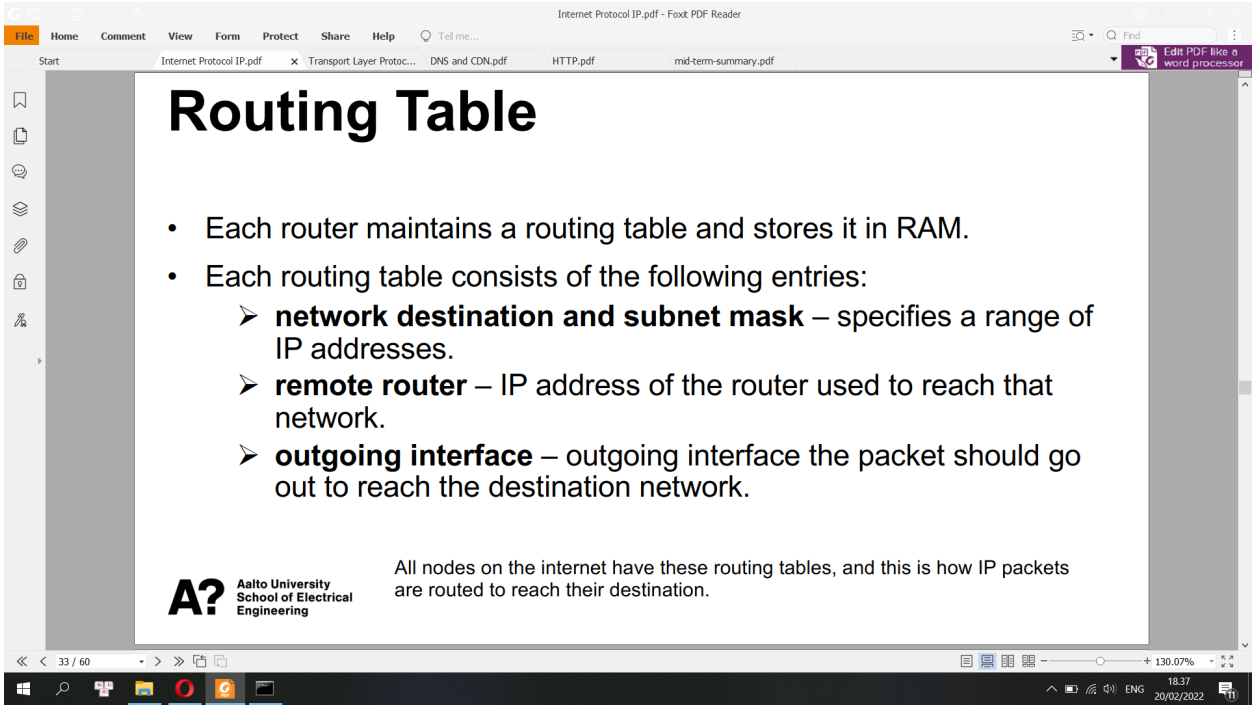
• Routers examine the destination IP address of a packet , determine the next-hop address, and forward the packet.

• Routers use routing tables to determine a next hop address to which the packet should be forwarded.

Port forwarding

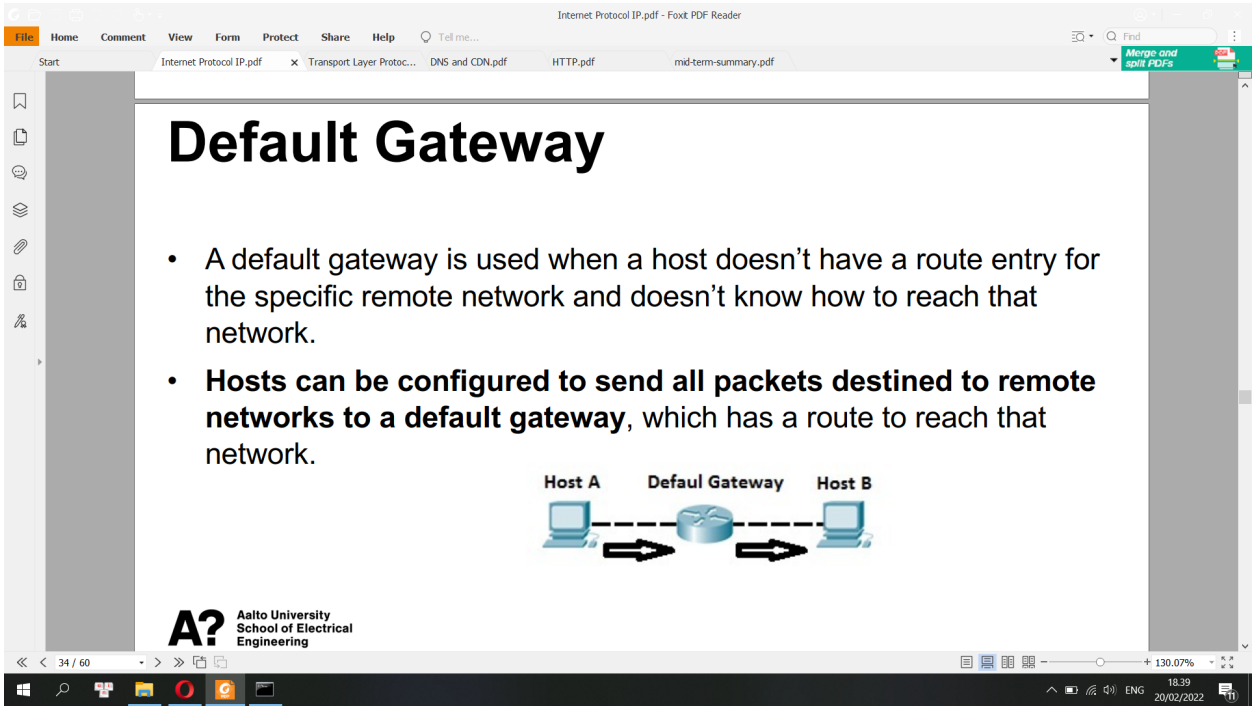


Routing Table



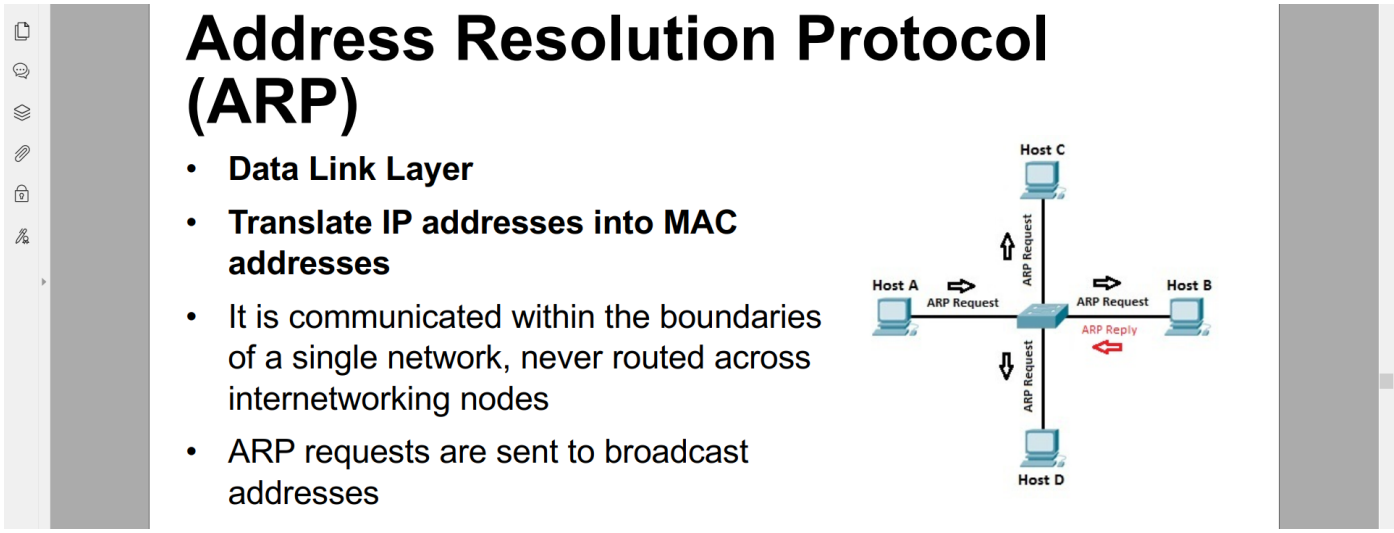
Default Gateway

A default gateway is the node in a computer network using the Internet protocol suite that serves as the forwarding host (router) to other networks when no other route specification matches the destination IP address of a packet.



Address Resolution Protocol (ARP)

The Address Resolution Protocol (ARP) is a communication protocol used for discovering the link layer address, such as a MAC address, associated with a given internet layer address, typically an IPv4 address. This mapping is a critical function in the Internet protocol suite.



ARP caches

• All operating systems maintain ARP caches that are checked before sending an ARP request message.

• The addresses will stay in the cache for a couple of minutes.
• Command line %arp -a

Process of IP Routing

Host 1:

1) Create an IP packet with its own IP address (192.168.1.1) as the source, and H2 (192.168.2.2) as the destination

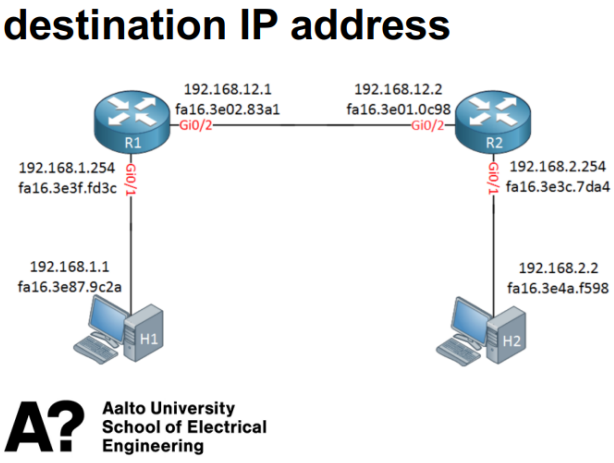
```
C:\Users\H1>arp -a

Interface: 192.168.1.1 --- 0x4
  Internet Address      Physical Address      Type
  192.168.1.254         fa-16-3e-3f-fd-3c     dynamic
  192.168.1.255         ff-ff-ff-ff-ff-ff     static
  224.0.0.22            01-00-5e-00-00-16     static
  224.0.0.251           01-00-5e-00-00-fb     static
  224.0.0.252           01-00-5e-00-00-fc     static
  239.255.255.250       01-00-5e-7f-ff-fa     static
```

2) Check if the destination is local or remote (e.g. check own IP, subnet mask, destination IP)
3) Build an Ethernet frame, enter its own source MAC address, and check the destination MAC address of the default gateway
4) Send an ARP request if MAC address is not known. After that, form an Ethernet frame that carries an IP packet and send it to Router 1

Router 1:

1) Check if FCS (Frame Check Sequence) of the Ethernet frame is correct or not. If FCS is incorrect, the frame is dropped.
2) If FCS is correct, check if the destination MAC address is
• The address of the interface of the router, or
• A broadcast address of the subnet that the router interface is connected to, or
• A multicast address that the router listens to
3) If MAC address matches in previous step, de-encapsulate the IP packet out of the Ethernet frame
4) Check if the header checksum of the IP packet is correct. If not, drop the IP packet.
5) Check its routing table to see if there is a match for the destination IP address




```
R1#show ip route
Codes: L - local, C - connected, S - static, R - RIP, M - mobile, B - BGP
       D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area
       N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
       E1 - OSPF external type 1, E2 - OSPF external type 2
       i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2
       ia - IS-IS inter area, * - candidate default, U - per-user static
route
       o - ODR, P - periodic downloaded static route, H - NHRP, l - LISP
       a - application route
       + - replicated route, % - next hop override, p - overrides from PfR

Gateway of last resort is not set

      192.168.1.0/24 is variably subnetted, 2 subnets, 2 masks
C        192.168.1.0/24 is directly connected, GigabitEthernet0/1
L        192.168.1.254/32 is directly connected, GigabitEthernet0/1
S     192.168.2.0/24 [1/0] via 192.168.12.2
      192.168.12.0/24 is variably subnetted, 2 subnets, 2 masks
C        192.168.12.0/24 is directly connected, GigabitEthernet0/2
L        192.168.12.1/32 is directly connected, GigabitEthernet0/2
```

6) Decrease the TTL field by one and recalculate the header checksum
7) Check its ARP table to see if there is an entry for 192.168.12.2. If not, send an ARP req

```
R1#show ip arp
Protocol  Address          Age (min)  Hardware Addr   Type   Interface
Internet  192.168.1.1            58   fa16.3e87.9c2a  ARPA
GigabitEthernet0/1
Internet  192.168.1.254           -   fa16.3e3f.fd3c  ARPA
GigabitEthernet0/1
Internet  192.168.12.1            -   fa16.3e02.83a1  ARPA
GigabitEthernet0/2
Internet  192.168.12.2           95   fa16.3e01.0c98  ARPA
GigabitEthernet0/2
```

8) Build a new Ethernet frame with its own MAC address of Gi0/2 as source and R2 as the destination. Encapsulate the IP packet into the new Ethernet frame

Router 2:

This Ethernet frame makes it to R2. Like R1 it will first do this:

- Check the FCS of the Ethernet frame.
- De-encapsulates the IP packet, discard the frame.
- Check the IP header checksum.
- Check the destination IP address.

```
R2#show ip arp
Protocol  Address          Age (min)  Hardware Addr   Type   Interface
Internet  192.168.2.2            121   fa16.3e4a.f598  ARPA
GigabitEthernet0/1
Internet  192.168.2.254            -   fa16.3e3c.7da4  ARPA
GigabitEthernet0/1
Internet  192.168.12.1           111   fa16.3e02.83a1  ARPA
GigabitEthernet0/2
Internet  192.168.12.2             -   fa16.3e01.0c98  ARPA
GigabitEthernet0/2
```

```
R2#show ip route
Codes: L - local, C - connected, S - static, R - RIP, M - mobile, B - BGP
       D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area
       N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
       E1 - OSPF external type 1, E2 - OSPF external type 2
       i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2
       ia - IS-IS inter area, * - candidate default, U - per-user static
route
       o - ODR, P - periodic downloaded static route, H - NHRP, l - LISP
       a - application route
       + - replicated route, % - next hop override, p - overrides from PfR

Gateway of last resort is not set

S     192.168.1.0/24 [1/0] via 192.168.12.1
      192.168.2.0/24 is variably subnetted, 2 subnets, 2 masks
C        192.168.2.0/24 is directly connected, GigabitEthernet0/1
L        192.168.2.254/32 is directly connected, GigabitEthernet0/1
      192.168.12.0/24 is variably subnetted, 2 subnets, 2 masks
C        192.168.12.0/24 is directly connected, GigabitEthernet0/2
L        192.168.12.2/32 is directly connected, GigabitEthernet0/2
```

Host 2:

Host 2 receives the Ethernet frame and will:

- Check the FCS
- Find its own MAC address as the destination MAC address.
- De-encapsulates the IP packet from the frame.
- Finds its own IP address as the destination in the IP packet.

Internet Control Message Protocol (ICMP):

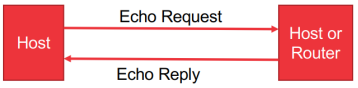
The Internet Control Message Protocol (ICMP) is a supporting protocol in the Internet protocol suite. It is used by network devices, including routers, to send error messages and operational information indicating success or failure when communicating with another IP address, for example, when an error is indicated when a requested service is not available or that a host or router could not be reached.[2] ICMP differs from transport protocols such as TCP and UDP in that it is not typically used to exchange data between systems, nor is it regularly employed by end-user network applications

• Used by network tools like Ping and Traceroute

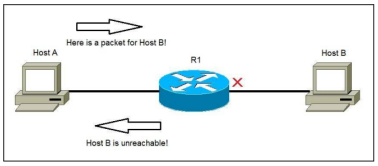
• Host-to-host protocol, used for sending IP layer error and status messages

ICMP messages

ICMP Messages

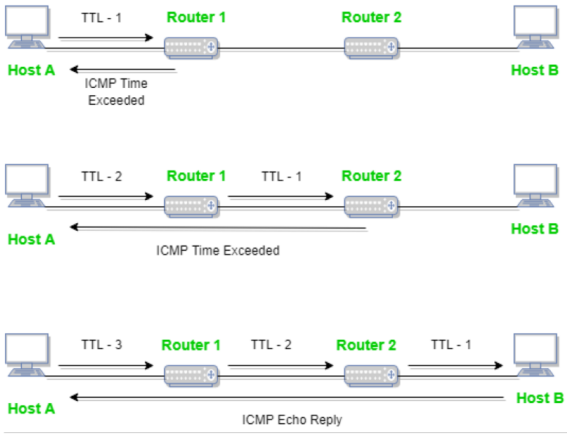


Host A will start the ping utility that will send **ICMP Echo Request** packets to Server. If Server is reachable, it will respond with **ICMP Echo Reply** packets.



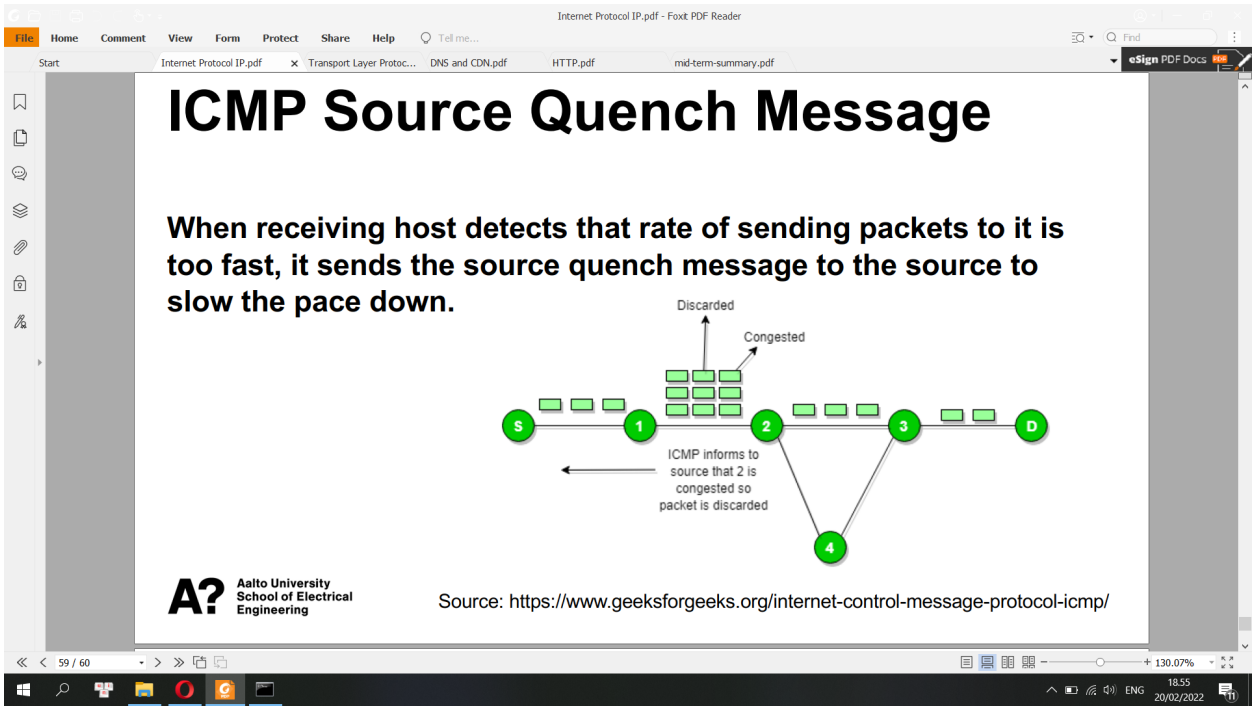
Host A sends a packet to Host B. Because the Host B is down, the router will send an **ICMP Destination host unreachable** message to Host A, informing it that the destination host is unreachable

ICMP Time Exceeded Messages



Aalto University
School of Electrical
Engineering

Source: https://www.geeksforgeeks.org/internet-control-message-protocol-icmp/



Self-test: IP (Internet Protocol)

For example, for the network 192.168.30.0 255.255.255.0, what would be the broadcast address?
The broadcast address for a network has all host bits on.
Answer: 192.168.30.255 will be the broadcast address
Example: IPv4 address 192.128.64.7/24

192.128.64.7 is the IP address and 24 is the subnet mask. The /24 corresponds to the subnet mask 255.255.255.0. The IP address consists of 4 decimals – called octets – which are separated by points. One octet contains 8 bits, which is why IPv4 is a 32-bit address. Each octet can represent a number between 0 and 255. In this case, the whole of the last octet consists of host bits. Therefore, in this example, the broadcast address would be 192.128.64.255 – so all host bits at 1.

What can 255.255.255.255 be used for?

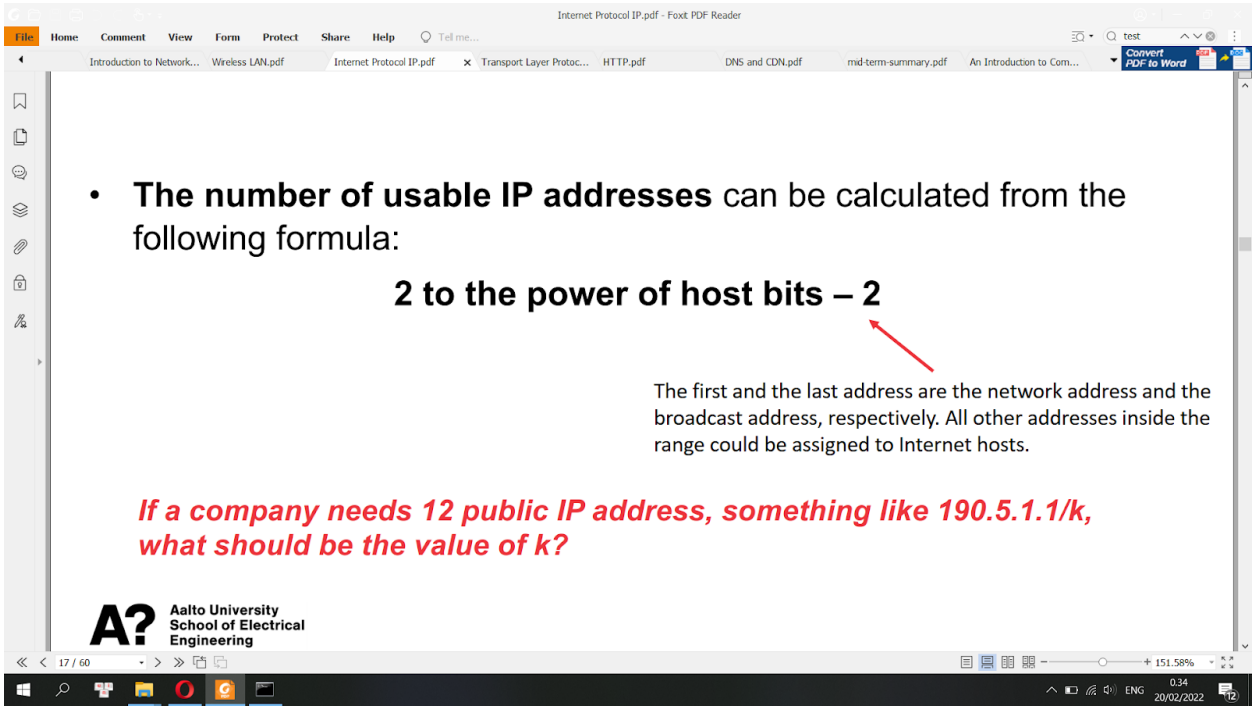
0.0.0.0 – Represents the “default” network, i.e. any connection
255.255.255.255 – Represents the broadcast address, or place to route messages to be sent to every device within a network
127.0.0.1 – Represents “localhost” or the “loopback address”, allowing a device to refer to itself, regardless of what network it is connected to
169.254.X.X – Represents a “self-assigned IP address”, which a device will assign itself if it is unable to receive an IP address from a DHCP server

Which actions could be taken by a router if a specific match is not made to a route in the routing table?

a) The packet will be discarded
b) The packet will be sent back to the source
c) The packet will be flooded out all interfaces
d) Neighbouring routers are polled to find the best path
e) The packet will be forwarded to a default route if one is present

=> answer (e) is correct. If no matching entries can be found, the packet is sent to the defined default gateway. If more than one match is found in the routing table entries, the metric is used and the route with the fewest hops typically is selected.

If a company needs 12 public IP address, something like 190.5.1.1/k, what should be the value of k?



12 public + 2 bit of network and broadcast = 14 < 16 = 2^4
=> needs last four bits => k = 32 - 4 = 28

Transport Control Protocols (TCP)

TCP/UDP Ports

• TCP/UDP extends host-to-host delivery to process-to-process delivery
• Port <-> server process
• Ports are identified for each protocol and address combination by 16-bit unsigned numbers, known as port number.
• One IP address à 65535 TCP Ports and another 65535 UDP Ports
• Socket: <host, port>
• Socket address: <IP address, port number>, e.g. 192.168.0.10:80
• Client must know server's port

• Port numbers 0 – 1023 are well-known ports. These are allocated to server services by the Internet Assigned Numbers Authority (IANA), e.g. web servers normally use port 80 and SMTP servers use port 25.
• Ports 1024-49151 – Registered Port: these can be registered for services with the IANA and should be treated as semi-reserved.
• Ports 49152-65535: these are used by client programs and you are free to use these in client programs.

IP provides unreliable service. It makes its best effort to deliver segments between communicating hosts, but it makes no guarantees.
• It does not guarantee segment delivery
• It does not guarantee orderly delivery of segments
• It does not guarantee the integrity of the data in the segments

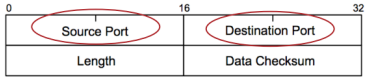
UDP

• UDP is unreliable, in that there is no UDP-layer attempt at timeouts, acknowledgment and retransmission; applications written for UDP must implement these.
• UDP is also unconnected, or stateless; if an application has opened a port on a host, any other host on the Internet may deliver packets to that ⟨host, port⟩ socket without preliminary negotiation.
• UDP packets can be dropped due to queue overflows either at an intervening router or at the receiving host.

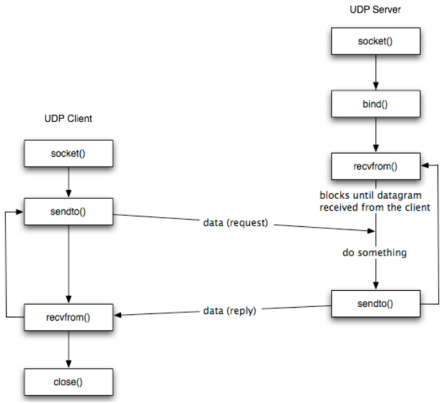
UDP Socket Flow

UDP

- **Support multicast and broadcast**
- **Low overhead**
- **Commonly used for real-time apps**



Socket(): socket creation
Bind(): binds the socket to specified address and port number

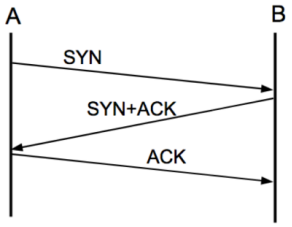


TCP extends IP with the following features
• Connection-orientation
• Stream-orientation
• Reliability
• Throughput management

Connection-oriented and Three-way handshake
A send SYN
B send Syn + Ack
A send Ack

- Before one application process can begin to send data to another, the two processes must "handshake" with each other
- The process that is initiating the connection is called the *client process*, while the other process is called the *server process*
- **Three-way Handshake**

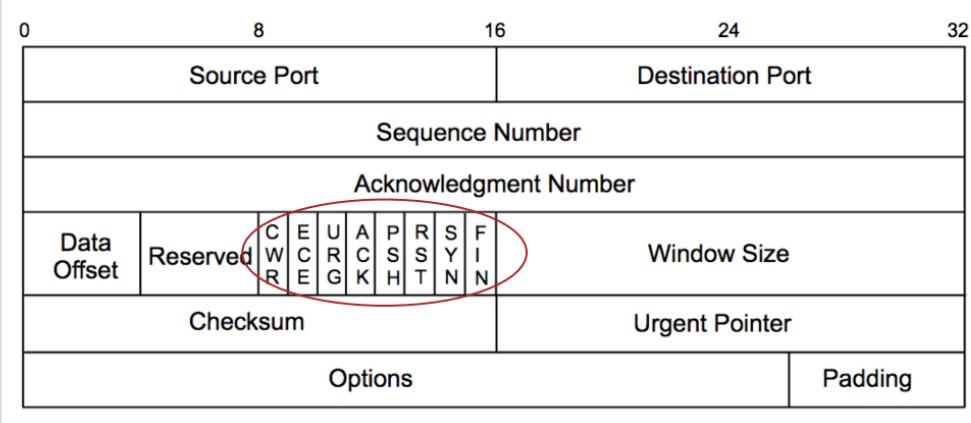
  1) A sends B a packet with the SYN bit set
  2) B responds with a SYN packet of its own; the ACK bit is also set.
  3) A responds to B's SYN with its own ACK.



TCP three-way handshake

Aalto-yliopisto
Sähkötekniikan korkeakoulu

TCP Header



- Source port: this is a 16 bit field that specifies the port number of the sender.
- Destination port: this is a 16 bit field that specifies the port number of the receiver.
- Sequence number: the sequence number is a 32 bit field that indicates how much data is sent during the TCP session. When you establish a new TCP connection (3 way handshake) then the initial sequence number is a random 32 bit value known as ISN (Initial Sequence Numbers). The receiver will use this sequence number and sends back an acknowledgment. Protocol analyzers like wireshark will often use a relative sequence number of 0 since it's easier to read than some high random number.
- Acknowledgment number: this 32 bit field is used by the receiver to request the next TCP segment. This value will be the sequence number incremented by 1.
- Data Offset: this is the 4 bit data offset field, also known as the header length. It indicates the length of the TCP header so that we know where the actual data begins.
- Reserved: these are 3 bits for the reserved field. They are unused and are always set to 0.
- Flags: there are 9 bits for flags, we also call them control bits. We use them to establish connections, send data and terminate connections:

- CWR and ECE: part of the explicit congestion notification mechanism
- URG: urgent pointer. When this bit is set, the data should be treated as priority over other data.
- ACK: used for the acknowledgment.
- PSH: this is the push function. This tells an application that the data should be transmitted immediately and that we don't want to wait to fill the entire TCP segment.
- RST: this resets the connection, when you receive this you have to terminate the connection right away. This is only used when there are unrecoverable errors and it's not a normal way to finish the TCP connection.
- SYN: we use this for the initial three way handshake and it's used to set the initial sequence number.
- FIN: this finish bit is used to end the TCP connection. TCP is full duplex so both parties will have to use the FIN bit to end the connection. This is the normal method how we end an connection.

- Window Size : the 16 bit window field specifies how many bytes the receiver is willing to receive. It is used so the receiver can tell the sender that it would like to receive more data than what it is currently receiving. It does so by specifying the number of bytes beyond the sequence number in the acknowledgment field.

- Checksum: 16 bits are used for a checksum to check if the TCP header is OK or not.
- Urgent pointer: these 16 bits are used when the URG bit has been set, the urgent pointer is used to indicate where the urgent data ends.
- Options: this field is optional and can be anywhere between 0 and 320 bits.

TCP Socket Flow

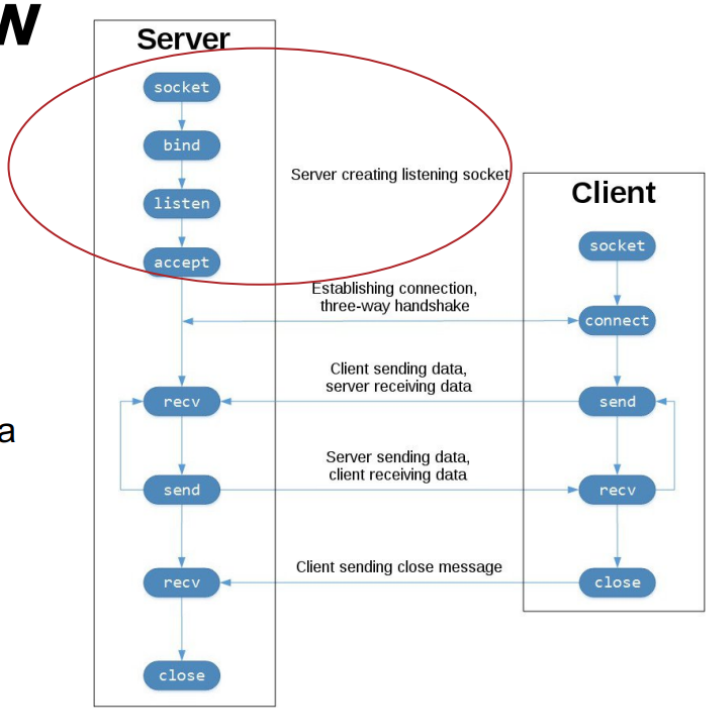
TCP Socket Flow

Socket(): socket creation

Bind(): binds the socket to specified address and port number

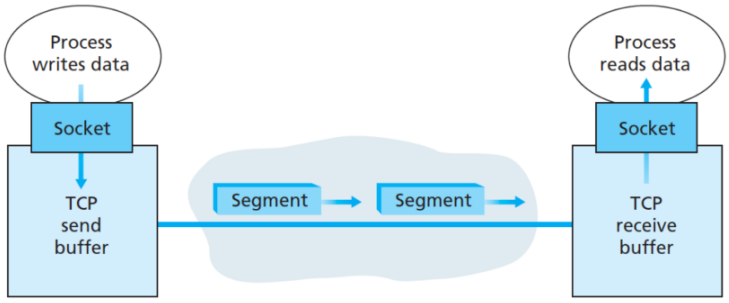
Listen(): puts the server socket in a passive mode, where it waits for the client to approach the server to make a connection

Accept(): extracts the first connection request on the queue of pending connections for the listening socket, creates a new connected socket



TCP Connection:
<Client IP, Client TCP Port, Server IP, Server TCP Port>
• Client port numbers are dynamically assigned, and can be reused once the session is closed.
• There can be multiple TCP connections between two hosts. These connections should use different TCP Ports.
• TCP protocol runs only in the end systems and not in the intermediate network elements (routers and link-layer switches), the intermediate network elements do not maintain TCP connection state
Data Exchange via TCP
• A TCP connection provides a full-duplex service
• A TCP connection is always point-to-point between a single sender and a single receiver

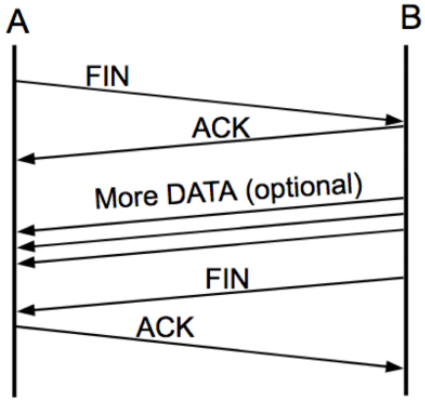
How to determine Maximum Segment Size (MSS)?
Do you remember what Maximum Transmission Unit (MTU) is?



How to close a TCP connection?

Two Two-way FIN/ACK Handshake

- A sends B a packet with the FIN bit set
- B sends A an ACK of the FIN
- B may continue to send additional data to A
- When B is also ready to cease sending, it sends its own FIN to A
- A sends B an ACK of the FIN; this is the final packet in the exchange

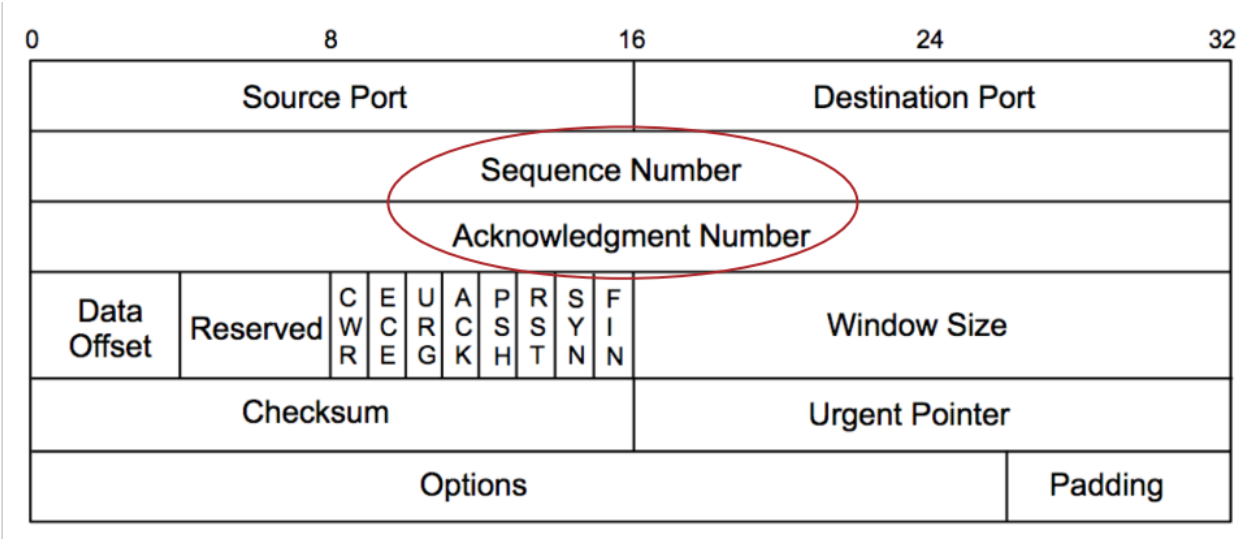


A typical TCP close

A: FIN
B: ACK
B: may send more data
B: FIN
A: ACK

Reliability

• TCP numbers each packet, and keeps track of which are lost and retransmits them after a timeout.



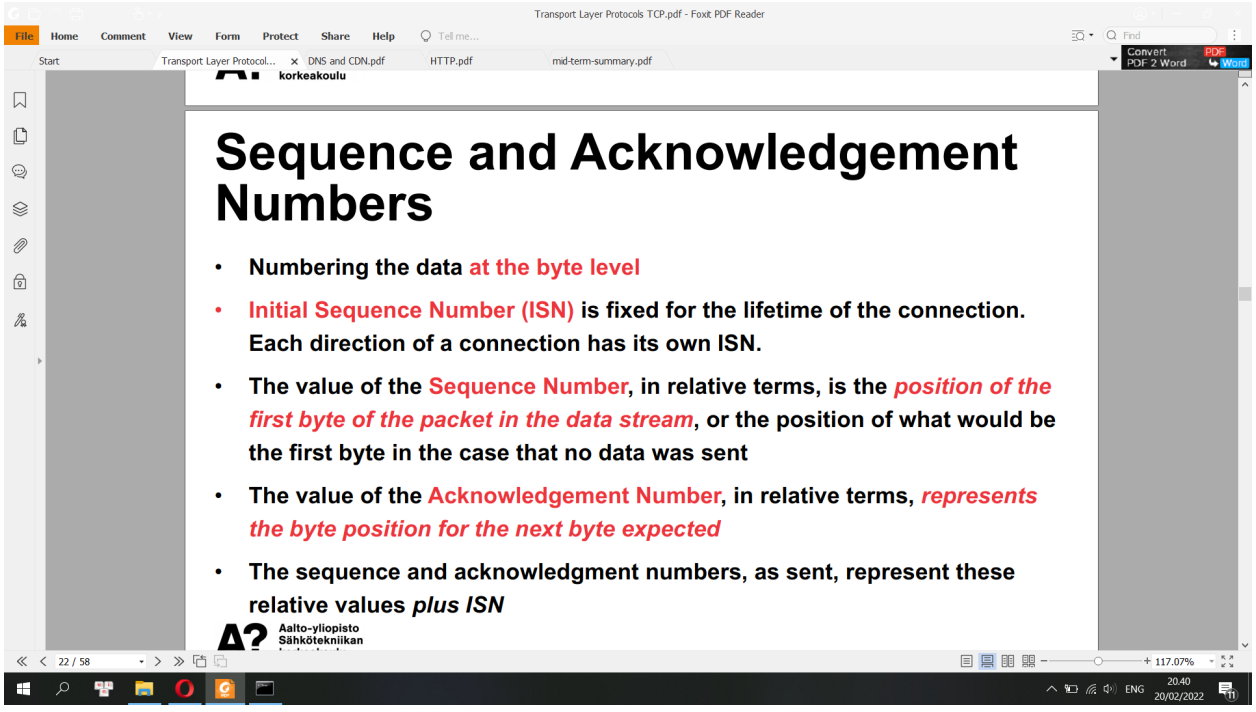
Sequence number:
every segment has a sequence number. But in the above examples, we can see that some packets don't have sequence numbers. That is because they are ack segments.

Note that the ACK segment does not consume any sequence numbers if it does not carry data. An ACK segment, if carrying no data, consumes no sequence number.

the TCP sequence number is 32 bits long
the most significant byte of the number is sent first
TCP sequence numbers count bytes rather than packets
the sequence number in the header is the sequence number of the first byte in the data
if there is no data, the sequence number is still set to the sequence number of the next byte that could be sent
since a TCP connection is bidirectional, a different initial sequence number (ISN) is used in each direction: each peer picks the ISN it will use in sending data

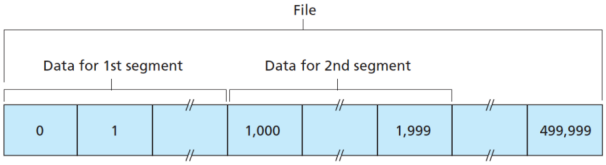
All bytes in a TCP connection are numbered, beginning at a randomly chosen initial sequence number (ISN). The SYN packets consume one sequence number, so actual data will begin at ISN+1. The sequence number is the byte number of the first byte of data in the TCP packet sent (also called a TCP segment). The acknowledgement number is the sequence number of the next byte the receiver expects to receive. The receiver ack'ing sequence number x acknowledges receipt of all data bytes less than (but not including) byte number x.

The sequence number is always valid. The acknowledgement number is only valid when the ACK flag is one. The only time the ACK flag is not set, that is, the only time there is not a valid acknowledgement number in the TCP header, is during the first packet of connection set-up.



Sequence Number

Example: MSS = 1000 Bytes, File size = 500,000 Bytes



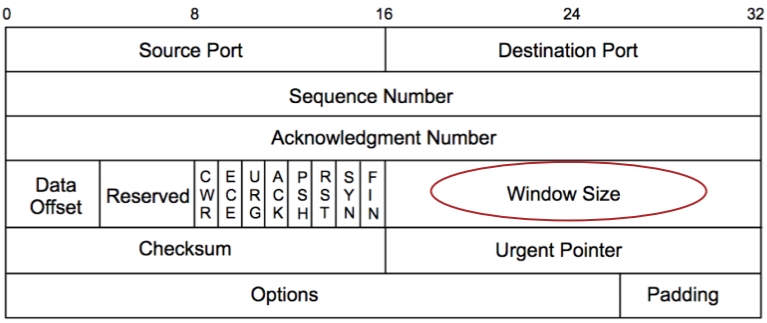
If ISN = 0, sequence number of each segment = 0, 1000, 2000, …

• Both sides of a TCP connection randomly choose an ISN. It may be any value between 0 and 4,294,967,295, inclusive.
• Protocol analyzers like Wireshark will typically display relative sequence and acknowledgement numbers in place of the actual values

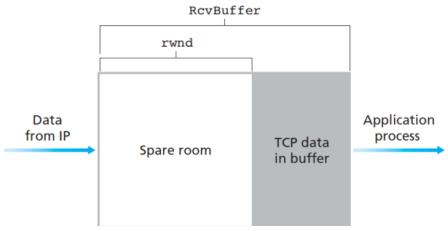
TCP provides cumulative acknowledgements (i.e. TCP only acknowledges bytes up to the first missing byte in the stream)

TCP Timeout and Retransmission
• When TCP sends a segment containing user data (this excludes ACK-only packets), it sets a timer. If that timer expires before the segment data is acknowledged, the segment is retransmitted.
• The length of timeout interval is adapted to RTT
• Acknowledgements are sent for every arriving data packet (unless Delayed ACKs are implemented)
In telecommunications, round-trip delay (RTD) or round-trip time (RTT) is the amount of time it takes for a signal to be sent plus the amount of time it takes for acknowledgement of that signal having been received. This time delay includes propagation times for the paths between the two communication endpoints.[1] In the context of computer networks, the signal is typically a data packet. RTT is also known as ping time, and can be determined with the ping command.

TCP Flow Control
• Window size indicates the number of bytes that a receiver is willing to accept.
• Flow-control service is used for eliminating the possibility of the sender overflowing the receiver's buffer



- Suppose that Host A is sending a large file to Host B over a TCP connection. Host B allocates a receive buffer to this connection; denote its size by ***RcvBuffer*** .
- ***LastByteRecv***: the number of the last byte in the data stream that has arrived from the network and has been placed in the receive buffer at B
- ***LastByteRead***: the number of the last byte in the data stream read from the buffer by the app process in B



Aalto-yliopisto
Sähkötekniikan korkeakoulu

Receive Window

- **Sender maintains a variable called the receive window**
- TCP is not permitted to overflow the allocated buffer

$$LastByteRecv - LastByteRead \leq RcvBuffer$$

- The receive window rwnd is set to the amount of spare room in the buffer:

$$rwnd = RcvBuffer - [LastByteRecv - LastByteRead]$$

Because TCP is full-duplex, the sender at each side of the connection maintains a distinct receive window.

TCP is not permitted to overflow the allocated buffer

$LastByteRecv - LastByteRead \leq RcvBuffer$

The receive window rwnd is set to the amount of spare room in the buffer:

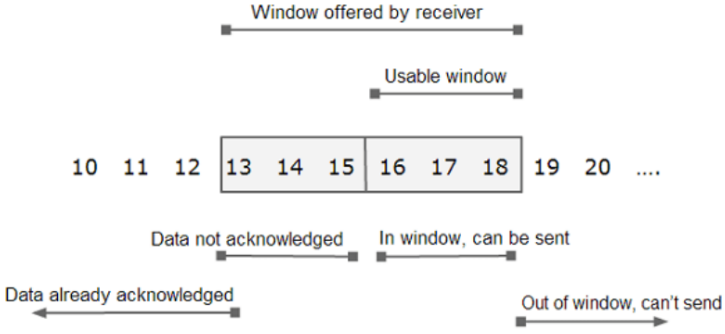
$rwnd = RcvBuffer - [LastByteRecv - LastByteRead]$

How does the connection use the variable rwnd to provide the flowcontrol service?
Sliding Window

Sliding Window

- Host B tells Host A how much spare room it has in the connection buffer by placing its current value of *rwnd* in the receive window field of every segment it sends to A. Initially, Host B sets rwnd = RcvBuffer.

This window slides towards right depending upon how fast receiver consumes data and sends acknowledgement and hence known as sliding window.



TCP Window Scale Option
• TCP window scale option is needed for efficient transfer of data when the bandwidth-delay product is greater than 64KB.
E.g. 1.5Mbps x 0.513s = 96 KB
• Effective window size = receive window size x window scale

Congestion control

• End-to-end congestion control, since IP layer provides no explicit feedback to end systems regarding congestion
• Each sender limits the rate at which it sends traffic into its connection as a function of perceived network congestion
If a TCP sender perceives that there is little congestion on the path between itself and the destination, then the TCP sender increases its send rate;
If the sender perceives that there is congestion along the path, then the sender reduces its send rate

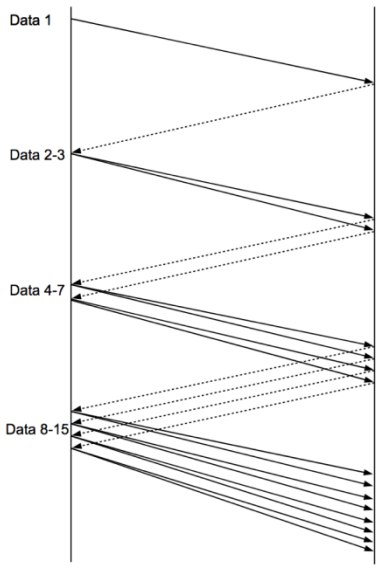
Congestion Window
• Sender maintains a variable called congestion window, denoted cwnd.
• The amount of unacknowledged data at a sender may not exceed the minimum of cwnd and rwnd
LastByteSent – LastByteAcked ≤ min{cwnd, rwnd}
• A lost segment implies congestion, and hence, the TCP sender's rate should be decreased when a segment is lost.
• Loss events: a timeout event, receipt of duplicated acks for a given segment
• TCP uses acknowledges to trigger (or clock) its increase in congestion window size

TCP Congestion-Control Algorithm
Three Major Components:
• Slow Start
• Congestion Avoidance
• Fast recovery (recommended for TCP senders, but not required)

Slow start

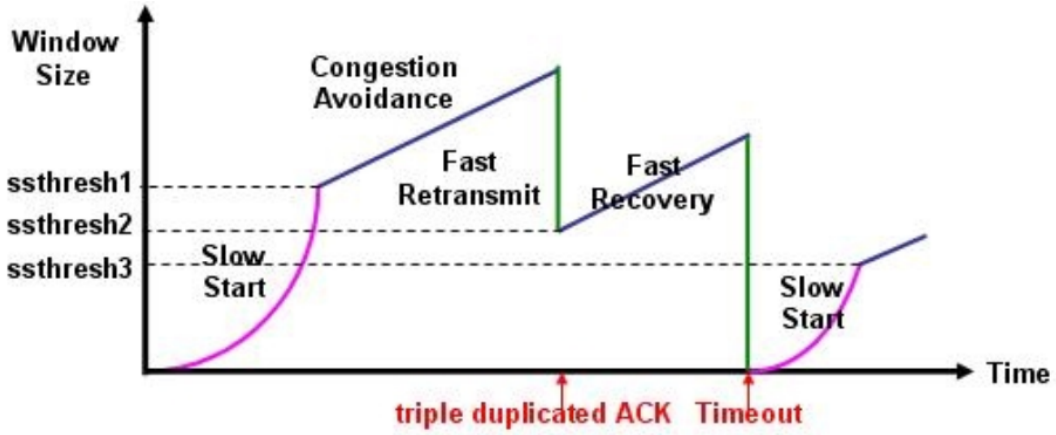
Slow Start

- **Set initial cwnd = 1**
- **cwnd = cwnd x 2 after receiving cwnd ACKs**
- **The TCP send rate starts slow but grows exponentially during the slow start phase**



Slow Start with discrete packet flights

When should the exponential growth end?
• If there is a loss event indicated by a timeout, the TCP sender sets the value of cwnd to 1 and begins the slow start anew. Also set ssthresh (‘slow start threshold’) to cwnd/2
When cwnd reaches or surpasses ssthresh, slow start ends and TCP transitions into congestion avoidance mode

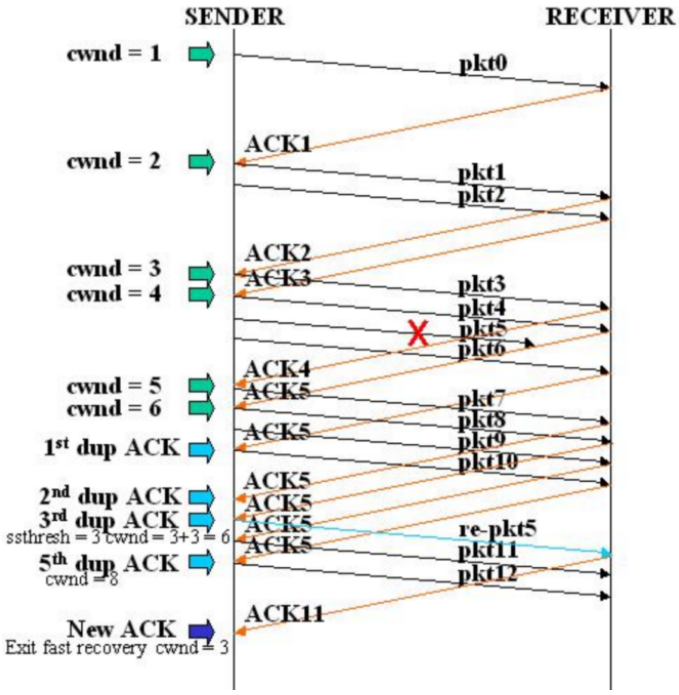


Congestion Avoidance
• cwnd = cwnd + 1 MSS whenever a new ACK arrives
• If congestion was indicated by a timeout, cwnd is reset to 1 segment, which automatically puts the sender into slow start mode
• If congestion was indicated by duplicate Acknowledgements the fast retransmit and fast recovery algorithms are invoked
Fast Recovery

Fast Recovery

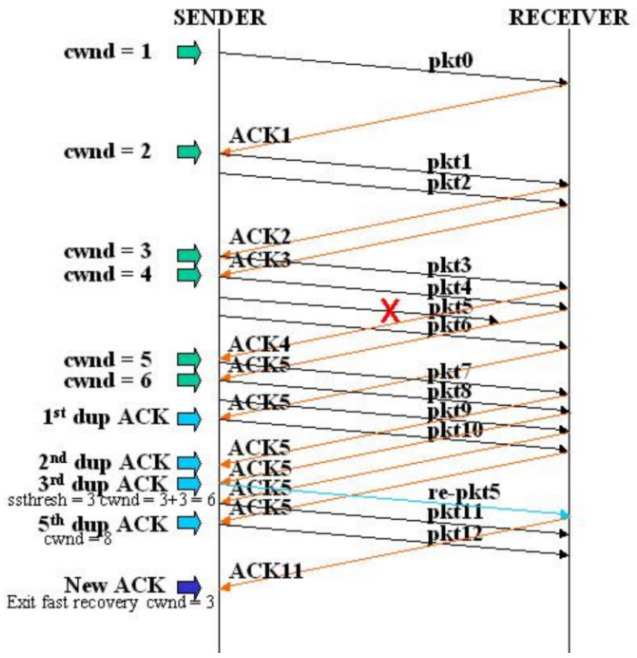
After receiving 3 duplicate ACKs in a row:

1) Set *ssthresh* = cwnd/2.
2) Retransmit the missing segment.
3) Set cwnd = ssthresh + 3.
4) Each time another duplicate ACK arrives, set cwnd = cwnd + 1. Then, send a new data segment if allowed by the value of cwnd.



Fast Recovery

5) Once receive a new ACK (an ACK which acknowledges all intermediate segments sent between the lost packet and the receipt of the first duplicate ACK), exit fast recovery. This causes setting *cwnd* to *ssthresh* (the ssthresh in step 1). Then, continue with linear increasing due to congestion avoidance algorithm.



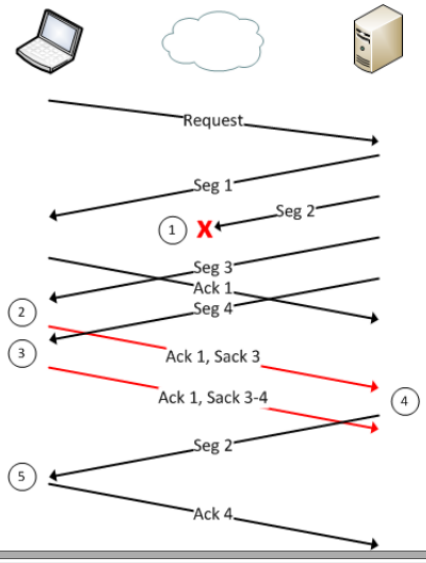
Selective Acknowledge (SACK)

Selective Acknowledge (SACK)

- **Selective ACK** (SACK) option is implemented at the receiver.
- The sender does not have to guess from dupACKs what has gotten through.

The receiver can send an ACK that says:

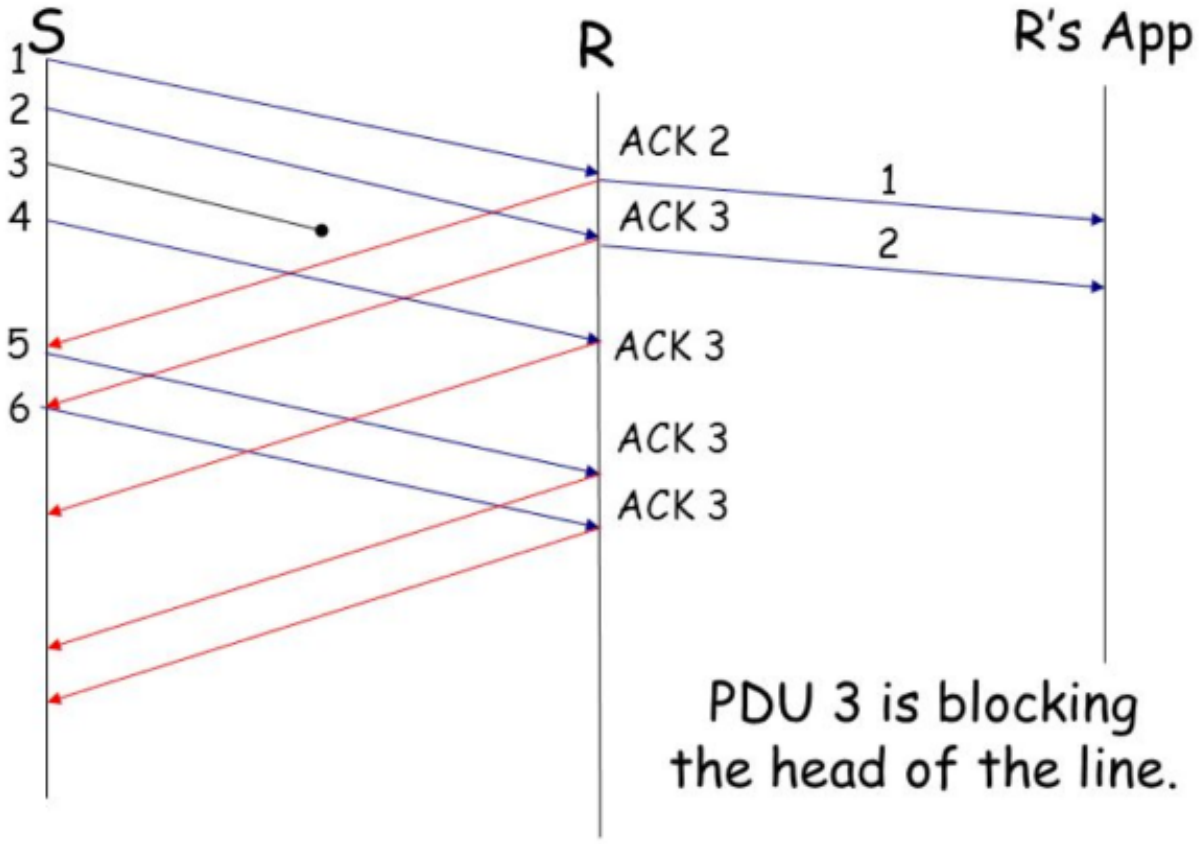
- All packets up through N have been received (the cumulative ACK)
- *All packets up through M have been received except for x, y, z.*



Source: http://packetlife.net/blog/2010/jun/17/tcp-selective-acknowledgments-sack/

Head-of-Line Blocking in TCP

Head-of-line blocking (HOL blocking) in computer networking is a performance-limiting phenomenon that occurs when a line of packets is held up by the first packet. Examples include input buffered network switches, out-of-order delivery and multiple requests in HTTP pipelining.



In telecommunications, a protocol data unit (PDU) is a single unit of information transmitted among peer entities of a computer network. A PDU is composed of protocol-specific control information and user data. In the layered architectures of communication protocol stacks, each layer implements protocols tailored to the specific type or mode of data exchange.

For example, the Transmission Control Protocol (TCP) implements a connection-oriented transfer mode, and the PDU of this protocol is called a segment, while the User Datagram Protocol (UDP) uses datagrams as protocol data units for connectionless communication.

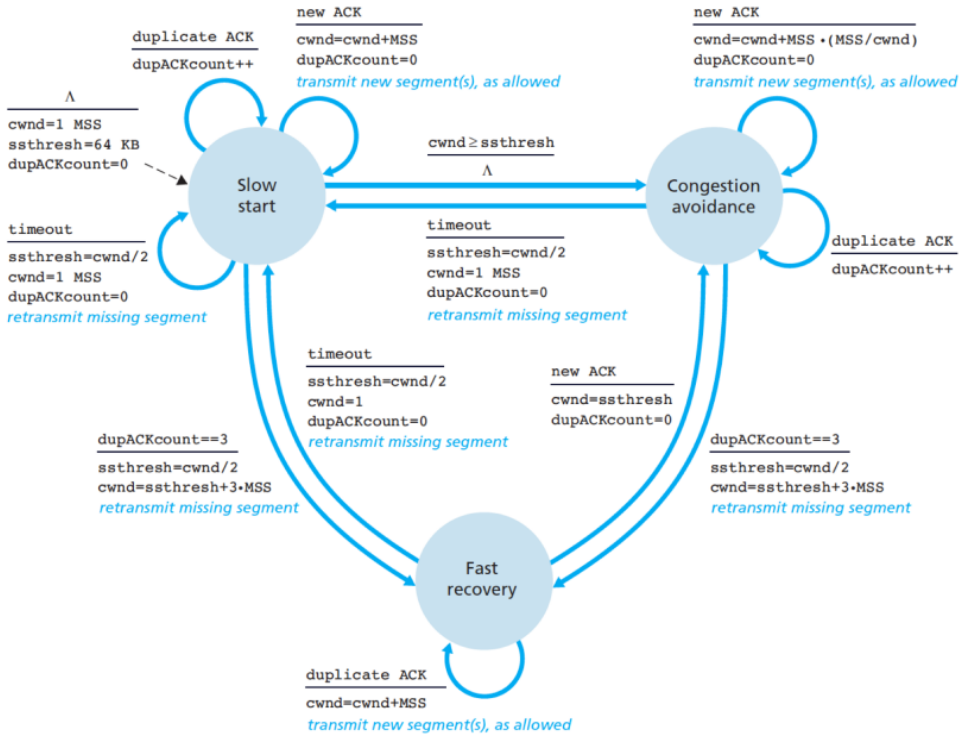
TCP vs UDP

TCP	UDP
Keeps track of lost packets. Makes sure that lost packets are re-sent	Doesn't keep track of lost packets
Adds sequence numbers to packets and reorders any packets that arrive in the wrong order	Doesn't care about packet arrival order
Slower, because of all added additional functionality	Faster, because it lacks any extra features
Does not support multicast or broadcast	Support multicast and broadcast
Example services that use TCP: - HTTP - FTP	Example services that use UDP: - VoIP - DHCP

Types of Sockets

• Datagram sockets: connectionless sockets which use UDP
• Stream sockets: connection-oriented sockets which use TCP, SCTP or DCCP
• Raw sockets: the transport layer is bypassed and the packet headers are made accessible to the application. There is no port number in address, only IP address

State machine for congestion control algorithm



Self-test: TCP

1) Suppose that Host A has received all bytes numbered 0 through 535 from B and suppose that it is about to send a segment to Host B. What should be the acknowledge number of the segment it sends to B?

In other words, host A is waiting for byte 536 and all the subsequent bytes in host B's data stream. So host A puts 536 in the acknowledgment number field of the segment it sends to B.

2) Suppose that Host A has received one segment from Host B containing bytes 0 through 535 and another segment containing bytes 900 through 1,000. For some reason Host A has not received bytes 536 through 899. In the next segment sent from A to B, what should be the value of the acknowledge number field?

In this example, host A is still waiting for byte 536 (and beyond) in order to recreate B's data stream. Thus, A's next segment to B will contain 536 in the acknowledgment number field. Because TCP only acknowledges bytes up to the first missing byte in the stream, TCP is said to provide cumulative acknowledgements.

Host A received the third segment (bytes 900 through 1,000) before receiving the second segment (bytes 536 through 899). Thus, the third segment arrived out of order. The subtle issue is: What does a host do when it receives out-of-order segments in a TCP connection?

Interestingly, the TCP RFCs do not impose any rules here, and leave the decision up to the people programming a TCP implementation. There are basically two choices: either (i) the receiver immediately discards out-of-order bytes; or (ii) the receiver keeps the out-of-order bytes and waits for the missing bytes to fill in the gaps. Clearly, the latter choice is more efficient in terms of network bandwidth, whereas the former choice significantly simplifies the TCP code. However, we focus on the former implementation, that is, we assume that the TCP receiver discards out-of-order segments.

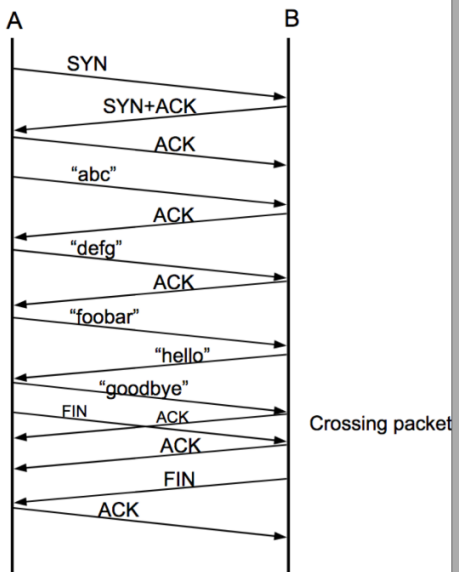
3) Assume that ISN = 0 on both sides. In terms of the sequence and acknowledge numbers, SYNs count as 1 byte, as do FINs. Can you calculate the seg and ack values of each segment?

Exercise

In terms of the sequence and acknowledgment numbers, SYNs count as 1 byte, as do FINs.

Assume that ISN = 0 on both sides

Can you calculate the seq and ack values of each segment?



	A sends	B sends
1	SYN, **seq=0**	
2		SYN+ACK, seq=0, **ack=1** (expecting)
3	ACK, **seq=1**, ack=1 (ACK of SYN)	
4	"abc", **seq=1**, ack=1	
5		ACK, seq=1, **ack=4**
6	"defg", **seq=4**, ack=1	
7		seq=1, **ack=8**
8	"foobar", **seq=8**, ack=1	
9		seq=1, **ack=14**, "hello"
10	**seq=14**, ack=6, "goodbye"	
11,12	**seq=21**, ack=6, FIN	seq=6, **ack=21** ;; ACK of "goodbye", crossing packets
13		seq=6, **ack=22** ;; ACK of FIN
14		seq=6, **ack=22**, FIN
15	**seq=22**, ack=7 ;; ACK of FIN	

What if ISN is not 0

	A, ISN=1000	B, ISN=7000
1	SYN, **seq=1000**	
2		SYN+ACK, seq=7000, **ack=1001**
3	ACK, **seq=1001**, ack=7001	
4	"abc", **seq=1001**, ack=7001	
5		ACK, seq=7001, **ack=1004**
6	"defg", **seq=1004**, ack=7001	
7		seq=7001, **ack=1008**
8	"foobar", **seq=1008**, ack=7001	
9		seq=7001, **ack=1014**, "hello"
10	**seq=1014**, ack=7006, "goodbye"	

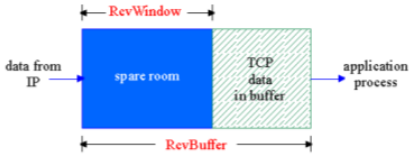
4) What is receive window rwnd? How does the connection use the variable rwnd to provide the flow control service?

The Receiver Window (rwnd) is a variable that advertises the amount of data that the destination side can receive. It is the window size part in TCP header

eecs489 COMPUTER NETWORKS

Lecture 30:
Flow Control, Reliable Delivery

The receiver side of a TCP connection maintains a receiver buffer:

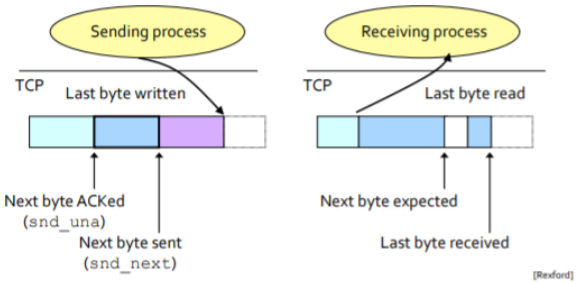


application process may be slow at reading from the buffer

Flow control ensures that sender won't overflow receiver's buffer by transmitting too much, too fast

Sliding Window

TCP uses sliding window flow control: allows a larger amount of data "in flight" than has been acknowledged
- allows sender to get ahead of the receiver
- but not too far ahead

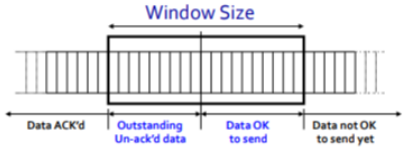


TCP Receiver Window

Receiver window size (`rwnd`)
- amount that can be sent without acknowledgment
- receiver can buffer this amount of data

Receiver continually advertises buffer space available to sender by including the current value of `rwnd` in TCP header

Sender limits unACKed data to `rwnd`
⇒ guarantees receiver buffer wouldn't overflow



5) What is congestion window cwnd? How to tune the value of cwnd in slow start and congestion avoidance modes, respectively? When should the connection transition from slow start mode into congestion avoidance mode, and vice versa?
What would happen if senders collectively send too fast or too slowly?

Congestion Window (cwnd) is a TCP state variable that limits the amount of data the TCP can send into the network before receiving an ACK.

- How to tune the value of cwnd in slow start and congestion avoidance modes:

- In slow start:
+ Set initial cwnd = 1
+ cwnd = cwnd x 2 after receiving cwnd ACKs
+ The TCP send rate starts slow but grows exponentially during the slow start phase
- In congestion avoidance mode:
+ cwnd = cwnd + 1 MSS whenever a new ACK arrives
+ If congestion was indicated by a timeout, cwnd is reset to 1 segment, which automatically puts the sender into slow start mode

+ If congestion was indicated by duplicate Acknowledgements the fast retransmit and fast recovery algorithms are invoked.
- Fast recovery mode:
After receiving 3 duplicate ACKs in a row:
1) Set ssthresh = cwnd/2.
2) Retransmit the missing segment.
3) Set cwnd = ssthresh + 3.
4) Each time another duplicate ACK arrives, set cwnd = cwnd + 1. Then, send a new data segment if allowed by the value of cwnd.
5) Once receive a new ACK (an ACK which acknowledges all intermediate segments sent between the lost packet and the receipt of the first duplicate ACK), exit fast recovery. This causes setting cwnd to ssthresh (the ssthresh in step 1). Then, continue with linear increasing due to congestion avoidance algorithm.

- When should the connection transition from slow start mode into congestion avoidance mode, and vice versa:

- When cwnd reaches or surpasses ssthresh, slow start ends and TCP transitions into congestion avoidance mode
- If congestion was indicated by a timeout, cwnd is reset to 1 segment, which automatically puts the sender into slow start mode

6) What are the limitations of TCP?

TCP and UDP limitations

- TCP limitations:
  - TCP keeps strict order: head-of-line blocking may be a problem (data flow blocked until recovering a lost segment)
  - Byte-oriented nature of TCP: must use PSH to ensure data goes to app
  - No multi-home IP hosts
  - Relatively vulnerable to some attacks (SYN flooding)
- UDP limitations
  - Not reliable
  - No data order
  - No congestion control
- Solution: SCTP

How does the connection use the variable rwnd to provide the flow control service?
It tells the sender the maximum speed at which the data can be sent to the receiver device. The sender adjusts the speed as per the receiver's capacity to reduce the frame loss from the receiver side

What would happen if senders collectively send too fast or too slowly?

Overview of Flow control and congestion control
I. Flow Control in TCP
Flow control deals with the amount of data sent to the receiver side without receiving any acknowledgment. It makes sure that the receiver will not be overwhelmed with data. It's a kind of speed synchronization process between the sender and the receiver. The data link layer in the OSI model is responsible for facilitating flow control.

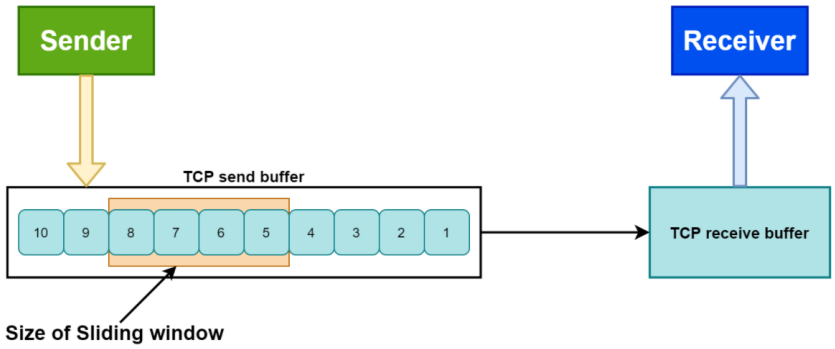
Let's take a practical example to understand flow control. Jack is attending a training session. Let's assume he's slow in grasping concepts taught by the teacher. On the other hand, the teacher is teaching very fast without taking any acknowledgment from the students.

After some time, every word that comes out from his teacher is overflowing over Jack's head. Hence, he doesn't understand anything. Here, the teacher should be having information about how many concepts a student can handle at a time.

After some time, Jack requested the teacher to slow down the pace as he was overwhelmed with the data. The teacher decided to teach some of the concepts first and then wait for acknowledgment from the students before proceeding to the following concepts.

Similar to the example, the flow control mechanism tells the sender the maximum speed at which the data can be sent to the receiver device. The sender adjusts the speed as per the receiver's capacity to reduce the frame loss from the receiver side.

Flow control in TCP ensures that it'll not send more data to the receiver in the case when the receiver buffer is already completely filled. The receiver buffer indicates the capacity of the receiver. The receiver won't be able to process the data when the receiver buffer is full:

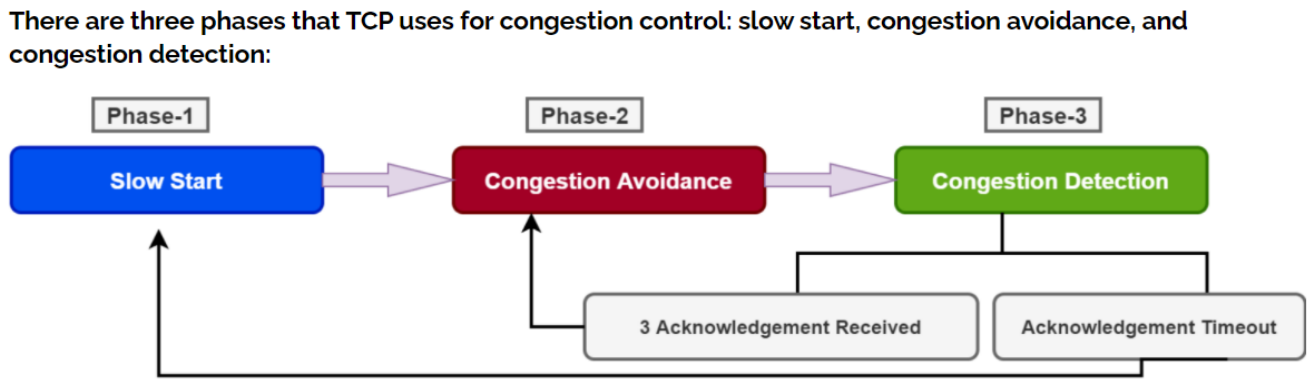


II Congestion Control

Congestion control is a mechanism that limits the flow of packets at each node of the network. Thus, it prevents any node from being overloaded with excessive packets. Congestion occurs when the rate of flow of packets towards any node is more than its handling capacity.

When congestion occurs, it slows down the network response time. As a result, the performance of the network decreases. Congestion occurs because routers and switches have a queue buffer, holding the packets for processing. After the holding process completes, they pass the packet to the next node, resulting in congestion in the network.

There are three phases that TCP uses for congestion control: slow start, congestion avoidance, and congestion detection:



TCP Congestion Control

1) How does a TCP sender limit the rate at which it sends traffic into its connection?: The congestion window indicates the maximum amount of data that can be sent out on a connection without being acknowledged. If a TCP sender perceives that there is little congestion on the path between itself and the destination, then the TCP sender increases its send rate; if the sender perceives that there is congestion along the path, then the sender reduces its send rate.

2) How does a TCP sender perceive that there is congestion on the path between itself and the destination?: TCP uses a congestion window in the sender side to do congestion avoidance. The congestion window indicates the maximum amount of data that can be sent out on a connection without being acknowledged. TCP detects congestion when it fails to receive an acknowledgement for a packet within the estimated timeout.

3) What algorithm should the sender use to change its send rate as a function of perceived end-to-end congestion?: slow start, congestion avoidance and fast recovery algorithms

How to determine Maximum Segment Size (MSS)?

In order to notify MSS to the other end, an inter-layer communication is done as follows:

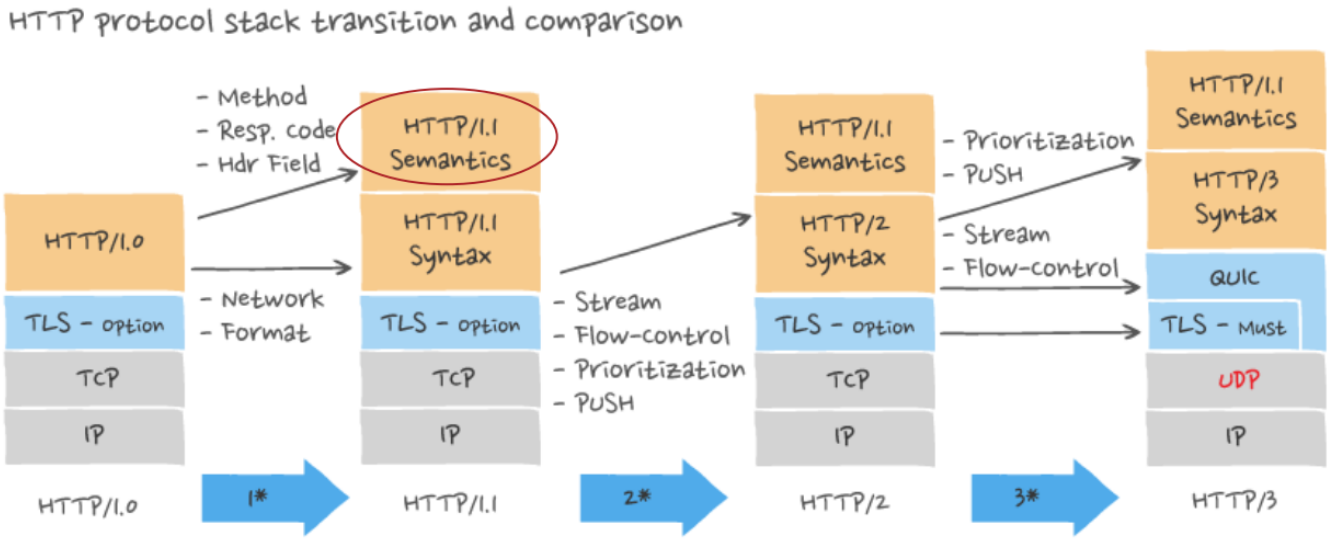
- The Network Driver (ND) or interface should know the Maximum Transmission Unit (MTU) of the directly attached network.
- The IP should ask the Network Driver for the Maximum Transmission Unit.
- The TCP should ask the IP for the Maximum Datagram Data Size (MDDS). This is the MTU minus the IP header length (MDDS = MTU - IPHdrLen).
- When opening a connection, TCP can send an MSS option with the value equal to: MDDS - TCPHdrLen. In other words, the MSS value to send is: MSS = MTU - TCPHdrLen - IPHdrLen

Do you remember what Maximum Transmission Unit (MTU) is?

A maximum transmission unit (MTU) is the largest packet or frame size, specified in octets (eight-bit bytes) that can be sent in a packet- or frame-based network such as the internet. The internet's transmission control protocol (TCP) uses the MTU to determine the maximum size of each packet in any transmission.

(Hypertext Transfer Protocol) HTTP

HTTP is an application layer protocol for distributed, collaborative, hypermedia information system.



Transport Layer Security (TLS), the successor of the now-deprecated Secure Sockets Layer (SSL), is a cryptographic protocol designed to provide communications security over a computer network. The protocol is widely used in applications such as email, instant messaging, and voice over IP, but its use in securing HTTPS remains the most publicly visible.

The TLS protocol aims primarily to provide cryptography, including privacy (confidentiality), integrity, and authenticity through the use of certificates, between two or more communicating computer applications. It runs in the application layer and is itself composed of two layers: the TLS record and the TLS handshake protocols.

The familiar Hypertext Transfer Protocol (HTTP) is most often associated with the World Wide Web. This is one of the transfer protocols in use on the internet. Others include File Transfer Protocol (FTP) and Simple Mail Transfer Protocol (SMTP). HTTP is by far the most used transport protocol for web services, and it plays a crucial role in REST.

The HTTP specification describes messages that represent requests from a client to a server and responses from a server to a client. A message from a client to a server indicates a method (i.e., an action) that is desired for a specific resource designated by a URI, (Universal Resource Identifier). URIs are simply formatted strings which identify by name, location or some other characteristic, a resource located on the internet.

HTTP methods include GET, PUT, POST, DELETE, HEAD, TRACE, and CONNECT. To create web services using BIS, only the first four methods are important.

HTTP Semantics

“This document defines the semantics of HTTP/1.1 messages, as expressed by request methods, request header fields, response status codes, and response header fields, along with the payload of messages (metadata and body content) and mechanisms for content negotiation.”

Resources

• Resources are identified with Uniform Resource Identifiers (URIs)
• A URI is a sequence of characters that identifies a logical or physical resource.
• Multiple URIs may refer to same resource
• Separate from their representation(s)

URI

Two types of URIs:
• Uniform Resource Locators (URLs) e.g., http://www.google.com
• Uniform Resource Names (URNs)
• URN does not state which protocol should be used to locate and access the resource. It labels the resource with a persistent, location-independent unique identifier. The generic form of any URI is scheme:[//[user:password@]host[:port]][/]path[?query][#fragment]

REST Resource Naming Guide

• A resource can be a singleton or a collection
/customers /customers/{customerId}
• A resource may contain sub-collection resources
/customers/{customerId}/accounts/{accountId}

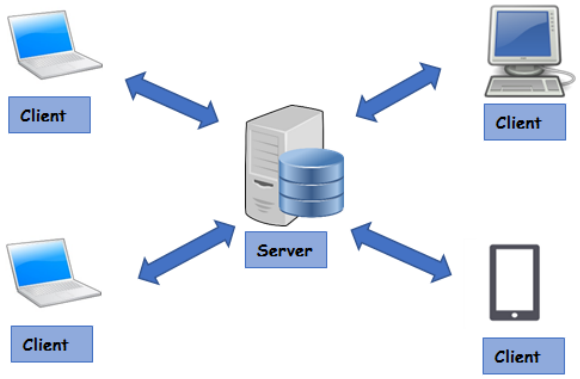
REST (Representational State Transfer)

REST is an architectural style with six constraints:
• Client-server architecture
• Uniform interface
• Stateless
• Layered System
• Cacheable
• Code on Demand (optional)

Representation

• A "representation" is information that is intended to reflect a past, current, or desired state of a given resource, in a format that can be readily communicated via the protocol, and that consists of a set of representation metadata and a potentially unbounded stream of representation data.
• Example:
• Resource: Person
Service: contact information (GET)
Representation:
+ Name, address, phone number
+ JSON or XML format

Client-server architecture



Uniform Interface
• Defines the interface between client and server
• Simplifies and decouples the architecture
• For us this means:
+ HTTP verbs
+ URIs (resource name)
+ HTTP response

HTTP is stateless
• Server contains no client state
• Each request contains all of the information necessary to understand the request, and all session state data should then be returned to the client at the end of each request
• Server will not store anything about latest HTTP request client made. It will treat each and every request as new.
• Statelessness ensures that each request can be treated independently

Layered System
• Client can't assume direct connection to server
• Allow software or hardware intermediaries (e.g. proxies, gateways, and firewalls) between client and server
• Improves scalability

Cacheable
• Server responses (representations) are cacheable
• Services must be designed to produce accurate cache control metadata and return it in response messages.
• Response messages are marked as cacheable or non-cacheable, either with explicit message metadata or as part of the contract definition.
• An optional consumer-side or intermediary cache repository enables the consumer to reuse cacheable response data for later request messages.
• Request messages must be comparable to determine whether or not they are equivalent.

Code on demand (the only optional constraint)

• REST allows client functionality to be extended by downloading and executing code in the form of applets or scripts

• Allow logic within clients (such as Web browsers) to be updated independently from server-side logic

• For example: Jave applets, JavaScript

HTTP Request Methods

Method	Description
GET	Transfer a current representation of the target resource
HEAD	Same as GET, but only transfer the status line and header section
POST	Perform resource-specific processing on the request payload
PUT	Replace all current representations of the target resource with the request payload
DELETE	Remove all current representations of the target resource
CONNECT	Establish a tunnel to the server identified by the target resource
OPTIONS	Describe the communication options for the target resource
TRACE	Perform a message loop-back test along the path to the target resource

Request Header Fields

“A client sends request header fields to provide more information about the request context, make the request conditional based on the target resource state, suggest preferred formats for the response, supply authentication credentials, or modify the expected request processing.”

-- RFC7231

• Controls (e.g. Cache-Control)

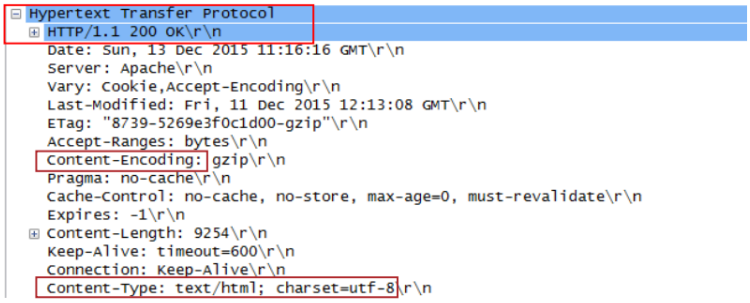
• Conditionals (e.g. if-match)

• Content Negotiation(e.g. Accept, Accept-Encoding)

• Authentication Credentials

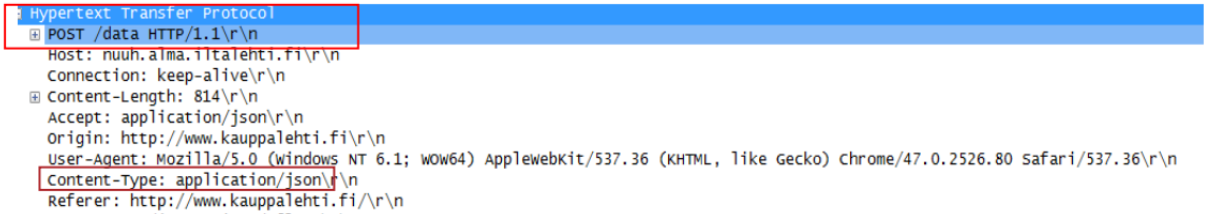
• Request Context (e.g. from, user-agent)

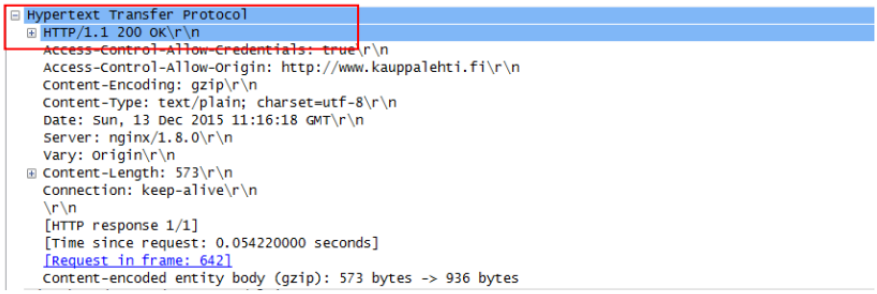
HTTP GET



HTTP/1.1 is a plain text protocol

HTTP POST





- The POST request method requests that a web server accepts the entity enclosed in the body of the request message as a new subordinate of the web resource identified by the URI
- POST changes the state of the server
- Examples: submitting a web form

POST vs PUT example

GET /device-management/devices : Get all devices

POST /device-management/devices : Create a new device

GET /device-management/devices/{id} : Get the device information identified by "id"

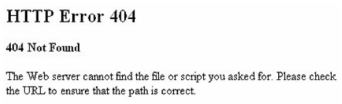
PUT /device-management/devices/{id} : Update the device information identified by "id"

DELETE /device-management/devices/{id} : Delete device by "id"

• The PUT method requests that the enclosed entity be stored under the supplied URI
• If the URI refers to an already existing resource, it is modified; if the URI does not point to an existing resource, then the server can create the resource with that URI

HTTP Response

- **The first line of the response is called the *status* line and has a numeric status code and a text-based reason phrase**



- **HTTP status codes are primarily divided into five groups:**

Informational 1XX, Successful 2XX, Redirection 3XX, Client Error 4XX, and Server Error 5XX

Response Header Fields

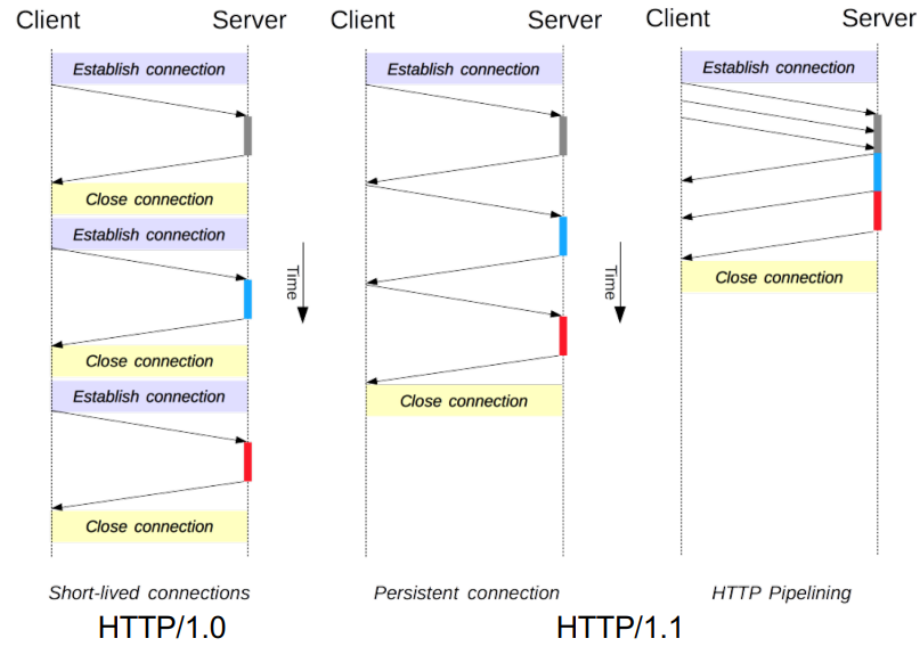
These header fields give information about the server, about further access to the target resource, or about related resources.
• Control Data (e.g. directs caching, or instructs the client where to go next)
• Validator Header Fields
• Authentication Challenges
• Response Context

HTTP/1.1 vs. HTTP/2

• Both HTTP/1.1 and HTTP/2 run on top of TCP
• HTTP/2 maintains high compatibility for methods, status code, URI's and header fields with HTTP/1.1
• HTTP/2 supports all the core features of HTTP/1.1, but aims to be more efficient in several ways
• HTTP/2 is a binary protocol, instead of a plain text protocol
Plain text protocols are pretty simple to debug. You print out their contents and you can see exactly what happened.

Binary protocols are a bit harder, simply because they need translation before you read them. But they have one huge advantage in that they are much more compact, and thus far more efficient.

HTTP Pipelining

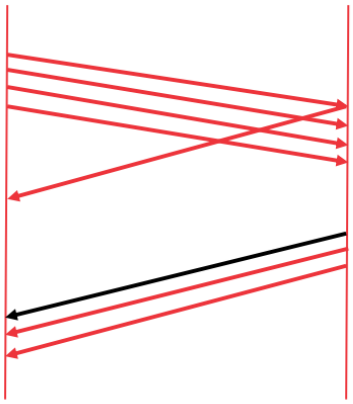


Pipelining is the process to send successive requests, over the same persistent connection, without waiting for the answer.

Pipelining is the process to send successive requests, over the same persistent connection, without waiting for the answer.

Head of Line Blocking

- **First-in-first-out:** The server must send its responses in the same order that the requests were received.



A large or slow response can block others behind it

Primary Goals of HTTP/2

• Reduce latency by enabling full request and response multiplexing. Multiplexing allows your Browser to fire off multiple requests at once on the same connection and receive the requests back in any order.

• Minimize protocol overhead via efficient compression of HTTP header fields

• Add support for request prioritization and server push.

HTTP/2 prioritization is requested by the client (browser) and it is up to the server to decide what to do based on the request.

HTTP/2 Server Push allows an HTTP/2-compliant server to send resources to a HTTP/2-compliant client before the client requests them. Server Push is a performance technique aimed at reducing latency by loading resources preemptively, even before the client knows they will be needed

Self-Test HTTP

• HTTP request methods: GET, PUT, POST

• HTTP response status line

• Is HTTP stateless or not?

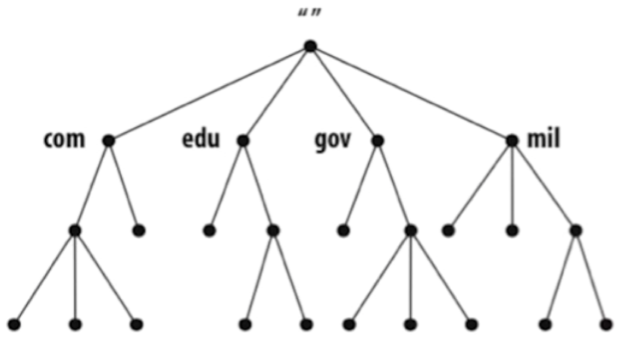
• What is head of line blocking?

<Check all above>

Domain Name System (DNS) and Content Delivery Network (CDN)

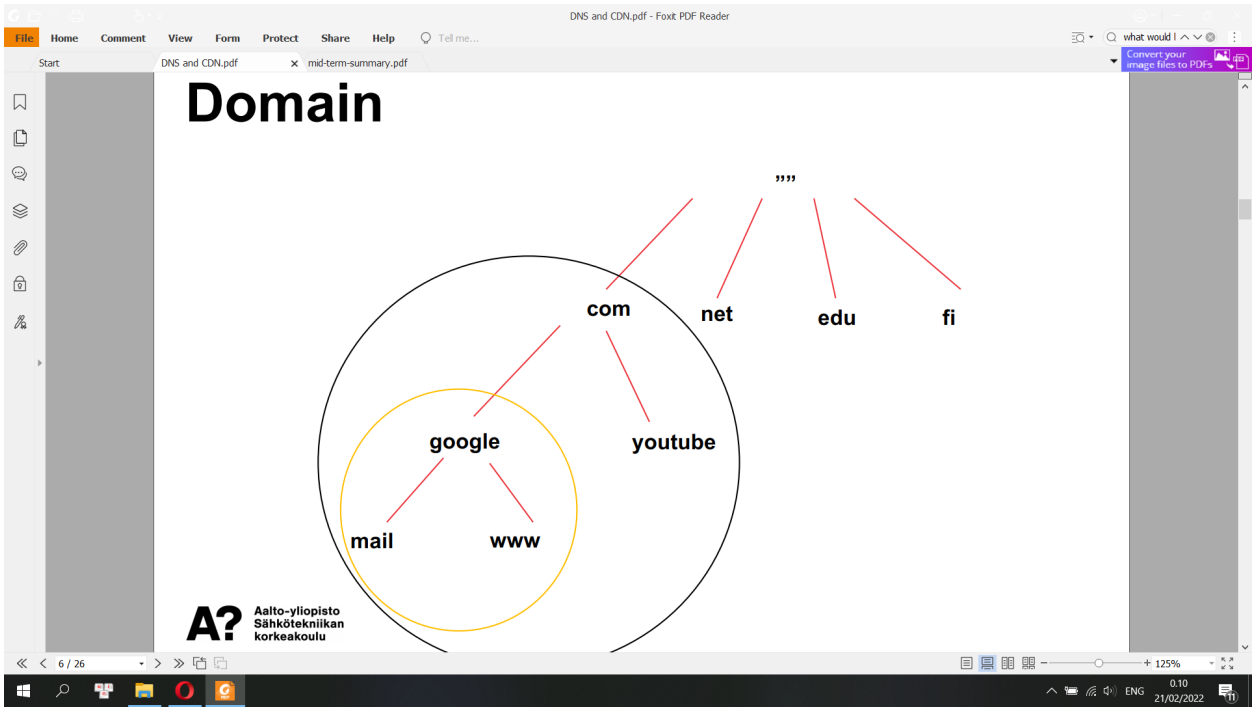
Domain Names

- Domain names -> resources (e.g., computers, websites)
- Paths in an inverted tree
- An inverted tree is made up of nodes and links between nodes
- Each node is connected to its parent by a single link
- One node can be parent to arbitrarily many children



Labels

- Each node has a label, between 0 and 63 bytes in length
- The root node (at the top) has a special, reserved label, written “” (for a zero-length label)
- Otherwise, the main restriction is that all of the children of a node have different labels
- Domain name: the series of labels from the node to the root, read with dots separating the labels



Domain Name System (DNS)

The Domain Name System (DNS) is the Internet's system for mapping alphabetic names to numeric Internet Protocol (IP) addresses

• Application layer protocol, responsible for mapping domain names into IP addresses

• DNS protocol relies on UDP by default, but can also work over TCP as a fallback when firewalls block UDP.

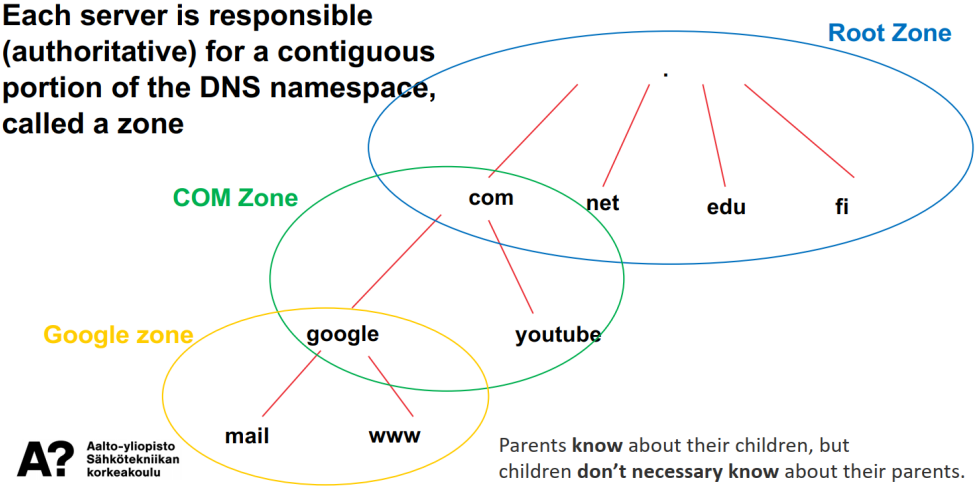
• DNS is a distributed database

• Resolver: A DNS client that sends DNS messages to obtain information about the requested domain name space

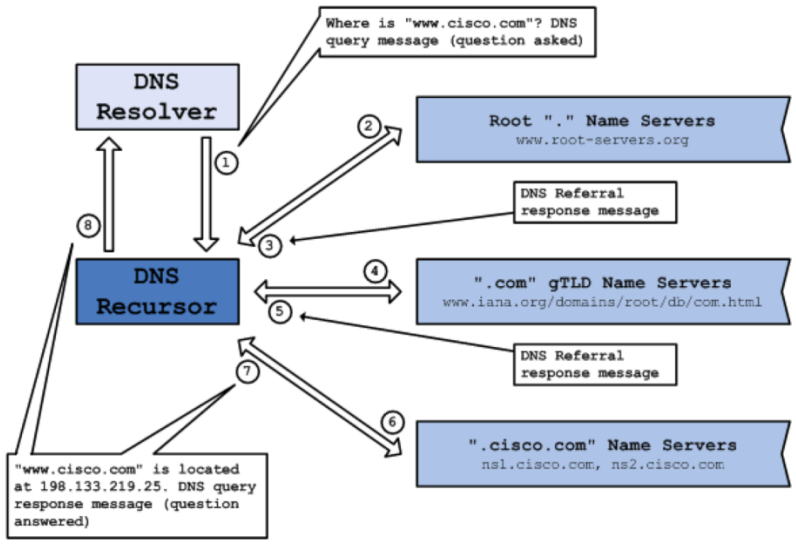
• A DNS name server is a server that stores the DNS records for a domain; A DNS name server responds with answers to queries against its database

DNS Zones

Each server is responsible (authoritative) for a contiguous portion of the DNS namespace, called a zone



Recursive Query



- **Authoritative Server:** A DNS server that responds to query messages with information stored in resource records for a domain name space stored on the server.
- **Recursive Resolver:** A DNS server that recursively queries for the information asked in the DNS query.

Aalto-yliopisto
Sähkötekniikan

gTLD: generic top-level domain
Source: Cisco

DNS Messages

DNS protocol is composed of three types of messages: **queries, responses**, and updates.

The first sixteen bits are for the **Transaction ID**, used to match the response to the query, and is created by the client on the query message and returned by the server in the response.

The **question** is present in both the query and the response, and should be identical.

Header

Transaction ID: 0xd7da									
QR: 1	Opcode: 0	AA: 0	TC: 0	RD: 0	RA: 0	Z	AD: 0	CD: 0	Rcode: 0
Number of Questions: 1									
Number of Answer RRs: 0									
Number of Authority RRs: 13									
Number of Additional RRs: 16									

Example of Query and Answer

```
Queries
  www.google.com: type A, class IN
    Name: www.google.com
    Type: A (Host address)
    Class: IN (0x0001)
```

```
Answers
  www.google.com: type A, class IN, addr 74.125.131.147
      Name: www.google.com
      Type: A (Host address)
      Class: IN (0x0001)
      Time to live: 5 minutes
      Data length: 4
      Addr: 74.125.131.147 (74.125.131.147)
  www.google.com: type A, class IN, addr 74.125.131.103
  www.google.com: type A, class IN, addr 74.125.131.104
  www.google.com: type A, class IN, addr 74.125.131.106
  www.google.com: type A, class IN, addr 74.125.131.99
  www.google.com: type A, class IN, addr 74.125.131.105
```

Answer: in the response from recursive resolver to an end user's computer, or in the response from the authoritative name server of the domain to the recursive resolver.

Answer RR

```
Domain Name System (response)
  [Request In: 3]
  [Time: 0.014981000 seconds]
  Transaction ID: 0xccf9
  Flags: 0x8000 Standard query response, No error
  Questions: 1
  Answer RRs: 0
  Authority RRs: 4
  Additional RRs: 4
  Queries
  Authoritative nameservers
    google.com: type NS, class IN, ns ns2.google.com
      Name: google.com
      Type: NS (Authoritative name server)
      Class: IN (0x0001)
      Time to live: 2 days
      Data length: 6
      Name Server: ns2.google.com
    google.com: type NS, class IN, ns ns1.google.com
    google.com: type NS, class IN, ns ns3.google.com
    google.com: type NS, class IN, ns ns4.google.com
  Additional records
```

Domain name in both the RR name and RR data fields

Authority RR

```
Additional records
  ns2.google.com: type A, class IN, addr 216.239.34.10
      Name: ns2.google.com
      Type: A (Host address)
      Class: IN (0x0001)
      Time to live: 2 days
      Data length: 4
      Addr: 216.239.34.10 (216.239.34.10)
  ns1.google.com: type A, class IN, addr 216.239.32.10
  ns3.google.com: type A, class IN, addr 216.239.36.10
  ns4.google.com: type A, class IN, addr 216.239.38.10
```

Additional records

When a name server does not have the answer to the query (as is not authoritative), it will not send answer records.

Resource Record (RR)

• A resource record is a format used in DNS message that is composed of NAME, TYPE, CLASS, TTL, RDLENGTH, and RDATA.

- A **resource record** is a format used in DNS message that is composed of NAME, TYPE, CLASS, TTL, RDLENGTH, and RDATA.

Type	Meaning	Value
SOA	Start of Authority	Parameters for this zone
A	IP address of a host	32-Bit integer
MX	Mail exchange	Priority, domain willing to accept email
NS	Name Server	Name of a server for this domain
CNAME	Canonical name	Domain name
PTR	Pointer	Alias for an IP address
HINFO	Host description	CPU and OS in ASCII
TXT	Text	Uninterpreted ASCII text

AAAA, IPv6 address record

DNS Caching

• Temporarily storing the results of recently browsed websites' DNS queries on a local file for faster retrieval

• DNS clients and DNS server both use caching to speed up the domain name lookup process and to ease traffic on the root servers

• In the context of a DNS record, Time to Live (TTL) is a numerical value that determines how long a DNS cache server can serve a DNS record before reaching out to the authoritative DNS server and getting a new copy of the record

Host Names

• A hostname is a domain name that has at least one associate IP address. (e.g. www.aalto.fi)

• A valid domain name may not necessarily be valid as a hostname

Anycast Addressing

• Anycast addresses are allocated from the unicast address space. Assigning a unicast address to more than one interface makes a unicast address an anycast address. Example:
ipv6 address 2002:0db8:6301::/128 anycast

• Anycast Address can not be used as source address of IPv6 packet

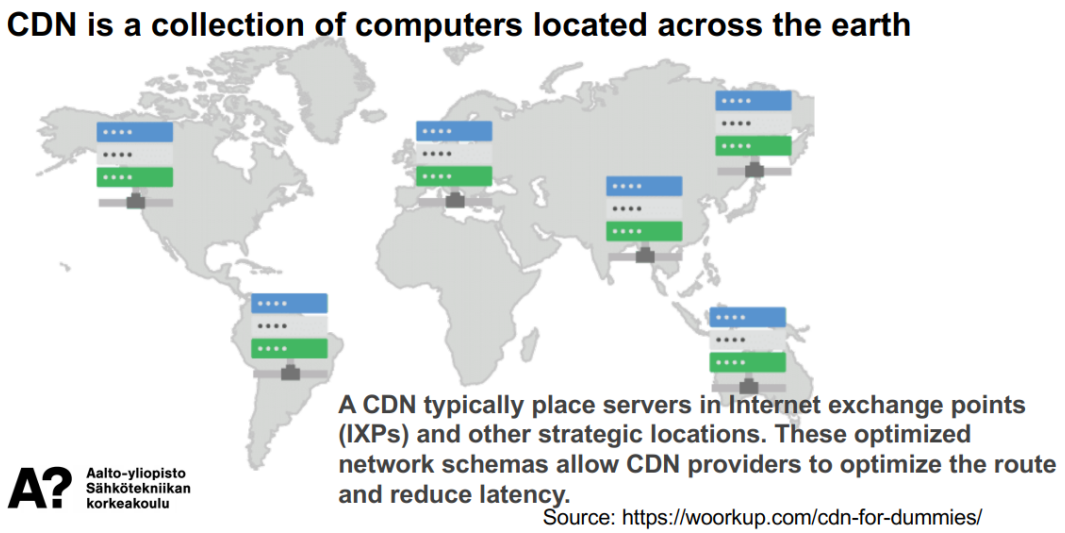
• The routing algorithm selects the single receiver from the group based on least-expensive routing metric (e.g. #hops, distance, latency, efficiency, cost)

An anycast address is an address allocated to a set of interfaces that typically belong to different routers. When a packet is destined to an anycast address, it is delivered to the closest interface that has this anycast address, where the term "closest" is determined by the routing protocol. An anycast address must be assigned to a router not a host and cannot be used as a source address

Content Delivery Network (CDN): is a group of geographically distributed servers that speed up the delivery of web content by bringing it closer to where users are. ... CDNs cache content

like web pages, images, and video in proxy servers near to your physical location. CDN makes your website load fast all around the world

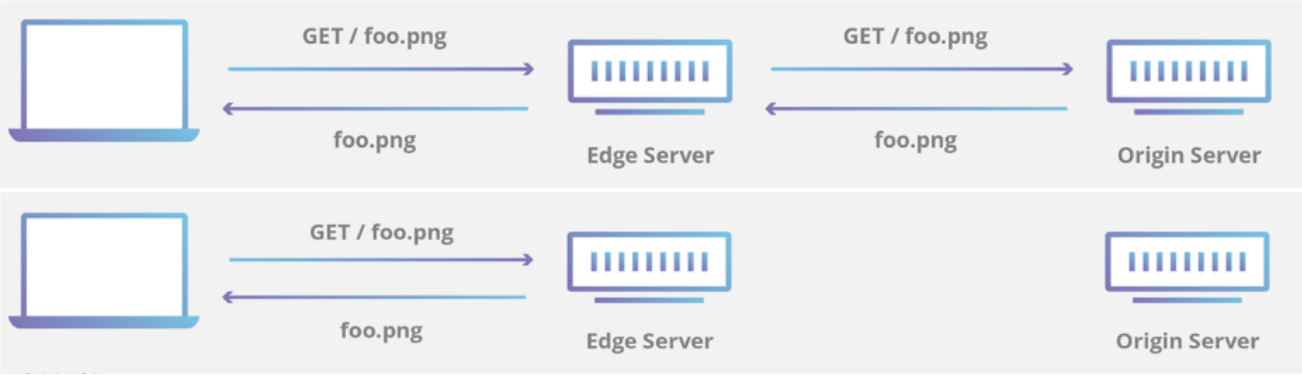
Web Hosting + CDN



Benefits of CDN

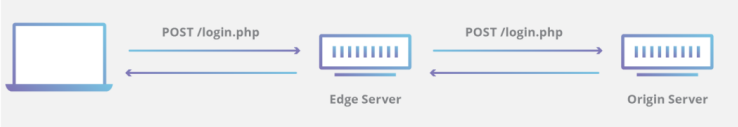
• With a CDN requests to your website are always routed to the nearest available server.

• A CDN improves the latency by pulling static content files from the origin server into the distributed CDN network in a process called caching.



Edge Server vs. Original Server

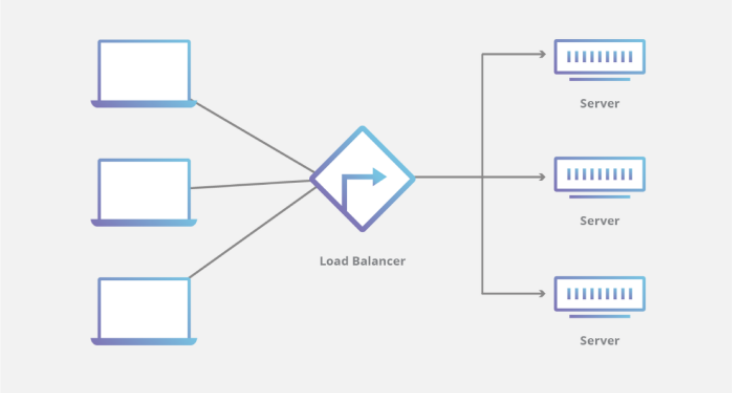
• This interplay between edge servers handling static content and origin servers serving up dynamic content is a typical separation of concerns when using a CDN.



Important server-side code such as the database of hashed client credentials used for authentication is typically maintained inside an origin server.

Aalto-yliopisto
Sähkötekniikan korkeakoulu

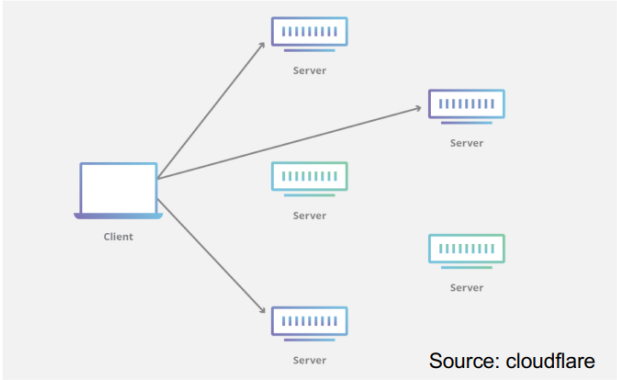
- Load balancing distributes network traffic evenly across several servers, making it easier to scale rapid boosts in traffic.



Aalto-yliopisto
Sähkötekniikan
korkeakoulu

Source: cloudflare

- In the event that an entire data center is having technical issues, Anycast routing transfers the traffic to another available data center, ensuring that no users lose access to the website.

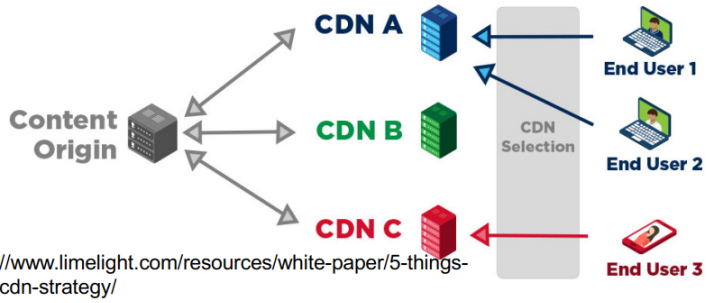


Aalto-yliopisto
Sähkötekniikan
korkeakoulu

Source: cloudflare

Multi-CDN

- The goal of deploying multiple CDNs is to distribute load among two or more CDNs
- Multi-CDN can minimize single points of failure by providing alternate delivery options in the event of a CDN outage.

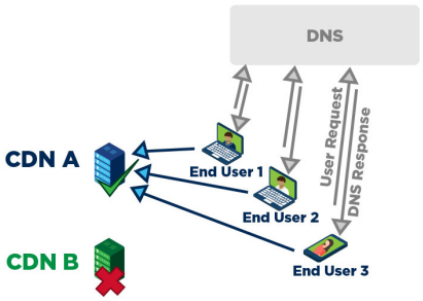


Aalto-yliopisto
Sähkötekniikan
korkeakoulu

https://www.limelight.com/resources/white-paper/5-things-multi-cdn-strategy/

Selection of CDN

- Selecting an optimal CDN could be based on a number of criteria – availability, geographic location, traffic type, capacity, cost, performance or combinations of the above.
- CDN selection may be carried out manually, or it may be automated using techniques like DNS or commercial decision engines.



Aalto-yliopisto
Sähkötekniikan korkeakoulu

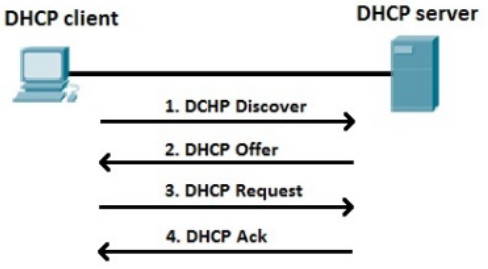
Self-Test DNS and CDN

• What is DNS responsible for?
• Can you describe the process of recursive query?
• What are the benefits of CDN?
• What is anycast?
<Check above>

Dynamic Host Configuration Protocol (DHCP)

- DHCP is a network management protocol that is used to assign an IP address and various network parameters (e.g., default gateway, domain name, name servers, time server) to a device.
- Application layer protocol
- It runs on top of UDP/IP

DHCP uses a well-known UDP port number 67 for the DHCP server, and the UDP port number 68 for the client.



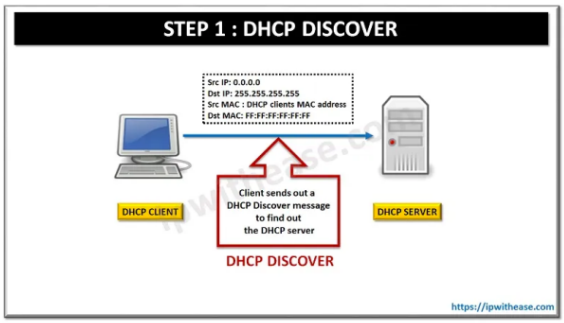
Aalto University

A DHCP client-to-server communication consists of three kinds of interaction between the two peers:

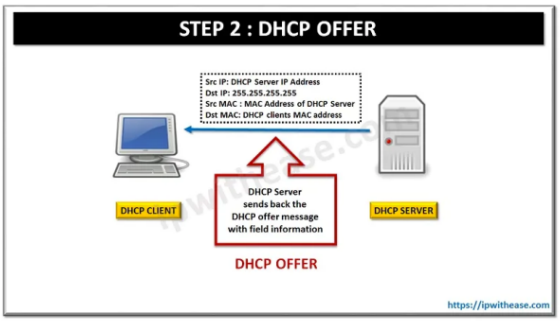
Broadcast-based DORA (Discover, Offer, Request, Acknowledgement). This process consists of the following steps:

The DHCP client sends a DHCP Discover broadcast request to all available DHCP servers within range.
A DHCP Offer broadcast response is received from the DHCP server, offering an available IP address lease.
The DHCP client broadcast Request asks for the offered IP address lease and the DHCP broadcast Acknowledgement at the end.
If the DHCP client and server are located in different logical network segments, a DHCP relay agent acts a forwarder, sending the DHCP broadcast packets back and forth between peers.

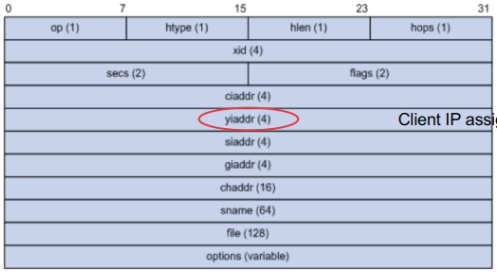
Example of the DORA process



Layer 3 broadcast + Layer 2 broadcast

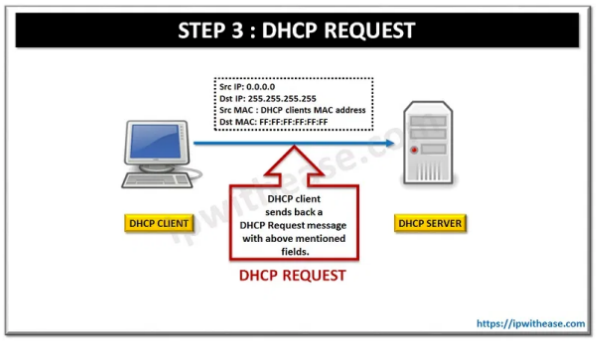


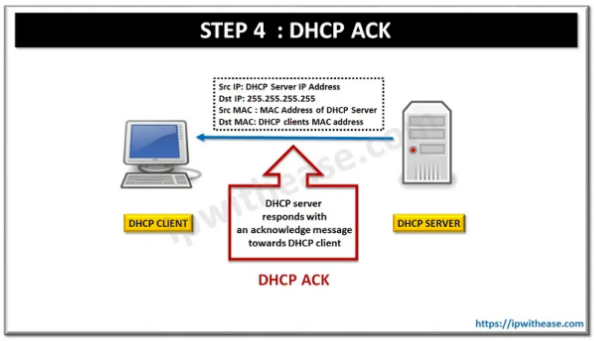
Layer 3: Still Broadcast as Client still has no IP Address
Layer 2: Unicast



Client IP assigned by the server

Aalto University
School of Electrical
Engineering

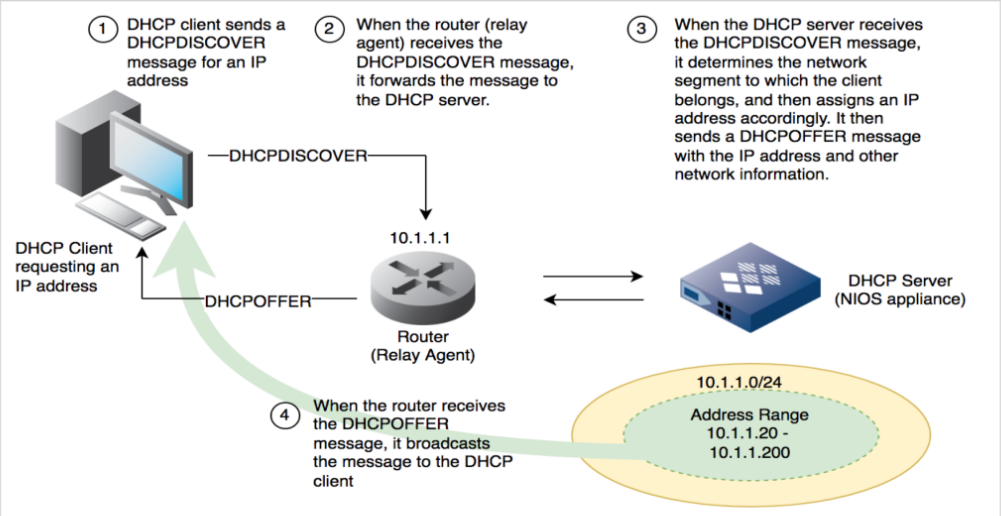




Layer 3: Still Broadcast as Client must have received Offer from more than one DHCP server in their domain and the DHCP client accepts the Offer that its receives the earliest and by doing a broadcast it intimates the other DHCP server to release the Offered IP address to their available pool again

DHCP
1: A DHCP client sends a broadcast packet (DHCP Discover) to discover DHCP servers on the LAN segment.
2: The DHCP servers receive the DHCP Discover packet and respond with DHCP Offer packets, offering IP addressing information.
3: If the client receives the DHCP Offer packets from multiple DHCP servers, the first DHCP Offer packet is accepted. The client responds by broadcasting a DHCP Request packet, requesting the network parameters from the server that responded first.
4: The DHCP server approves the lease with a DHCP Acknowledgement packet. The packet includes the lease duration and other configuration information.

DHCP with relay agent



Aalto University
School of Electrical Engineering

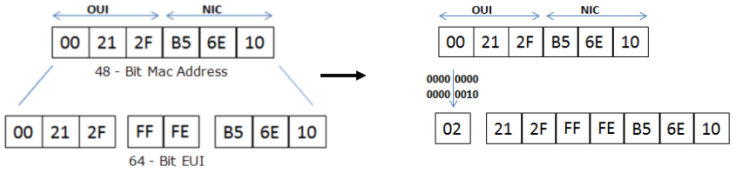
https://docs.infoblox.com/display/NAG8/IPv4+DHCP+Protocol+Overview

DHCP v6

- **SLAAC** (**S**tate**L**ess **A**ddress **A**uto **C**onfiguration): The preferred method of assigning IP addresses in an IPv6 network.
  - SLAAC devices send the router a request for the network prefix, and the device uses the prefix and its own MAC address to create an IP address. After the IP is computed, it checks to see if a duplicate IP was previously created.
- If the router does not implement SLAAC and no network prefix is received, the device sends a request to the DHCPv6 server, which responds with an IP address similar to the DHCP in IPv4.

IPv6 Interface Identifier

- Most IPv6 addresses can be divided into a **64-bit network prefix** and a 64-bit "host" portion. These host-portion bits are known officially as the **interface identifier.**
- **Modified EUI-64 Identifier (EUI: extended unique identifier)**
  - Given an Ethernet address (48 bits), insert 0xfffe between the first 3 bytes and the last 3 bytes, to get 64 bits in all.
  - Set the 7th bit of the first byte to 1



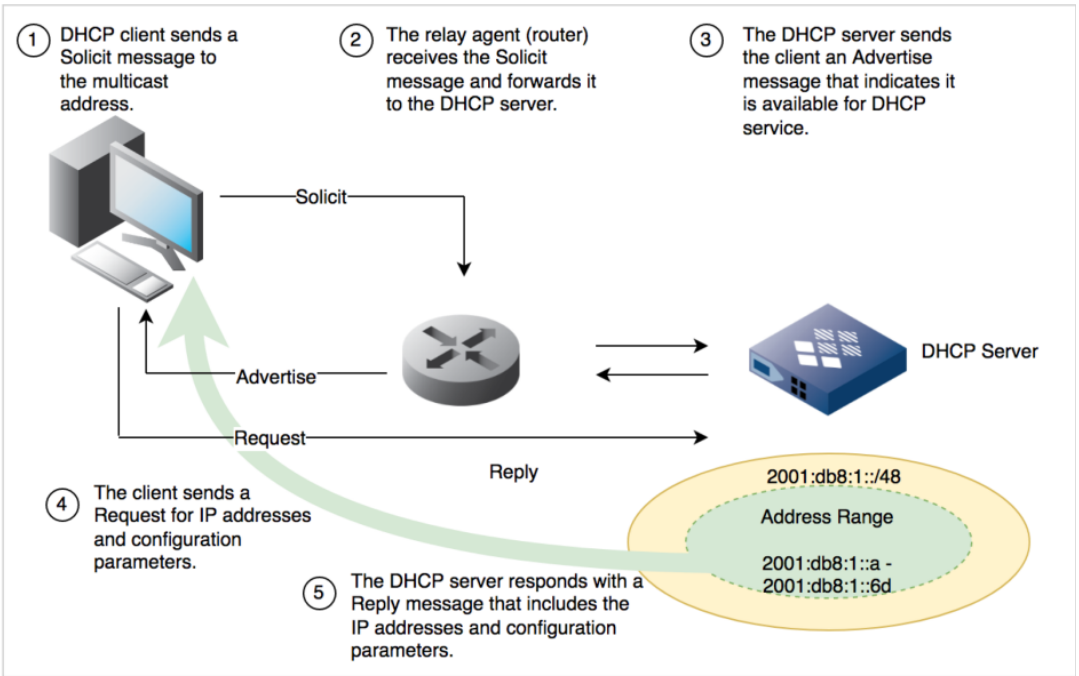
Source: Cisco

Stateful DHCPv6

Stateful DHCPv6 uses a DHCPv6 Server to centrally manage IPv6 address and prefix assignment.

- DHCPv6 Clients get IPv6 address or prefix information from the DHCPv6 Server.
- DHCPv6 Clients can obtain configuration information that is not available from other protocols, such as DNS.

Stateless DHCPv6: does not require a DHCPv6 Server to maintain any dynamic state for clients, such as DNS Server addresses.

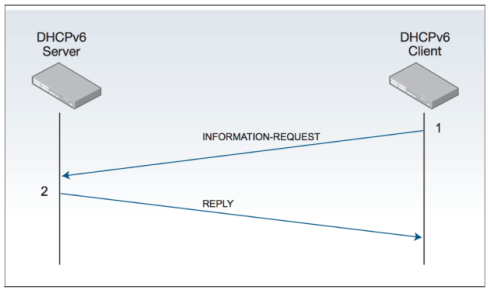


Normal DHCPv6 Message Exchange

1. **Solicit** - sent by a DHCPv6 Client to locate DHCPv6 Servers.
2. **Advertise** - sent by a DHCPv6 server to a DHCPv6 Client in answer to the solicit message as an affirmative message that DHCPv6 Server services are available to a DHCPv6 Client.
3. **Request** - sent by a DHCPv6 Client to a DHCPv6 Server to request configuration parameters.
4. **Reply** - sent by a DHCPv6 Server to a DHCPv6 Client with configuration information.
5. **Renew** - sent by a DHCPv6 Client to a DHCPv6 Server requesting an extension to the address lifetime.

Stateless DHCP v6

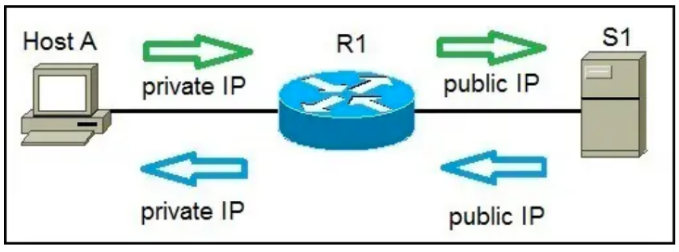
- The Client starts by sending an **INFORMATION-REQUEST** message to the Server. This request specifically excludes the assignment of any IPv6 address.
- The Server sends a **REPLY** message back to the Client to finish.



NAT (Network Address Translation)

NAT (Network Address Translation) is a process of changing the source and destination IP addresses and ports.

- Address translation reduces the need for IPv4 public addresses and hides private network address ranges.
- This process is usually done by routers or firewalls.



There are three types of address translation:

Static NAT – translates one private IP address to a public one. The public IP address is always the same.

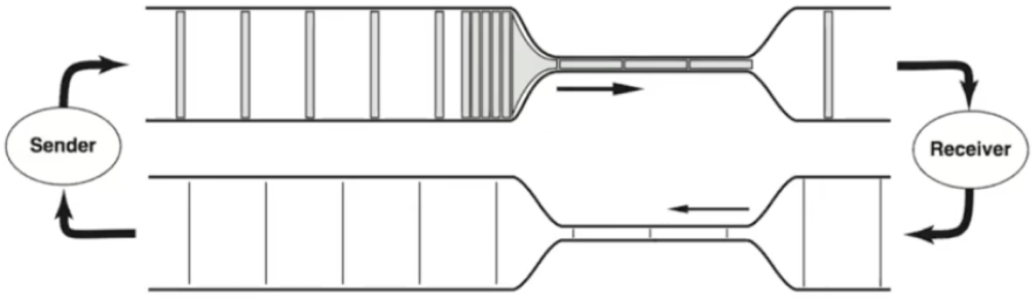
Dynamic NAT – private IP addresses are mapped to the pool of public IP addresses.

Port Address Translation (PAT) – one public IP address is used for all internal devices, but a different port is assigned to each private IP address. Also known as **NAT Overload**.

TCP Congestion Control: How to perceive congestion?

• Bottleneck

- It determines the connection's maximum data-delivery rate.
- It is where persistent queues form



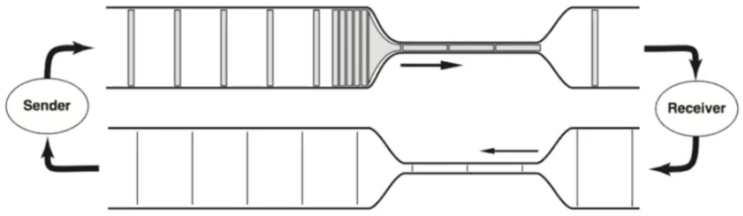
- **Implicit end-to-end feedback**
  - IP layer provides no explicit feedback to end systems regarding congestion
  - Presence of congestion inferred by the end systems based only on observed network behavior (e.g. packet loss and delay)
- **Success Event (ACK received) vs. Loss Event (timeout or duplicated ACKs)**
- **Earlier versions of TCP interprets packet loss as congestion. Does this assumption always hold?**

- **Sources of errors in wireless links**
  - Pauses due to handoff between cells
  - Mobile host out of reach of other receivers (little or no overlaps between cells)
  - Packet losses due to transmission errors in wireless links

- **In wireless lossy links, the sporadic losses are not due to congestion → unnecessary window and transmission rate reduction if loss is interpreted as congestion**

What has changed since earlier versions of TCP were invented?

- **NIC evolves from Mbps to Gbps and memory chips from KB to GB**



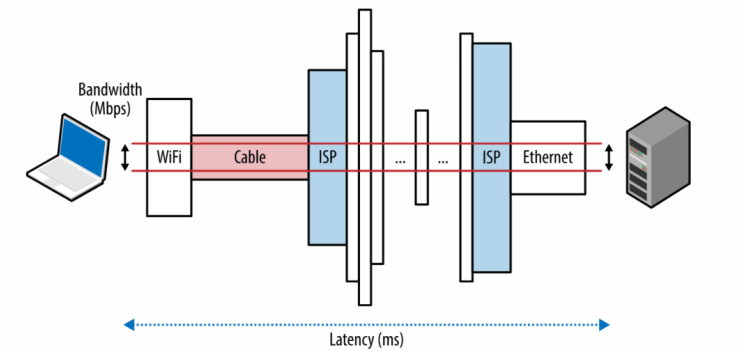
What would happen to loss-based congestion control if bottleneck buffers are large?

What is Bufferbloat?

"Bufferbloat is the undesirable latency that comes from a router or other network equipment buffering too much data."

Latency and Bandwidth

Latency: The time from the source sending a packet to the destination receiving it
Bandwidth: Maximum throughput of a logical or physical communication path



• Propagation delay: Amount of time required for a message to travel from the sender to receiver, which is a function of distance over speed with which the signal propagates.
• Transmission delay: Amount of time required to push all the packet's bits into the link, which is a function of the packet's length and data rate of the link.
• Processing delay: Amount of time required to process the packet header, check for bit-level errors, and determine the packet's destination.
• Queuing delay: Amount of time the packet is waiting in the queue until it can be processed. If the packets are arriving at a faster rate than the router is capable of processing, then the packets are queued inside an incoming buffer.

• Congestion wasn't reported until after a prolonged period of queueing, resulting in greater packet loss
• Packet loss alone is not a good proxy to detect congestion

TCP BBR Bottleneck Bandwidth and Round-trip propagation time
Design Criteria of BBR
• Make network full but buffer empty à Maintain the data rate at the bottleneck bandwidth and the pipe is full
• BBR looks at the path's bottleneck bandwidth and an estimate of the RTT to determine congestion in a network