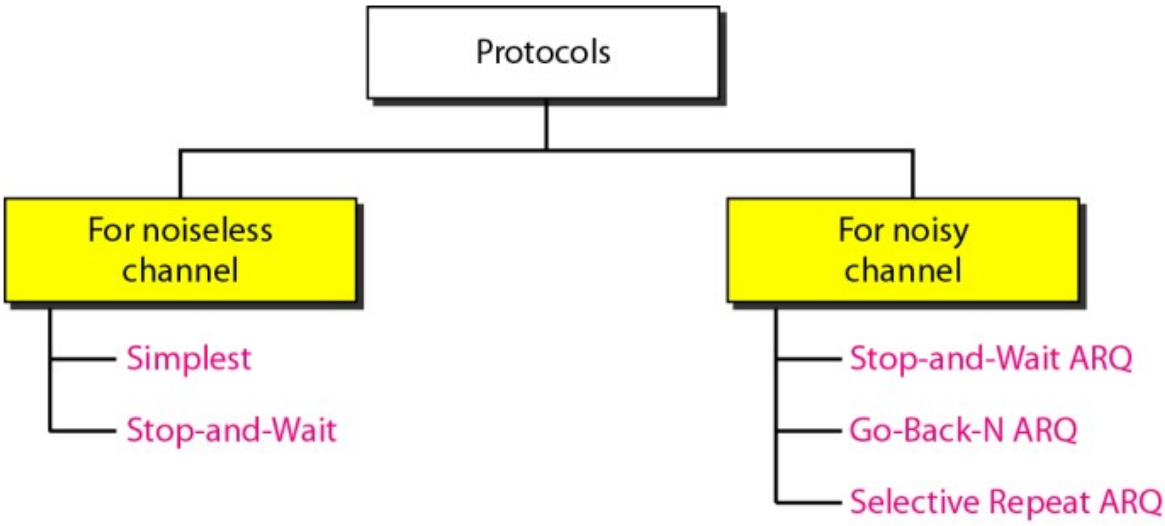
Flow Control

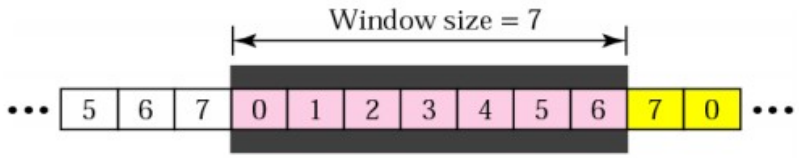
Flow Control

- Forouzan's Definition: Flow control refers to a set of procedures used to restrict the amount of data that the sender can send before waiting for acknowledgment.

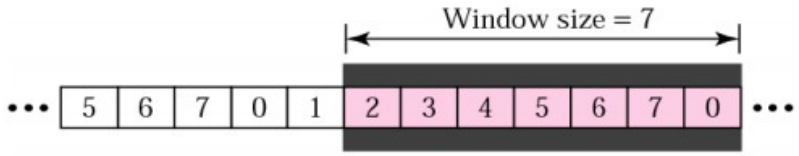


Sliding Window

- m: Size of the sequence number field in bits
- 1...2^m: Sequence numbers
- Send window: Box of size $2^m - 1$



a. Before sliding



b. After sliding two frames

* Figure is courtesy of B. Forouzan

Window Size for Go-Back-N

- Depends on size of max. frame number
  - Frame # needs to be included in every frame
  - e.g. 4 bits – 2^4 = 16 frame numbers
- Trade-off between window size and frame size

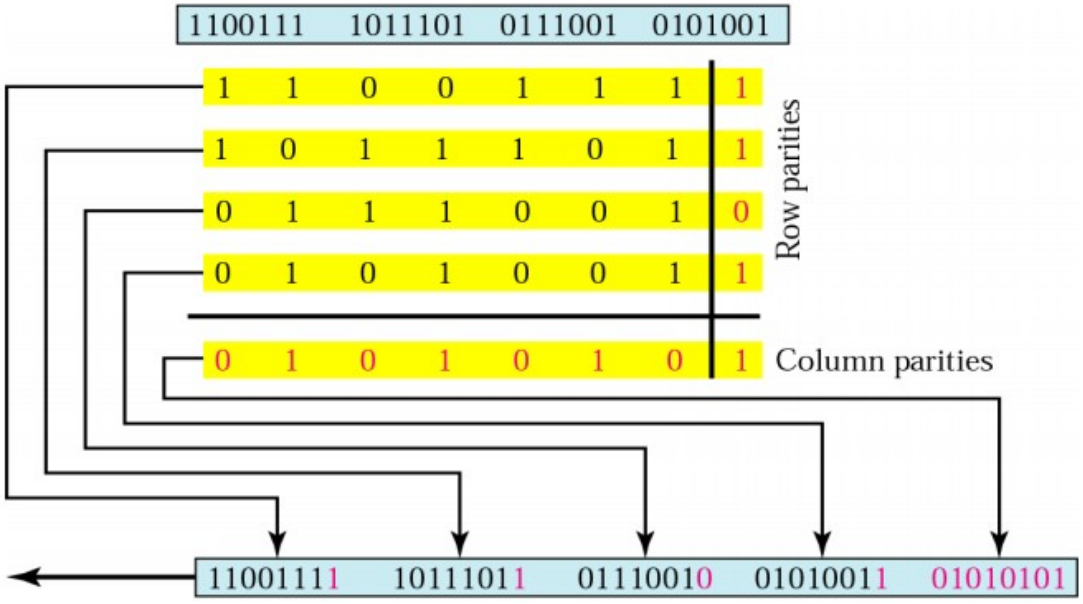
Error Checking

Summary: Flow Control

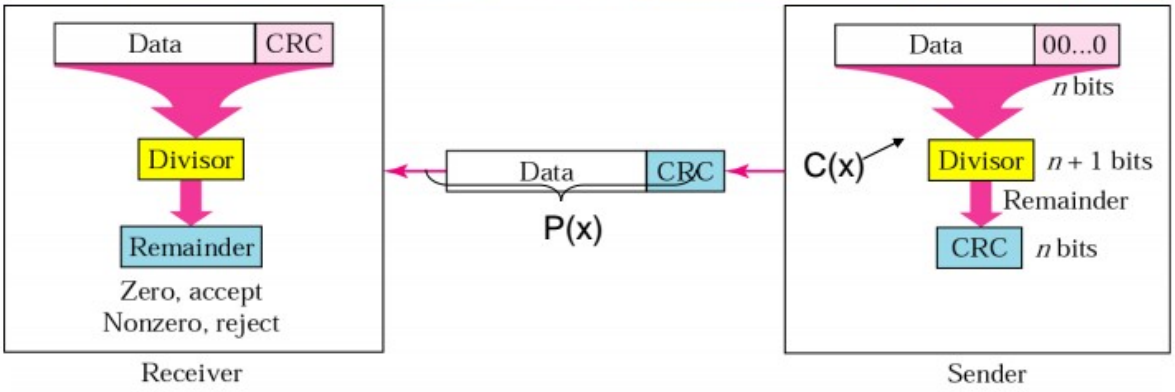
- Flow Control:
  - Stop-and-Wait
  - Sliding Window
- Error Control
  - Stop-and-Wait ARQ
  - Go-back-N ARQ
  - Selective Repeat ARQ

Two-Dimensional Parity Check

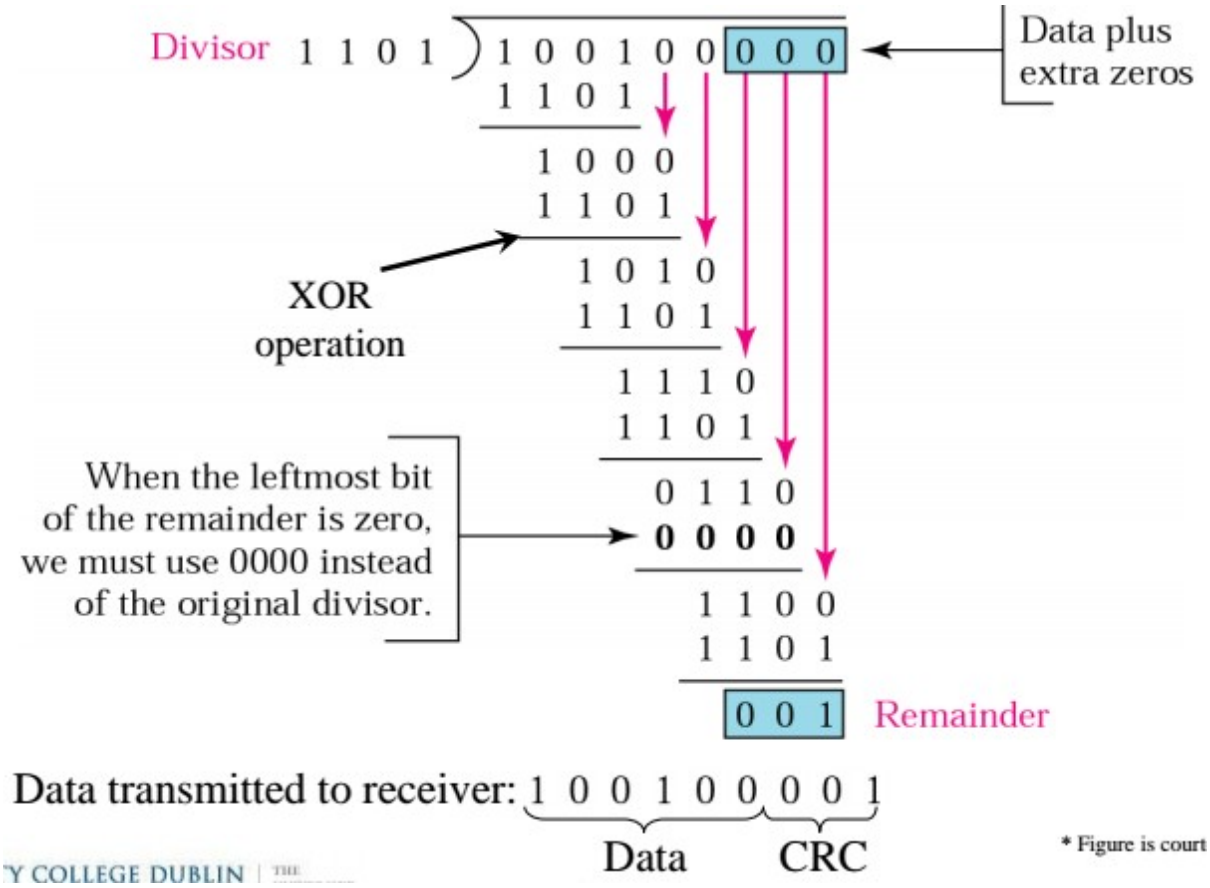
In two-dimensional parity check, a block of bits is divided into rows and a redundant row of bits is added to the whole block.



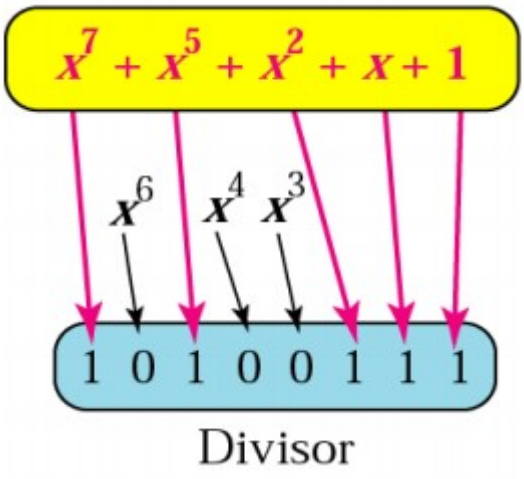
Cyclic Redundancy Check (CRC)



- P(x) divided by C(x) = 0
- (P(x)+remainder) divided by C(x) should be != 0



* Figure is court



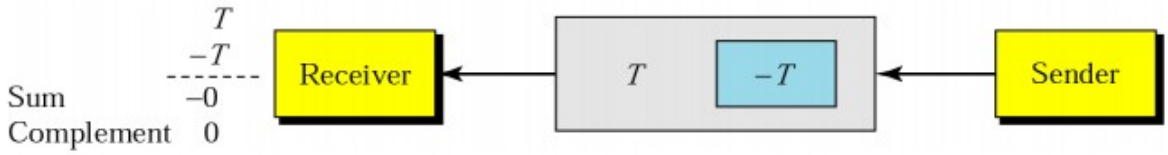
Sender:

The unit is divided into k sections, each of n bits.

All sections are added using one's complement to get the sum.

The sum is complemented and becomes the checksum.

The checksum is sent with the data.



Receiver:

The unit is divided into k sections, each of n bits.

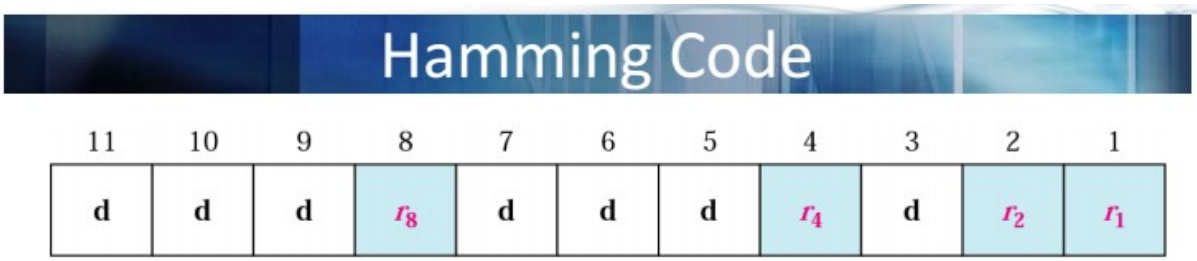
All sections are added using one's complement to get the sum.

The sum is complemented.

If the result is zero, the data are accepted: otherwise, rejected.

Hamming Code

https://www.youtube.com/watch?v=373FUw-2U2k (Good Short Visual Tutorial)



- Redundancy bits distributed throughout data bits
- Individual redundancy bits work as parity bits for specific data bits
  - e.g. r_1 is the parity bit for all odd numbers

    3 = binary 0011 7= binary 0111

    5 = binary 0101 9= binary 1001

Summary: HDLC

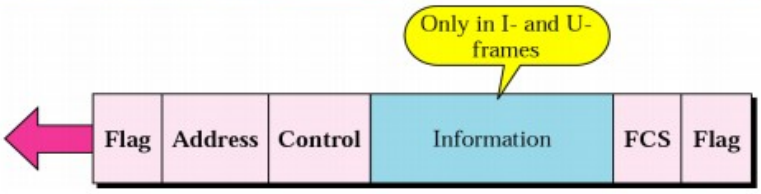
- Three station types
  - Primary station
  - Secondary station
  - Combined station
- Operation modes
  - Normal response mode
  - Asynchronous response mode
- Three frame types
  - I-Frame: Information Transfer Format
  - S-Frame: Supervisory Format – Flow Control
  - U-Frame: Unnumbered Format – Connection setup/term./etc

- Bit-Stuffing - to avoid confusion of data and flag

HDLC Modes

- Three modes:
  - Normal Response Mode (NRM)
  - Asynchronous Response Mode (ARM)
  - Asynchronous Balanced Mode (ABM)

HDLC frame



- Flag= 01111110
  - specifies beginning and end of frame
- Address
  - specifies secondary station
  - as either sender or receiver
- Control
  - specifies type of frame and seq.&ack. number
- Frame Check Sequence (FCS)
  - either 16- or 32-bit CRC

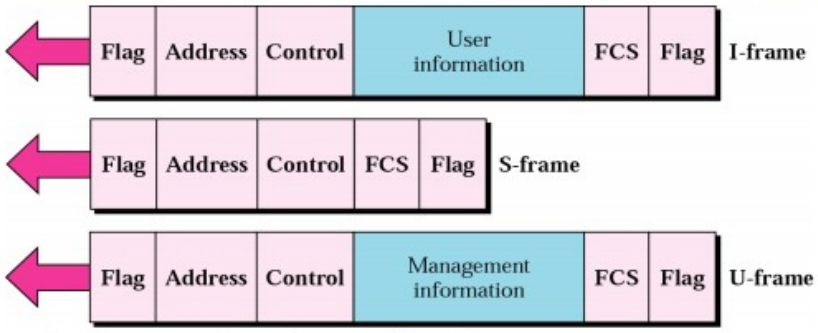
* Figure is courtesy of B. Forouz

Bit stuffing used to avoid confusion with data containing same combination as flag 01111110

- 0 inserted after every sequence of five 1s
- If receiver detects five 1s
  - it checks next bit
  - If 0, it is deleted
  - If 1 and seventh bit is 0, accept as flag
  - If sixth and seventh bits 1, sender is indicating abort

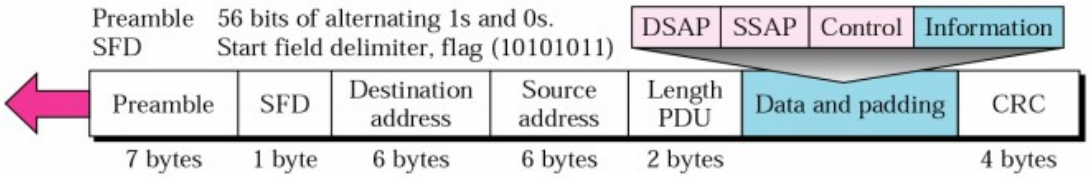
* Figure is courtesy of B. Fo

HDLC Frame Types



- I-Frame: Information Transfer Format
  - Control= 0 ? ? ? ? ? ? ?
- S-Frame: Supervisory Format
  - Control= 1 0 ? ? ? ? ? ?
- U-Frame: Unnumbered Format
  - Control= 1 1 ? ? ? ? ? ?

802.3 MAC Frame

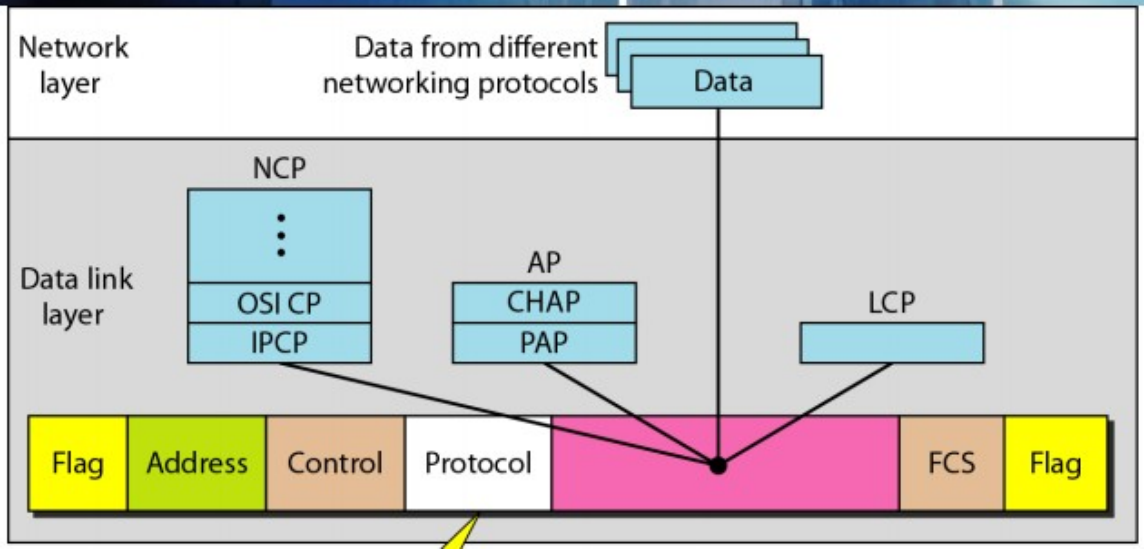


Point-to-Point Protocol (PPP)

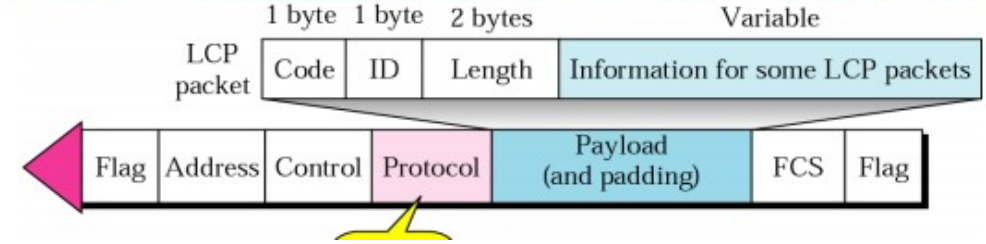
- Used for any kind of serial point to point connection e.g. dial-up, serial x-wire

- Based on HDLC

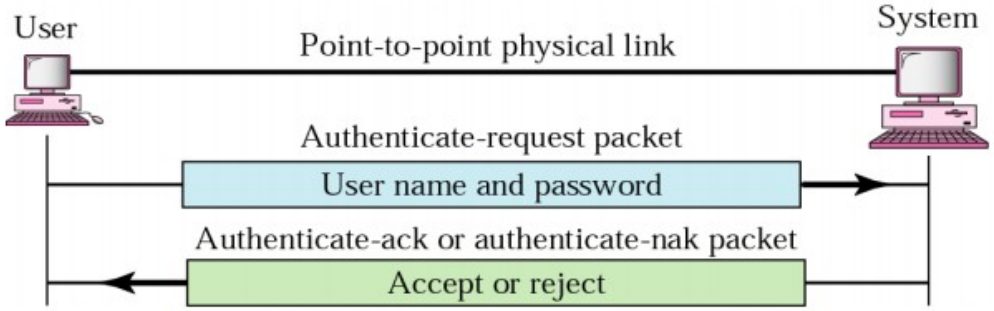
PPP Components



LCP Packet

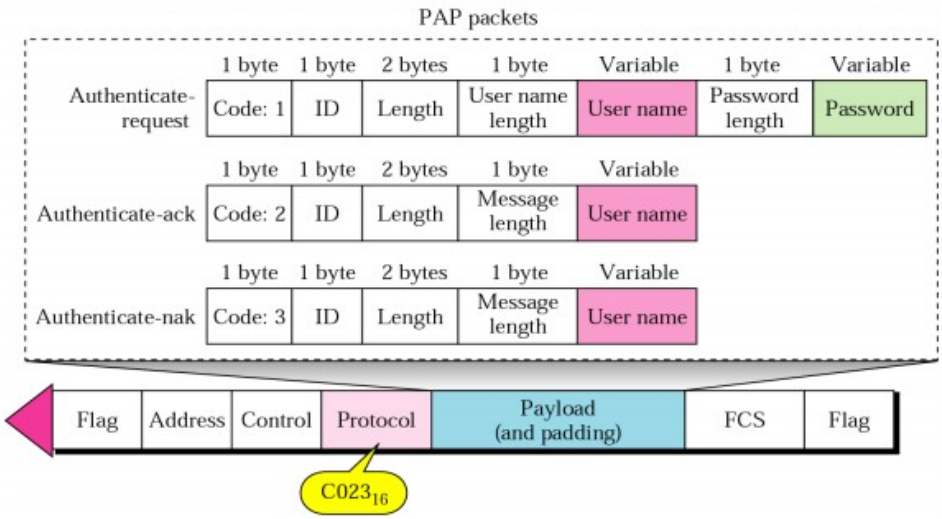


Password Authentication Protocol (PAP)

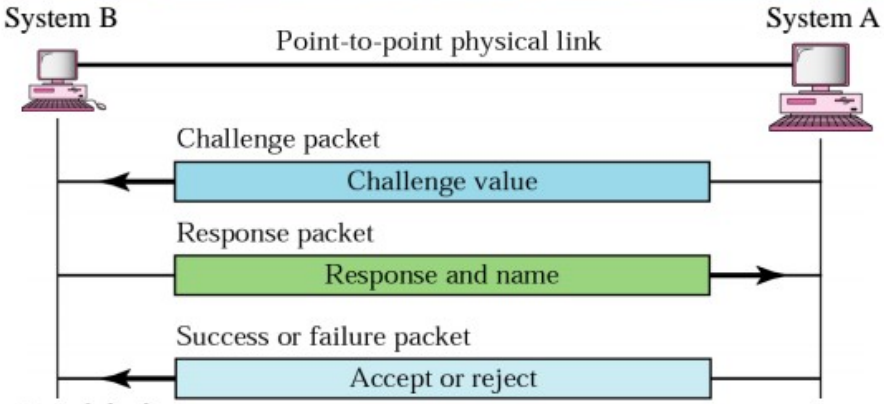


- 2-Way Handshake
- Password transmitted in clear text

PAP Packets



CHAP



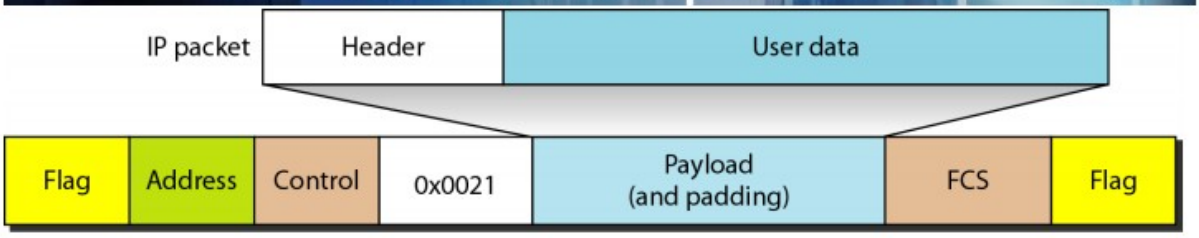
- **3-Way Handshake**

  1. A creates challenge ⇨ challenge value

  2. B processes challenge value with password ⇨ response

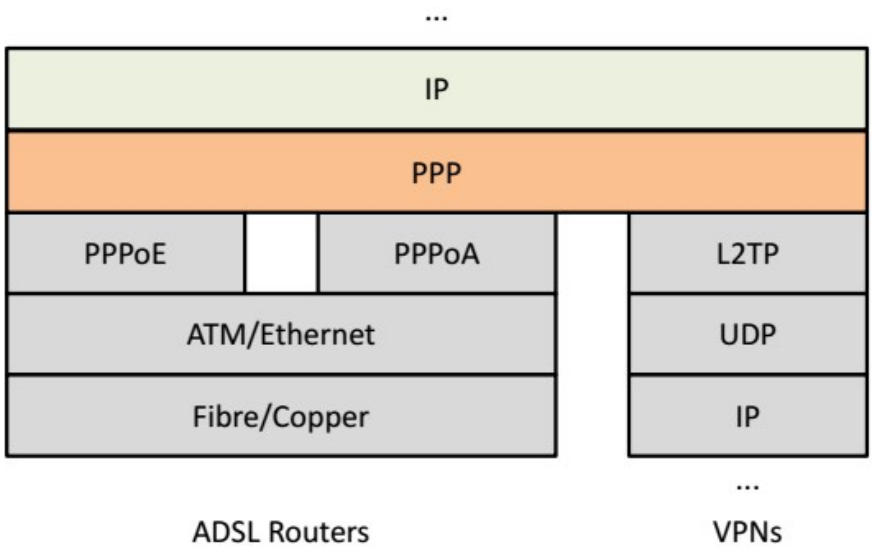
  3. A compares response with own calculation ⇨ accepts or rejects response

IP Packet Encapsulation

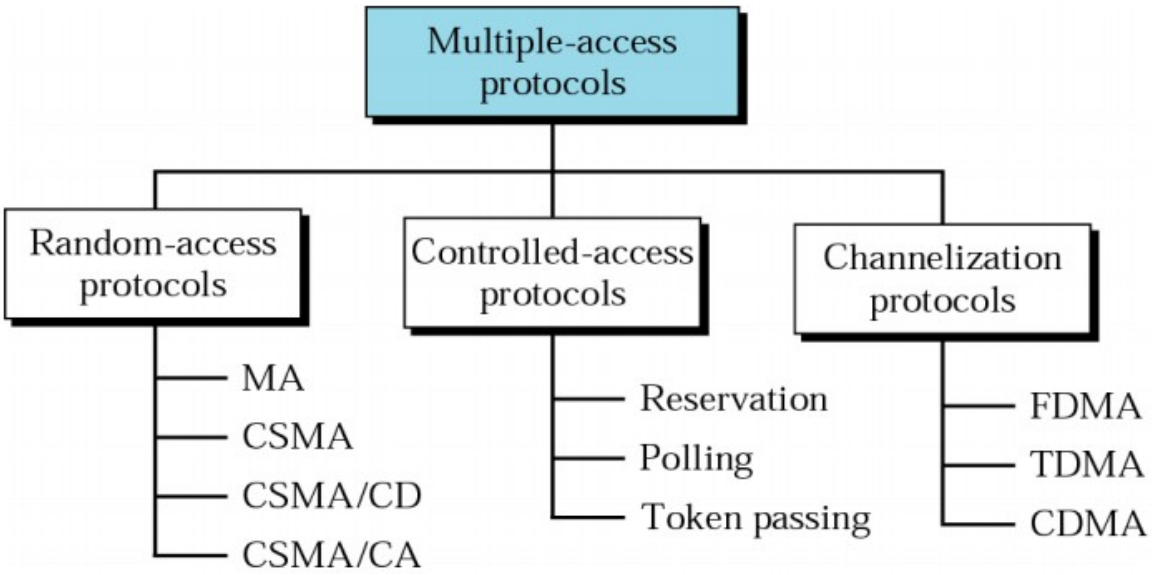


- PPP frame carries IP Packet as payload

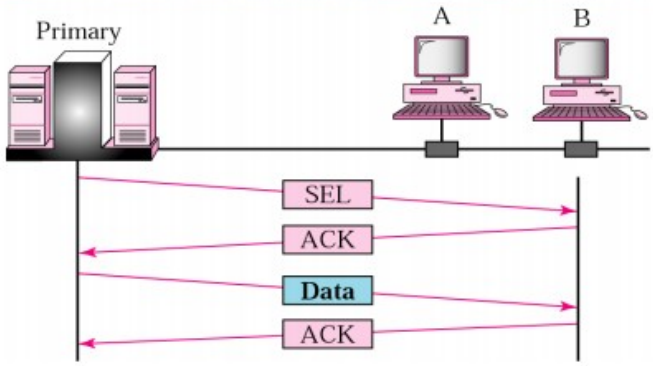
PPP – As Foundation for IP



Multiple-Access Protocols

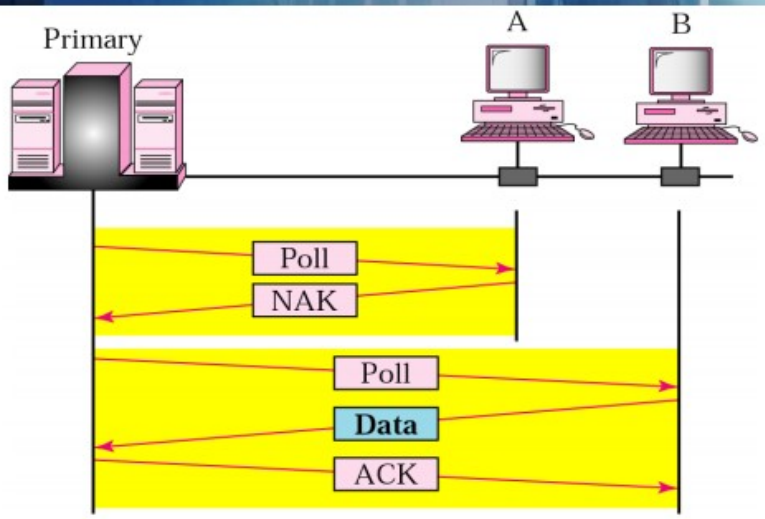


Select / Push



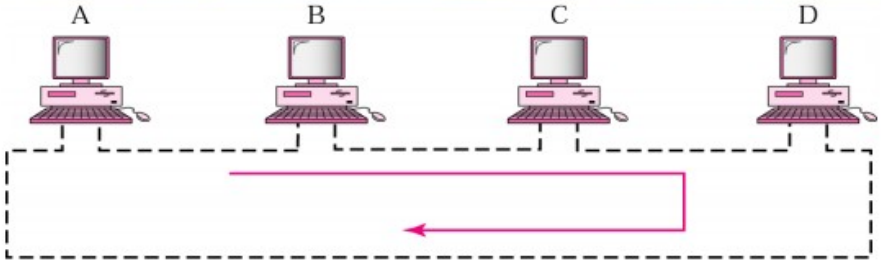
- Primary co-ordinates all communication
- Primary selects station that is destination then transmits data

Poll



- Primary contacts stations to determine if they have data to transmit

Token-Passing Network



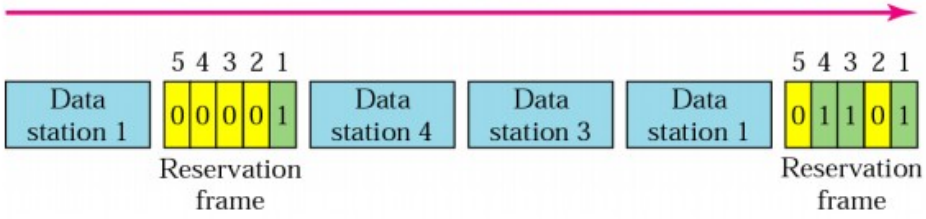
- Token passes around a network
- Machine with token is allowed to transmit data

Static Channel Allocation

- Frequency Division Multiplexing (FDM)
  - N users get 1/N of the total bandwidth
  - << N users ⇨ wasted bandwidth
  - > N users ⇨ denial of service
  - Bursts cannot be accommodated

- Time Division Multiplexing (TDM)
  - N users get full bandwidth 1/N of the time
  - Same arguments apply

Reservation Access Method



- Station that wants to transmit data
  - transmits 1 during its slot in the reservation frame

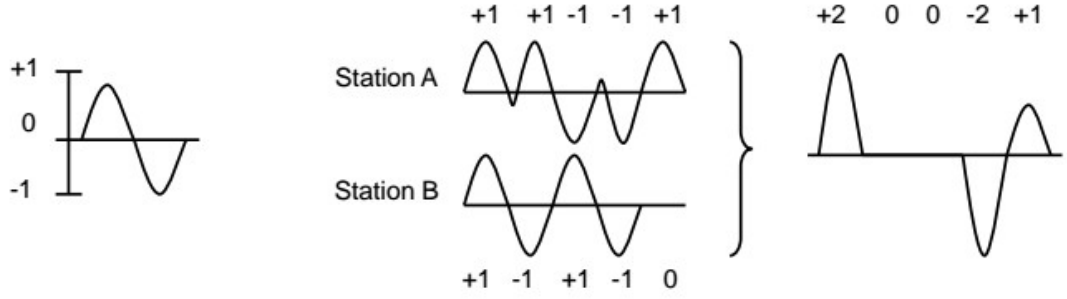
- All stations are informed about all planned communication

- Limited number of pre-allocated slots/stations

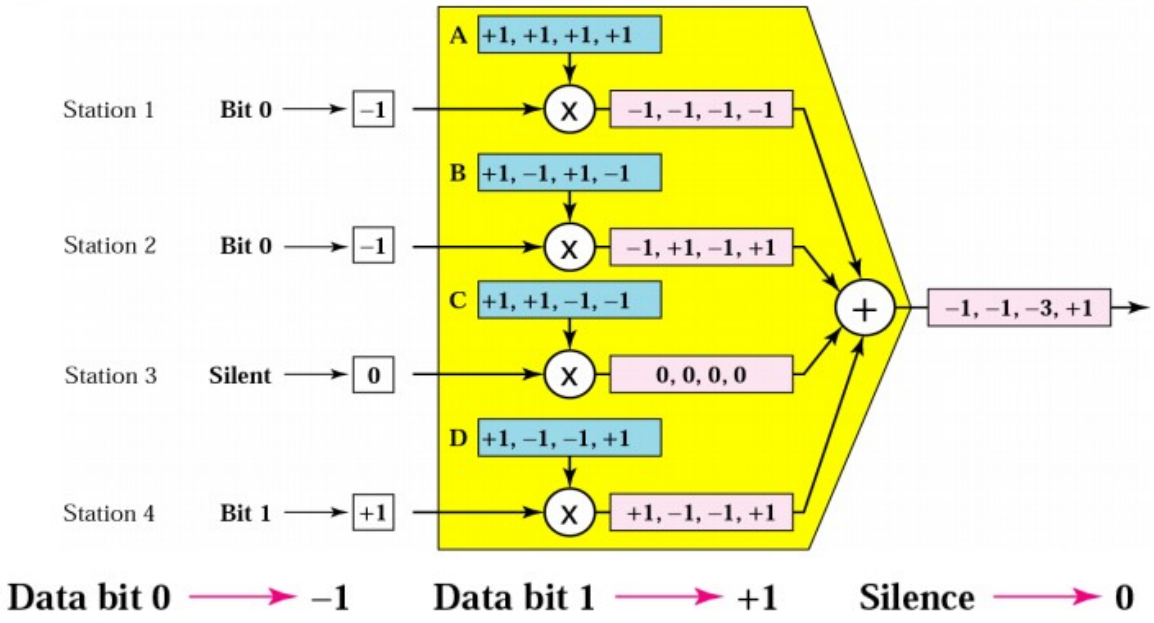
Code Division Multiple Access (CDMA)

- Makes use of physical properties of interference
  - If two stations send signals in phase, they will "add up" to give twice the amplitude
  - If the signals are out of phase, they will "subtract" and give a signal that is the difference

- Difficult to implement because control of exact power strength is essential

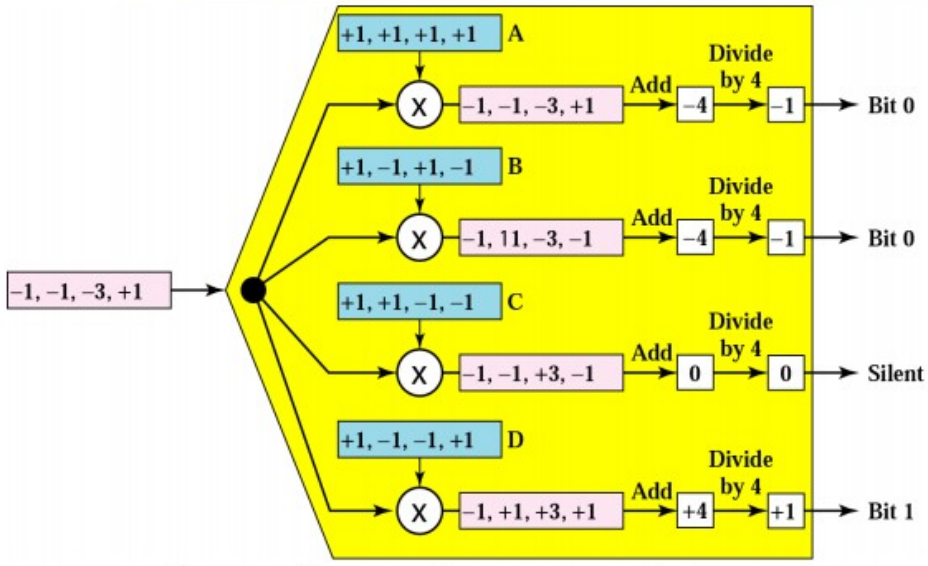


CDMA Multiplexer



Data bit 0 → **–1** **Data bit 1** → **+1** **Silence** → **0**

CDMA De-Multiplexer



Decoding of received signal

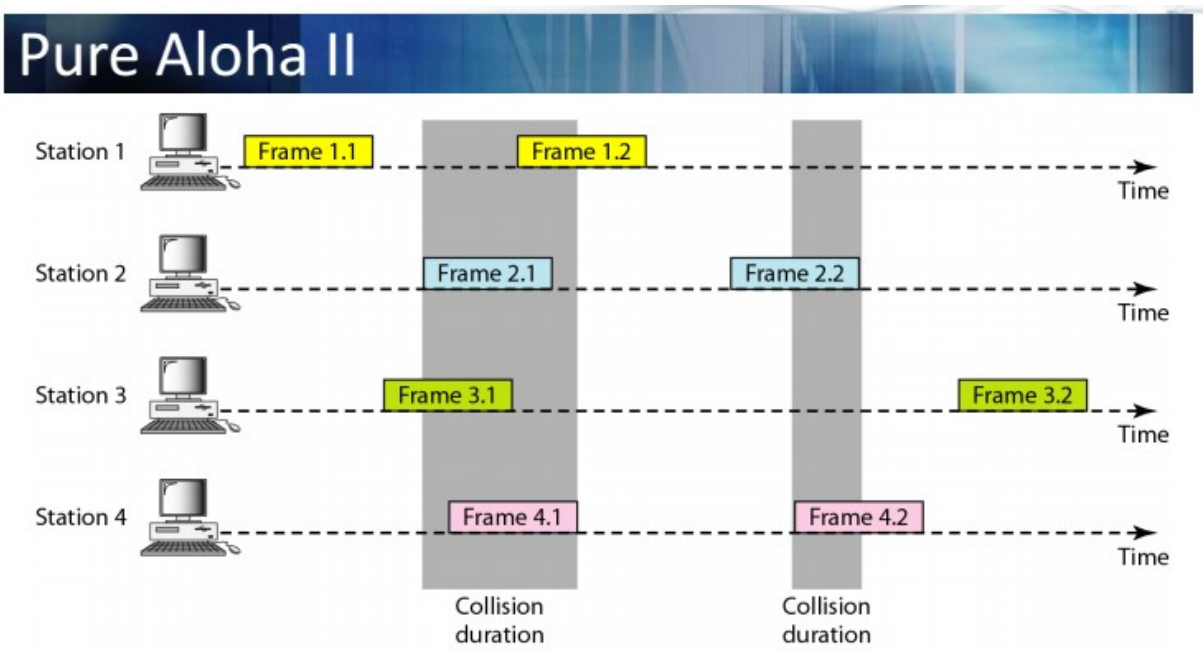
Walsh Tables

$$W_1 = \begin{bmatrix} +1 \end{bmatrix} \qquad W_{2N} = \begin{bmatrix} W_N & W_N \\ W_N & \overline{W_N} \end{bmatrix}$$

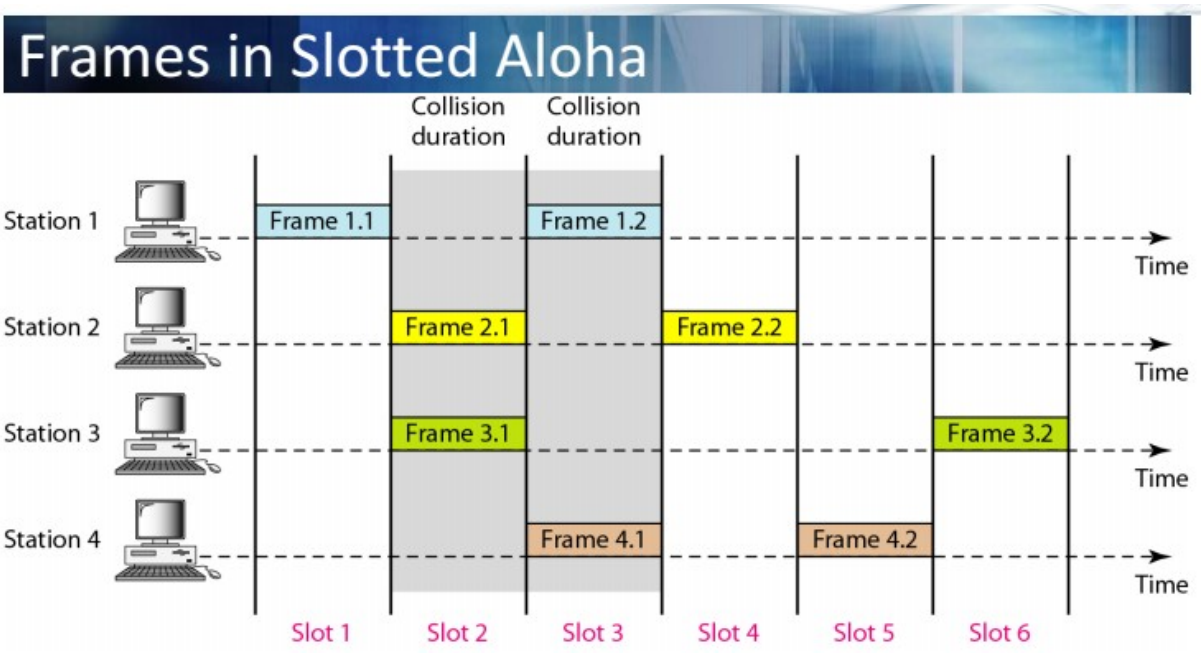
$$W_1 = \begin{bmatrix} +1 \end{bmatrix}$$

$$W_2 = \begin{bmatrix} +1 & +1 \\ +1 & -1 \end{bmatrix}$$

$$W_4 = \begin{bmatrix} +1 & +1 & +1 & +1 \\ +1 & -1 & +1 & -1 \\ +1 & +1 & -1 & -1 \\ +1 & -1 & -1 & +1 \end{bmatrix}$$

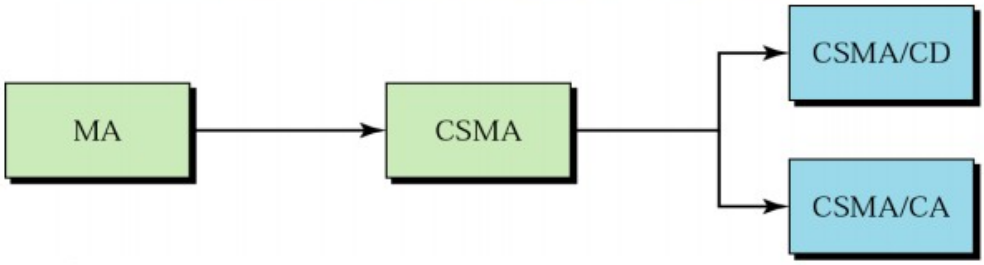


- Collision occurs when frames are transmitted by stations at the same time



Max utilisation is 2x Pure Aloha

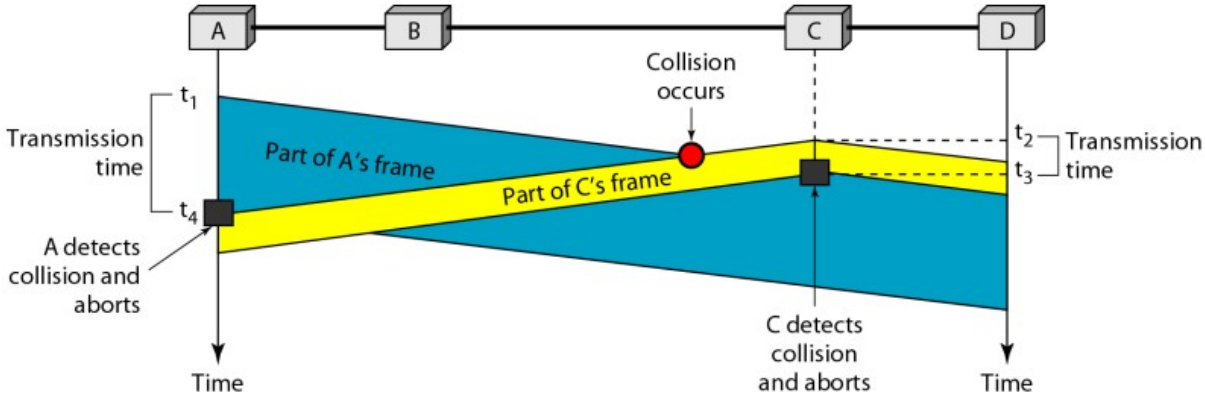
Random-Access Methods



- CS ⇨Carrier Sense
- MA ⇨Multiple Access

- CD ⇨ Collision Detection
- CA ⇨ Collision Avoidance

Collision in CSMA/CD

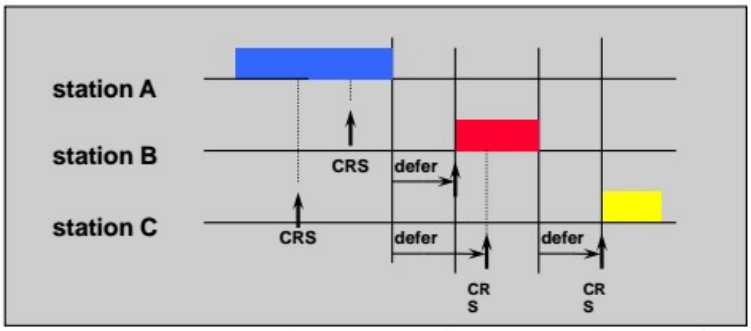


- Both stations will realize that a collision has taken place
- Backoff and attempt to send later

- 1-persistent CSMA
  - if medium idle send imidiately
- p-persistent CSMA
  - if medium available station may send depdending on probability
  - reduces chance of collison and improves efficiency

CSMA in Wireless Media

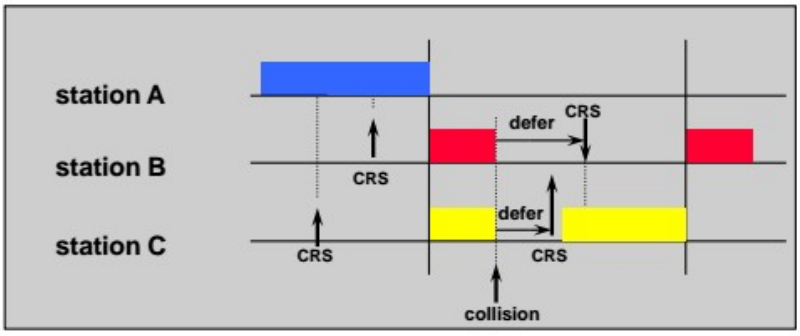
Collision is at the receiver !!!



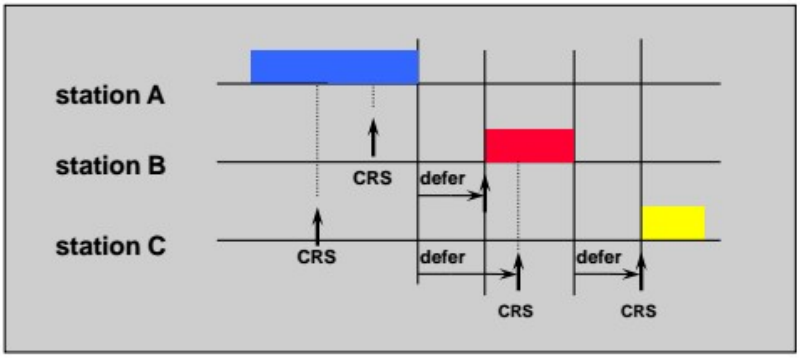
- Sense carrier to determine if medium is free
- Once free pick a random number
  - then start sending

* Figure is courtesy of Avaya Communications Inc

- CSMA/CD



- CSMA/CA

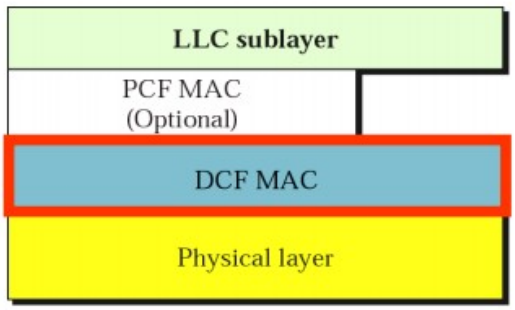


802.11

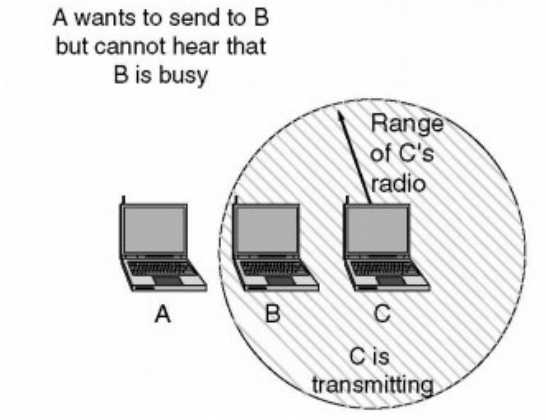
- DCF ⇨ Distributed Coordination Function
  - Stations compete for access to the medium
  - Hidden Station / Expose Station Problem
  - CSMA/CA + RTS/CTS
- PCF ⇨ Point Coordination Function
  - Access point polls stations
- IFS ⇨ Inter-Frame Space
  - Time between frames

Distributed Coordination Function (DCF)

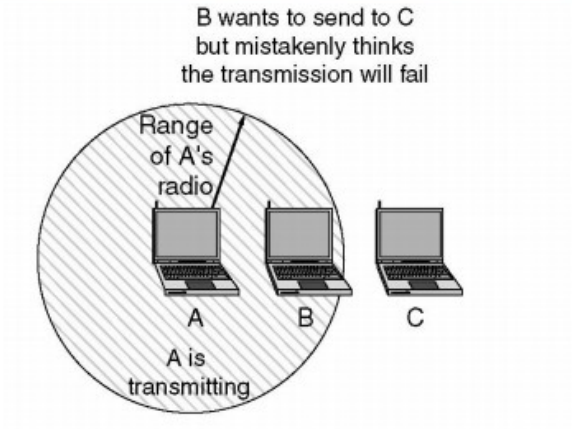
- Stations compete for access to the medium
- Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA)
- No collision detection
  - Not practical on wireless network
  - Transmitting station cannot distinguish incoming weak signals from noise and effects of own transmission



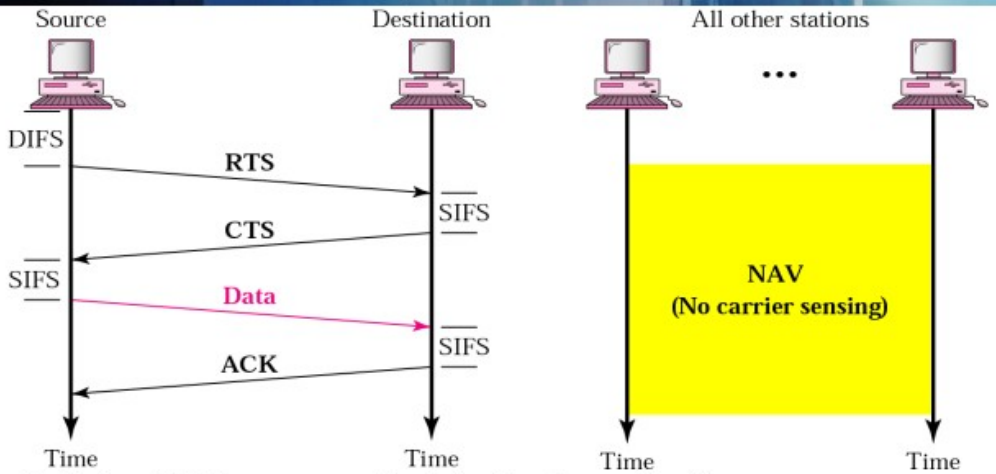
Hidden Station Problem



Exposed Station Problem



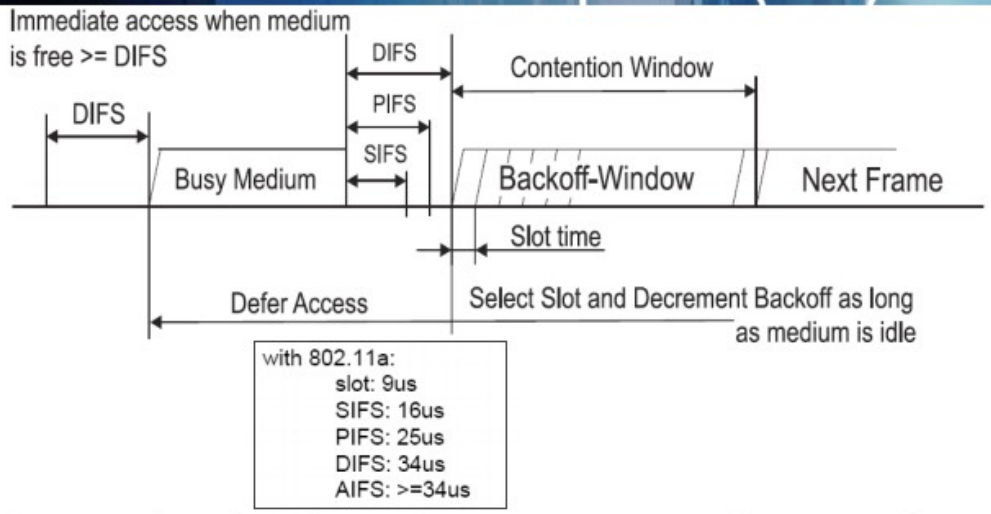
CSMA/CA and NAV



- Ready-To-Send (RTS) announces the intention to send traffic
- Clear-To-Send (CTS) announces that the receiving station is ready
- SIFS is the smallest possible Inter-Frame Space that separates two transmissions
- The Network Allocation Vector (NAV) as part of RTS/CTS announces the length of the subsequent transmission

* Figure is courtesy of B. Forouzan

Inter-Frame Space (IFS)

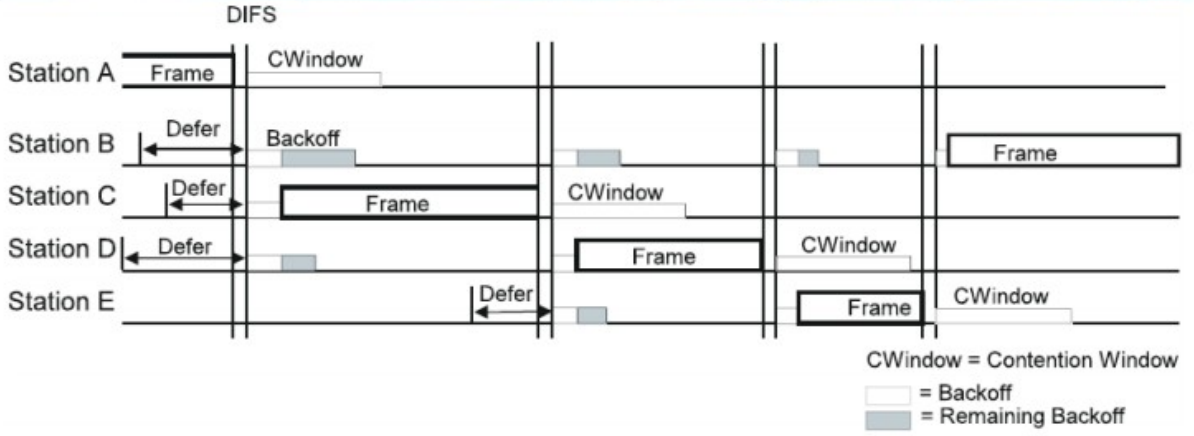


- Short IFS (SIFS) defines the minimum time between frames
- DCF IFS (DIFS) defines the time between the end of one transmission and the beginning of a subsequent transmission

802.11 MAC

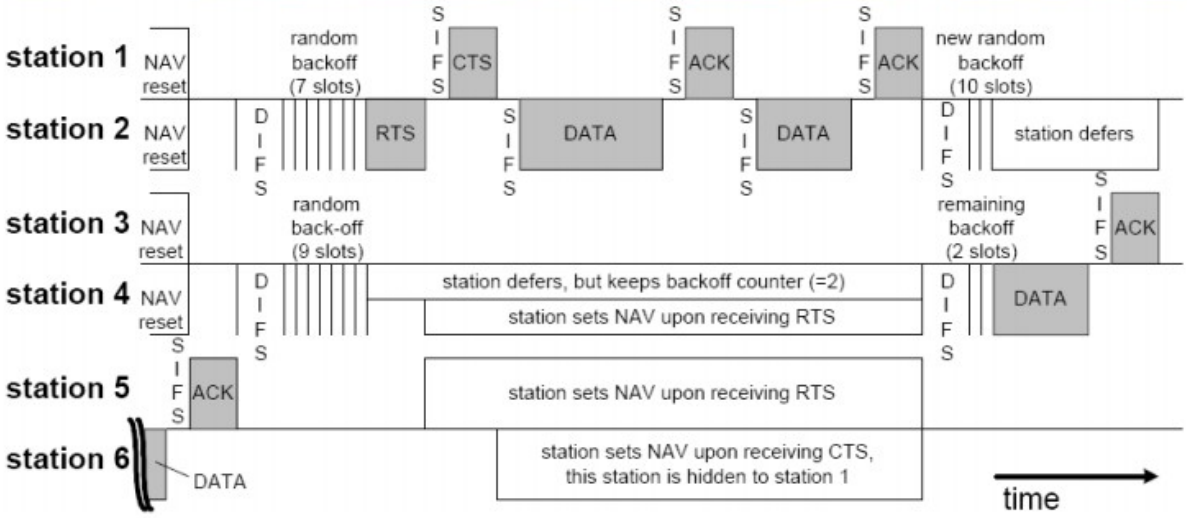
- Station with frame senses medium
  - If idle, wait to see if remains idle for one IFS.
  - If so, may transmit immediately
- If busy - either initially or becomes busy during IFS - station defers transmission
  - Continue to monitor until current transmission is over
- Once current transmission over, delay for another IFS
  - If remains idle, back off random time and again sense
  - If medium still idle, station may transmit
  - During backoff time, if becomes busy, backoff timer is halted and resumes when medium becomes idle

Contention & Backoff

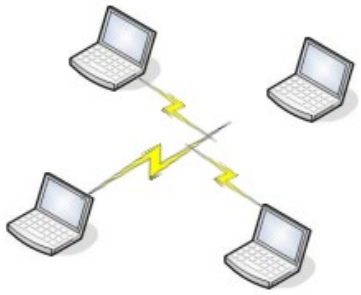


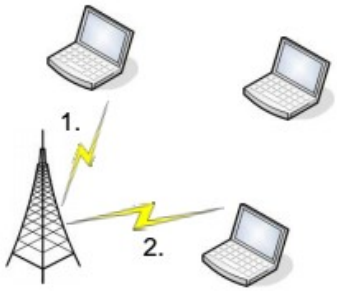
- DIFS defines the minimum time between the end of one transmission and the beginning of a subsequent transmission
- All stations that want to send sense the medium
- Once the sending station is silent all stations start their DIFS timer
- After the DIFS timer every station starts a random exponential backoff

DCF & RTS/CTS



DCF vs. PCF





- Stations compete for the medium

- Point coordinator polls stations

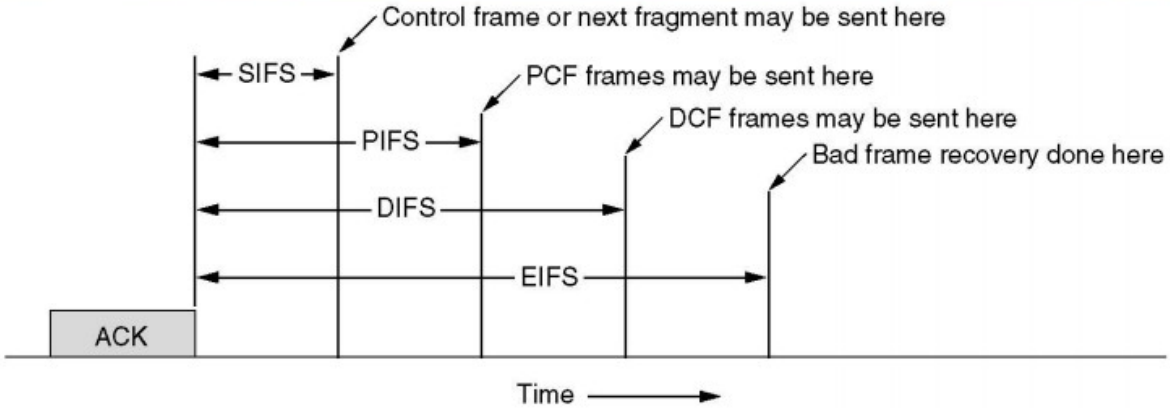
Point Coordination Function (PCF)

- Used by access points
- Polling by centralized polling master – or point coordinator
- Uses PIFS
  - PIFS smaller than DIFS
  - Gives coordinator priority over individual stations
- Point coordinator polls in round-robin to stations configured for polling
  - When poll issued, polled station may respond within SIFS
  - If point coordinator receives response, it issues another poll

PCF

- Time= Contention Period + Contention Free Period
  - Contention Period: All stations compete for the medium
  - Contention Free Period: The AP coordinates communication

IFS in 802.11



- SIFS influences replies
- PIFS gives PCF priority over DCF
- DIFS is the time between two DCF communications

Control Frames

- Assist in reliable data delivery
- Power Save-Poll (PS-Poll)
  - Sent by any station to station that includes AP
  - Request AP transmit frame buffered for this station while station in power-saving mode
- Request to Send (RTS)
  - First frame in four-way frame exchange
- Clear to Send (CTS)
  - Second frame in four-way exchange
- Acknowledgment (ACK)
- Contention-Free (CF)-end
  - Announces end of contention-free period part of PCF
- CF-End + CF-Ack:
  - Acknowledges CF-end
  - Ends contention-free period and releases stations from associated restrictions

Data Frames – Data Carrying

- Eight data frame subtypes, in two groups
- First four carry upper-level data from source station to destination station
- Data
  - Simplest data frame
  - May be used in contention or contention-free period
- Data + CF-Ack
  - Only sent during contention-free period
  - Carries data and acknowledges previously received data
- Data + CF-Poll
  - Used by point coordinator to deliver data
  - Also to request station send data frame it may have buffered
- Data + CF-Ack + CF-Poll
  - Combines Data + CF-Ack and Data + CF-Poll

Management Frames

- Used to manage communications between stations and APs
- E.g. management of associations
  - Requests, response, reassociation, dissociation, and authentication

Summary: 802.11

- Hidden Station / Expose Station

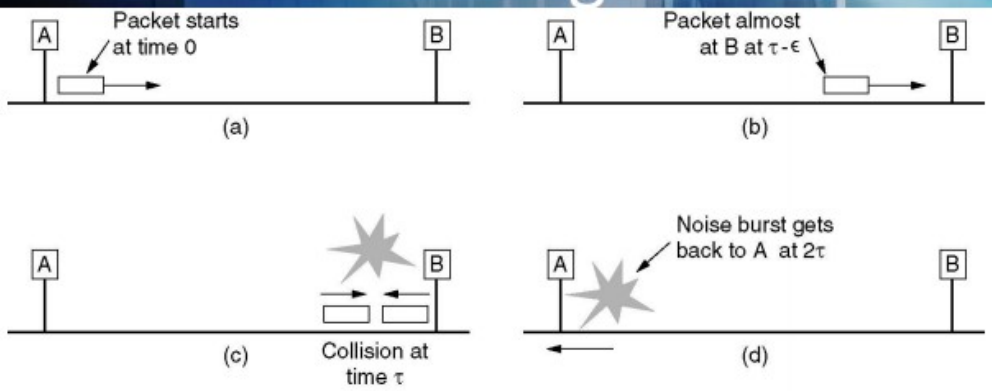
- DCF ⇨ Distributed Coordination Function
  - Stations compete for access to the medium
- CSMA/CA + RTS/CTS

- PCF ⇨ Point Coordination Function
  - Access point polls stations

- IFS ⇨ Inter-Frame Space
  - Time between frames

- Three types of frames:
  - Control frames
  - Data frames
  - Management frames

Frame Length II



- It takes at 2τ to detect a collision
- Roundtrip time = 100μsec
- 10 Mbit/s ⇒ 500 bits

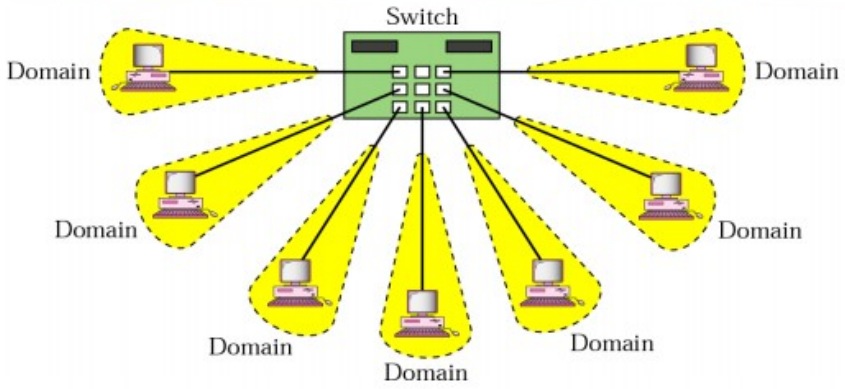
  ~512 bits or 64 bytes

Ethernet Addresses

- Types of Addresses:
  - Unicast – delivered to one station
    - 00-10-4B 3Com 3C905-TX PCI
    - 00-A0-C9 Intel (PRO100B and PRO100+)
  - Multicast – delivered to a set of stations
    - 01-80-C2-00-00-00 Spanning tree (for bridges)
    - 03-00-00-00-00-01 NETBIOS
  - Broadcast – delivered to all stations

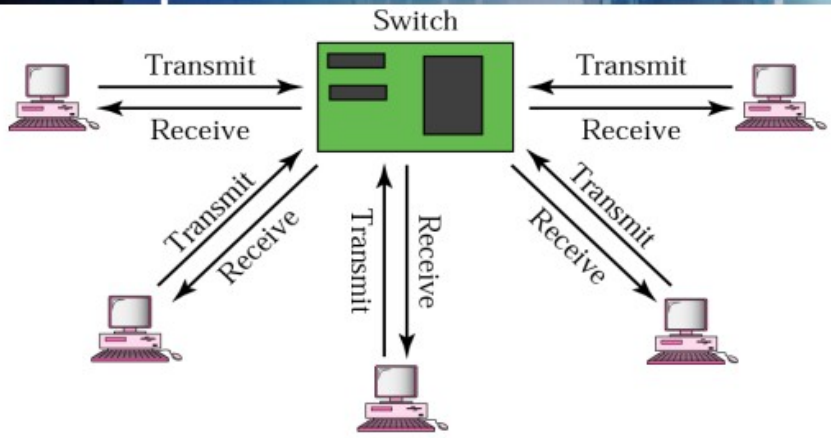
- Extension of Networks:
  - Repeaters, Hubs - Physical Layer
  - Bridges, Switches - Data Link Layer
  - Routers - Network Layer
- Collision domains:
  - Collision affects all machines in one segment

Switched Ethernet



- Switch delivers packets to individual machines
  - Without affecting communication with other machines
- Collisions only occur on individual links

Full-duplex Switched Ethernet

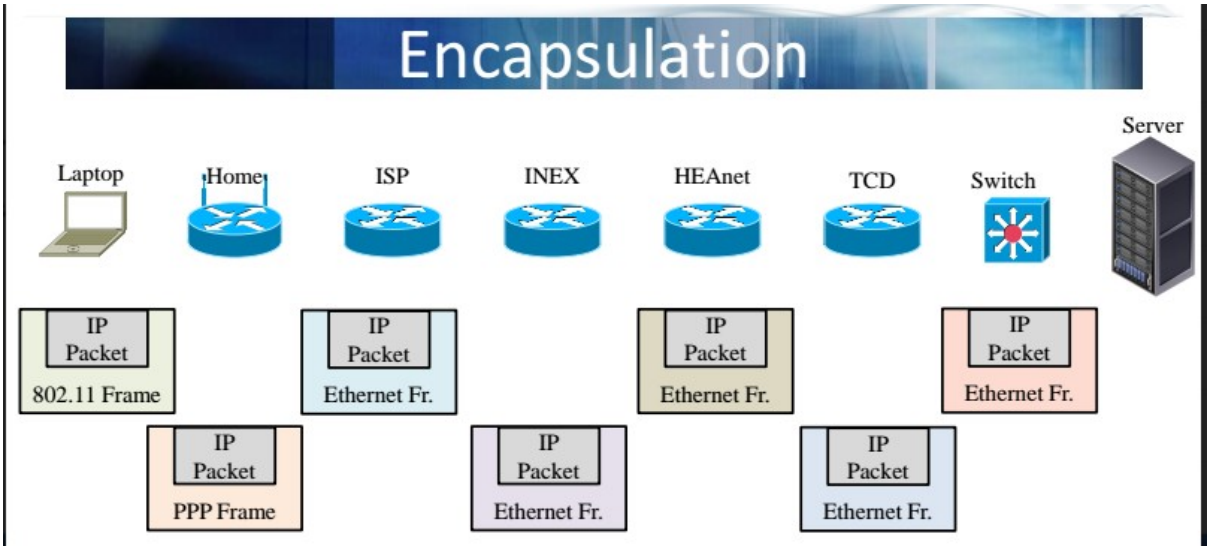


- No collisions
  - One channel to send
  - One channel to transmit

Summary: Ethernet

- Ethernet frame
  - Preamble to signal start of frame
  - MTU & minimum frame size
  - Addressing
- CSMA/CD
- Collision Domains
- Switched Networks

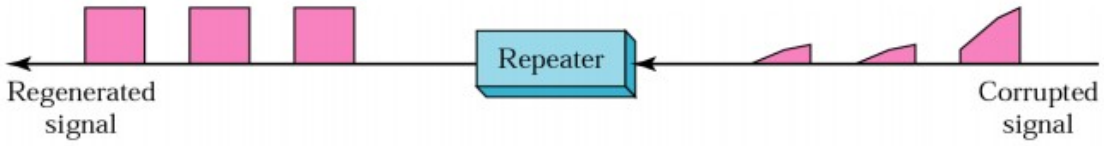
Encapsulation



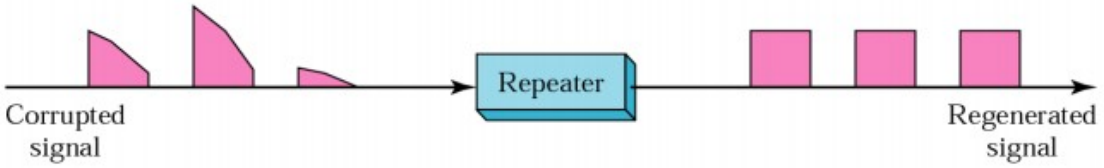
- A repeater connects segments of a LAN
- A repeater forwards every frame; it has no filtering capability

* Figure is courtesy of B. For

Function of a Repeater

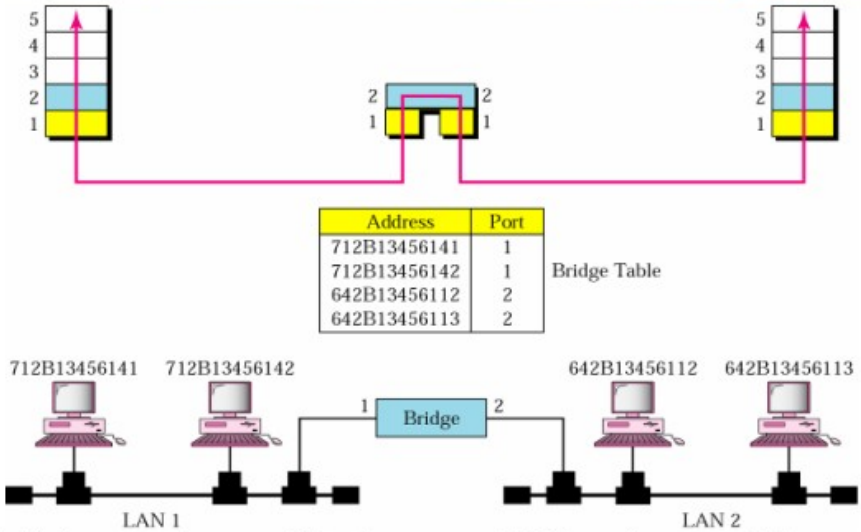


a. Right-to-left transmission.



b. Left-to-right transmission.

Functions of a Bridge

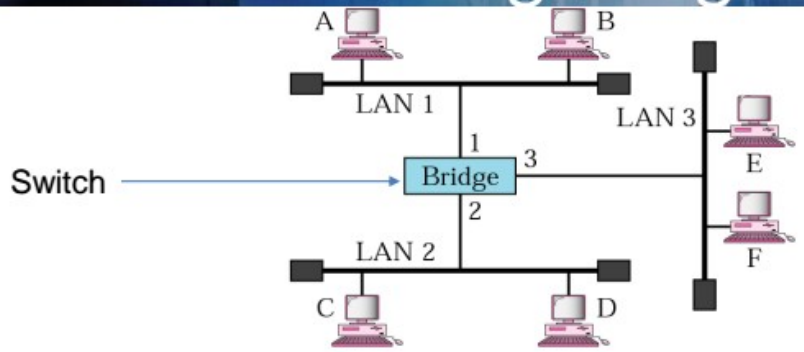


Address	Port
712B13456141	1
712B13456142	1
642B13456112	2
642B13456113	2

Bridge Table

- Read all frames transmitted on one LAN and accept those address to any station on the other LAN
- Using MAC protocol for second LAN, retransmit each frame
- Do the same the other way round

Learning Bridges



Address	Port

a. Original

Address	Port
A	1

b. After A sends a frame to D

Address	Port
A	1
E	3

c. After E sends a frame to A

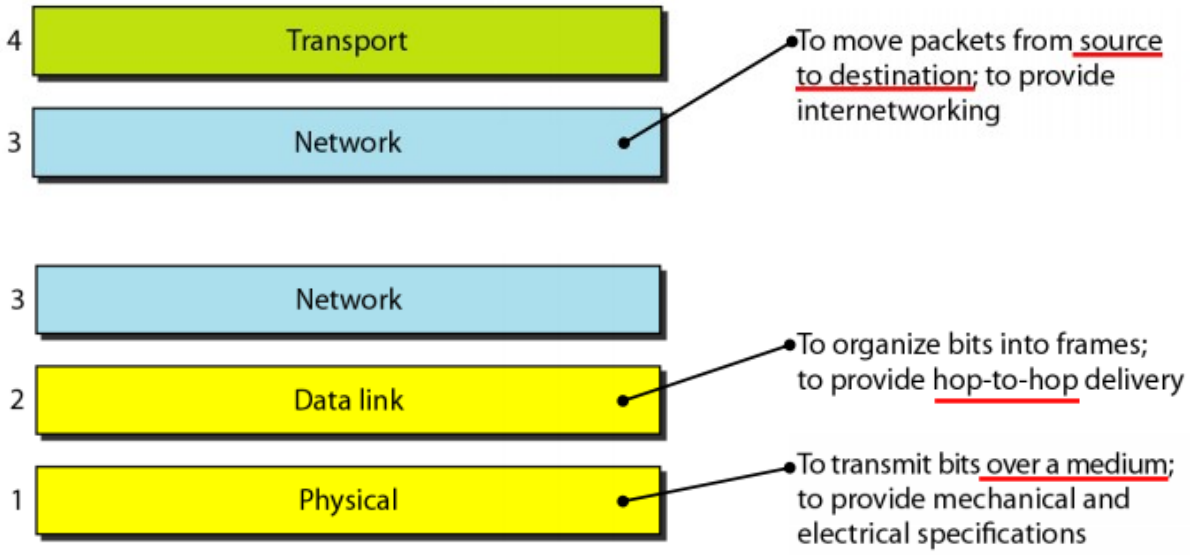
Address	Port
A	1
E	3
B	1

d. After B sends a frame to C

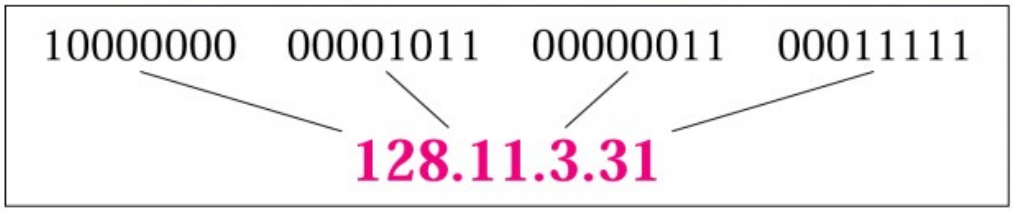
- Initially bridge forwards frames on all segments except incoming port
- Learns addresses from frames that pass through

Types of Layer 2 Switches

- Store-and-Forward switch
  - Accepts frame on input line
  - Buffers it briefly,
  - Then routes it to appropriate output line
  - Delay between sender and receiver
  - Boosts integrity of network
- Cut-Through switch
  - Takes advantage of destination address appearing at beginning of frame
  - Switch begins repeating frame onto output line as soon as it recognizes destination address
  - Highest possible throughput
  - Risk of propagating bad frames
    - Switch unable to check CRC prior to retransmission

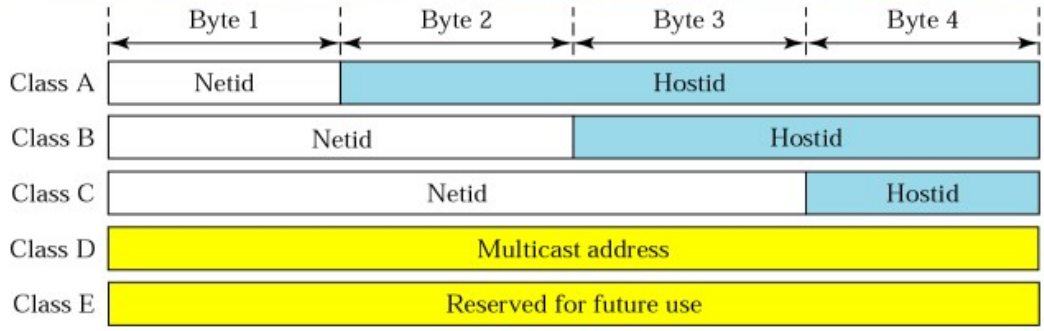


IP Addresses

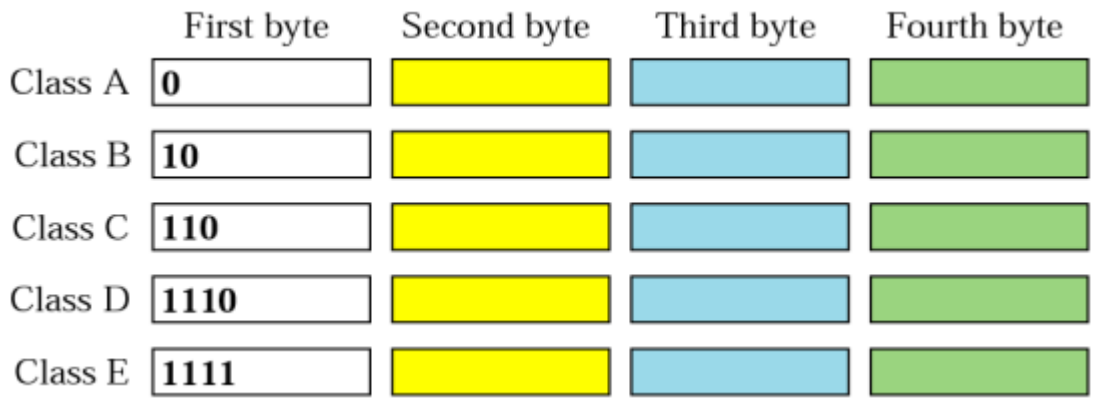


- 32-bit number
  - 4.294.967.296 addresses
- IP addresses are unique and universal
  - with some exceptions
- Dotted decimal notation:
  - Bytes of binary notation represented as decimal separated by dot

Network ID and Host ID



- Network ID: Used to find a particular network
- Host ID: Identifies indentify individual nodes

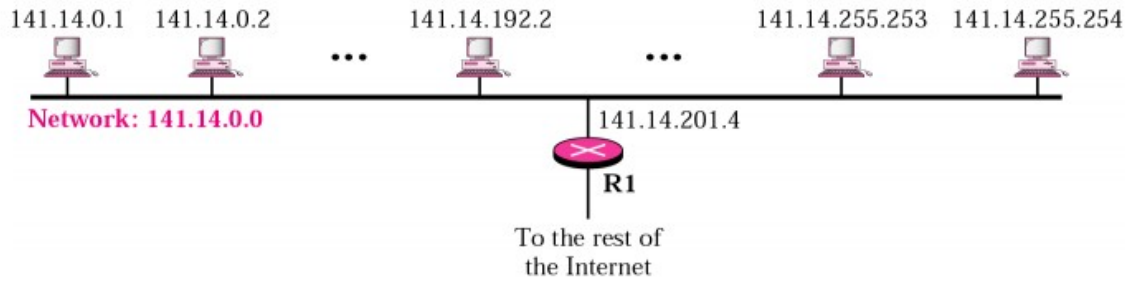


- Class A (international organisations)
  - 126 networks with 16,277,214 hosts each
- Class B (large companies)
  - 16,384 networks with 65,354 hosts each
- Class C (smaller companies)
  - 2,097,152 networks with 254 hosts each

Special Addresses

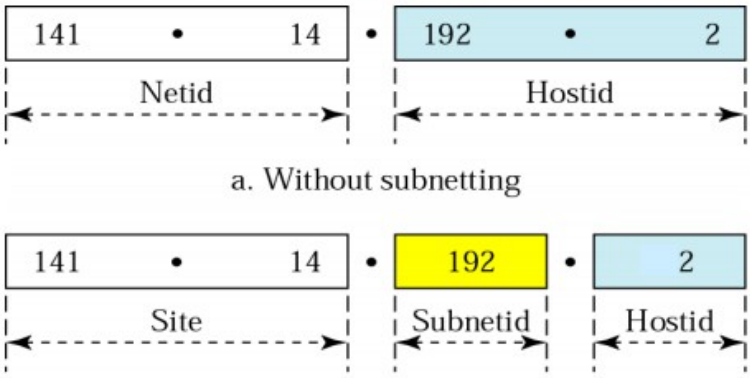
- Loopback device: 127.x.x.x
  - e.g. 127.0.0.1 = localhost
- This network: 0.0.0.x (all zero's)
  - e.g. 0.0.0.54 = host 54 on this network
- Broadcast: x.x.255.255 (all one's)
  - e.g. 134.226.36.255 = all nodes in this network

2-Level Hierarchy with Classful Addresses



- 1st level: NetworkID within the Internet
- 2nd level: HostID within the network

Subnetting

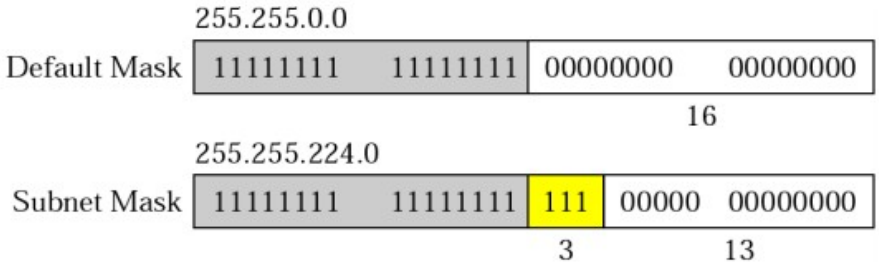


- Add another level to address hierarchy: *subnet*
- Splitting a class B network into a number of class C subnets

Default Masks

Class	In Binary	Dotted-Decimal	Using Slash
A	11111111 00000000 00000000 00000000	255.0.0.0	/8
B	11111111 11111111 00000000 00000000	255.255.0.0	/16
C	11111111 11111111 11111111 00000000	255.255.255.0	/24

- Subnet Masks



- Inefficient use of Hierarchical Address Space
  - Class C with 2 hosts (2/254 = 0.78% efficient)
  - Class B with 256 hosts (256/65534 = 0.39% efficient)

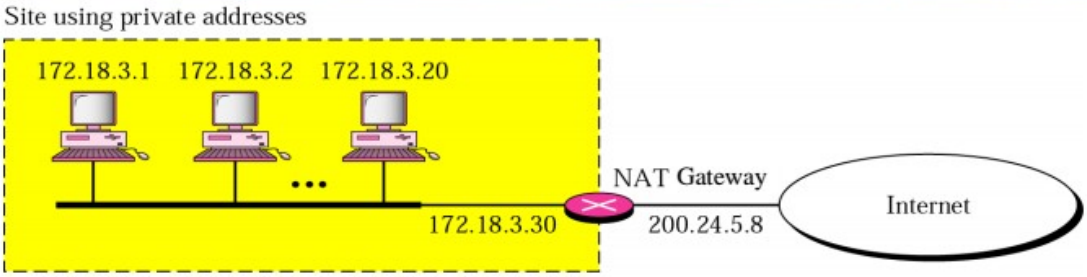
Classless Inter-Domain Routing(CIDR)

- Allow address space to be divided into blocks of addresses
  - only limited to the power of 2
- Notation as decimal number of the significant bits e.g. 134.226.36.0 /29
- 205.16.37.32/28
  - 32 bits – 28 bits are static - 4 bits are varied

	Dotted Decimal	32-bit binary equivalent
Lowest	128.211.168.0	10000000 11010011 10101000 00000000
Highest	128.211.175.255	10000000 11010011 10101111 11111111

= 128.211.168.0/21

Network Address Translation (NAT)



- NAT gateway translates traffic from the local network to the IP address of the gateway

- Involves processing of outgoing & incoming packet e.g. translation between addresses, recalculation of checksums, etc

Private Address	Private Port	External Address	External Port	Transport Protocol
172.18.3.1	1400	25.8.3.2	80	TCP
172.18.3.2	1401	25.8.3.2	80	TCP
...	...	...	...	...

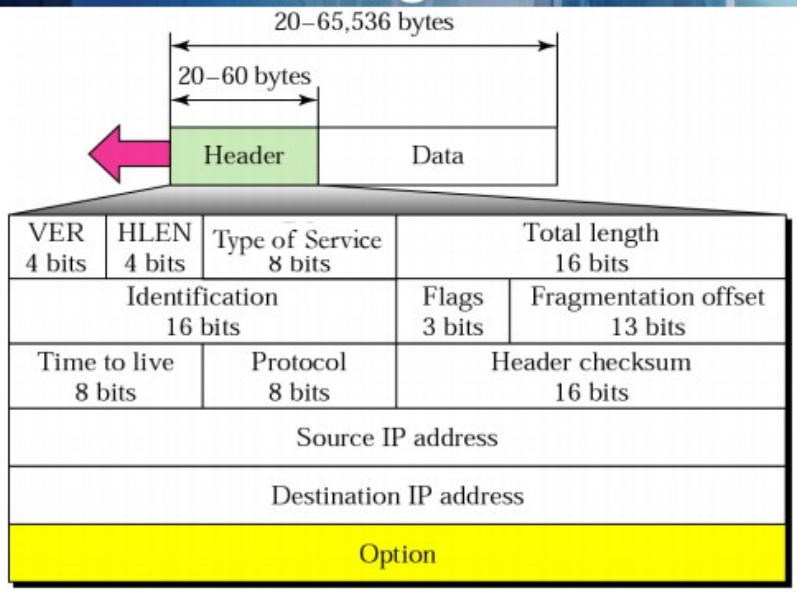
- Gateway maintains table to match incoming and outgoing packets; including IDs for applications

Summary: Addresses

- 32-bit number / Dotted decimal notation
- IP addresses are unique and universal
  - Exception: Private Addresses

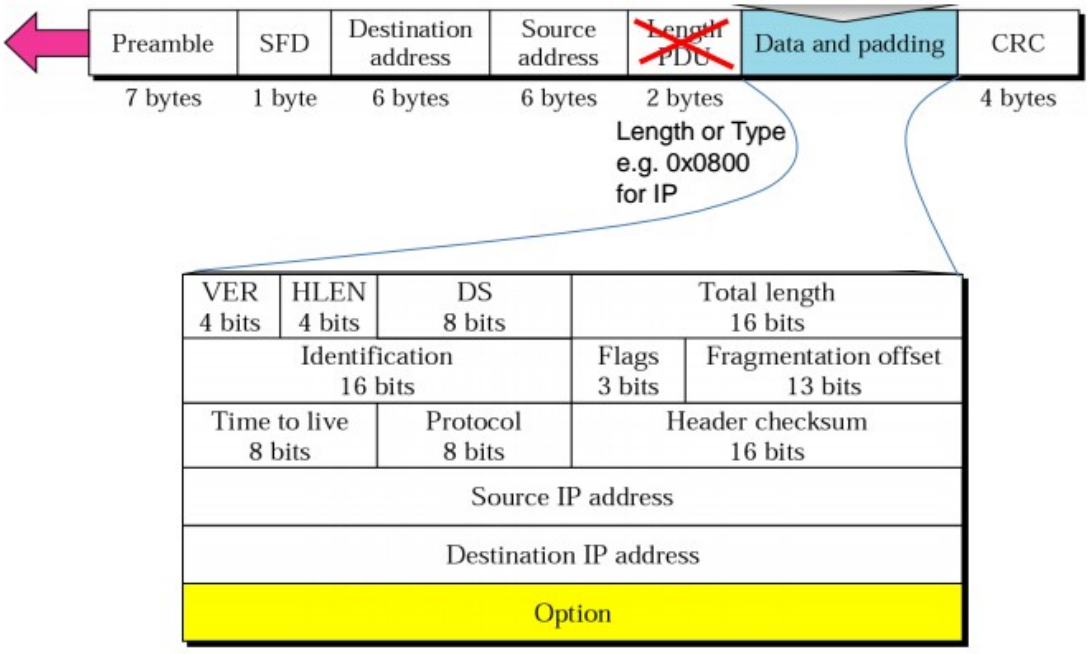
- Classful addresses
  - Classes A, B, and C for networks, D for multicast
  - Routing on Network IDs
- Subnetting + Netmasks
  - Dealing with scale in local networks
- Classless Inter-Domain Routing (CIDR)
  - / notation – significant bits of address
- Network address translation (NAT)

IP Datagram

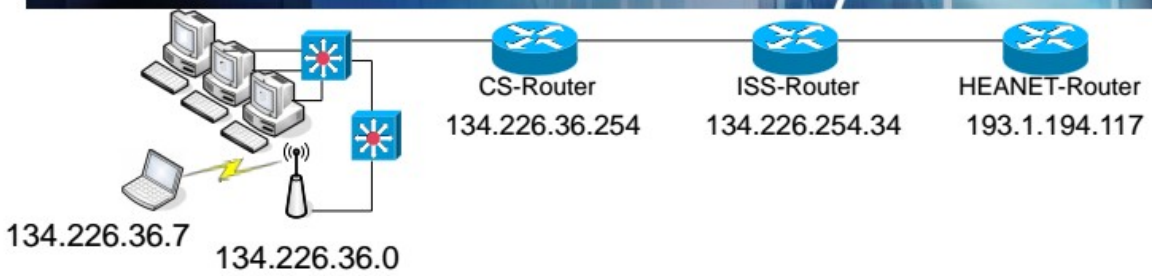


- The total length field defines the total length of the datagram including the header.

Ethernet & IP



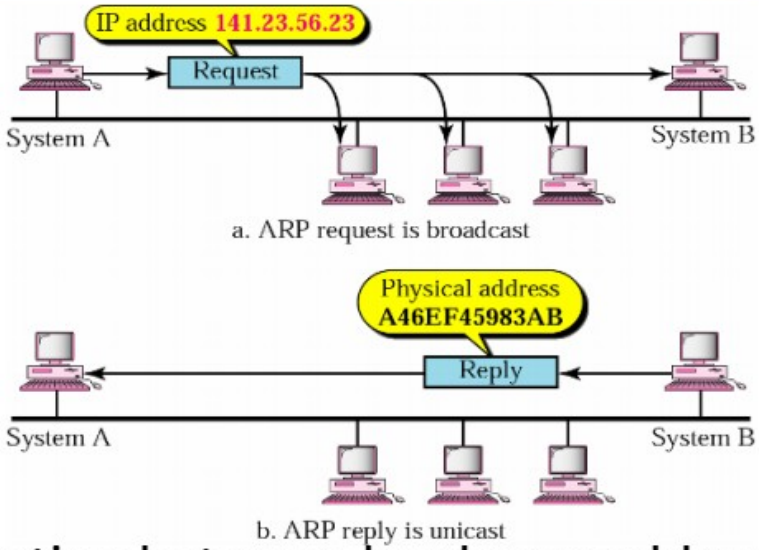
Default Gateway



Subnet	Gateway	Netmask	Interface
134.226.36.0	0.0.0.0	255.255.0.0	eth1
0.0.0.0	134.226.36.254	0.0.0.0	eth1

- All nodes within the subnet can communicate directly with each other
- All communication with nodes in other networks passes through the default gateway e.g. router 134.226.36.254

Address Resolution Protocol (ARP)

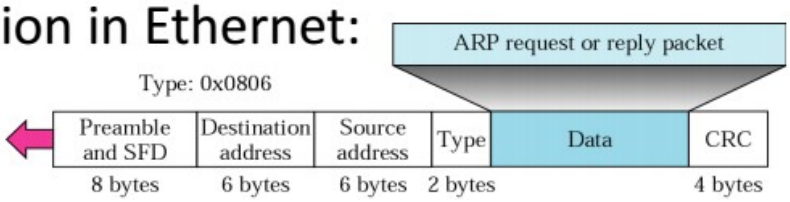


- Association between hardware address and IP address

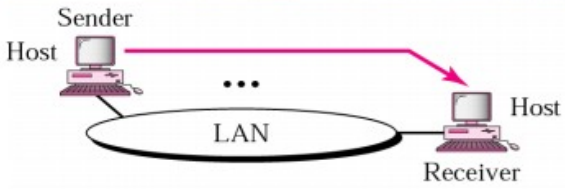
ARP Packet

Hardware Type		Protocol Type
Hardware length	Protocol length	Operation Request 1, Reply 2
Sender hardware address (For example, 6 bytes for Ethernet)		
Sender protocol address (For example, 4 bytes for IP)		
Target hardware address (For example, 6 bytes for Ethernet) (It is not filled in a request)		
Target protocol address (For example, 4 bytes for IP)		

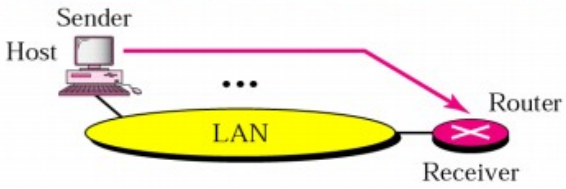
- Encapsulation in Ethernet:



4 Cases for ARP

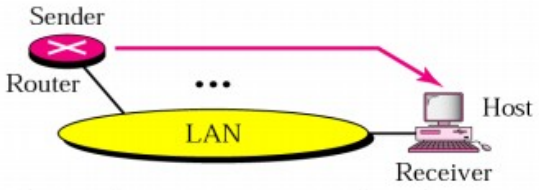


Case 1. A host has a packet to send to another host on the same network.



Case 2. A host wants to send a packet to another host on another network.
It must first be delivered to the appropriate router.

Case 3. A router receives a packet to be sent to a host on another network.
It must first be delivered to the appropriate router.

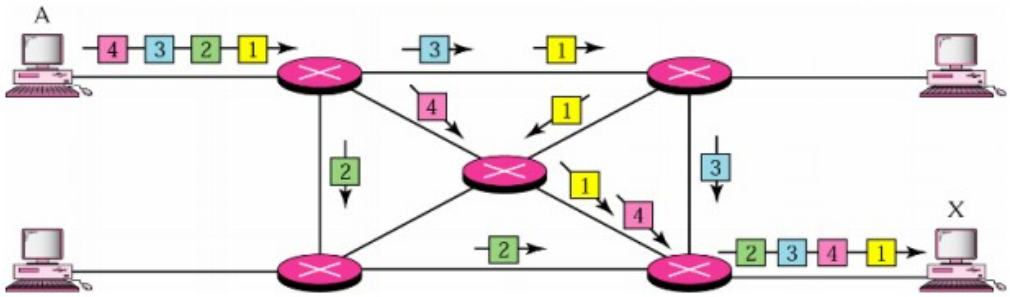


Case 4. A router receives a packet to be sent to a host on the same network.

Summary: Network Layer&Addresses

- Doted-decimal notation
- Classful addresses
  - Classes A, B, and C for networks, D for multicast
- Subnetting
- Classless Inter-Domain Routing (CIDR)
  - / notation = significant bits in subnet mask
- Network address translation (NAT)

IP Service Model

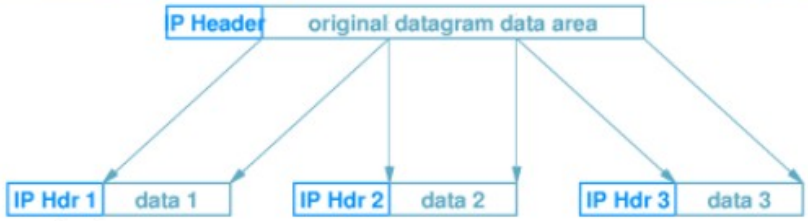


- Connection-less Communication
  - No state is kept about individual packets
- Order not guaranteed
- Best-effort delivery (unreliable service)
  - Packets may be lost
  - Packets may be delivered out of order
  - Duplicate copies of a packet may be delivered
  - Packets can be delayed for a long time

Maximum Transmission Unit (MTU)

- Maximum size of a data unit depends on underlying hardware architecture
- Maximum frame size determines *Maximum Transmission Unit (MTU)*

Fragmentation



- Possible techniques:
  - Limit datagram size to smallest MTU of any network
  - Adjust datagram size as packet progresses through networks
    - Router detects datagram larger than network MTU and splits into pieces

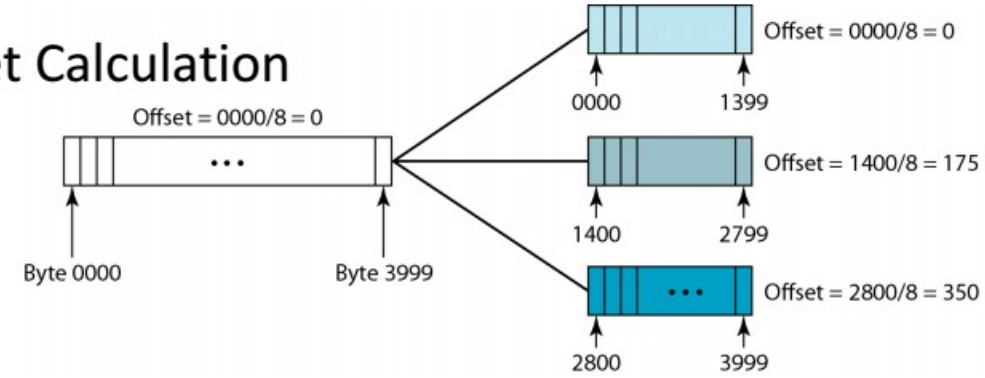
- IP Strategy
  - Fragment when necessary (Datagram > MTU)
  - Try to avoid fragmentation at source host
  - Re-fragmentation is possible
  - Delay reassembly until destination host
  - Do not recover from lost fragments
    - If one fragment is lost all fragments are discarded

- Each fragment is an independent datagram
  - Includes all header fields
  - Bit in header indicates datagram is a fragment
  - Other fields have information for reconstructing original datagram
  - FRAGMENT OFFSET gives original location of fragment

Header Fields

- "Do not fragment"-Request
- More Fragments

- Offset Calculation

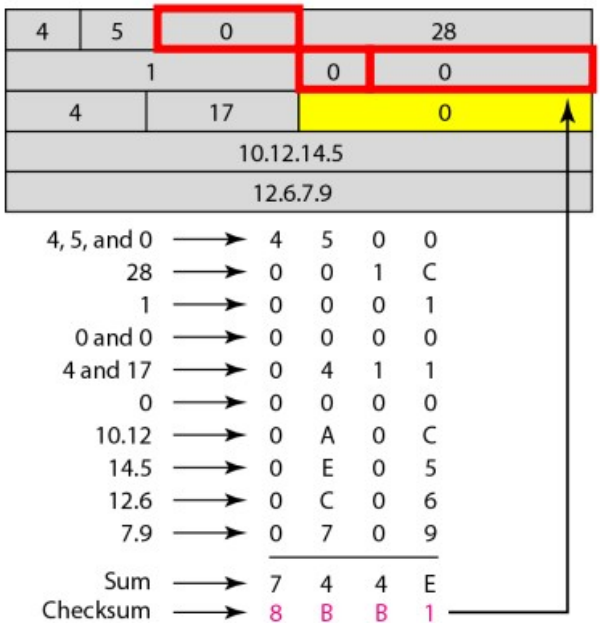


Fragment loss

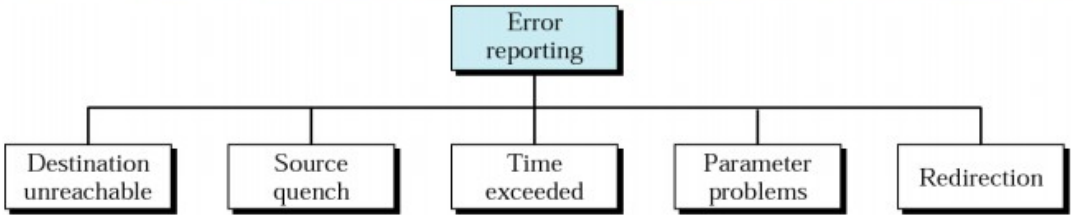
- A fragment may be lost/dropped in transfer
- What happens to original datagram?
  - Destination drops entire original datagram
- How does destination identify lost fragment?
  - Sets timer with each fragment
  - If timer expires before all fragments arrive, fragment assumed lost
  - Datagram dropped
- Source is assumed to retransmit

Checksum in IPv4 Header

- Calculation omits:
  - Service field
  - Fragment fields and offset
  - Checksum field



Internet Control Message Protocol



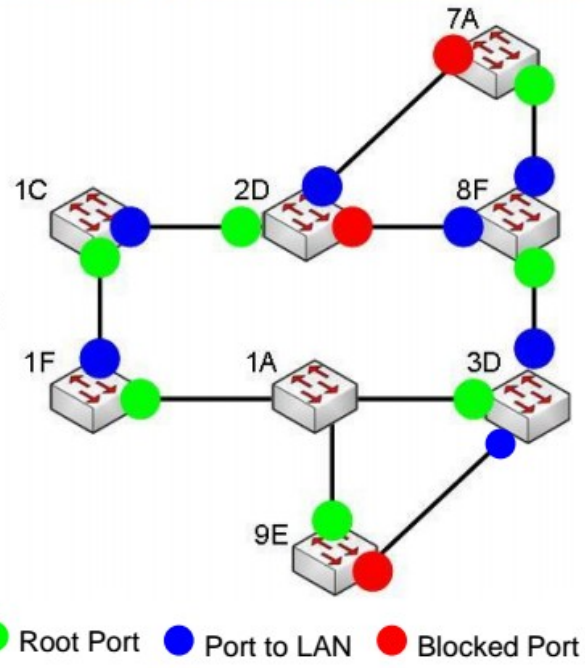
- Destination unreachable
- Source quench
  - Mechanism for destination and intermediate nodes to limit traffic from source
- Time exceeded
  - Send when datagram discarded due TTL value of 0
- Parameter problems
  - Send when datagram discarded due to parameter ambiguity
- Redirection
  - Send from router to update routing table of source

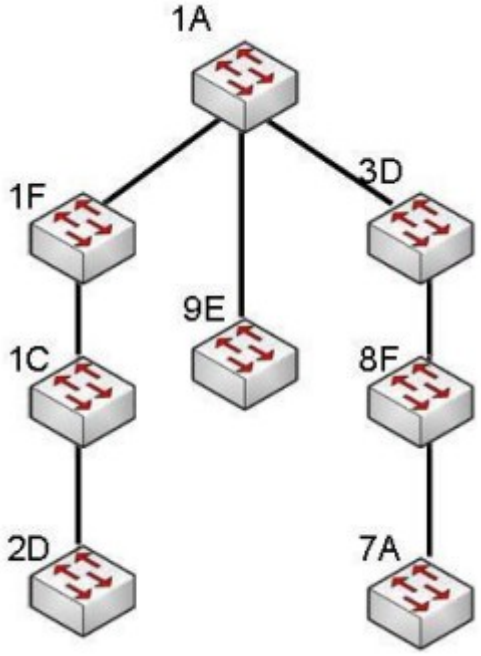
Summary: Internet Protocol

- IP Service Model
  - Connection-less, no order guaranteed
- IP Header
  - 20 bytes + options
- Fragmentation
  - Datagrams split into fragments to fit MTUs
  - Only re-assembled at destination
- Internet Control Message Protocol (ICMP)
  - Error Reporting e.g. source quench
  - Querying e.g. Ping

Spanning Tree Algorithm

1. Bridge with smallest ID is selected as root bridge
2. Mark port on each bridge with least-cost to root bridge as root port
3. Select designated bridge for each LAN that has root port with least-cost to root bridge – if two bridges with same cost select bridge with lowest ID
4. Mark root ports and designated ports as forwarding ports; other ports as blocking ports




1A
1F
3D
9E
1C
8F
2D
7A

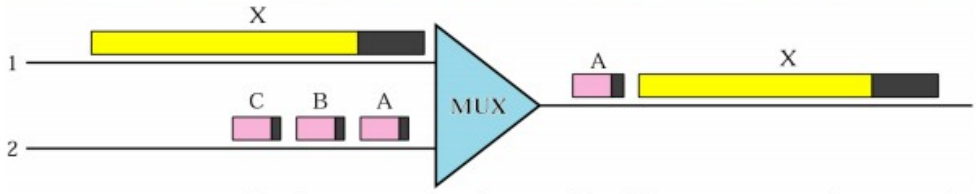
- Packet Switching:
  - Switching decisions are made on individual packets
- Virtual Circuit Switching:
  - A circuit is setup explicitly for individual connections

Packet switching	Virtual Circuits
• Frames can be transferred over different paths in the network • Reliability is generally delegated to higher layers • Order is not necessarily maintained	• Connection-oriented communication • Connection is established before communication • The network maintains order • Three phases – Connection setup – Data Transfer – Connection termination

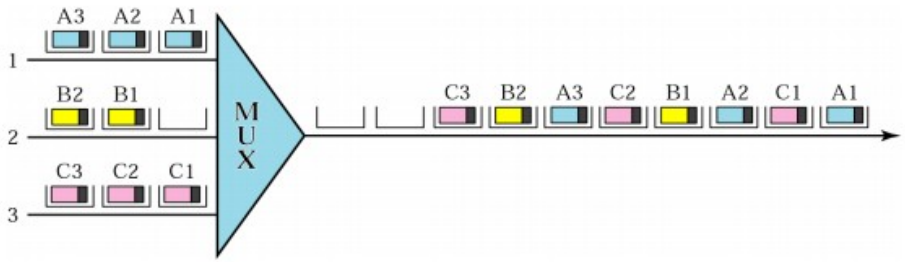
Asynchronous Transfer Mode (ATM)

- Example of virtual circuit switching
  - Cell-Switching
- Similarities between ATM and packet switching
  - Transfer of data in discrete chunks
  - Multiple logical connections over single physical interface
- In ATM flow on each logical connection is in fixed sized packets called cells
- Minimal error and flow control
  - Reduced overhead

Motivation for ATM

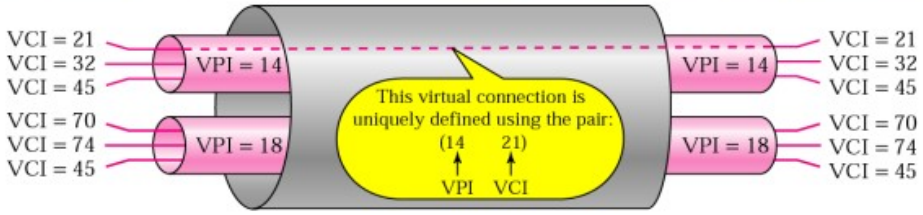


- Frames at a switch may be handled in any order and occupy switch for underspecified time

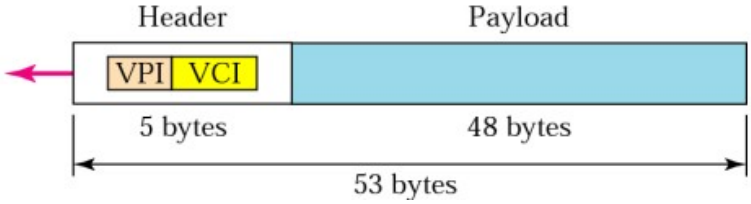


- Small, fixed-size frames allow simple, fast switches

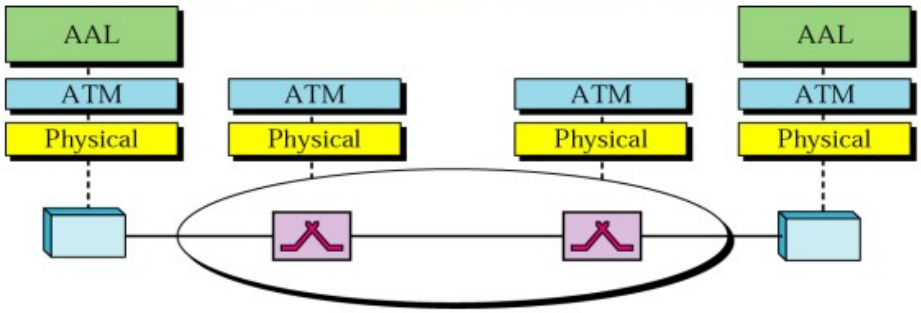
ATM Packet



- Connection is specified by combination of Virtual Path ID and Virtual Circuit ID



Application Adaptation Layer (AAL)



- ATM defined a number of AALs for various purposes (each has its own header format):

ATM – It Didn't Happen

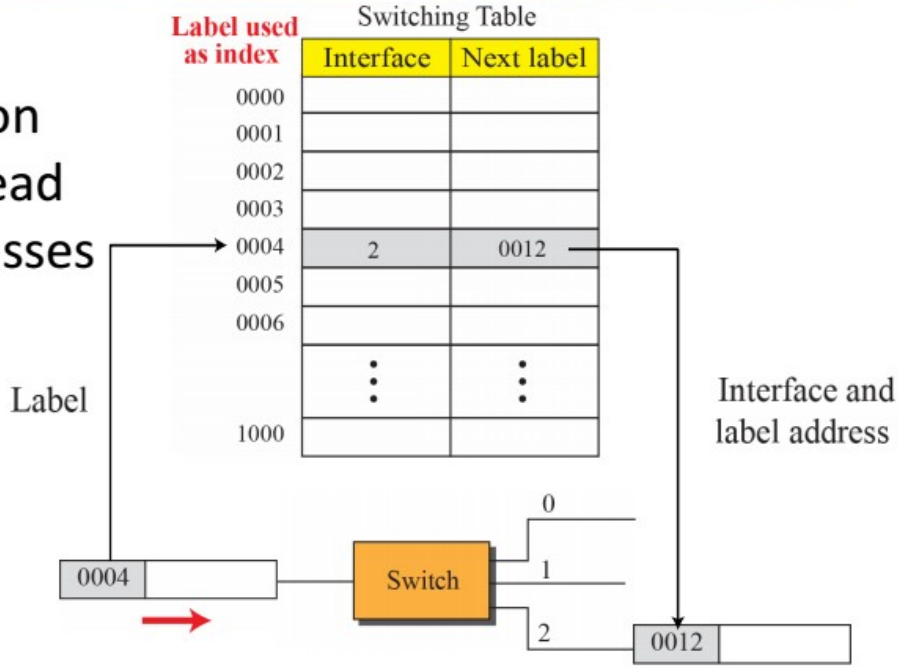
- From Tanenbaum:

  "ATM was going to solve all the world's networking and telecommunications problems by merging voice, data, cable television, telex, telegraph, carrier pigeons, …"

- It didn't happen:
  - Bad Timing
  - Technology
  - Implementation
  - Politics

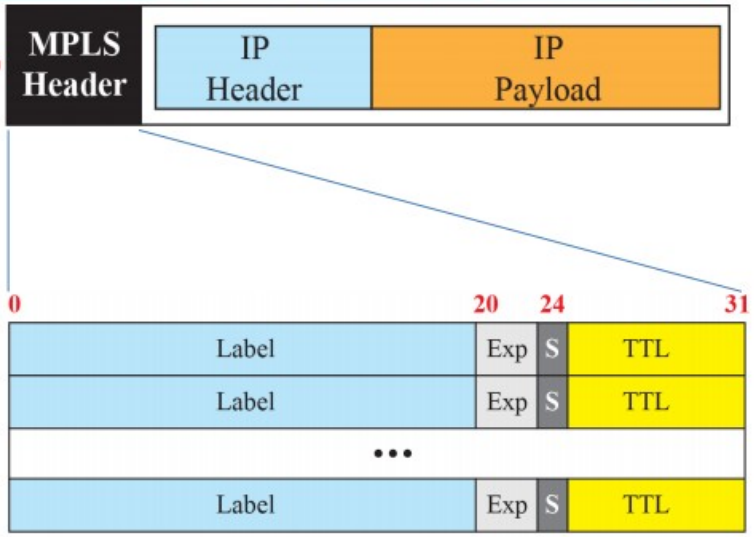
Multiprotocol Label Switching (MPLS)

- Enables switching on labels instead of IP addresses



MPLS Header

- MPLS header as stack of labels



Used in creating a virtual network

Summary:Virtual Circuit Switching – ATM

- Virtual Circuit Switching
  - Preplanned route established before any frames sent
  - Call request and call accept frames establish connection (handshake)
  - Each frame contains a virtual circuit identifier instead of destination address
  - No routing decisions required for each frame
  - Clear request to drop circuit
  - Not a dedicated path
- Asynchronous Transfer Mode (ATM)
  - Example for virtual circuit switching
  - Cells consist of 5-byte header and 48-byte payload
  - Circuits identified by virtual circuit ID and virtual path ID
  - Application adaptation layer (AAL) for specific application areas

- Distance Vector Routing
- Link State Routing
- Multicast Routing
  - Dense Mode (DM)
  - Sparse Mode (SM)

Routers

- One Main Interest
  Forwarding Packets

- Important Aspects
  Queue Length
  Routing Table

Destination	Gateway	Interface
$IP\ Range_1$ $IP\ Range_2$	G_1 G_2	IF_1 IF_2

- Distance Vector routing
  - Routes propagate through exchange of routing tables
  - Based on communication with neighbours

- Link State routing
  - Establishing view of complete topology
  - Makes use of Dijkstra's Shortest-Path Algorithm

- Each node maintains a set of triples
  - **(Destination, Cost, NextHop)**

Destination	Hop Count	Next Router
163.5.0.0	7	172.6.23.4
197.5.13.0	5	176.3.6.17
189.45.0.0	4	200.5.1.6
115.0.0.0	6	131.4.7.19

- Exchange updates with directly connected neighbors
  - periodically (on the order of several seconds)
  - whenever table changes (called *triggered* update)

- Each update is a list of all pairs in the routing table:
  - (**Destination, Cost**)

RIP Updating Algorithm

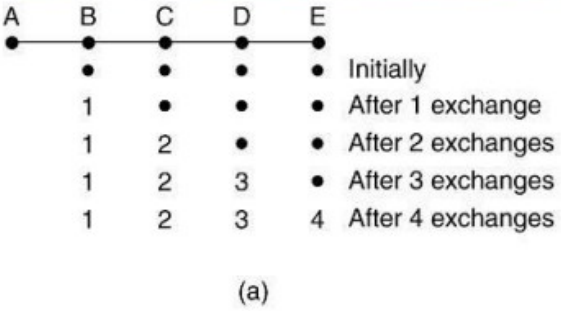
Receive: A response RIP message

- Add one hop to the hop count for each advertised destination.
- Repeat the following steps for each advertised destination:

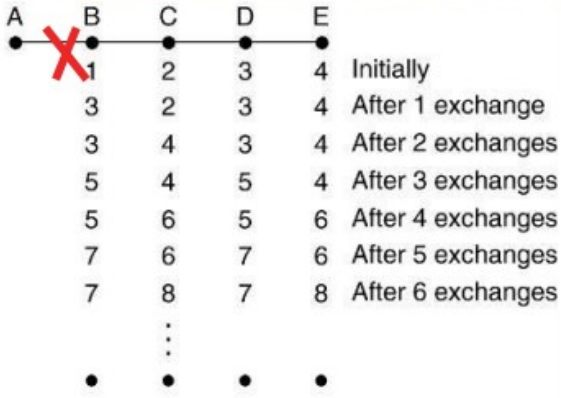
a) If (destination not in the routing table)
Add the advertised information to the table.

b) Else
If (next-hop field is the same)
Replace entry in the table with the advertised one.
Else
If (advertised hop count smaller than one in the table)
Replace entry in the routing table.

Count-to-Infinity Problem



(a)



(b)

- Routers exchange updates
- After 4 steps the network converges
- Every router knows how to get to router A

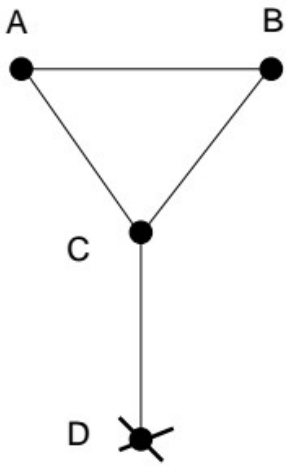
- Link between A and B breaks
- B receives update from C advertising route to A

Solutions for Count-to-Infinity

- Possible solutions:
  1. Impose upper bound on maximum distance
  2. *Split horizon*
     - C should not send to node B its new distance to node A, if node B is C's next-hop towards A
  3. *Split horizon with poisoned reverse*
     - C should tell node B that its distance to node A is ∞, when node B is C's next-hop towards A

- Unfortunately, none of these solutions can deal with arbitrary topology cycles

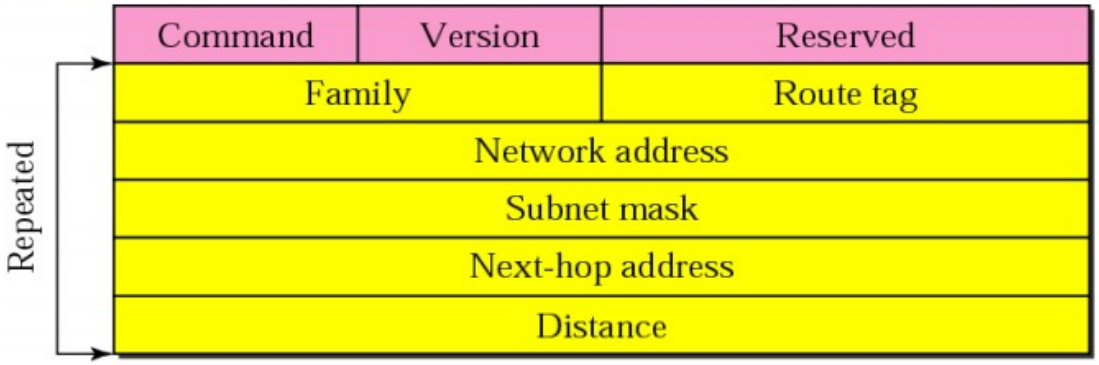
Split Horizon Failure



If D goes down, A and B will still count to infinity.

Split-Horizon infinity messages are sent from A->C and B->C, not A<->B

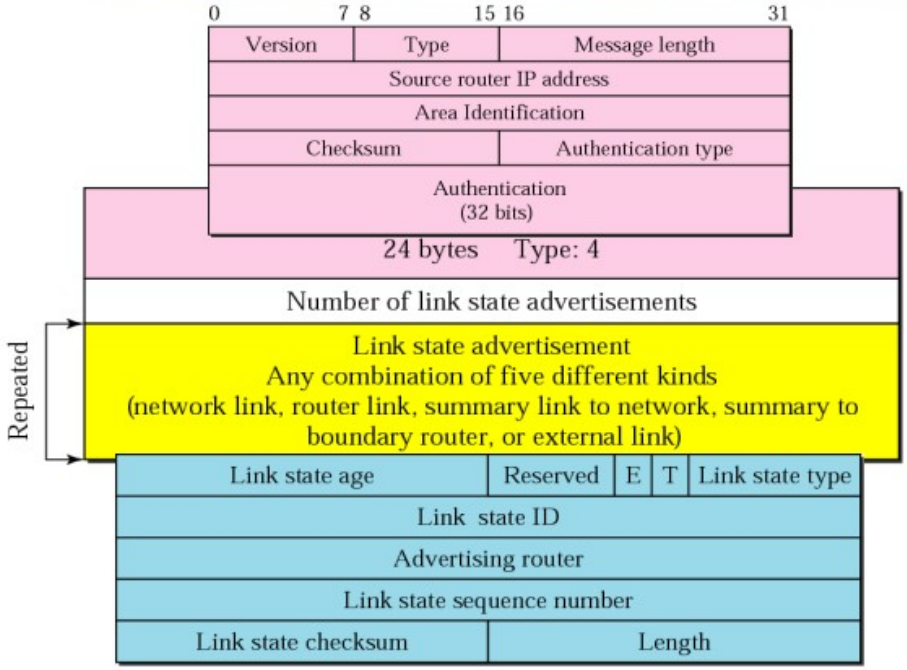
RIPv2 Message Format



Dijkstra's Algorithm

1. Start with the local node (router): the root of the tree.
2. Assign a cost of 0 to this node and make it the first permanent node.
3. Examine each neighbour node of the node that was the last permanent node.
4. Assign a cumulative cost to each node and make it tentative.
5. Among the list of tentative nodes
   a) Find the node with the smallest cumulative cost and make it permanent.
   b) If a node can be reached from more than one direction
      I. Select the direction with the shortest cumulative cost.
6. Repeat steps 3 to 5 until every node becomes permanent.

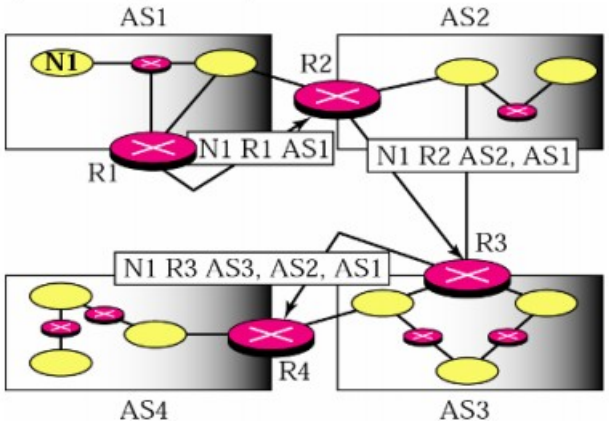
OSPF Link State Advertisement



Border Gateway Protocol (BGP)

- Uses Path Vector Routing
- Advertisements include complete path to destination
- Router that forwards advertisement adds itself to the list
- Path can be checked for loops
- Policies are applied when incorporating new routes

Network	Next Router	Path
N01	R01	AS14, AS23, AS67
N02	R05	AS22, AS67, AS05, AS89
N03	R06	AS67, AS89, AS09, AS34
N04	R12	AS62, AS02, AS09



BGP-4: Border Gateway Protocol

- AS Types
  - Stub AS: has a single connection to one other AS
    - Carries local traffic only
  - Multihomed AS: has connections to more than one AS
    - Refuses to carry transit traffic
  - Transit AS: has connections to more than one AS
    - Carries both transit and local traffic

- Each AS has:
  - One or more border routers
  - One BGP *speaker* that advertises:
    - Local networks
    - Other reachable networks (transit AS only)
    - Gives *path* information

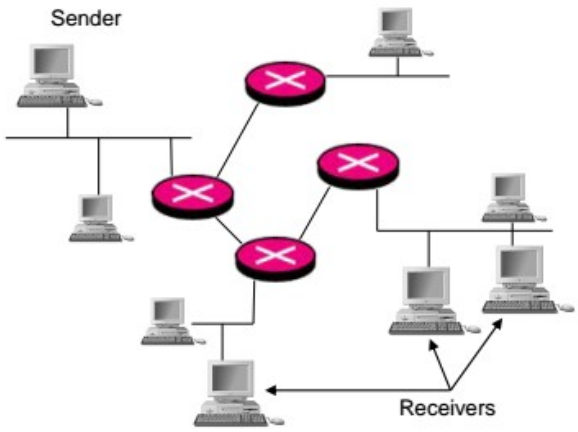
Protcol	Underlying Protcol	Protocol ID or Port
OSPF	IP	89
RIPv2	UDP	520
BGP	TCP	179

Summary: Routing

- Autonomous Systems
  - Stub network
  - Transient network
  - Point-to-point link
- Distance Vector routing
  - Share complete information with neighbours
  - Count-to-Infinity problem
  - Example: Routing Information Protocol (RIP)
- Link State routing
  - Share information about neighbours with everyone
  - Dijkstra's Shortest-Path Algorithm
  - Example: Open Shortest Path First (OSPF)

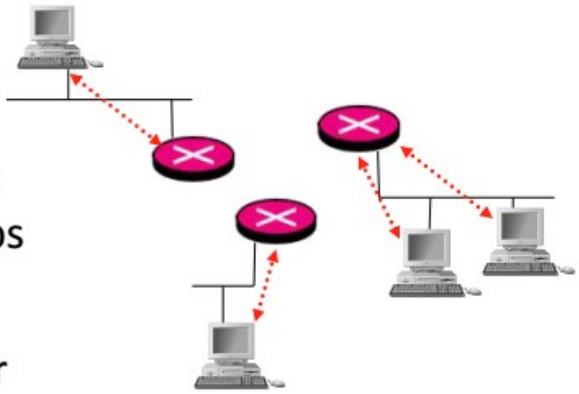
Multicast Overview

- Multicast requires group management
- Receivers join&leave multicast groups
- Multicast Addresses:
  224.0.0.0 – 239.255.255.255
  or 224.0.0.0/4



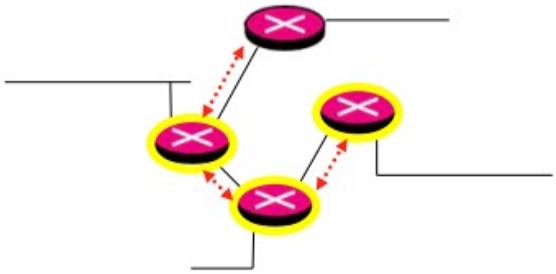
Internet Group Management Protocol (IGMP)

- Defines communication between hosts and router
- Specifies messages for hosts for joining and leaving groups
- Specifies query messages for routers

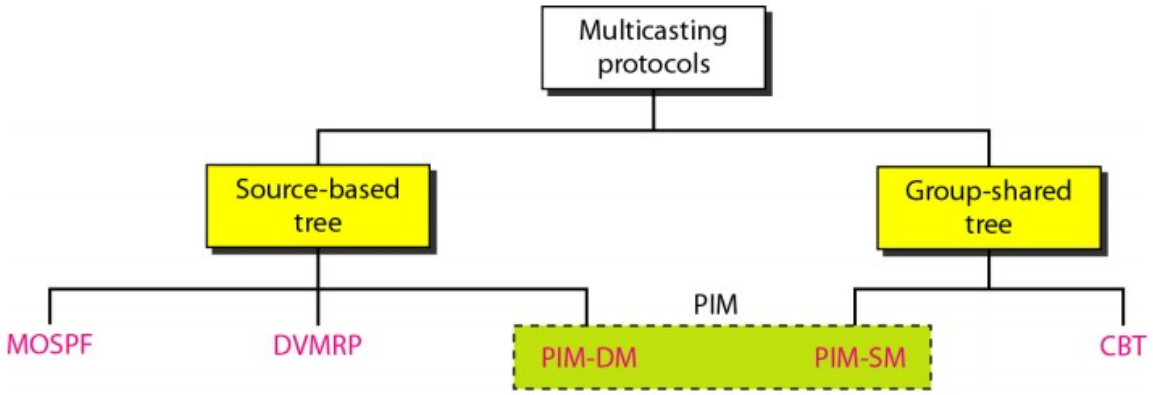


Network-Layer Multicast Protocols

- Distance Vector Multicast Routing Protocol (DVMRP)
- Multicast Open Shortest Path First protocol (MOSPF)
- Protocol Independent Multicast (PIM)



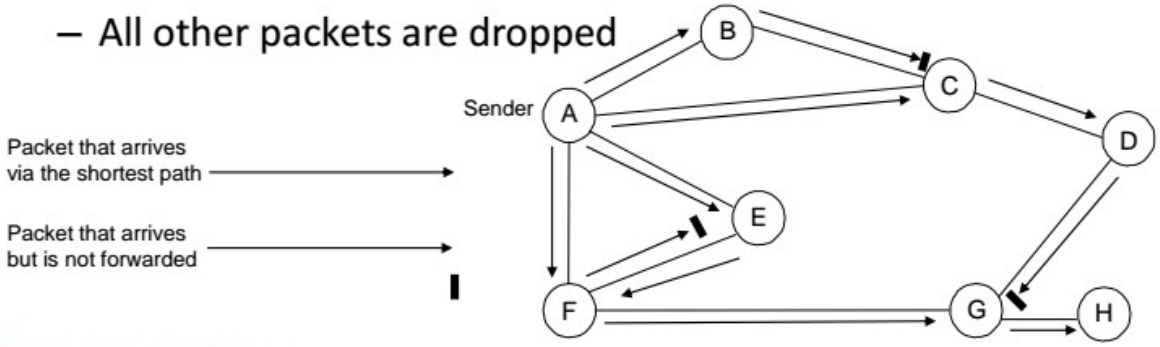
Multicast Routing Protocols



- Intra-AS
  - MOSPF
  - DVMRP
  - PIM
    - Sparse mode
    - Dense mode
- Inter-AS
  - MBGP + MSDP
  - BGMP + MASC

Reverse-Path Forwarding (RPF)

- Reverse-path forwarding simulates spanning tree routing without keeping state in the router
  - Each router knows shortest path to destination
  - Packets from A arriving on next hop to A are presumed to have followed shortest route from A, so they are forwarded on all other links
  - All other packets are dropped



PIM – Dense Mode (DM)

- When it is likely that many routers are involved in multicast routing

- Source tree created on demand based on RPF rule
- If the source goes inactive, the tree is torn down
- Branches that don't want data are pruned
- Grafts are used to join existing source tree

PIM – Sparse Mode (SM)

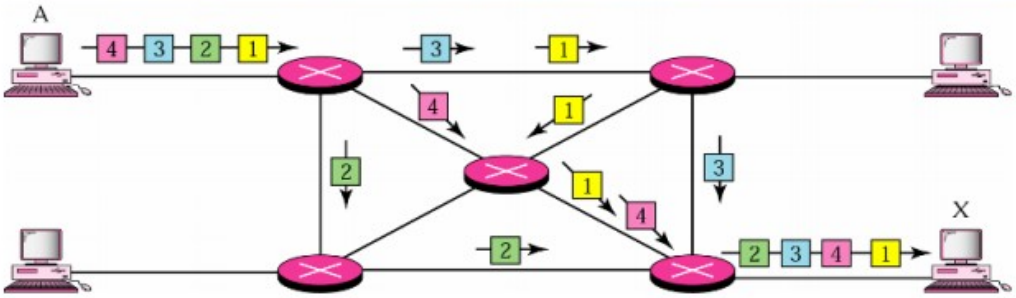
- When it is likely that many routers are involved in multicast routing

- One Rendez-Vous Point (RP) per group
- Explicit Join Model
  - Receivers send Join towards the RP
  - Sender Register with RP
  - Last hop routers can join source tree if the data rate warrants by sending joins to the source
- Dedicated "All-PIM-Routers" (224.0.0.13, ff02::d) multicast group

Summary: Multicast Routing

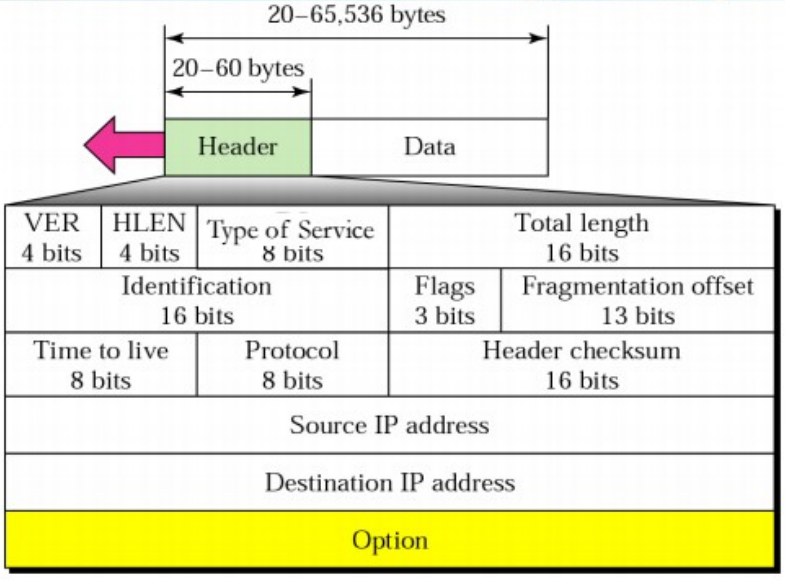
- Internet Group Management Protocol (IGMP)
  - Join&leave messages from hosts to routers
- Most protocols based on source trees
  - Reverse-Path Forwarding/Broadcast
  - Prune – remove subtree from tree
  - Graft – join subtree to tree
- Protocol Independent Multicast (PIM)
  - Dense Mode (DM)
  - Sparse Mode (SM)

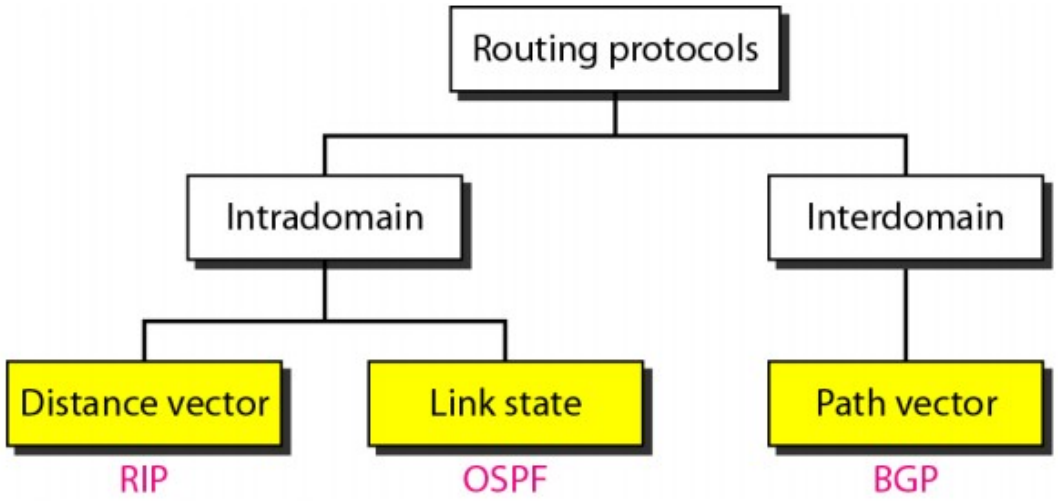
IP Service Model



- Connection-less Communication
  - No state is kept about individual packets
- Order not guaranteed
- Best-effort delivery (unreliable service)
  - Packets may be lost
  - Packets may be delivered out of order
  - Duplicate copies of a packet may be delivered
  - Packets can be delayed for a long time

IP Datagram





- Distance Vector routing
  - Routes propagate through exchange of routing tables

- Link State routing
  - Establish view of topology & run algorithm e.g. Dijkstra's Shortest-Path

Why IPv6? (Current Business Reasons)

2002 ☺

- Demand from particular regions
  - Asia, EU
  - technical, geo-political, and business reasons
  - demand is now
- Demand for particular services
  - cellular wireless (especially 3GPP standards)
  - Internet gaming (e.g., Sony Playstation 2)
  - use is >= 1.5 years away (but testbeds needed now)
- Potential move to IPv6 by Microsoft?
  - IPv6 included in Windows XP, but not enabled by default
  - to be enabled by default in next major release of Windows
  - use is >= 1.5 years away

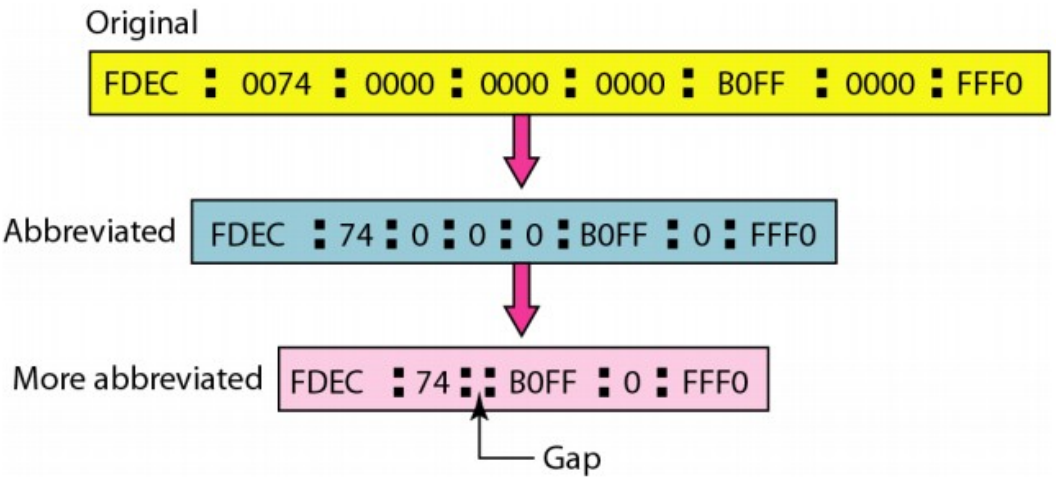
- Only compelling reason: **more IP addresses!**
  - For billions of new users (Japan, China, India,…)
  - For billions of new devices (mobile phones, cars, appliances,…)
  - For always-on access (cable, xDSL, ethernet-to-the-home,…)
  - For applications that are difficult, expensive, or impossible to operate through NATs (IP telephony, peer-to-peer gaming, home servers,…)
  - To phase out NATs to improve the robustness, security, performance, and manageability of the Internet

- Settled on fixed-length, 128-bit addresses

- Standard representation is set of eight 16-bit values separated by colons

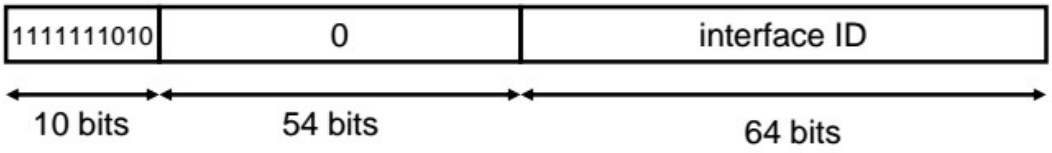
Address Abbreviation

- Sequences of zeros can be replaced with series of colons

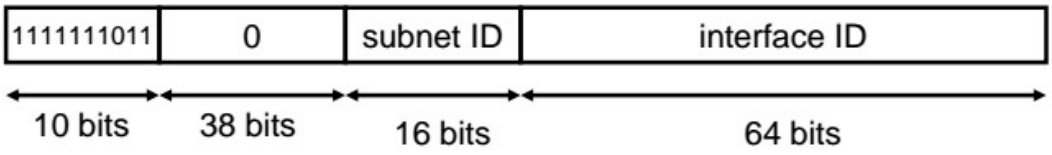


Non-Global Unicast Addresses

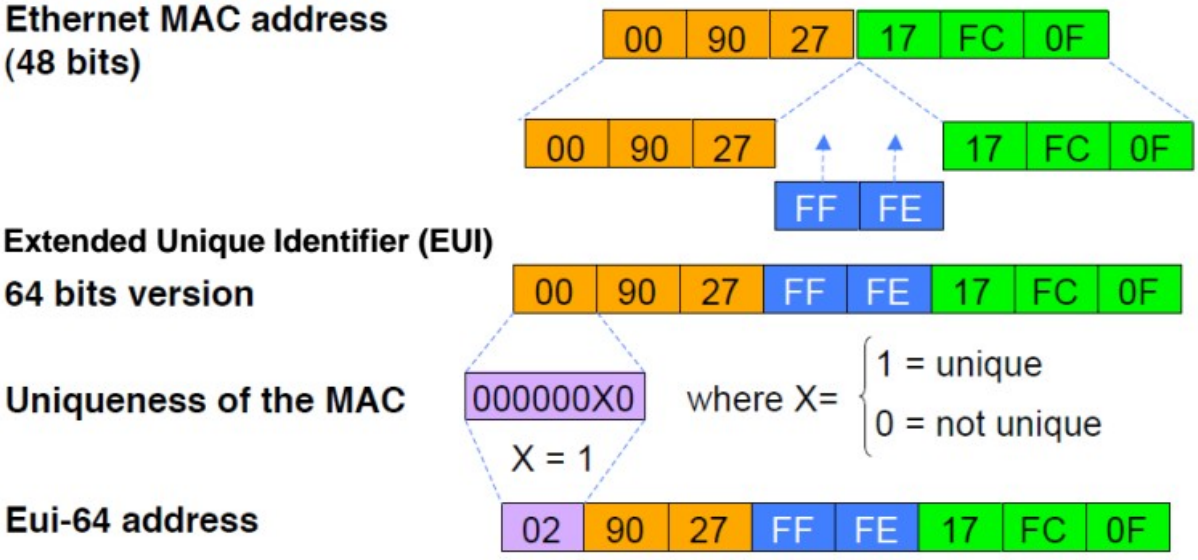
- Link-local unicast addresses are meaningful only in a single link zone, and may be re-used on other links



- Site-local unicast addresses are meaningful only in a single site zone, and may be re-used in other sites

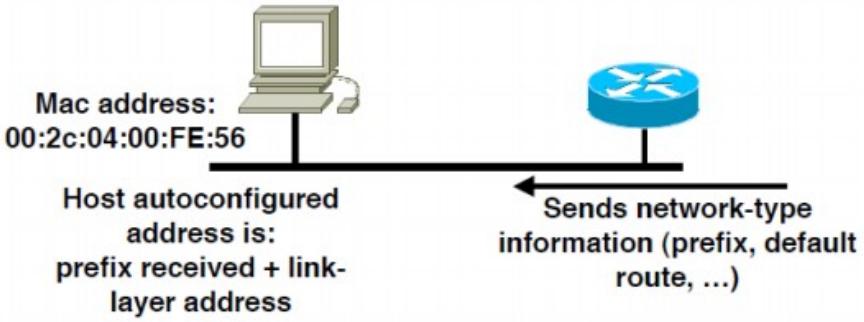


64-bit EUI Address



- EUI-64 address is formed by inserting FFFE and OR'ing a bit identifying the uniqueness of the MAC address

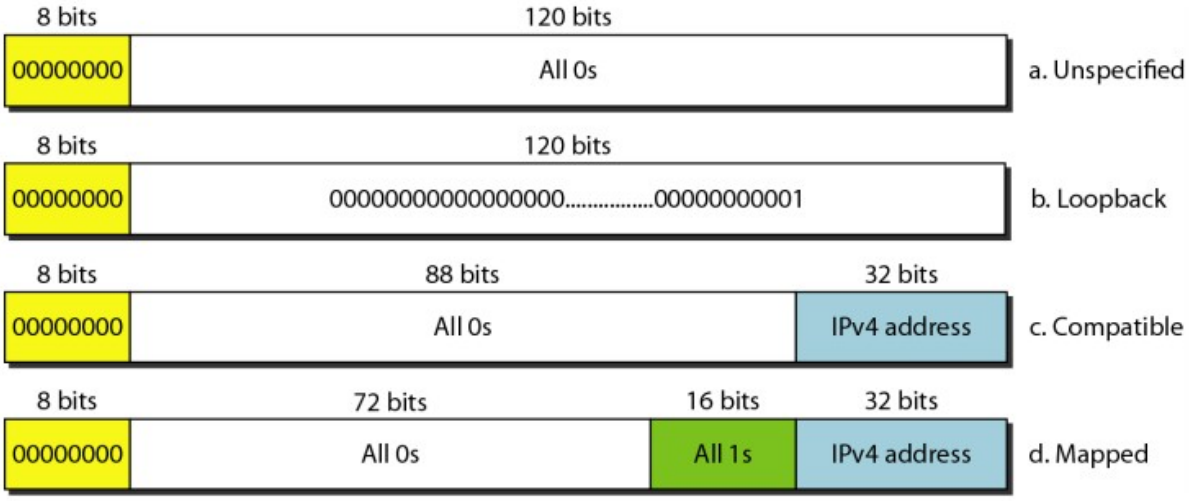
IPv6 Autoconfiguration



- Client sends router solicitation (RS) messages
- Router responds with router advertisement (RA)
  This includes prefix and default route
- Client configures its IPv6 address by concatenating prefix received with its EUI-64 address

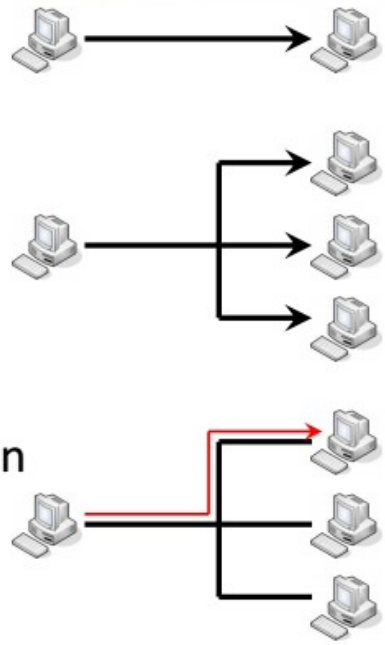
Multicast Addresses

- Low-order flag indicates permanent / transient group; three other flags reserved
- Scope field:
  - 1 - interface-local (for multicast loopback)
  - 2 - link-local (same as unicast link-local)
  - 3 - subnet-local
  - 4 - admin-local
  - 5 - site-local (same as unicast site-local)
  - 8 - organization-local
  - B - community-local
  - E - global (same as unicast global)
  - (all other values reserved)



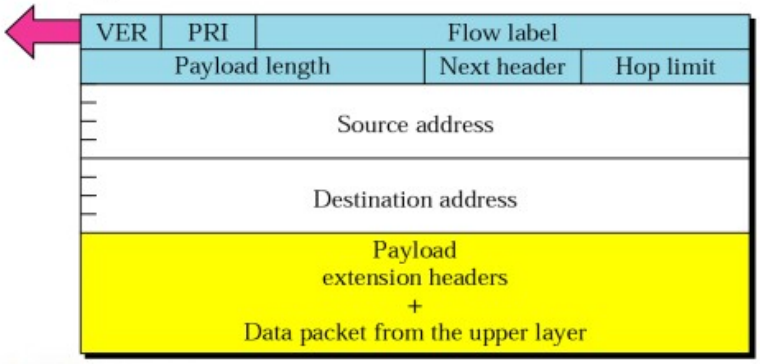
Communication Types

- Unicast
  - One-to-one communication
- Multicast
  - One-to-many communication
- Anycast
  - One-to-nearest communication
  - Delivered to any one interface

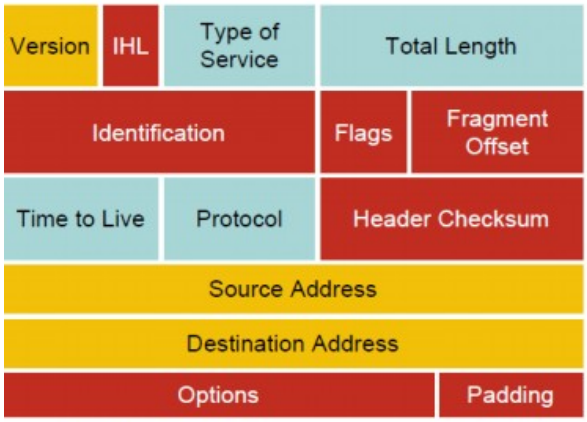


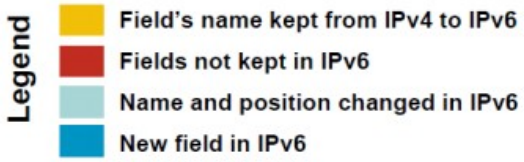
IPv6 Header

- Fixed length of all fields, header length irrelevant
- Remove Header Checksum – other layers are responsible
- No hop-by-hop fragmentation – fragment offset irrelevant
  - MTU discovery before sending or minimum MTU=1280
- Extension headers – next header type
- Basic Principle: Routers along the way should do minimal processing



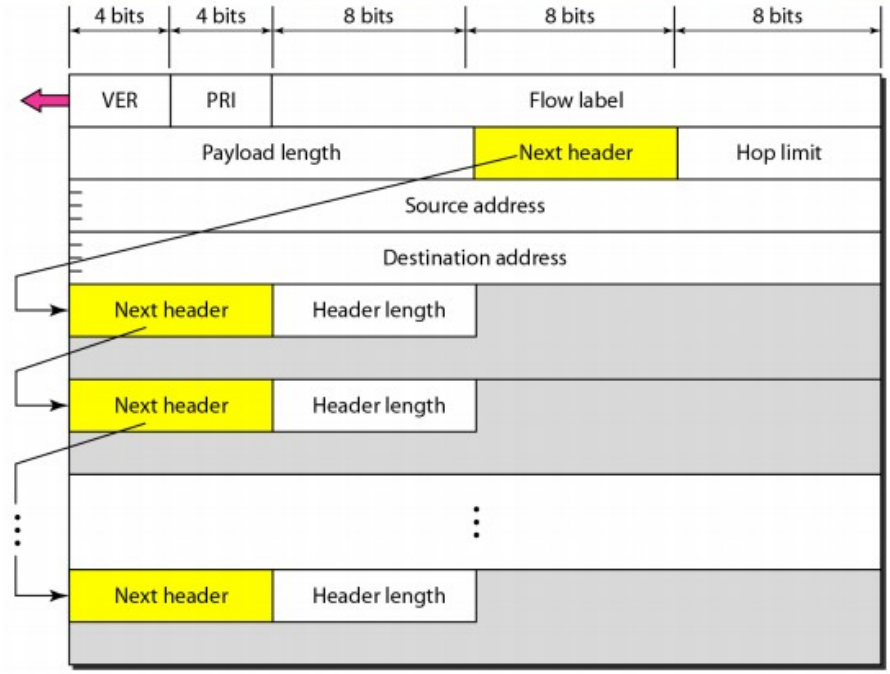
IPv4 Header





IPv6 Header

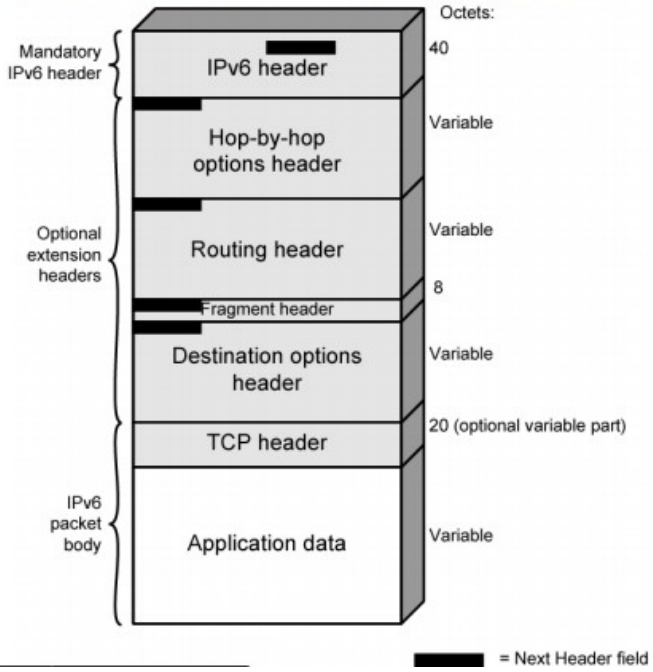
Extension Headers



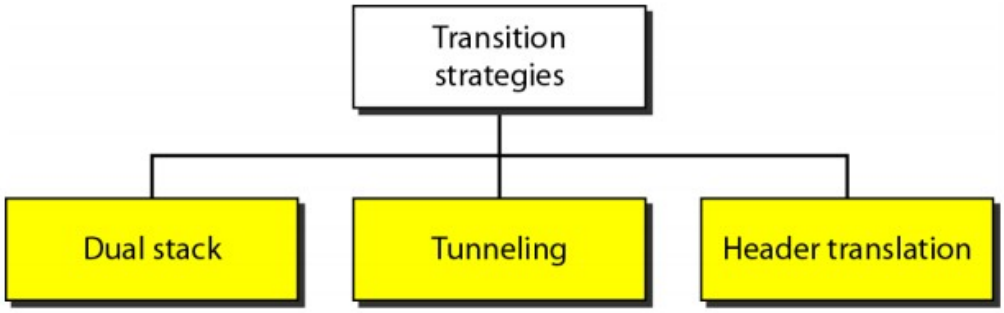
Extension header	Description
Hop-by-hop options	Miscellaneous information for routers
Destination options	Additional information for the destination
Routing	Loose list of routers to visit
Fragmentation	Management of datagram fragments
Authentication	Verification of the sender's identity
Encrypted security payload	Information about the encrypted contents

IPv6 Structure

- Every additional header is identified by “next header” field
  - including TCP and UDP header

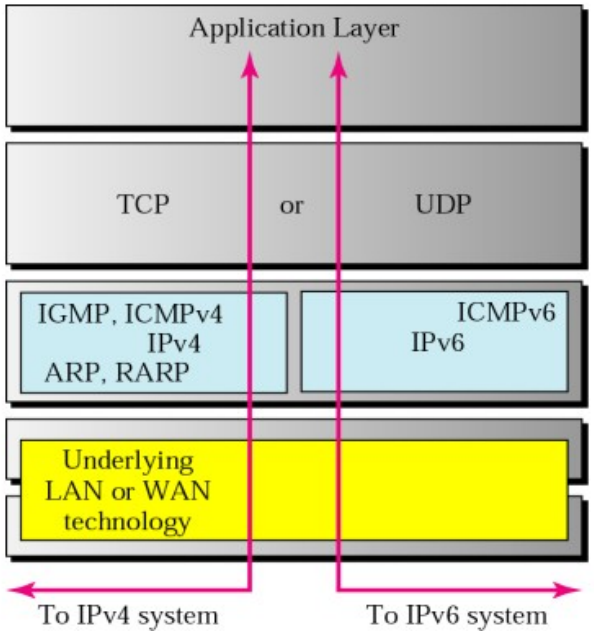


Transition from IPv4 to IPv6



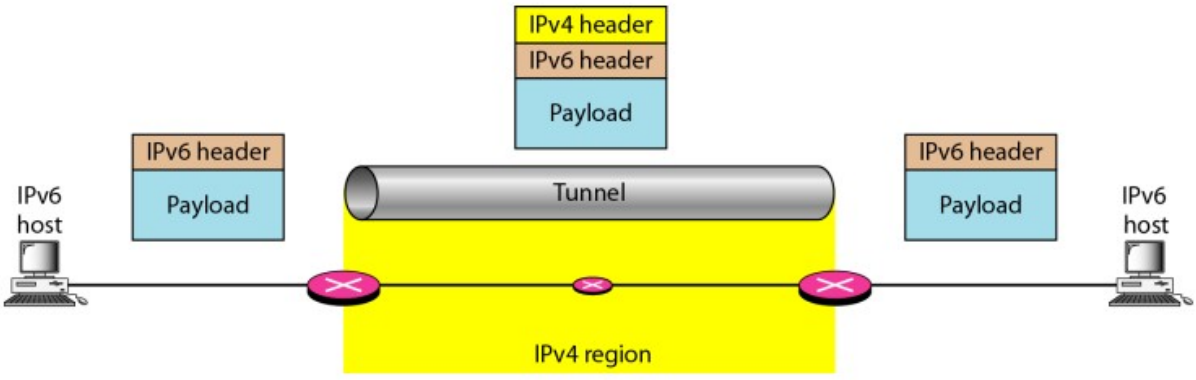
IPv4-IPv6 Transition

- Dual-Stack
  - Stack implements support for both IPv4 and IPv6
  - Allow IPv4 and IPv6 to co-exist in the same devices and networks



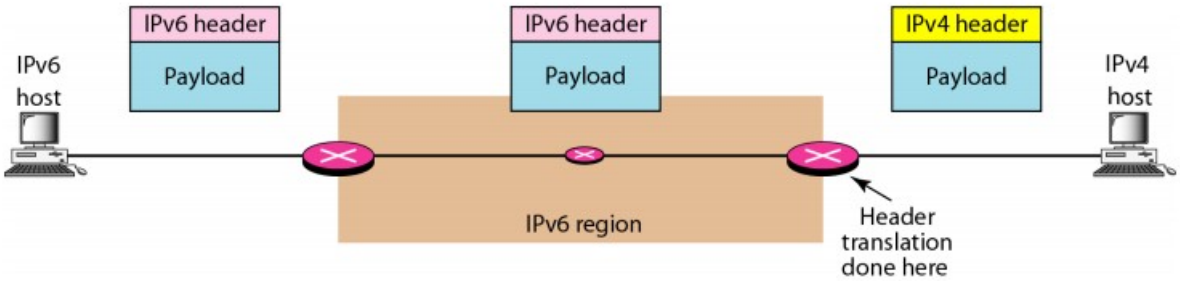
Tunneling

- Encapsulate IPv6 in IPv4 traffic
- Avoid order dependencies when upgrading hosts, routers, or regions



Header Translation

- Headers are translated into other format at "gateway"
- Allow IPv6-only devices to communicate with IPv4-only devices



Summary: IPv6

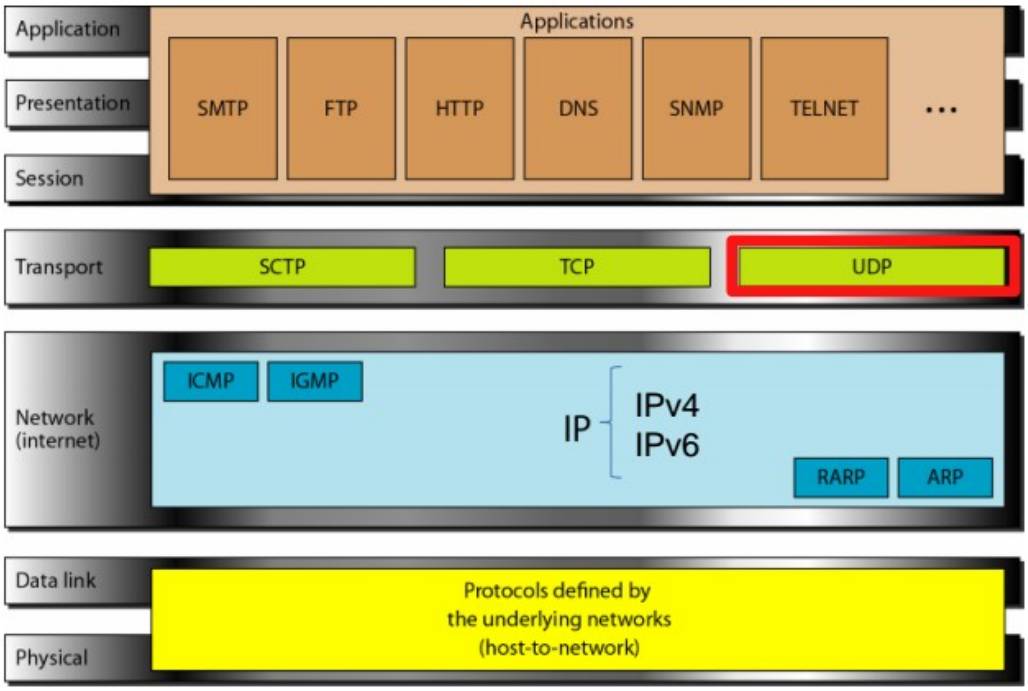
- Longer addresses: 128 bit
- Simpler, fixed-sized header

- Types of communication
  - Unicast
  - Multicast
  - Anycast

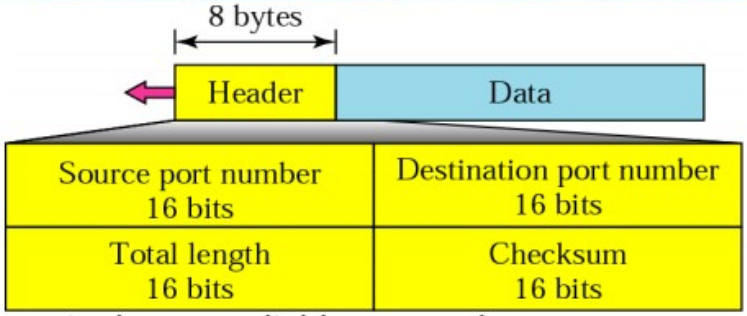
- Extension headers
  - Hop-by-Hop Options, Routing, Fragmentation, Authentication

- Techniques for transition
  - Dual-Stack
  - Tunneling
  - Translation

Protocols in the OSI Model

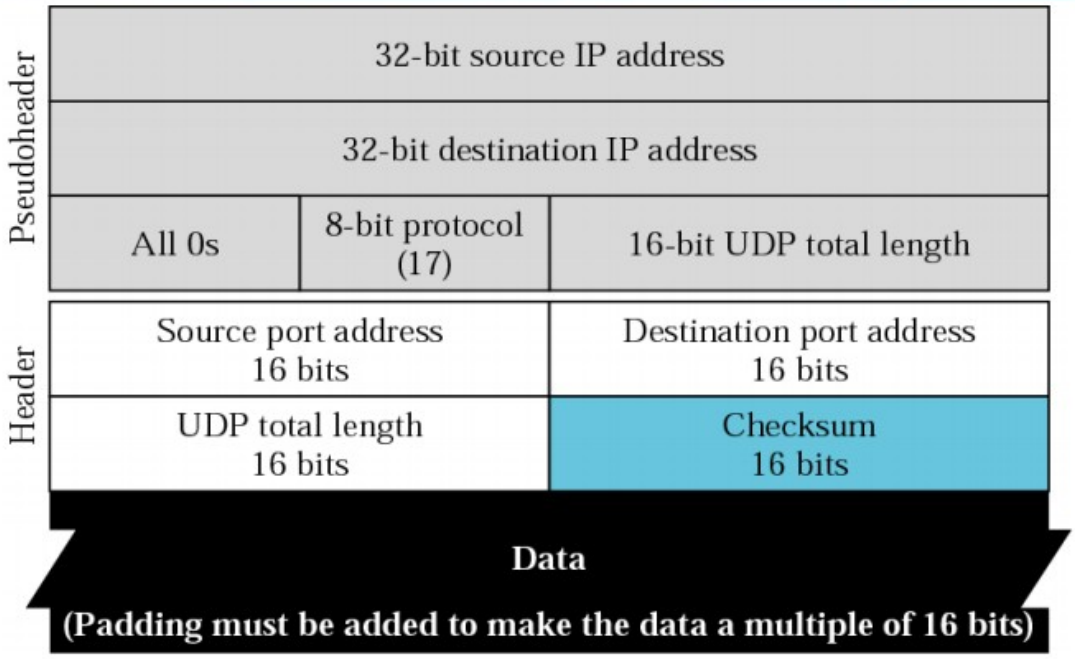


User Datagram Protocol (UDP)



- UDP is a connectionless, unreliable protocol
  - No flow and error control
  - Port numbers are used to multiplex data
- Calculation of checksum & its inclusion in datagram are optional.
- Convenient transport-layer protocol for applications that provide their own flow and error control
  - Also used by multimedia applications.

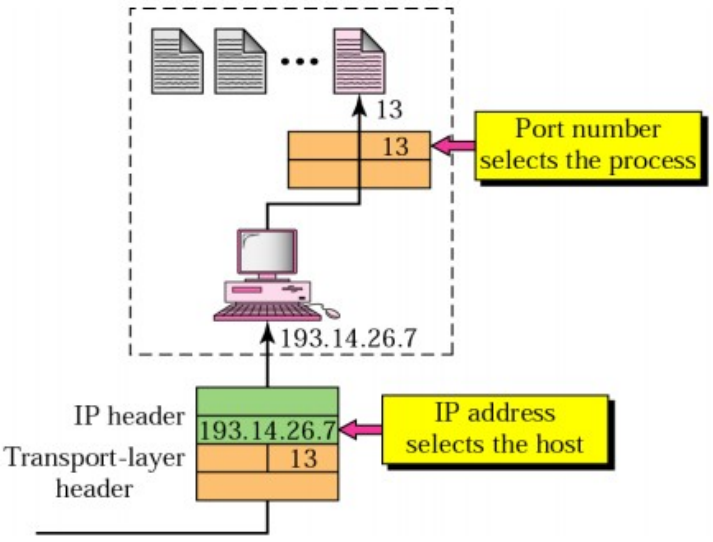
Pseudo-Header for Checksum



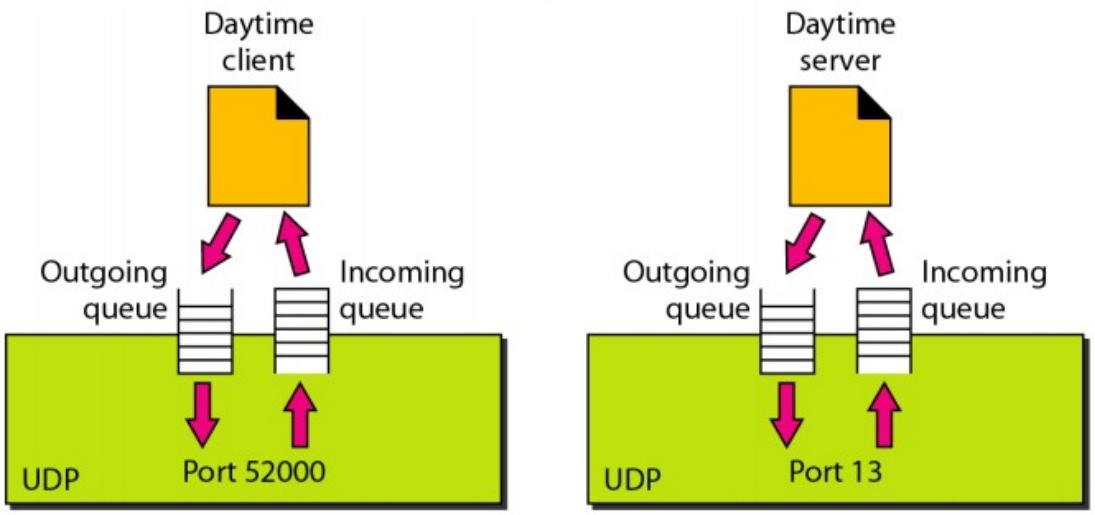
- 3 Categories of Ports:
  - Well-known Ports: 0 – 1023 (restricted access)
  - Registered Ports: 1024 – 49151
  - Dynamic/Private Ports: 49152 – 65535

- IP Addresses determine the host

- Port Numbers determine the application

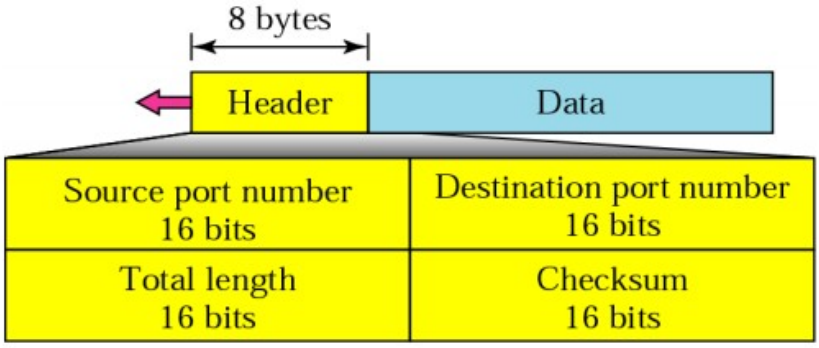


Queuing in UDP



User Datagram Protocol (UDP)

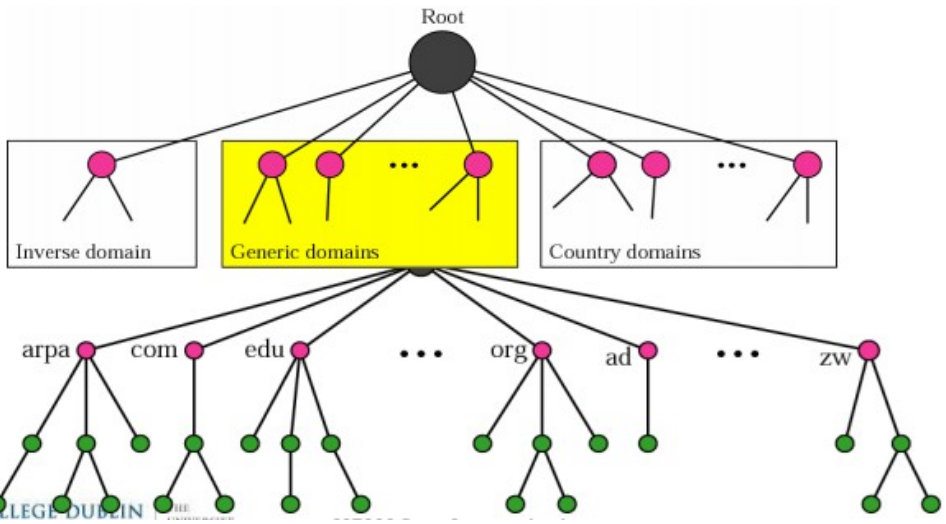
- Connectionless
- Unreliable
  - No flow or error control
- Small Header:



Domain Name Space

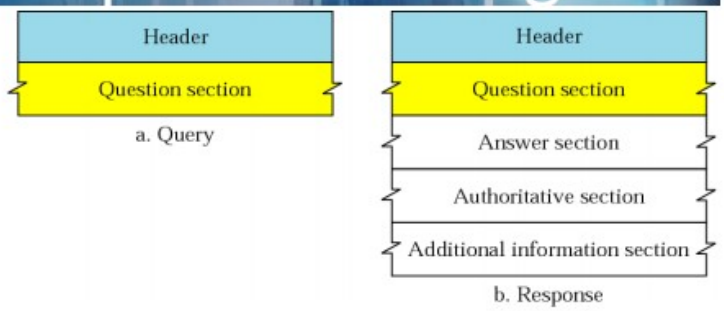
- Association between names and IP addresses

  www.dsg.scss.tcd.ie - 134.226.36.14

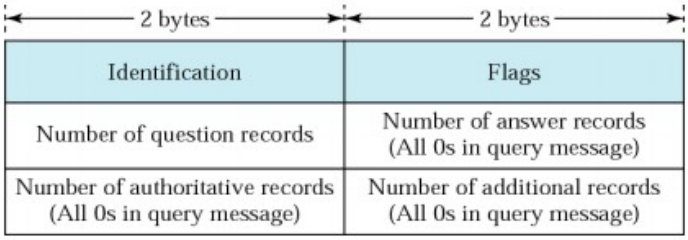


Query and Response Messages

- Two types of replies:
  - Authoritive answers
  - Cached or unauthoritive answers



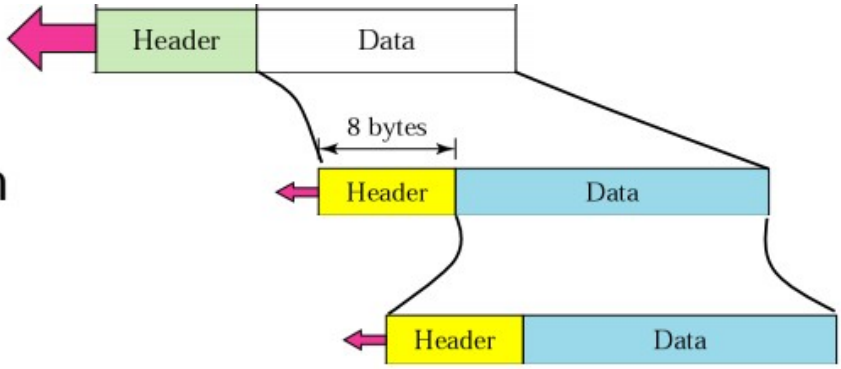
- 6-byte header

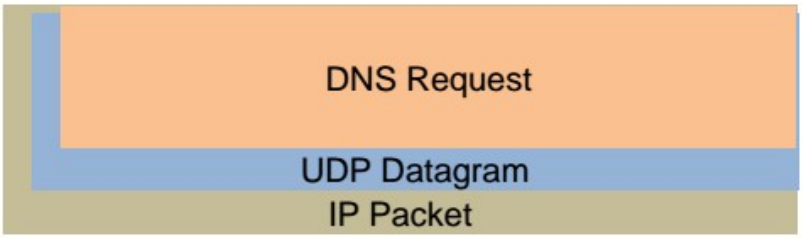


IP Packet

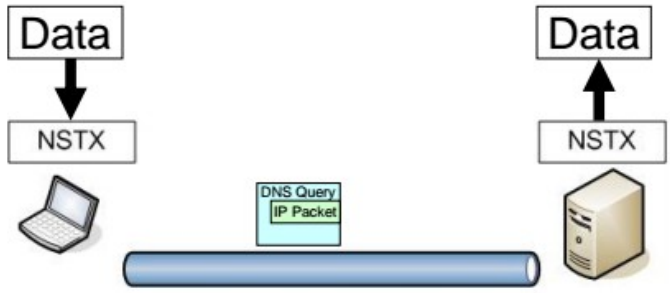
UDP Datagram

DNS Request

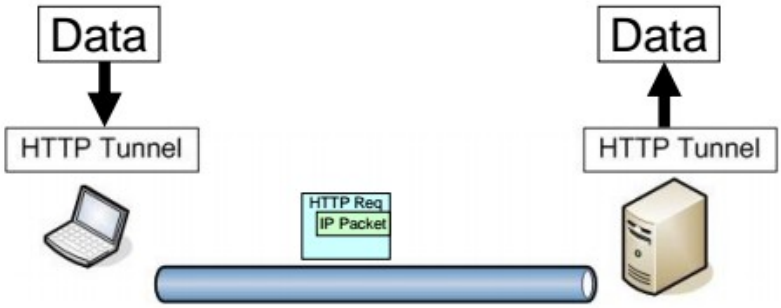




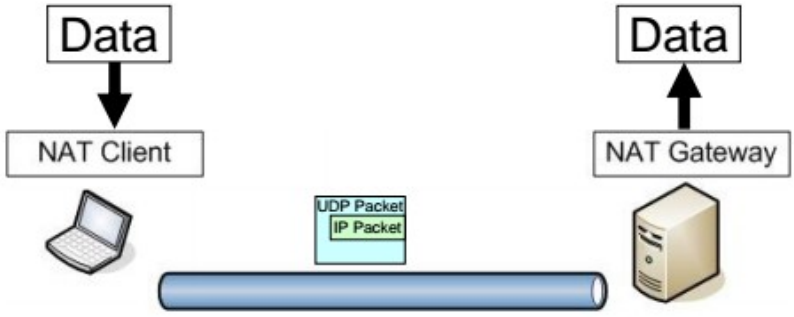
Tunnelling I



- Machine with access to Internet
- Machine with restricted access
- Both run programs that can pack and unpack data to be tunnelled

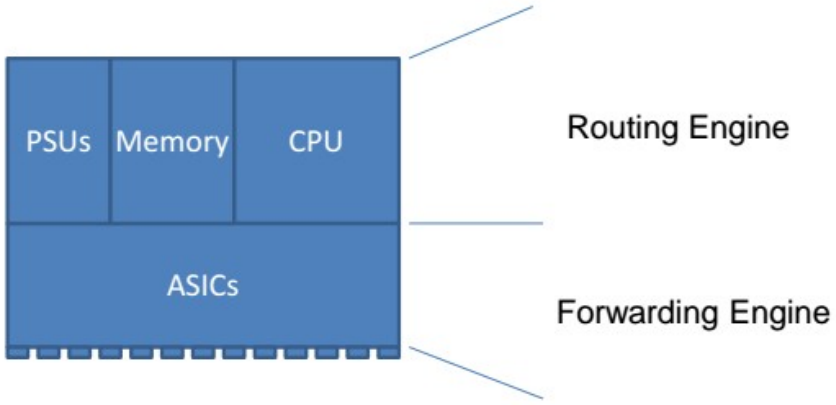


- Same process: Tunnel runs on both machines to pack and unpack data
- HTTP Request can transverse proxies etc

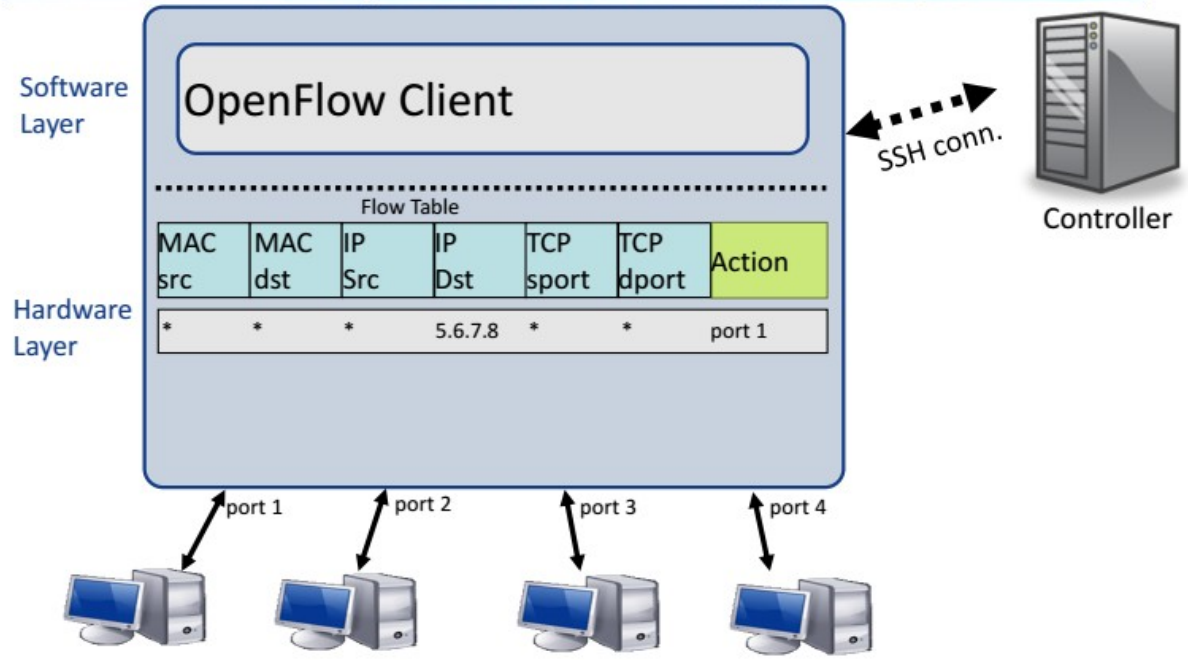


- Same process: Client and gateway should know how to pack and unpack data
- Gateway could be any machine in College with access to machines outside

Switches/Routers

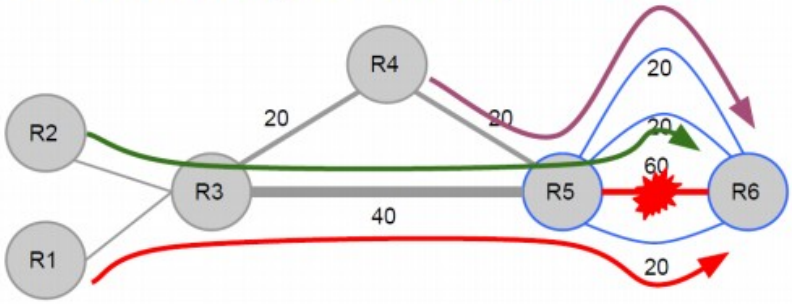


Openflow Switch



Traditional Net. Example

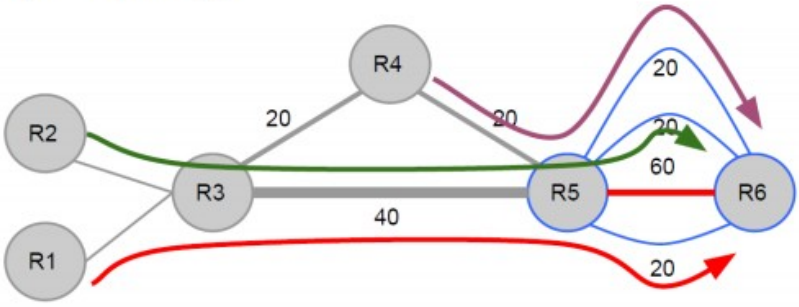
- Flows: R1->R6: 20; R2->R6: 20; R4->R6: 20



- R5-R6 link fails
  - R1, R2, R4 *autonomously* try for next best path
  - R1 wins, R2, R4 retry for next best path
  - R2 wins this round, R4 retries again
  - R4 finally gets third best path

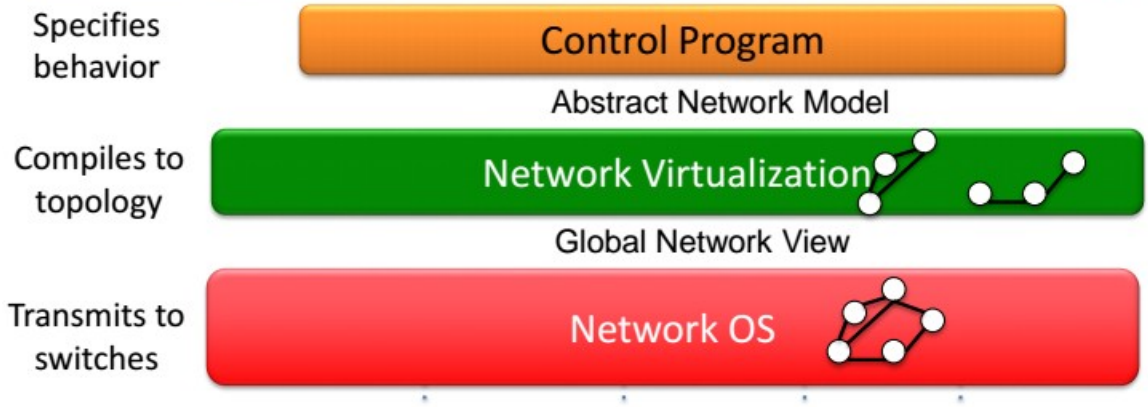
Topology with Central TE

- Simple topology



- Flows:
  - R1->R6: 20; R2->R6: 20; R4->R6: 20
- R5-R6 link fails
  - R5 informs TE, which programs routers in one shot
  - Leads to faster realization of target optimum

SDN Overview



Openflow SDN Architecture