Multicast

RP – Rendezvous Point

DVMRP- Distance Vector Multicast Routing Protocol

IGMP – Internet Group Management Protocol (IPv4)

MLD – Multicast Listener Discovery (IPv6)

SSM – Source Specific Multicast

PIM – Protocol Independent Multicast

Rendezvous Point: Is a router in multicast network domain that acts as a shared root for multicast tree.

Internet Group Management Protocol: is a communication protocol used by hosts and adjacent routers on IPV4 to establish multicast group membership. Used by host for routers.

Ex: online streaming, gaming.

Multicast Listener Discovery: Is a component of IPv6 suite. MLD is used by IPv6 for discovering multicast listeners on a directly attached link.

Source Specific Multicast: Method of delivering multicast packets in which the only packets delivered to the receiver are those originating from a specific source address requested by the receiver.

Protocol Independent Multicast: Is a family of multicast routing protocols for Internet protocol networks that provide one to many and many to many distributions of data over a LAN, WAN or a MAN.

It is called PIM since it does not include it own topology discover mechanism instead uses routing information supplied by routing protocols.

PIM Sparse Mode: Builds unidirectional shared trees rooted at rendezvous point per group. Optionally creates shortest path trees per source. Good scaling.

PIM Dense Mode: Floods the multicast traffic domain wide, and then pruning branches of the tree where no receivers are present. Poor scaling.

Anycast - A single destination address that has multiple routing paths to two or more destination end points.

Broadcast – One to all transmission.

Multicast – Addresses use one to many or many to many associations.

Multicast Considerations:

1. **Multicast is UDP**
2. **Best effort delivery**: Drops are to be expected.
3. **No congestion avoidance**: Lack of TCP windowing and "slow start" mechanisms can result in network congestion.
4. **Duplicates**: Duplicates are expected, and receivers should be designed to accept.
5. **Out of order**: some protocol mechanism may also result in out of order delivery of packets.

Multicast Addressing:

1. Class D address range: **224.0.0.0 – 239.255.255.255**
2. Multicast Group addresses are not in the unicast route table.
3. A separate multicast route table is maintained for active multicast trees.
4. Multicast trees are initiated by receivers signaling their request to join the group.
5. Sources need not join just need to send.

Address Allocation

1. Reserved Link local address – 224.0.0.0 – 224.0.0.255
2. IANA reserved address – 224.0.1.0 – 224.0.1.255
3. Administratively Scooped address – 239.0.0.0 – 239.255.255.255
4. Source Specific Multicast – 232.0.0.0 – 232.255.255.255

Constructing Multicast Mac address

IPv4 Multicast address

Ex: 224.1.10.10 calculate mac address

Step 1:

Take last 28 bits and covert to binary.

0000 0000 0001 0000 1010 0000 1010

Step 2:

Remove 5 MSB

000 0001 0000 1010 0000 1010

Step 3:

Prepend the value from step 2 with 01 00 5e 0xxx

01 00 5e 01 0a 0a

The above is the mac address associated with multicast ipv4 address 224.1.10.10

1 and 129 value in the octet have same mac address overlap

Address Scoping:

1. **Campus Scope**
2. **Regional Scope**
3. **Enterprise scope** – Identifies application for IP Multicast communication between any sites within the organization.

Host – Router Signaling

1. **IGMPv3 is the current version.**
2. **Uses 224.0.0.22 as link local multicast address**

   All IGMP host send membership reports to this address.

   All IGMP routers listen to this address.

   Hosts do not listen or respond to this address.
3. **Membership Reports**

   Sent by hosts

   Contains list of multicast pairs to include/exclude.
4. **Membership queries**

   Sent by routers to refresh/ maintain list of multicast traffic to deliver

Membership Query Type

0x01 and 0x02 – current state records used to report the current filter state.

0x03, 0x04 – Filter change record to change a state from include to exclude.

0x05, ox06 – source list change records indicate the change to the sources being handled.

Working:

1. A source ready to send multicast traffic send a join group Allow New report to the router. It send group id and source id
2. Router sends periodic general queries to all hosts. All IGMP members respond.

   Reports contain multiple group state records.
3. A source leaving the cast send Block Old membership report.
4. Router send group source query.
5. The remaining member hosts will send the report.
6. The remaining host will leave by sending block old membership report.
7. State times out and group source flow pruned.

Multicast routing is backwards from unicast routing.

Only when a tree is completely built from receiver backwards to the source can source traffic flow down the tree to the receivers.

Attacks

Mainly 2 types

1. DoS
2. Resource Utilization

1. **DoS**

   Sending IGMP or MLD to subscribe to a large number of high BW multicast groups causing bandwidth exhaustion.

2. **Resource utilization**

   The attacker is chosen as the querier and can advertise his own robustness variable. If the maximum response time is made low, hosts will have to respond to very short intervals and load will be increased on CPU.

3. **Fragmented IGMP** packets to induce crashes on OS.

Domain Name System

For Scalability DNS name are hierarchical.

Application's point of view, access to the DNS is through an application library called a Resolver.

The top of the search is

1. Root server
2. TLD Sever
3. Name Server

Generic TLD - .com, .org, .net, .info

Country code TLD - .AC, .AD, .AE

Internationalized Country Code TLD.

DNS Naming Syntax

The longest domain name is limited to 63 characters.

Functions

1. Distributed system whose primary function is to provide name to address mappings.
2. Adding Zones which means administrative delegation.

    a) A zone is a subtree of the DNS name space that can be administered separately from other zones.

    b) Whenever a new record is added to the zone the DNS administrator of the zone allocates a name and additional information (IP address) for the new entry and enters it in to its name server database.

    c) For example, in sjsu.edu other domains like engr.sjsu.edu.

Caching

1. Most name servers also cache zone information they learn. Only up to a time limit (TTL)
2. This reduces the DNS traffic.
3. While returning the server indicates whether the reply is from it cache or administrative zone.

4. When it returns from the cache it indicates this by including the domain name of the name servers that can be contacted to retrieve the authoritative information.

Root servers from A to M

Root servers hold data for gTLDs (.com,.net,.org,) also ccTLDs (.au,.uk,.us).

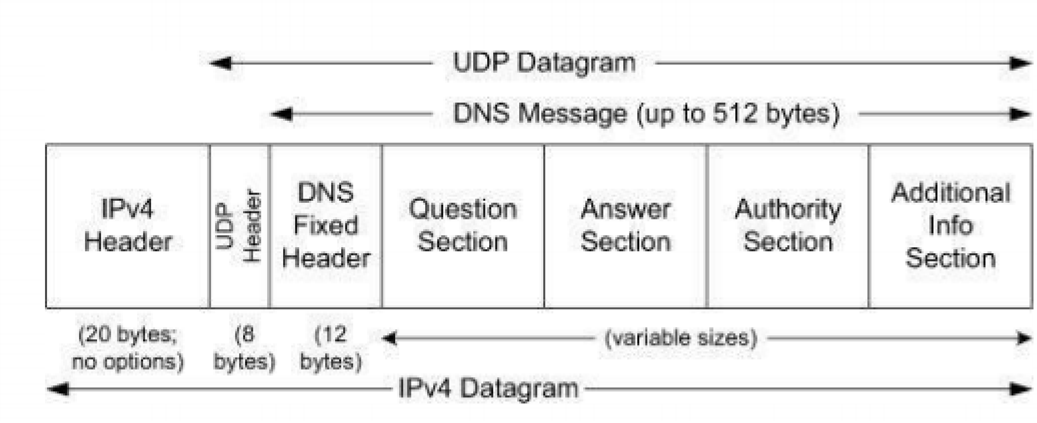
Once a new domain is registered the root DNS server update servers around the world.

DNS primarily uses **UDP** and Port **53**

DNS queries consists of a single query from the client and a single reply from the server.

It uses TCP when response data size exceeds 512 bytes or for task of zone transfers.

DNS Packet



1. When a resolver issues a query, the response comes back with a TC bit field set.
2. Only the 512 bytes are sent, this make the client to establish a TCP connection.

Question section

The question section consists of 3 fields:

1. Query name
2. Query Type
3. Query Class

The next 3 section RRs

Popular Resource records

A: Address record for IPv4

NS: Name server

SOA: Start of authority

PTR: provides address to name mapping.

CNAME: Canonical name introduce an alias for a single domain name in the naming system.

Ex: www.sjsu.edu or www.w3.sjsu.edu

Reverse DNS Queries: PTR Records

Used for reverse mapping

Need: A server receiving incoming TCP connection request can detect the source IP address of the connection from the incoming IP datagram, but names corresponding to that address is not carried in the connection itself.

Start of Authority (SOA): In DNS each zone has an authoritative record

Identify name of the host providing official database.

MX record: provides name of the mail exchanger. A host willing to engage in SMTP (simple mail transfer protocol) to receive incoming email on behalf of users associated with a domain name.

Zone Transfer

1. A zone transfer is used to copy a set of RRs for a zone from one server to another.
2. This is to keep multiple servers providing DNS service be in sync. Provides a backup if one of the servers shut down.
3. Latency can be improved by placing servers close to client.
4. Refresh is attempted periodically
5. The entire zone contents are refreshed with in the expire interval

Incremental Zone transfers

To improve the efficiency of zone transfers. Use incremental zone transfers which updates only the changes.

Attacks of DNS

1. **DOS –** overloading of important DNS servers, such as root or TLD servers.

2. **Reply with bogus resource records –** masquerading contents and reply with bogus records

User Datagram Protocol

UDP does not provide

1. Error correction
2. Sequencing
3. Duplicate elimination
4. Flow control
5. Congestion control

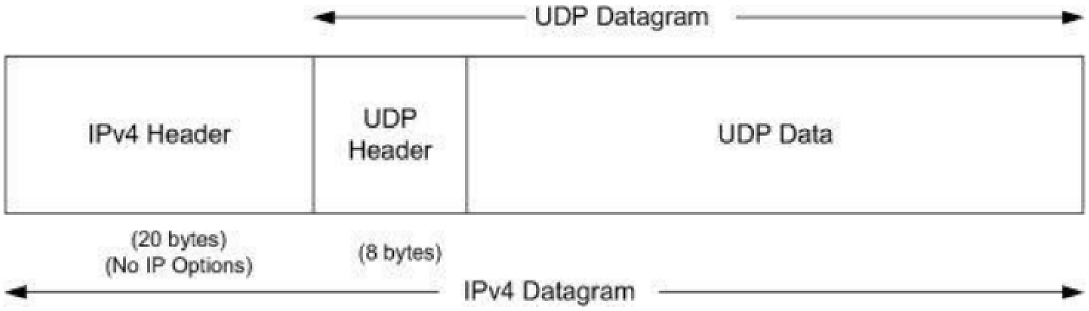
It can provide error detection and it includes the first end to end checksum at the transport layer we have encountered.

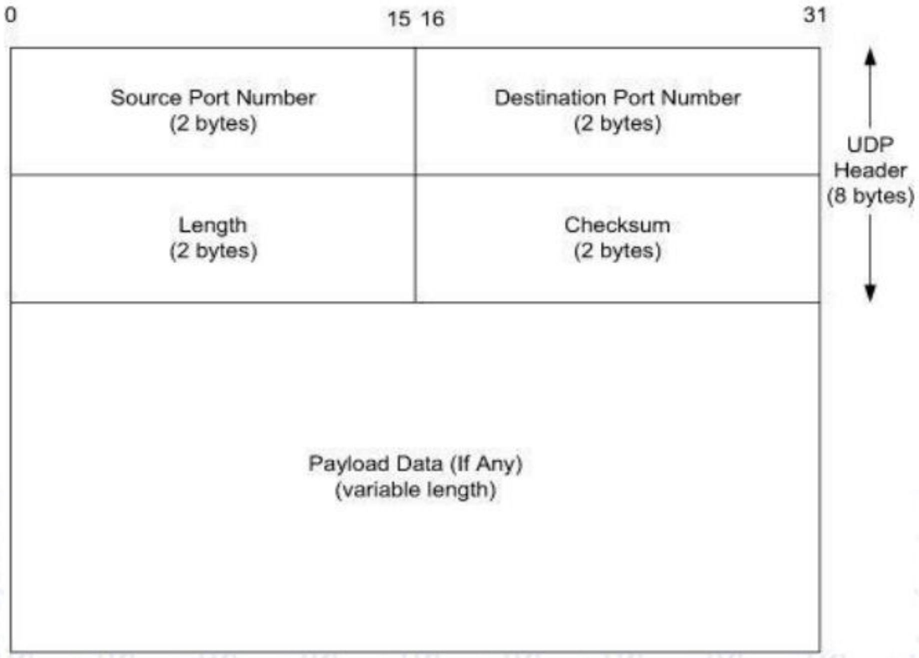
Advantages

Less overhead

Broadcast and multicast are much more straight forward.

UDP packet:





Port Numbers:

In UDP source port number is optional.

UDP length field is the length of UDP header and UDP data in bytes.

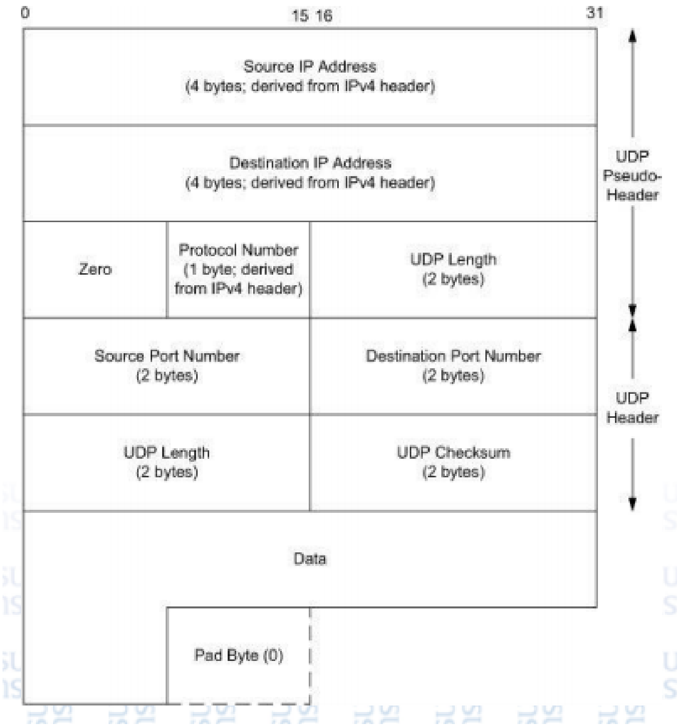
The minimum value is 8.

UDP checksum covers UDP header, UDP data and Pseudo Header.

It is computed at the sender and checked at the receiver. It is not modified in transit.

The length of the UDP datagram can be odd number of bytes, where as checksum algorithm adds 16-bit words. The procedure is to append a pad byte of 0 to the end of odd length datagrams.

Pseudo Header includes source and destination address, protocol and next header field. Its purpose is to let UDP layer know the data has arrived at the correct destination.



If the value of check sum calculate is 0x0000 it is stored as 0xFFFF.

This causes layering issues violation

Since UDP protocol is directly processing bits owned by IP layer.

Problem of UDP at NAT especially when packets are fragmented.

When a UDP packet passes through NAT not only IP layer checksum should be modified, but UDP pseudo header check sum should also be modified.

UDP lite

UDP lite computes the checksum of only payload in the datagram.

UDP lite has its own IPv4 and IPv6 next header field.

It modifies the header length field with a check sum coverage field. Which hold the number of bytes covered by the checksum.

The value 0 indicate entire payload is covered by checksum.

IP fragmentation

When an IP datagram is fragmented it is not re assembled until it reaches final destination.

2 reasons

1. Not performing reassembly with in the network alleviates the forwarding software in routers from implementing this feature.
2. It is possible different fragments take different path to reach the destination.

When a fragment of a TCP segment is lost, TCP retransmits the entire TCP segment, which corresponds to an entire IP datagram. There is no way to resend only one fragment of a datagram.

Problems of IP fragmentation.

1. Fragmentation causes inefficient use of resources.
2. Loss of fragments leads to degraded performance.
3. Efficient reassembly is hard.

Steps avoiding fragmentation

1. Send small datagrams
2. Guess minimum MTU of path.
3. Discover actual MTU of path.
4. Guess and discover MTU and backtrack if wrong.

Path MTU discovery using UDP

ICMP PTB helps the largest packet size along the routing path.

The messages are typically processes below the UDP layer and are not directly visible to the application. So IP layer can perform PMTUD without the application knowing.

IP fragmentation and ARP

When multiple fragments arrive at a router how many ARP requests should be made.

Sock command

Packet processing at receiver end.

Theoretically, the maximum size of an IPv4 datagram is 65,535 bytes, with an optionless IPv4 header of 20 bytes and UDP header of 8 bytes.

This leaves a maximum of 65,507 bytes for user data

2 reason why a full payload cannot be delivered.

1. The system's local protocol implementation may have some limitation.

2. The receiving application may not be prepared to handle such large datagrams.

When application receives a UDP message it will be stripped of the IP header and UDP header. The application must be told by the operating system on who sent the message someway. If it intends to furnish a reply. This allows UDP server to handle multiple clients.

Attacks involving UDP

1. Exhaustion of some resource – buffer, link, capacity
2. Exploitation of bugs in protocol implementation – system crash or failure.

1. Generate maximum amount of traffic as fast as possible.
2. **Magnification attack** – attacker sending a small amount of traffic that induces other systems to generate more

Transmission Control Protocol

ARQ – Automatic Repeat Request – The process of trying to send again until the information is finally received.

Ways to determine packet loss

(1) Whether the receiver has received packet.

(2) Whether the packet received is the same the sender sent.

Approaches:

1. Stop and Wait
2. Window of Sliding window

Flow Control

When a receiver is slow to accept packets and way to force the sender to slow down. Flow control.

Window advertisement or a window update.

Congestion Control

Sender slowing down not to overwhelm the network between itself and receiver.

Explicit signaling – Flow control

Implicit signaling – Congestion control

How Log to wait to decide whether concluding packet is lost and resend

Round Trip Time Estimation

TCP

Connection oriented: Two entities must establish a TCP connection before exchanging data.

TCP before providing a byte stream interface must convert a sending application stream of bytes in to packets. This is called packetization.

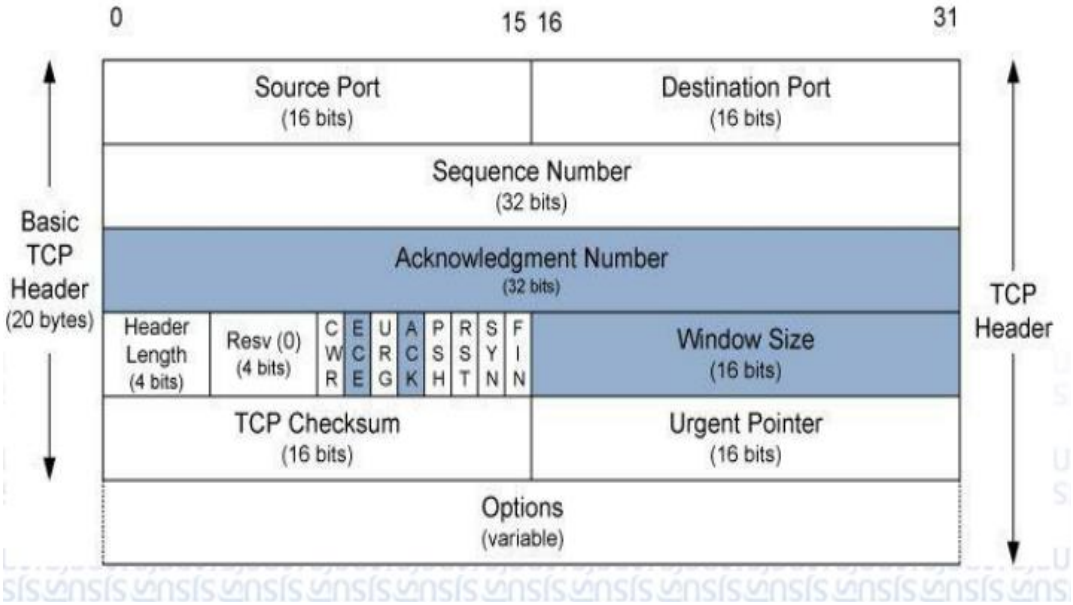
The application data is broken into size for which the TCP believes best size of chunks to send.

The chunk of data passing from TCP to IP is known as segment.

When TCP sends a group of segments it uses a single retransmission timer. And waits for other end to acknowledge the reception.

TCP provides Full duplex service to the application layer.

TCP never delivers out of order packets to the application.



Sequence Number: Identifies the byte in stream of data from the sending TCP to the receiving TCP.

Acknowledgement Number: The next sequence number acknowledgement the sender expects to receive. Is valid if only ack bit field is ON.

SYN: when a new connection is established this bit field is set, to indicate first segment is sent from client to server

1. CWR (Congestion window reduced)
2. ECE (ECN echo)

3. URG (Urgent)
4. ACK (Acknowledgement)
5. PSH (PUSH) The receiver should push the data immediately to the application
6. RST (RESET) Connection abort
7. SYN (Synchronize) sequence numbers to initiate the connection.
8. FIN (Finish) The sender has finished sending data.

Window Size: TCPs flow control is provided by each end advertising a window size using the window size filed.

Checksum: covers both header and data fields.

Urgent Pointer: to indicate the last byte of urgent data. Added along with sequence number.

Option field:

The **MSS** option specifies the maximum size segment that the sender of the option is willing to receive in the reverse direction.

TCP operations

Listen: (server) represents waiting for a connection request from any remote TCP and port.

SYN-SENT: (client) represents waiting for a matching connection request after having sent a connection request.

SYN-RECEIVED: (server) represents waiting for a confirming connection request acknowledgment after having both received and sent a connection request.

ESTABLISHED: (server and client) represents an open connection, data received can be delivered to the user. The normal state for the data transfer phase of the connection.

FIN-WAIT-1: represents waiting for a connection termination request from the remote TCP, or an acknowledgment of the connection termination request previously sent.

FIN-WAIT-2: represents waiting for a connection termination request from the remote TCP.

CLOSE-WAIT: represents waiting for a connection termination request from the local user.

CLOSING: represents waiting for a connection termination request acknowledgment from the remote TCP.

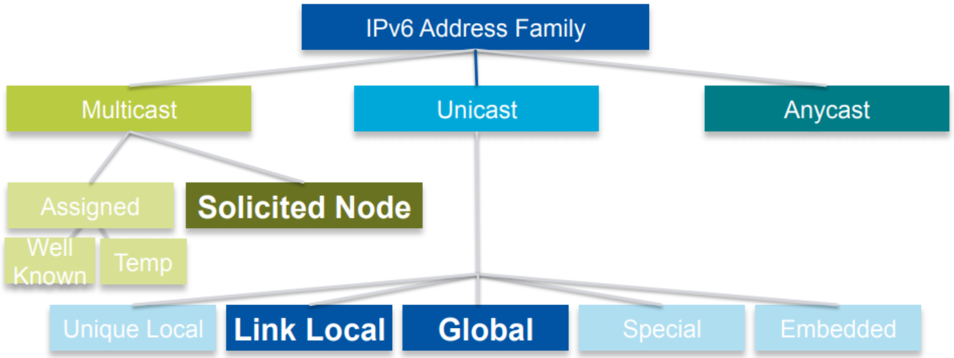
LAST-ACK: represents waiting for an acknowledgment of the connection termination request previously sent to the remote TCP.

TIME-WAIT: represents waiting for enough time to pass to be sure the remote TCP received.

CLOSED: represents no connection at all.

IPv6

1. 128 bits
2. Each nibble mapped as a Hex character.
3. A total of 32 hex characters.
4. 8 groups or (quads).
5. First 64 bits network id and next 64 interface id or host id.
6. 64 bits of network id = 48 bits global routable prefix + 16 bits of subnet id.
7. Leading zeros can be omitted. 00a4 → a4 and 0000 → 0
8. Double colon can appear only once. :0:0:0 → ::



Link Local: Non-routable exists on single layer2 domain.

fe80:0000:0000:0000:XXXX: XXXX: XXXX: XXXX

Unique local: Routable within administrative domain.

fc00:qqqq:qqqq:ssss:XXXX: XXXX: XXXX: XXXX -- fd00:qqqq:qqqq:ssss:XXXX: XXXX: XXXX: XXXX

Global: Routable across the internet.

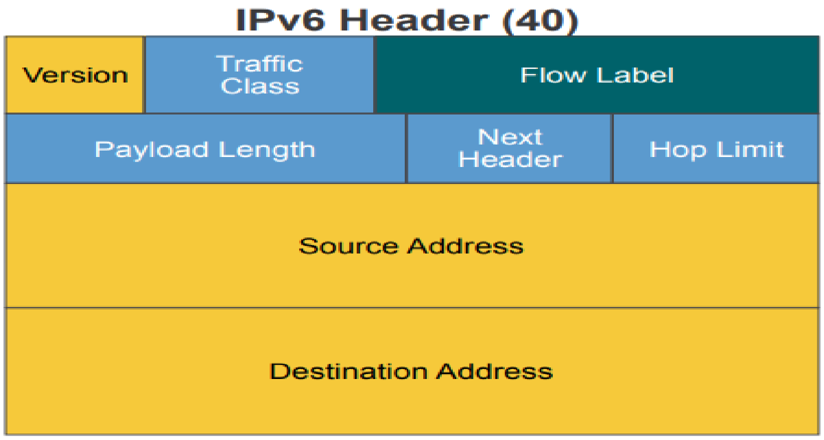
2000:NNNN:NNNN:SSSS:HHHH:HHHH:HHHH:HHHH -- 3ffff:NNNN:NNNN:SSSS:HHHH:HHHH:HHHH:HHHH

Global address assignment

IANA	/3
Registries	/12
ISP/ORG	/32
Entity	/48

Protocol ID: **86DD**

Multicast: 33:33:xx:xx:xx:xx



Version (4 bits): stores value 6.

Traffic Class (6+2 bits): MSB 6 bits hold differentiated services (related to QOS). 2 bits for Explicit congestion.

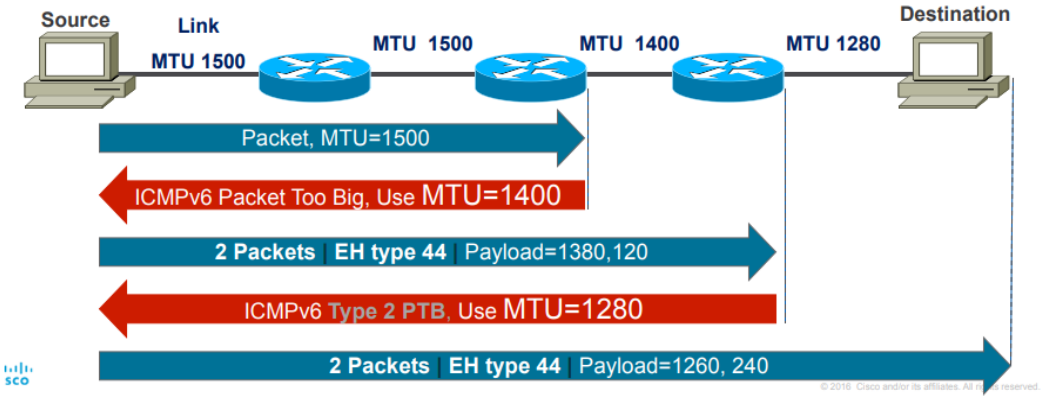
Flow Label (20 bits): for server load balancing. Set by host not altered on transit.

Payload Length: (16 bits) size of payload in octets.

Next Header: (8 bits) specifies the transport layer protocol used by the payload.

Hop Limit: (8 bits) similar to TTL of IPv4 but here the destination node must process the packet even if the value is zero.

Path MTU discovery


Source
Link
MTU 1500
MTU 1500
MTU 1400
MTU 1280
Destination
Packet, MTU=1500
ICMPv6 Packet Too Big, Use MTU=1400
2 Packets | EH type 44 | Payload=1380,120
ICMPv6 Type 2 PTB, Use MTU=1280
2 Packets | EH type 44 | Payload=1260, 240
© 2016 Cisco and/or its affiliates. All rights reserved.

IPV6 Host Portion Address Assignment

SLAAC – Stateless Address Automatic Configuration

1. Any client connecting to a router sends a router solicitation message.
2. Router in response to the message sends Router Advertisement.

   It contains network information.

   For ex: router advertises 2001:db8:1::/64 -- Prefix information.
3. The client can now automatically configure its own IPv6 address without DHCP.
4. It does not tell which DNS server to contact for name resolution.
5. To resolve the DNS, IPv6 uses DHCPv6.

Extended Unique Identifier

00 90 27 17 fc 0f

00 90 27 **ff fe** 17 fc 0f first byte 0000 00U0

If universal one U → 1 else U → 0

02 90 27 **ff fe** 17 fc 0f

Routing Protocols

Why we need Routing Protocols?

If a router does not have any routes in its IP routing table for a destination address of a packet. The router discards the packet.

Functions of routing protocol:

1. Learn routing information about IP subnets from neighboring routers.
2. Advertise routing information about IP subnets to neighboring routers.
3. If more than one route present to reach a subnet, pick the best route based on metric.
4. If the network topology changes due to link failure, react by advertising that the route and failed and pick a new currently best route.

Interior Gateway Protocols and Exterior Gateway Protocol

Interior Gateway Protocol (IGP): A networking protocol designed and intended to use inside a single autonomous system. (EIGRP, OSPF)

Exterior Gateway Protocol (EGP): A routing protocol that was designed and designated to use between different autonomous systems. (BGP is the only EGP used now)

Autonomous System: Network under administrative control of a single organization.

Routing Protocol

IGRP	- Interior Gateway Routing Protocol
EIGRP	- Enhanced Interior Gateway Routing Protocol
OSPF	- Open Shortest Path First
IS-IS	- Intermediate System to Intermediate System
BGP	- Border Gateway Protocol

IGP Routing Protocol Algorithms

Three main branches:

1. Distance Vector (IGRP)
2. Advanced Distance Vector (EIGRP)
3. Link-state (OSPF, IS-IS)

Distance Vector (EIGRP)

1. Advertise their routing table to all directly connected neighbors at regular frequent intervals using a lot of bandwidth and take time to converge.
2. When a router becomes unavailable all router tables must be updated with new information.
3. Each router having to advertise that new information to its neighbors, takes a long time for all routers to have a current accurate view of the network.
4. Uses fixed length subnet masks which are not scalable.

Link State (OSPF, IS-IS)

1. Advertise routing update only when changes occur.
2. Does not advertise routing table which makes convergence faster.

3. The routing protocol will flood the network with link state advertisement to all neighbor routers per area to converge the network with new routing information.
4. They use variable length subnet masks, which are scalable and use addressing more efficiently.

IGRP – (Interior Gateway Routing Protocol)

1. IGRP is a distance vector routing protocol for routing multiple protocols across small and medium sized networks.
2. IGRP will route IP, IPX, Decnet and AppleTalk which makes it versatile for clients running many different networks.
3. IGRP does recognize assignment of different autonomous systems and automatically summarizes at network class boundaries.
4. There is an option of load balancing traffic across equal or unequal metric cost paths.

Advantages:

1. Multicast and Unicast Instead of Broadcast address.
2. Support for authentication.
3. Manual summarization at any network interface.
4. 100% loop free classless routing.

EIGRP – Enhanced Interior Gateway Routing Protocol

Router using EIGRP does the below functions

1. Send a Hello Packet and receive an acknowledge for EIGRP.

2. Calculate metrics for all connected interfaces. Factors involved Bandwidth, Delay, Load, Reliability.

3. Advertise the metrics.

4. Build a neighbor table. A table storing labels of all directly connected neighbors.

5. Build a Topology Table. The topology table consists of label, Interface, Feasible distance and Advertised distance. The economical path is marked as successor.

Feasible distance - Cost to the neighbor + Metrics advertised by neighbor

Advertised distance - Metrics advertised by neighbor.

If a destination has several advertisements, then the cost effective one is chosen as successor.

6. Routing table which has entries of successors.

Topology

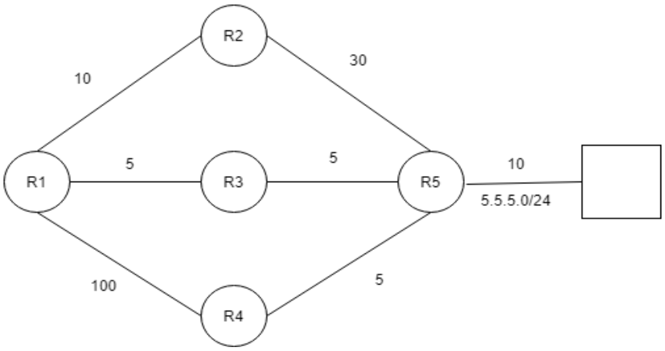


Table on R1

Neighbor Table

R2

R3

R4

Topology Table

Label	Interface	Feasible Distance	Advertised Distance
5.5.5.0/24	R2	50	40

5.5.5.0/24	R3	20	15	(Successor)
5.5.5.0/24	R4	115	15	(Feasible Successor)

Routing Table

5.5.5.0/24	R3	20	15

If the link in the routing table fails. The routing table will be replaced with feasible successor values.

Successor – The path with lowest Feasible distance

Feasible Successor – The path where FD of successor > AD of feasible successor.

OSPF – Open Shortest Path First

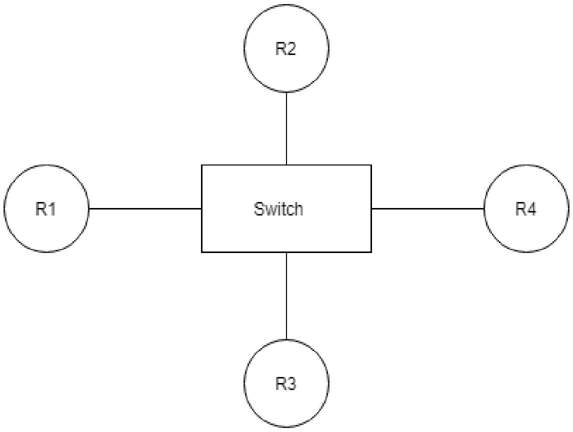
Working of OSPF

1. All routers will determine its directly connected neighbors
2. Every router will generate a LSA (Link State Advertisement) for its interfaces.
3. Once LSA are generated every router will flood the network with LSAs
4. All LSAs are used to build a network picture LSDB (Link State Database).
5. From LSDB find the Shortest Path First (SPF) Dijkstra to all nodes.
6. Place the best shortest paths in the routing table.

Calculating Cost

Cost = Reference BW / Interface BW

Select the path with lowest cost.



Designated Router and Backup Designated Router

1. Consider the topology above. All routers are adjacent to each other.
2. All routers will initiate hello packets and all routes will flood the network with LSAs.
3. To counter this problem one router will be chosen to act as a designated router. Which will handle LSA forwarding to all routers in the network. In the event of failure Backup designated router will take up the task.
4. The designate router establishes adjacencies with all routers in the network segment.
5. The Designated router generates LSA for the network and has other special responsibilities in the running of protocol.
6. The designated router sends multicast packets to all routers it has established adjacencies with.
7. The Designated router is elected by Hello protocol.
8. This helps reduce the routing protocol traffic and size of link state database.

Electing a DR and BDR

Router ID

Highest active IP address

Loopback

A router's hello

https://www.noction.com/blog/bgp_and_ospf_interaction