Chapter 6

Internetworking

Internetworking Terms (1)

- Communications Network

  A facility that provides data transfer service among stations attached to the network

- An internet

  A collection of communications networks interconnected by bridges and/or routers

- Subnetwork

  Refers to a constituent network of an internet.

- Intranet

  Corporate internet operating within the organization

  Uses Internet (TCP/IP and http) technology to deliver documents and resources

Internetworking Terms (2)

a) End System (ES)
   - Device attached to one of the networks of an internet
   - Supports end-user applications or services

b) Intermediate System (IS)
   - Device used to connect two networks
   - Permits communication between end systems attached to different subnetworks

Internetworking Terms (3)

a) Bridge
   - IS used to connect two LANs using similar LAN protocols
   - Acts as an address filter passing on packets to the required network only
   - Does not modify the contents of a packets and does not add anything to the packet
   - Operates at OSI layer 2 (Data Link)

b) Router
   - IS used to connect two (possibly dissimilar) networks
   - Uses internet protocol present in each router and each end system
   - Operates at OSI Layer 3 (Network)

Requirements of Internetworking

a) Providing a link between networks
   - Minimum physical and link control layer is needed

b) Providing for the routing and delivery of data between processes on different networks

c) Providing an accounting service that keep track of the use of various networks and routers and that maintains status info

d) Independent of network architectures: must accommodate a number of differences among networks including:

Network Architecture Features

a) Different addressing schemes

b) Different maximum packet size (packets from one network may have be broken up into smaller pieces for another. This process is referred to as segmentation or fragmentation

c) Different network-access mechanisms.

d) Different timeouts. Internetwork timing procedure must allow for successful transmission that avoids unnecessary retransmissions.

e) Error recovery

f) Status reporting: info on internetworking activity to interested and authorized processes.

g) Routing techniques: Internetwork routing may depend on fault detection and congestion control techniques peculiar to each network; the internetworking facility must be able to coordinate these to adaptively route data between stations on different networks.

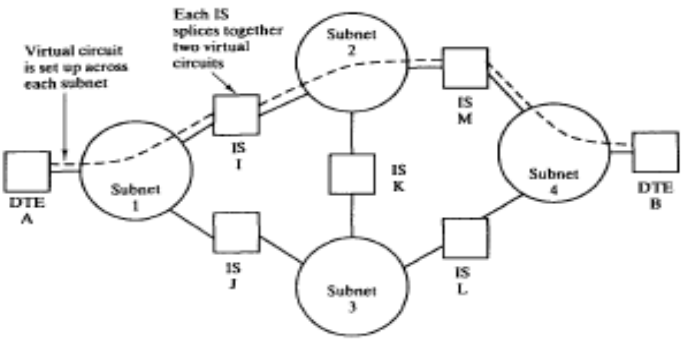
h) User- access control- each network will have its own user-access control technique that must be invoked by the internet facility as needed.

i) Connection based or connectionless service. Individual networks may provide connection-oriented (virtual circuits) or connectionless (datagram) service. It may be desirable for the internetwork service not to depend on the nature of the connection service of individual networks.

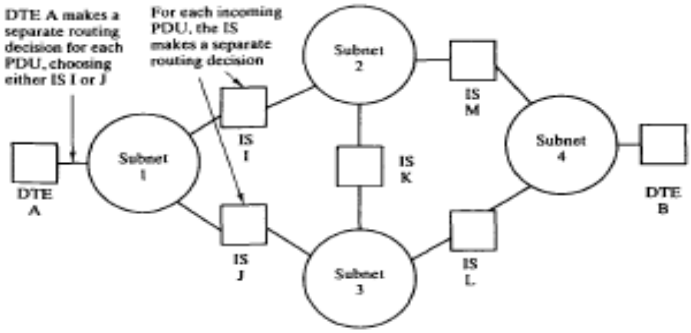
Architectural Approaches

a) Connection oriented
b) Connectionless

Architectural Approaches



(a) Connection-mode operation



(b) Connectionless-mode operation

Connection Oriented

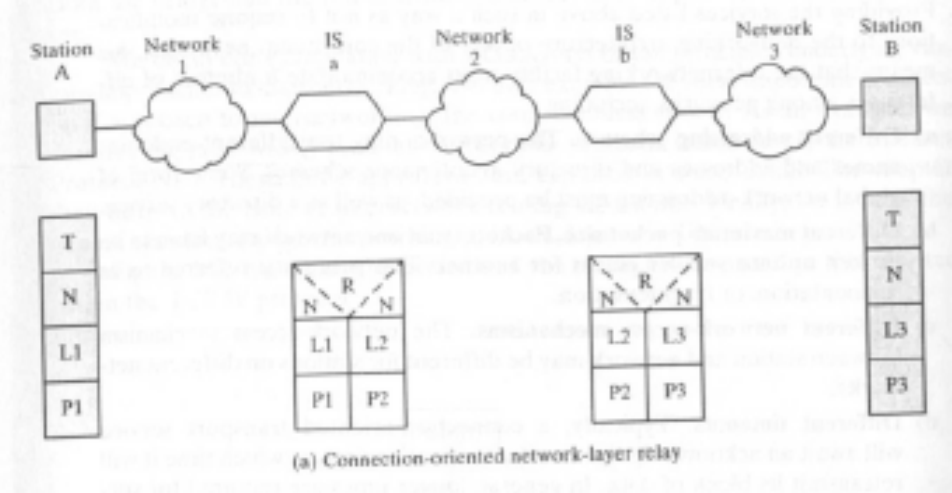
a) Assume that each network is connection oriented
b) IS connect two or more networks
   - IS appear as ES to each network
   - Logical connection set up between ESs
     - Concatenation of logical connections across networks
   - Individual network virtual circuits joined by IS
c) May require enhancement of local network services
   - 802, FDDI are datagram services

Connection-Mode Operation

Each subnetwork provides a connection-mode form of vervice

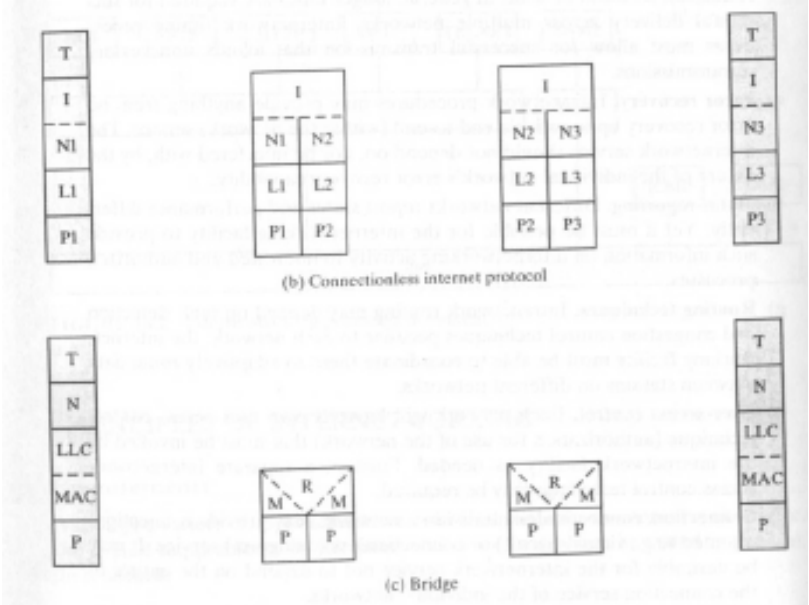
- ISs (referred as routers)used to connect 2 or more subnetworks;
- Connection Oriented IS Functions:
  - Relaying: data entities
  - Routing
  - e.g. X.75 used to interconnect X.25 packet switched networks
  - Connection oriented not often used
  - (IP dominant)

Internetwork Architecture



(a) Connection-oriented network-layer relay

Internetwork Architecture



(b) Connectionless internet protocol

(c) Bridge

Connection-Mode Operation

Connection oriented router

- Relaying
- Routing

All of the END systems share common protocols at layer 4 (Transport) and above.

Connectionless Operation

a) Corresponds to datagram mechanism in packet switched network

b) Network layer protocol common to all DTEs and routers

- Known generically as the internet protocol

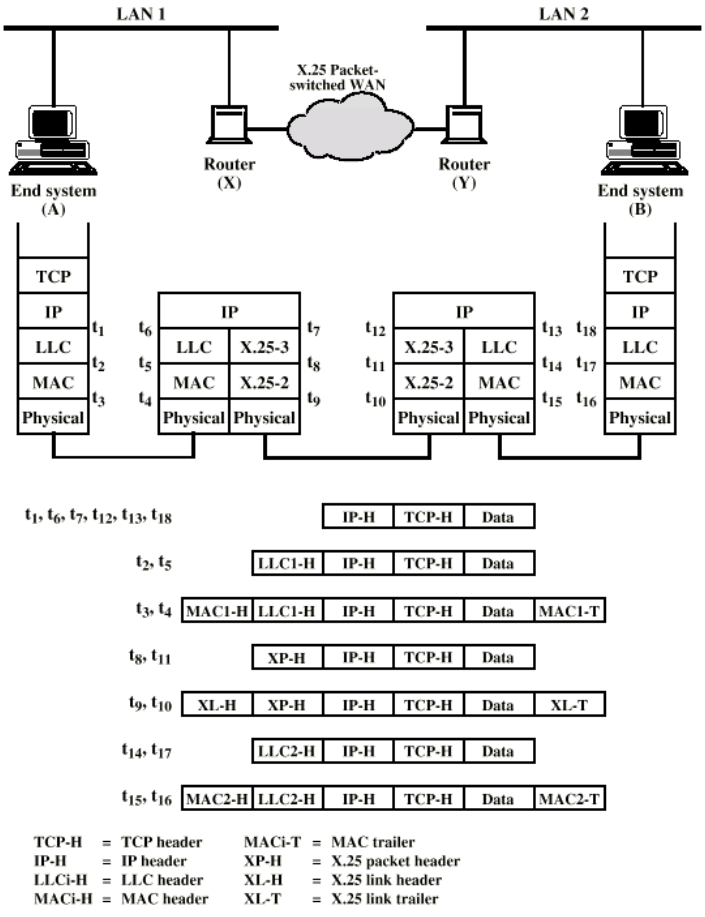
c) Upper layer protocol that provides the internetworking function - DTE

d) Lower layer protocol needed to access particular network - DTE

a) Advantages
   - Flexibility
   - Robust
   - No unnecessary overhead

b) Unreliable
   - Not guaranteed delivery
   - Not guaranteed order of delivery
     - Packets can take different routes
   - Reliability is responsibility of next layer up (e.g. TCP)

IP Operation



IP Operation

Condition:

- the End systems and routers must all share a common internet protocol
- the End system must share the same protocols above IP

The IP at A receives blocks of data to be sent to B from the higher layers software in A

IP attaches a header specifying among others things, the global internet address of B. that address is logically in 2 parts:

- Network identifier
- End system identifier

The result = internet protocol data unit = datagram

The datagram is encapsulated with the LAN protocol and sent to the router.

The router encapsulate the datagram with the X25 protocol fields and transmit it across the WAN to another router.

This router strips off the X25 fields and recovers the datagram which is then wraps in LAN fields appropriate to LAN2 and sends it to B

1) Destination address Y connected directly to one of the subnetworks to which the router is attached

   - router sends the datagram directly to the destination

2) To reach the destination one or more addiotional routers must be traversed. In this case a routing decision must be made: to which router should the datagram be sent?

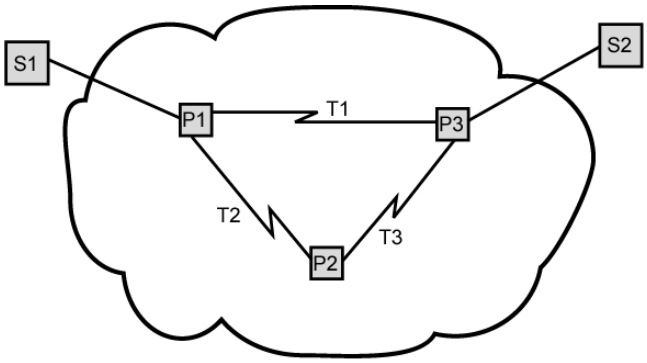
   In both cases the IP module sends the datagram down to the next lower layer with the destination subnetwork address.

3) Router does not know the destination address and returns an error message to the source of the datagram

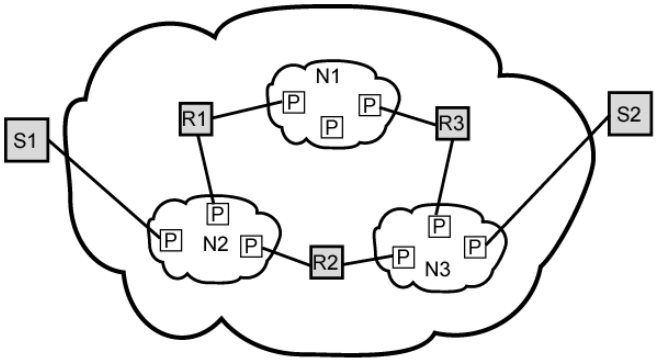
Design Issues

a) Routing
b) Datagram lifetime
c) Fragmentation and re-assembly
d) Error control
e) Flow control

The Internet as a Network



(a) Packet-switching network architecture



(b) Internetwork architecture

Routing

End systems and ***routers*** maintain routing tables

- Indicate next router to which datagram should be sent

a) Static
   - May contain alternative routes
- ***Dynamic***
   - Flexible response to congestion and errors

b) Source routing
- Source specifies route as sequential list of routers to be followed
- Security
- Priority

c) Route recording

- Record a route: router appends its internet address to a list of addresses in the datagram
- useful for testing and debugging purposes

Datagram Lifetime

a) Datagrams could loop indefinitely
   - Consumes resources
   - Transport protocol may need upper bound on datagram life

b) Datagram marked with lifetime
   - Time To Live field in IP
   - Once lifetime expires, datagram discarded (not forwarded)
   - Hop count
      - Decrement time to live on passing through a each router
   - Time count
      - Need to know how long since last router
      - Need some global clocking mechanism

Segmentation and Re-assembly

a) Different packet sizes for individual subnetw
b) Segmentation in the course of their travel
c) When to re-assemble
   - At destination
     - Results in packets getting smaller as data traverses internet
   - Intermediate re-assembly
     - Need large buffers at routers
     - Buffers may fill with fragments
     - All fragments must go through same router
       - Inhibits dynamic routing

Segmentation and Re-assembly(2)

a) IP re-assembles at destination only
b) Uses fields in header
   - Data Unit Identifier (ID)
     - Identifies end system originated datagram
       - Source and destination address
       - Protocol layer generating data (e.g. TCP)
       - Identification supplied by that layer
   - Data length
     - Length of user data in octets

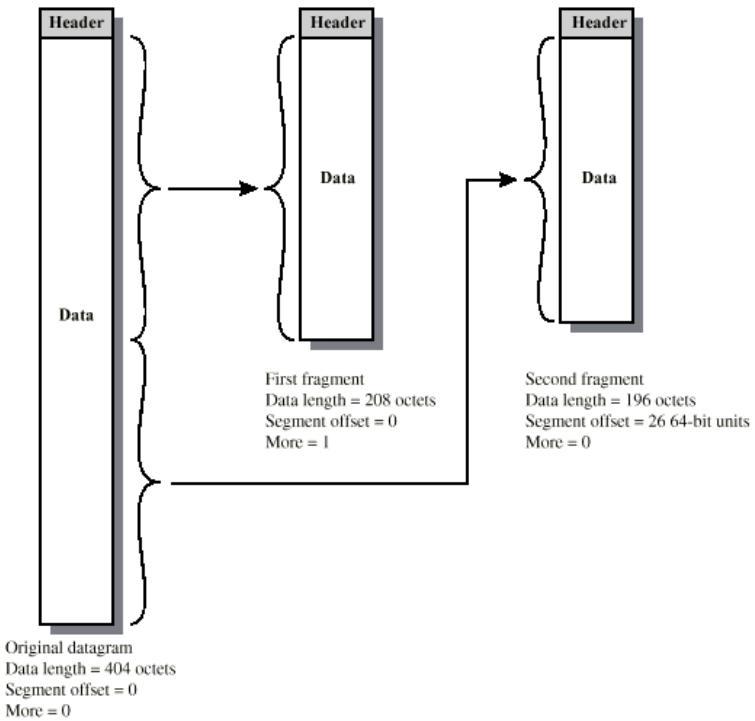
Segmentation and Re-assembly(3)

- Offset
  - Position of fragment of user data in original datagram
  - In multiples of 64 bits (8 octets)
- *More* flag
  - Indicates that this is not the last fragment

Source end system creates a datagram with:

- Data length = entire length of the data field
- Offset = 0
- *More* flag = 0

Fragmentation Example



Dealing with Failure

a) Re-assembly may fail if some fragments get lost
b) Need to detect failure
c) Re-assembly time out
   - Assigned to first fragment to arrive
   - If timeout expires before all fragments arrive, discard partial data
d) Use packet lifetime (time to live in IP)
   - If time to live runs out, kill partial data

Error Control

a) Not guaranteed delivery
b) Router should attempt to inform source if packet discarded
   - e.g. for time to live expiring
c) Source may modify transmission strategy
d) May inform high layer protocol
e) Datagram identification needed

Flow Control

a) Allows routers and/or stations to limit rate of incoming data
b) Limited in connectionless systems
c) Send flow control packets
   - Requesting reduced flow
d) e.g. ICMP

Internet Protocol (IP)

Part of TCP/IP

- Used by the Internet
- Similar to CLNP

IP is specified in 2 parts:

- the interface with higher layer e.g. TCP – services provided by IP
- Specifies protocol format and mechanisms

IP Services

a) Primitives
   - Functions to be performed
   - Form of primitive implementation dependent
     - e.g. subroutine call
   - Send
     - Request transmission of data unit
   - Deliver
     - Notify user of arrival of data unit

b) Parameters
   - Used to pass data and control info

Parameters (1)

a) Source address
b) Destination address
c) Protocol
   - Recipient e.g. TCP

d) Type of Service
   - Specify treatment of data unit during transmission through networks

e) Identification
   - Source, destination address and user protocol
   - Uniquely identifies PDU
   - Needed for re-assembly and error reporting
   - Send only

Parameters (2)

f) Don't fragment indicator
- Can IP fragment data
- If not, may not be possible to deliver
- Send only

g) Time to live
- Send only
- Measured in network hops

h) Data length

i) Option data – options requested by the IP user

j) Data - User data to be transmitted

Parameters (3)

a) Precedence
- 8 levels

b) Reliability
- Normal or high

c) Delay
- Normal or low

d) Throughput
- Normal or high

Options

a) Security
   - Security label

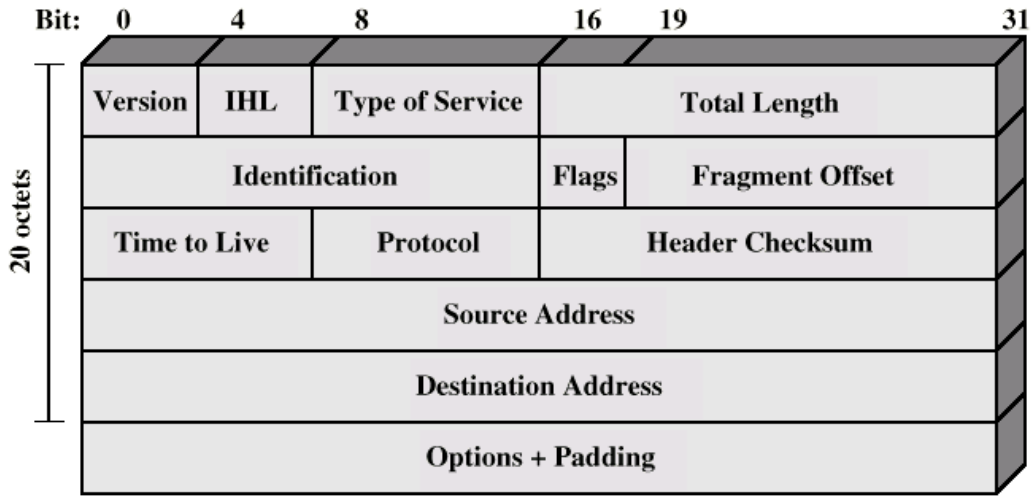
b) Source routing
   - Sequenced list of router addresses that specifies the route to be followed
     - Strict
     - Loose

c) Route recording
   - Record the sequence of routers visited by the datagram

d) Stream identification
   - Names reserved resources used for stream service

e) Timestamping
   - Timestamps are added to the data by the source IP entity and some intermediate routers

IP Protocol



Header Fields (1)

a) Version
   - Currently 4
   - IP v6 - see later

b) Internet header length (IHL)
   - In 32 bit words
   - Including options

c) Type of service
   - Specifies reliability, precedence, delay and throughput paarmeters

d) Total length
   - Of datagram, in octets

Header Fields (2)

e) Identification
   - Sequence number
   - Used with addresses and user protocol to identify datagram uniquely

f) Flags
   - More bit: used for segmentation, and reassembly
   - Don’t fragment: prohibits segmentation when set

g) Fragmentation offset
   - Where in the original datagram this fragment belongs

h) Time to live
   - Measured in router hops

i) Protocol

Header Fields (3)

j) Header checksum
   - Applied to the header only
   – Reverified and recomputed at each router
   – 16 bit ones complement sum of all 16 bit words in header
   – Set to zero during calculation

k) Source address
   - Coded to allow a variable allocation

l) Destination address

m) Options
   - Encodes the options requested by the sending user

Data Field

a) Carries user data from next layer up

b) Integer multiple of 8 bits long (octet)

c) Max length of datagram (header plus data) 65,535 octets

IP Addresses

a) Source and destination address in the IP header contain each a 32-bit global internet address consisting of:
   - A network identifier
   - A host identifier

b) Address is coded:
   - to allow a variable allocation of bits
   - Flexibility in assigning addresses to hosts
   - Allow a mix of network sizes on internet

IP Addresses - Class A

a) 32 bit global internet address

b) Network part and host part

c) Class A - few networks each with many hosts
   - Start with binary 0
   - All 0 reserved
   - 01111111 (127) reserved for loopback
   - Range 1.x.x.x to 126.x.x.x
   - All allocated

IP Addresses - Class B

Medium no of networks, each with a medium no of hosts

a) Start 10
b) Range 128.x.x.x to 191.x.x.x
c) Second Octet also included in network address
d) 2^{14} = 16,384 class B addresses
e) All allocated

IP Addresses - Class C

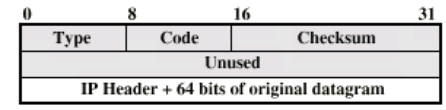
Many networks each with a few hosts

a) Start 110
b) Range 192.x.x.x to 223.x.x.x
c) Second and third octet also part of network address
d) 2^{21} = 2,097,152 addresses
e) Nearly all allocated
   - See IPv6

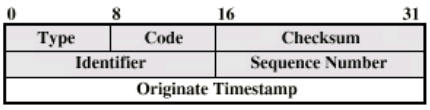
ICMP

a) Internet Control Message Protocol
b) RFC 792
c) Transfer of (control) messages from routers and hosts to hosts
d) Feedback about problems
   - e.g. time to live expired
e) Encapsulated in IP datagram
   - Not reliable

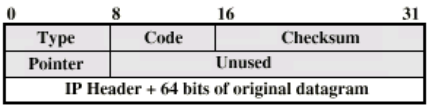
ICMP Message Formats



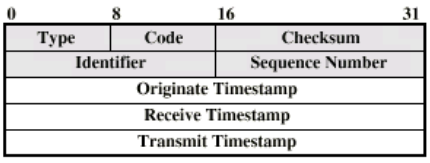
(a) Destination Unreachable; Time Exceeded; Source Quench



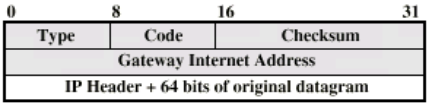
(e) Timestamp



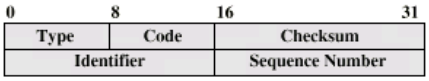
(b) Parameter Problem



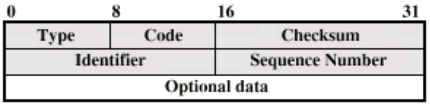
(f) Timestamp Reply



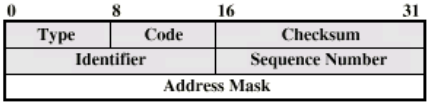
(c) Redirect



(g) Address Mask Request



(d) Echo, Echo Reply



(h) Address Mask Reply

Routing Protocols

I) Routers makes routing decision based on knowledge of topologyand on the condition of the internet

- simple internet – a fixed routing scheme
- complex internets dynamic cooperation of the routers

II)

a) Routing Information

– About topology and delays in the internet

b) Routing Algorithm

– Used to make routing decisions for a particular datagram, based on current routing information

III)

a) Routing between end systems and routers

b) Routing between routers

Autonomous Systems (AS)

a) Group of routers

b) Exchange information

c) Common routing protocol

d) Set of routers and networks managed by single organization

e) A connected network

– There is at least one route between any pair of nodes

Interior Router Protocol (IRP)

a) Passes routing information between routers within AS

b) May be more than one AS in internet

c) Routing algorithms and tables may differ between different AS

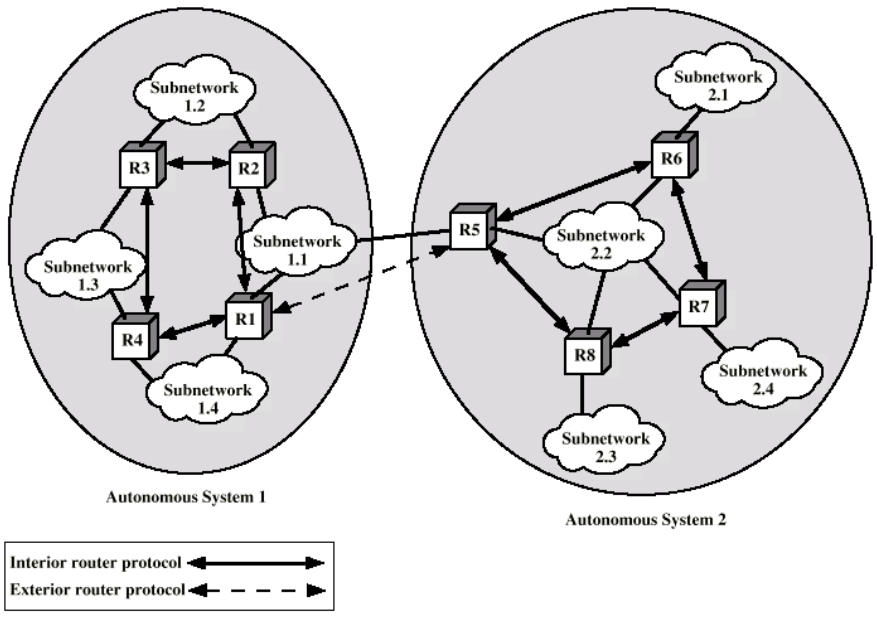
d) Routers need some info about networks outside their AS

e) Used exterior router protocol (ERP)

f) IRP needs detailed model

g) ERP supports summary information on reachability

Application of IRP and ERP



Border Gateway Protocol (BGP)

a) For use with TCP/IP internets

b) Preferred EGP of the Internet

c) Operates in terms of messages sent over TCP connections

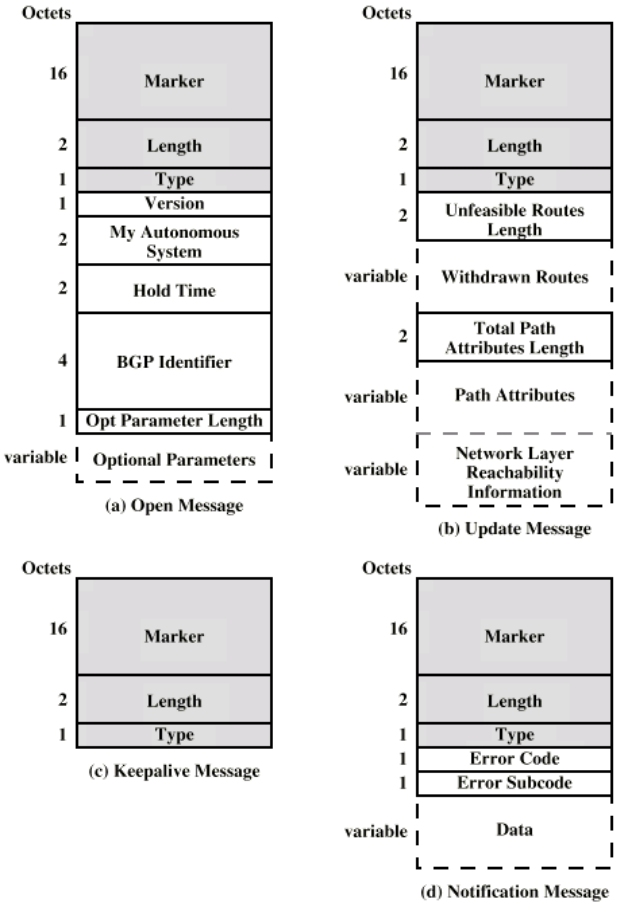
- Open: open a neighbor relationship with another router
- Update:
  - transmit info about a single route
  - list multiple routes to be withdraw
- Keep alive:
  - Ack an open message
  - Periodically confirm the neighbor relationship
- Notification: send when an error condition is detected

Border Gateway Protocol (BGP)

d) Procedures

- Neighbor acquisition
  - Request message and acceptance
  - Open message and Keep alive message
- Neighbor reachability
- Network reachability
  - Data base of the networks that it can reach
  - Update message - routers can build up and maintain routing info

BGP Messages



BGP Procedure

a) Open TCP connection to the neighbor of interest

b) Send Open message
   - Includes proposed hold time

c) Receiver selects minimum of its hold time and that sent in the Open message
   - Max time between Keep alive and/or update messages

Message Types

a) Keep Alive
   - To tell other routers that this router is still here

b) Update
   - Info about single routes through internet
     - Network layer reachability info (NLRI) field
     - Total path attributes length field
     - Path Attributes field: list of attribute that apply to this particular route
       - Origin (IGP or EGP)
       - AS_Path (list of AS traversed)
       - Next_hop (IP address of boarder router)
       - Multi_Exit_Disc (Info about routers internal to AS)
       - Local_pref (Inform other routers within AS)
       - Atomic_Aggregate, Aggregator (Uses address tree structure to reduce amount of info needed)
   - List of routes being withdrawn

Uses of AS_Path and Next_Hop

a) AS_Path
   - Enables routing policy
     - Avoid a particular AS
     - Security
     - Performance
     - Quality
     - Number of AS crossed

b) Next_Hop
   - Only a few routers implement BGP
     - Responsible for informing outside routers of routes to other networks in AS

Notification Message

a) Message header error
   - Authentication and syntax

b) Open message error
   - Syntax and option not recognized
   - Unacceptable hold time

c) Update message error
   - Syntax and validity errors

d) Hold time expired
   - Connection is closed

e) Finite state machine error

f) Cease
   - Used to close a connection when there is no error

BGP Routing Information Exchange

a) Within AS, router builds topology picture using IGP

b) Router issues Update message to other routers outside AS using BGP

c) These routers exchange info with other routers in other AS

d) Routers must then decide best routes

BGP Routing Info Exchange

Router - BGP

- internal routing OSPR-
  - routing info with all routers in AS1
  - build up a picture of the topology of subnetworks and routers in AS1, construct a routing table
- R1 issues an Update message to R5 in AS2:
  - AS-Path: the identity of AS1
  - Next_Hop: IP address of R1
  - NLRI: a list of all subnetworks in AS1

Suppose R5 also has a neighbor relationship with another router in anotherAS, say R9 in AS3. R5 will forward the info just received in a new Update message

- AS-Path: list of identifiers: {AS1,AS2}
- Next_Hop: IP address of R5
- NLRI: a list of all subnetworks in AS1

Informs R9 that all subnetworks listed in NLRI are reachable via R5 and that autonomous systems traversed are AS2 and AS1.

R9 decide the preferable route to the subnetworks listed

If R9 decides that the route provided in R5's Update message is preferable, then R9 incorporates the routing info.

Open Shortest Path First (1)

a) OSPF

b) IGP of Internet

c) Replaced Routing Information Protocol (RIP)

d) Uses Link State Routing Algorithm

- Each router keeps list of state of local links to network
- Transmits update state info
- Little traffic as messages are small and not sent often
- RFC 2328

e) Route computed on least cost ***based on user cost metric*** (cost expresses a function of delay, data rate, dollar cost, other factors)

Open Shortest Path First (2)

f) Topology stored as directed graph

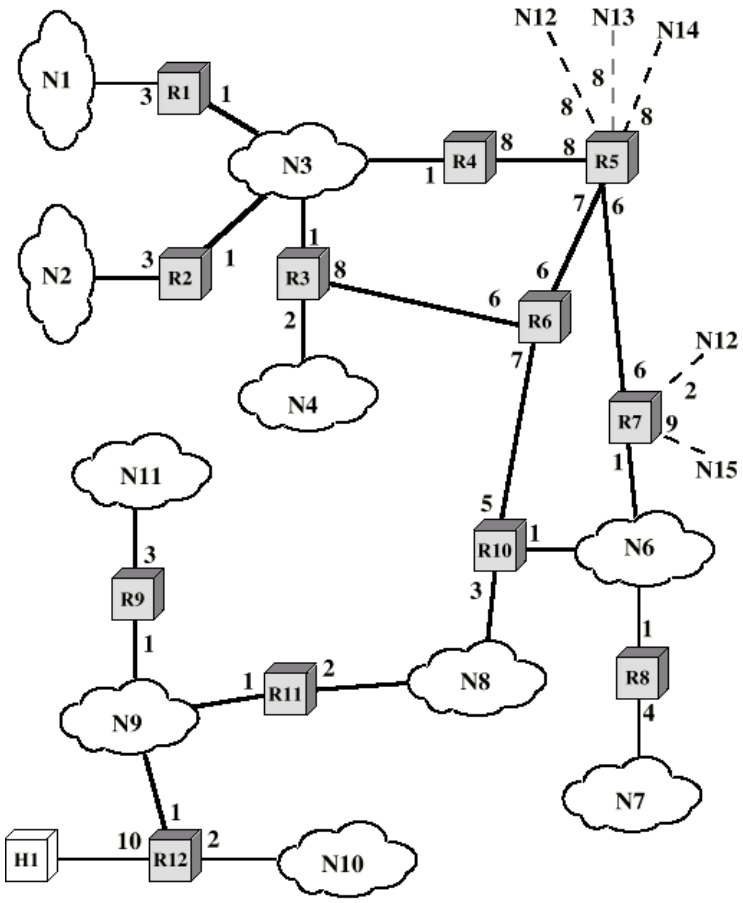
g) Vertices or nodes

- Router
- Network
  - Transit
  - Stub

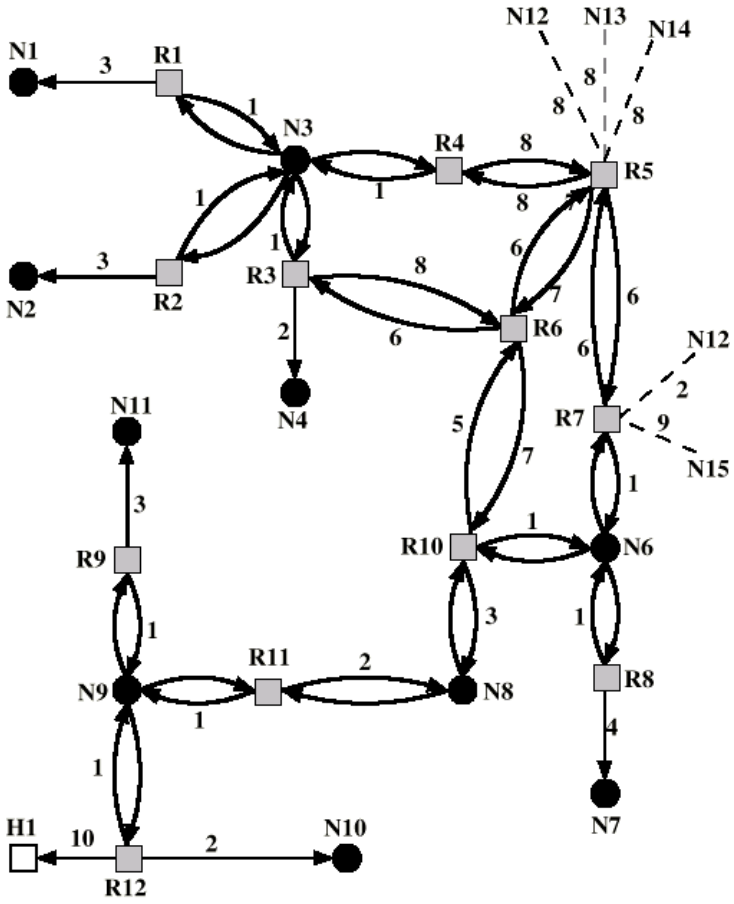
h) Edges

- Graph edge
  - Connect two router
  - Connect router to network

Sample AS



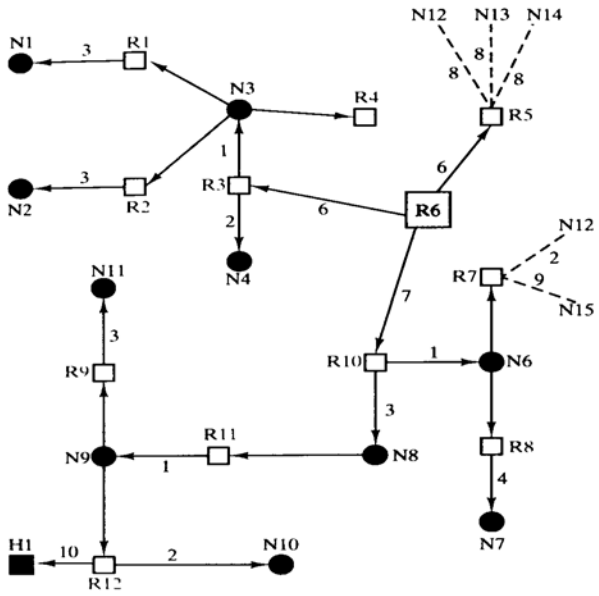
Directed
Graph of AS



Operation

- Dijkstra's algorithm used to find least cost path to all other networks
- Next hop used in routing packets

SPF for Router 6



Routing Table for RT 6

Destination	Next Hop	Distance
N1	RT3	10
N2	RT3	10
N3	RT3	7
N4	RT3	8
N6	RT10	8
N7	RT10	12
N8	RT10	10
N9	RT10	11
N10	RT10	13
N11	RT10	14
H1	RT10	21
RT5	RT5	6
RT7	RT10	8
N12	RT10	10
N13	RT5	14
N14	RT5	14
N15	RT10	17

IP v6 - Version Number

a) IP v 1-3 defined and replaced
b) IP v4 - current version
c) IP v5 - streams protocol
d) IP v6 - replacement for IP v4
   - During development it was called IPng
   - Next Generation

Why Change IP?

- Address space exhaustion
  - Two level addressing (network and host) wastes space
  - Network addresses used even if not connected to Internet
  - Growth of networks and the Internet
  - Extended use of TCP/IP
  - Single address per host
- Requirements for new types of service

IPv6 Enhancements

a) Expanded address space
   - 128 bit

b) Improved option mechanism
   - Separate optional headers between IPv6 header and transport layer header
   - Most are not examined by intermediate routes
     - Improved speed and simplified router processing
     - Easier to extend options

c) Address autoconfiguration
   - Dynamic assignment of addresses

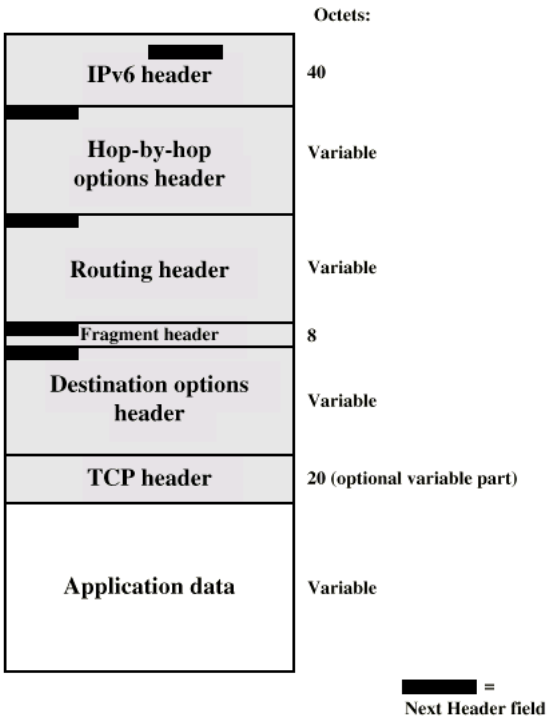
IPv6 Enhancements(2)

a) Increased addressing flexibility
   - Anycast - delivered to one of a set of nodes
   - Improved scalability of multicast addresses

b) Support for resource allocation
   - Replaces type of service
   - Labeling of packets to particular traffic flow
   - Allows special handling
   - e.g. real time video

c) Security capabilities
   - Features that support authentication and privacy

Structure



Extension Headers

a) Hop-by-Hop Options
   - Require processing at each router

b) Routing
   - Similar to v4 source routing

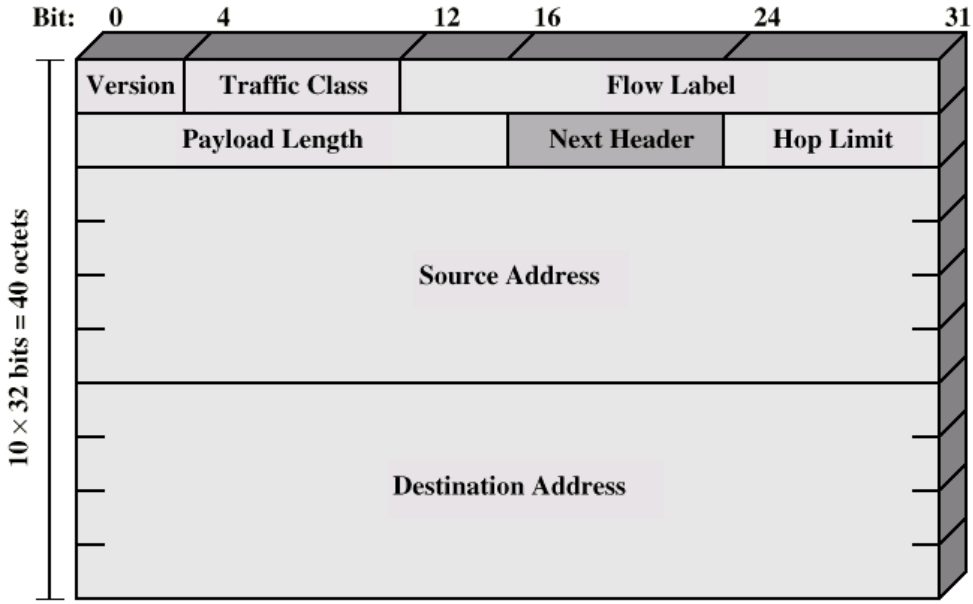
c) Fragment

d) Authentication

e) Encapsulating security payload

f) Destination options
   - For destination node

IP v6 Header



IP v6 Header Fields (1)

a) Version
   - 6

b) Traffic Class (Priority)
   - Classes or priorities of packet
   - Still under development
   - See RFC 2460

c) Flow Label
   - Used by hosts requesting special handling

d) Payload length
   - Includes all extension headers plus user data

IP v6 Header Fields (2)

e) Next Header
   - Identifies type of header
     - Extension or next layer up

f) Hop limit
   - Remaining number of allowable hops for this packet
   - Set to some desired value by source
   - Decremented

g) Source Address
   - Originator of the packet

h) Destination address
   - intended recipient
   - not the intended ultimate destination if a Routing header is present

Priority Field

Enables a source

- to identify the desired transmit and delivery priority of each packet relative to other packets from the same source
- to specify separate priority – related characteristics for each packet

Packets classification:

- being part of traffic for which the source is providing congestion control
- being part of traffic for which the source isnot providing congestion control

Packets are assigned one of eight levels of relative priority within each classification

IPv6 Priorities

Priority	Congestion-controlled traffic	Priority	Non congestion control traffic
0	Uncharacterized traffic	8	Most willing to discard(high – fidelity video)
1	“Filler” traffic (net news)	9	
2	Unattended data transfer(mail)		
3	(reserved)		
4	Attended bulk transfer (FTP, HTTP)		
5	(reserved)		
6	Interactive traffic (Telnet)		
7	Internet control traffic (routing protocols)	15	Least willing to discard(low – fidelity audio)

Flow label

IPv6 defines a flow as a sequence of packets sent from a particular (unicast or multicast) destination for which the source desires special handling by intervening routers.

Flow uniquely identified by:

- a combination of a source address and a non-zero 24-bit flow label

Source’s point of view

Router’s point of view

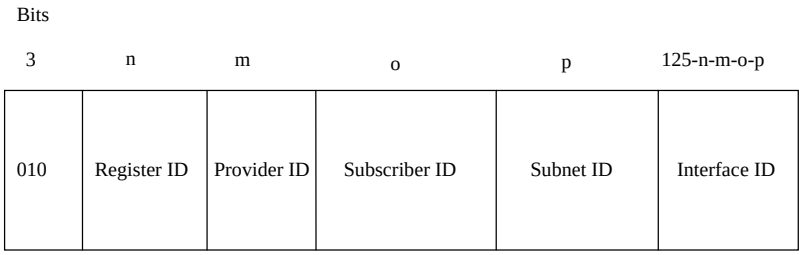
- path, resource allocation, discard requirements, accounting, security attributes

IPv6 Addresses

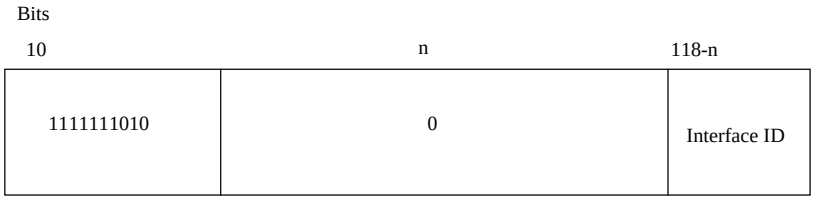
a) 128 bits long
b) Assigned to individual interfaces on nodes (include hosts + routers)
c) Single interface may have multiple unicast addresses
d) Combination of long addresses and multiple addresses per interface enables routing efficiency
e) Three types of address
   - Unicast
     - Single interface
   - Anycast
     - Set of interfaces (typically different nodes)
     - Delivered to any one interface
     - the “nearest”
   - Multicast
     - Set of interfaces
     - Delivered to all interfaces identified

Unicast addreses

(a) Provider based global unicast address

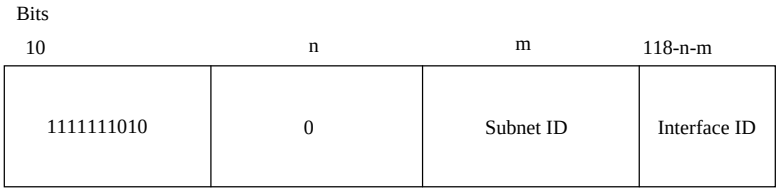


(b) Link-local address



Unicast addreses(2)

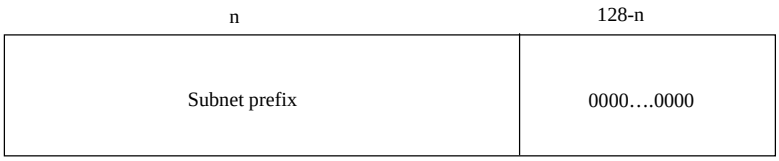
(c) Site-local address



Anycast addreses(2)

Enables a source to specify that it wants to contact anu one from a group of nodes

Subnet- router anycast address



In provider-based global address space, the subnet prefix is of the form:

010+registry ID+Provider ID+Subscriber ID+SubnetID

Addresses Autoconfiguration

Feature defined as part of IPv6 specification:

- enables a host to configure automatically one or more addresses per interface
- support “plug and play” capability- which allow a user to attach a host to a subnetwork and have IPv6 addresses atomatically assigned to its interfaces with no user intervention

3 Types

- local scope model
- stateless server model
- stateful server model

Addresses Autoconfiguration

Feature defined as part of IPv6 specification:

- enables a host to configure automatically one or more addresses per interface
- support “plug and play” capability- which allow a user to attach a host to a subnetwork and have IPv6 addresses atomatically assigned to its interfaces with no user intervention

3 Types

- local scope model
- stateless server model
- stateful server model

ICMPv6

RFC 1885

Key features:

- uses a new protocol number
- the same header format as IPv4
- some little used messages have been omitted from ICMPv6
- Larger maximum size (576 octets including IPv6 headers)

Provides a means for transferring error messages and informational messages among IPv6 nodes

Sent in response to an IPv6 packet, either by a router along the packet's path or by the intended destination node

ICMPv6(2)

Error messages

ICMPv6 includes 4 errors messages:

- destination unreachable
- packet to big
- time exceeded
- parameter problem

Each refers to a prior IPv6 packet and is sent to the originating source

Informational messages

- echo request
- echo reply
- group membership

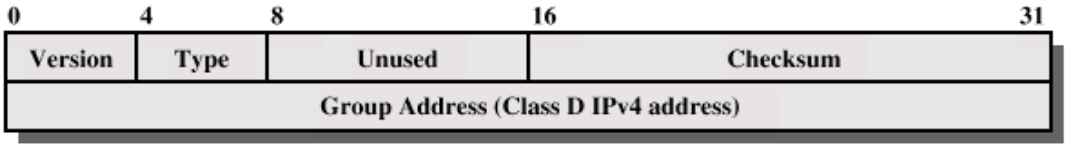
IGMP

Group- management message implements the procedure of the IGMP (Internet Group Management Protocol- definid in RFC 1112)

a) Host and router exchange of multicast group info

b) Use broadcast LAN to transfer info among multiple hosts and routers

IGMP format



IGMP(2)

a) Version
   - 1

b) Type
   - 1 - query sent by router
   - O - report sent by host

c) Checksum

d) Group address
   - Zero in request message
   - Valid group address in report message

IGMP Operation

- To join a group, hosts sends report message
  - Group address of group to join
  - In IP datagram to same multicast destination address
  - All hosts in group receive message
  - Routers listen to all multicast addresses to hear all reports
- Routers periodically issue request message
  - Sent to all-hosts multicast address
  - Host that want to stay in groups must read all-hosts messages and respond with report for each group it is in