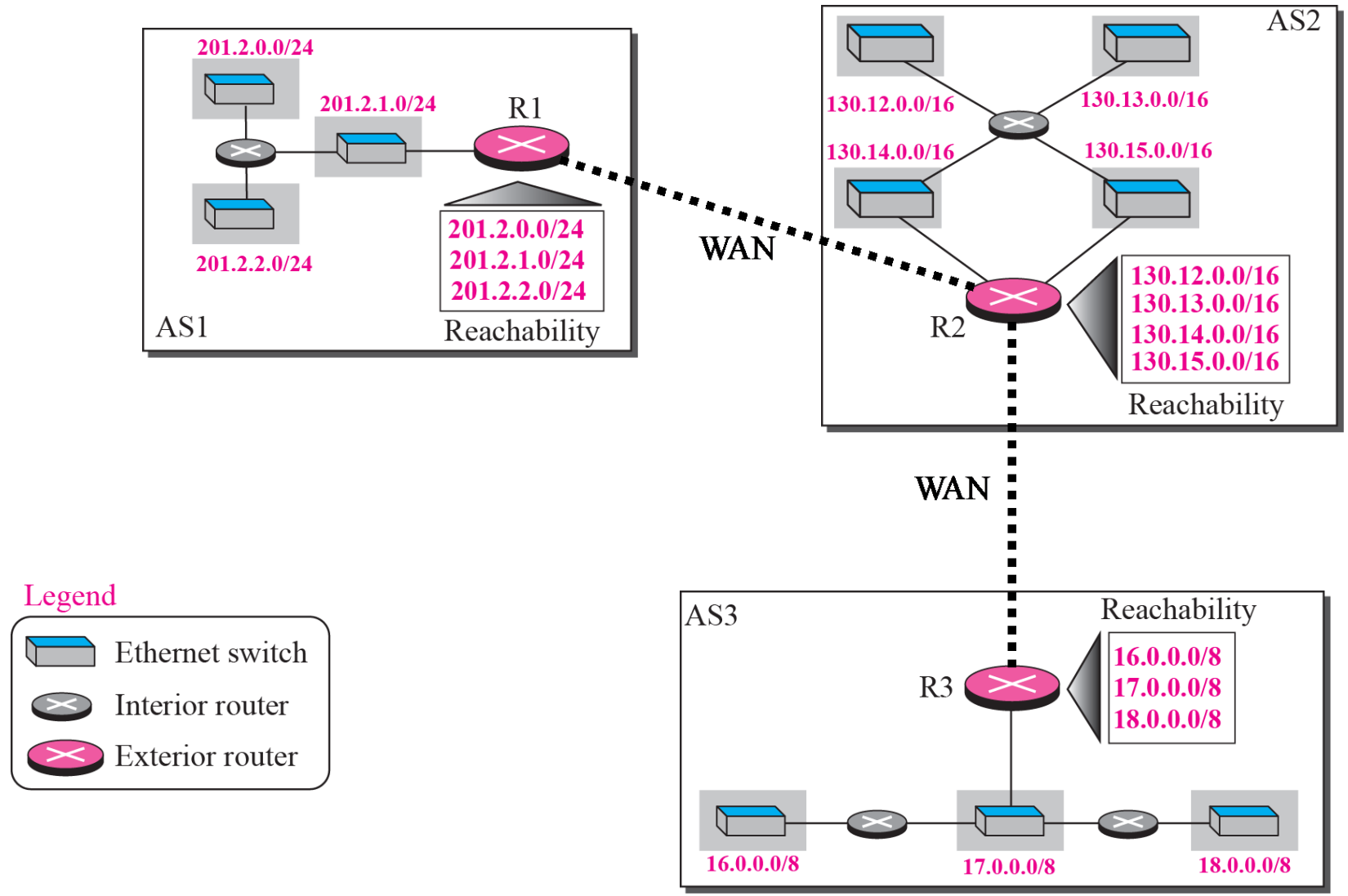
Network Layer – BGP and IPv6

Anand Baswade

anand@iitbhilai.ac.in

Reachability



Stabilized table for three autonomous system

R1

Network	Path
201.2.0.0/24	**AS1** (This AS)
201.2.1.0/24	**AS1** (This AS)
201.2.2.0/24	**AS1** (This AS)
130.12.0.0/16	AS1, AS2
130.13.0.0/16	AS1, AS2
130.14.0.0/16	AS1, AS2
130.15.0.0/16	AS1, AS2
16.0.0.0/8	AS1, AS2, AS3
17.0.0.0/8	AS1, AS2, AS3
18.0.0.0/8	AS1, AS2, AS3

Path-Vector Routing Table

R2

Network	Path
201.2.0.0/24	AS2, AS1
201.2.1.0/24	AS2, AS1
201.2.2.0/24	AS2, AS1
130.12.0.0/16	**AS2** (This AS)
130.13.0.0/16	**AS2** (This AS)
130.14.0.0/16	**AS2** (This AS)
130.15.0.0/16	**AS2** (This AS)
16.0.0.0/8	AS2, AS3
17.0.0.0/8	AS2, AS3
18.0.0.0/8	AS2, AS3

Path-Vector Routing Table

R3

Network	Path
201.2.0.0/24	AS3, AS2, AS1
201.2.1.0/24	AS3, AS2, AS1
201.2.2.0/24	AS3, AS2, AS1
130.12.0.0/16	AS3, AS2
130.13.0.0/16	AS3, AS2
130.14.0.0/16	AS3, AS2
130.15.0.0/16	AS3, AS2
16.0.0.0/8	**AS3** (This AS)
17.0.0.0/8	**AS3** (This AS)
18.0.0.0/8	**AS3** (This AS)

Path-Vector Routing Table

Routing tables after aggregation

R1

Network	Path
201.2.0.0/22	**AS1** (This AS)
130.12.0.0/18	AS1, AS2
16.0.0.0/6	AS1, AS2, AS3

Path-Vector Routing Table

R2

Network	Path
201.2.0.0/22	AS2, AS1
130.12.0.0/18	**AS2** (This AS)
16.0.0.0/6	AS2, AS3

Path-Vector Routing Table

R3

Network	Path
201.2.0.0/22	AS3, AS2, AS1
130.12.0.0/18	AS3, AS2
16.0.0.0/6	**AS3** (This AS)

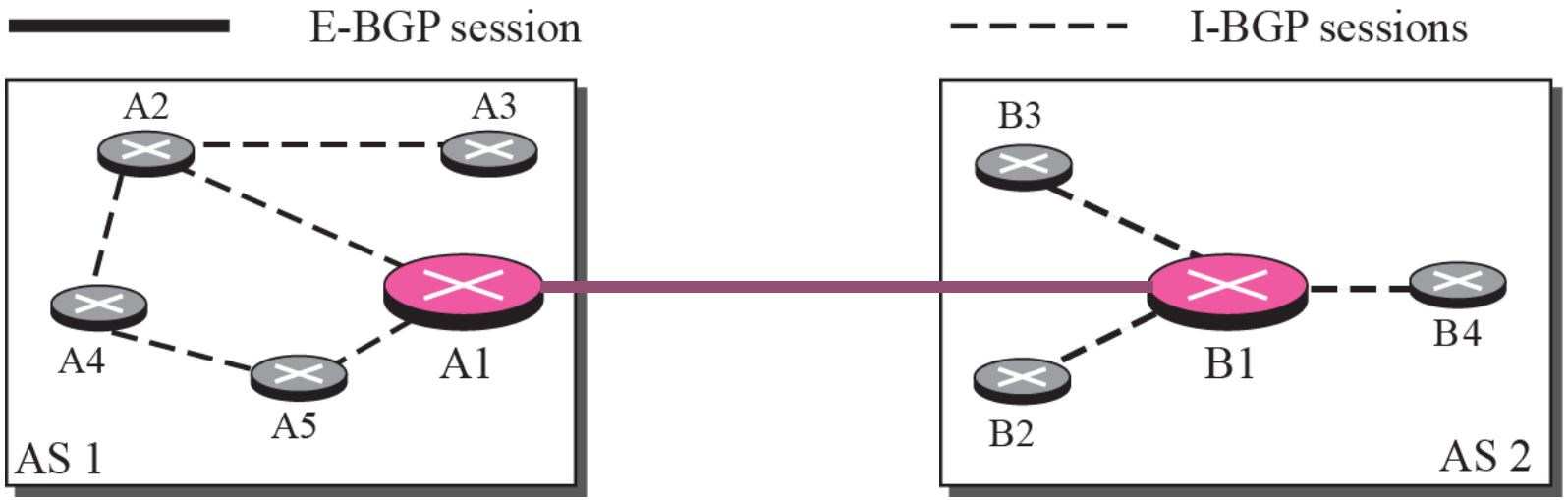
Path-Vector Routing Table

BGP

- Border Gateway Protocol (BGP) is an interdomain routing protocol using path vector routing. It first appeared in 1989 and has gone through four versions.

- BGP makes routing decision based on paths, network policies, or rule-sets configured by a network administrator and is involved in making core routing decisions.

Internal and external BGP sessions



A speaker node advertises the path, not the metric of the nodes, in its AS or other ASs.

Path Vector Routing (1)

- Sharing
  - A speaker in an AS shares its table with immediate neighbors
- Updating
  - Adding the nodes that are not in its routing table and adding its own AS and the AS that sent the table
  - The routing table shows the path completely

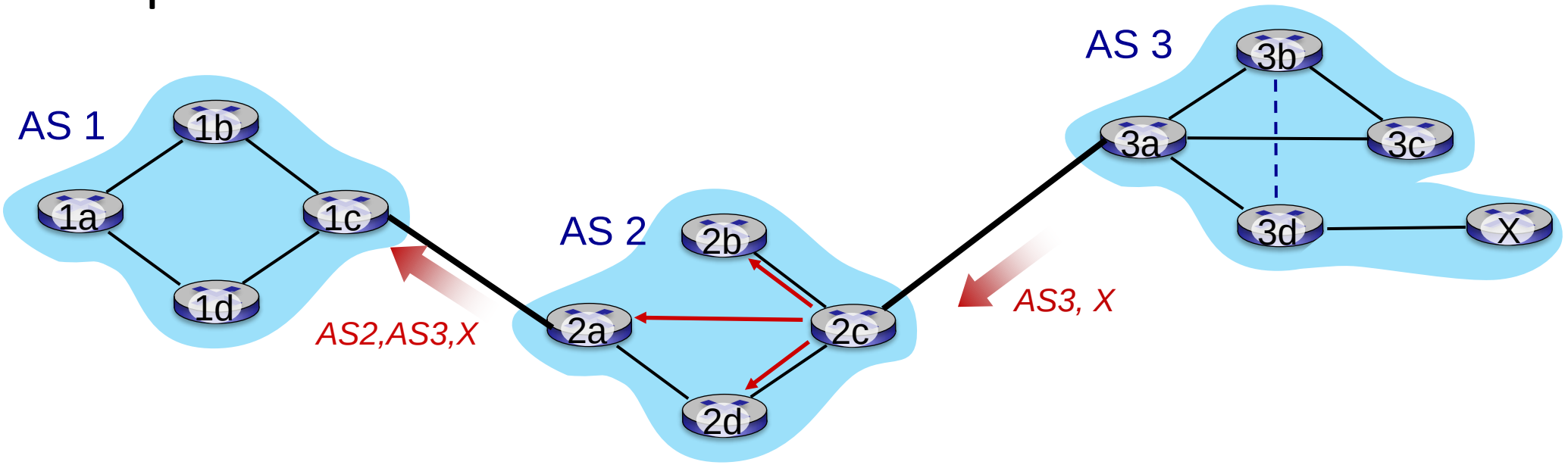
Path Vector Routing (2)

- Loop prevention
  - A route checks to see if its AS is in the path list to the destination
- Policy Based routing
  - Router receiving route advertisement to destination X uses policy to accept/reject a path (e.g., never route through AS P, or country Y)
  - Router uses policy to decide whether to advertise a path to neighboring AS Z (does router want to route traffic forwarded from Z to destined to X?)

Path Vector Routing (3)

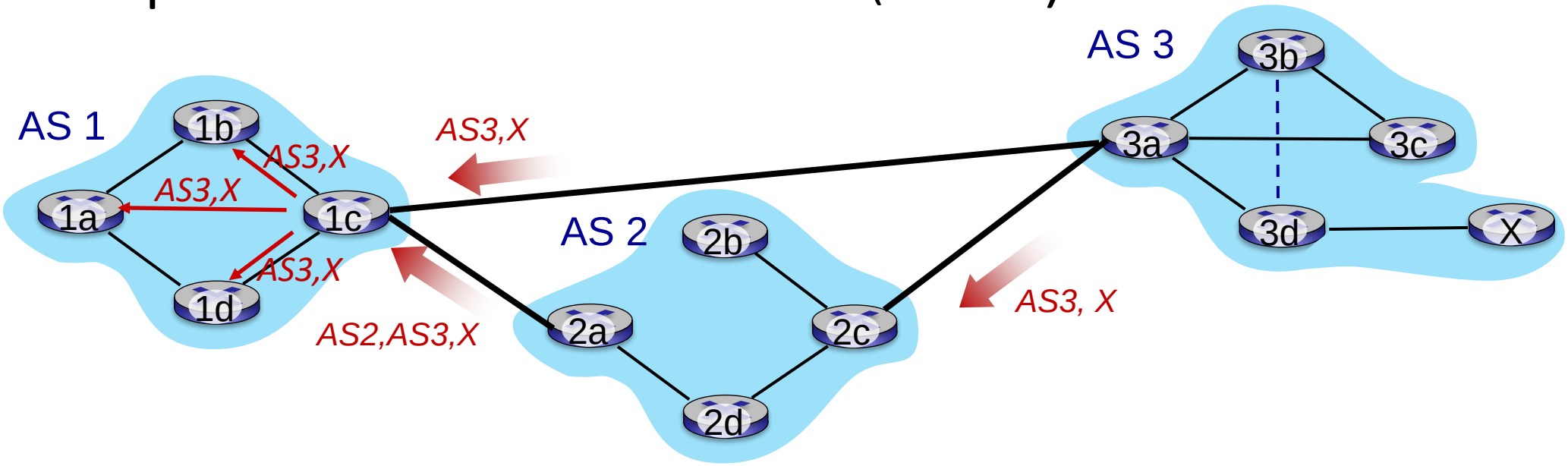
- Optimum path
  - Problem: each AS that is included in the path may use a different criteria for the metric
  - The optimum path is the path that fits the organization
  - criteria: number of ASs, security, safety, reliability, etc.

BGP path advertisement



- AS2 router 2c receives path advertisement AS3,X (via eBGP) from AS3 router 3a
- based on AS2 policy, AS2 router 2c accepts path AS3,X, propagates (via iBGP) to all AS2 routers
- based on AS2 policy, AS2 router 2a advertises (via eBGP) path AS2, AS3, X to AS1 router 1c

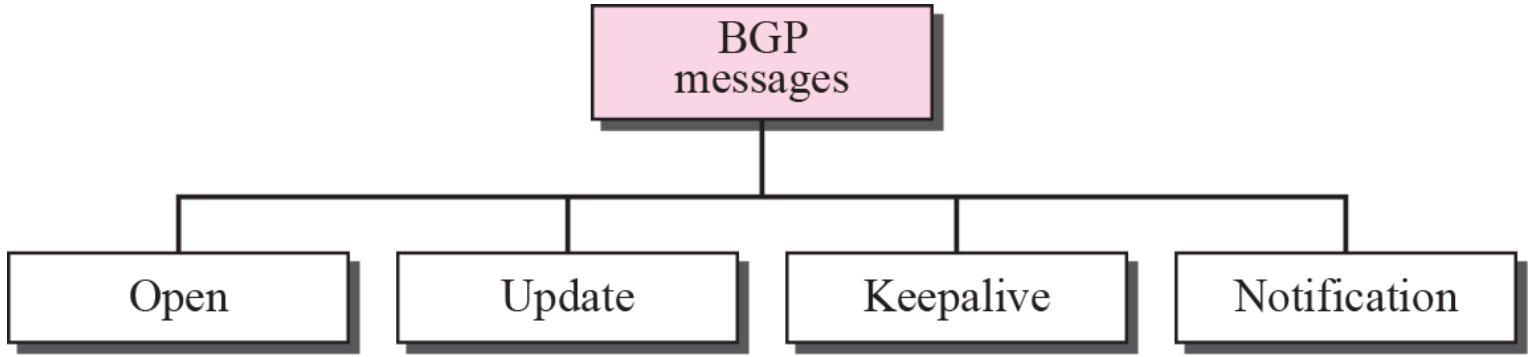
BGP path advertisement (more)



gateway router may learn about multiple paths to destination:

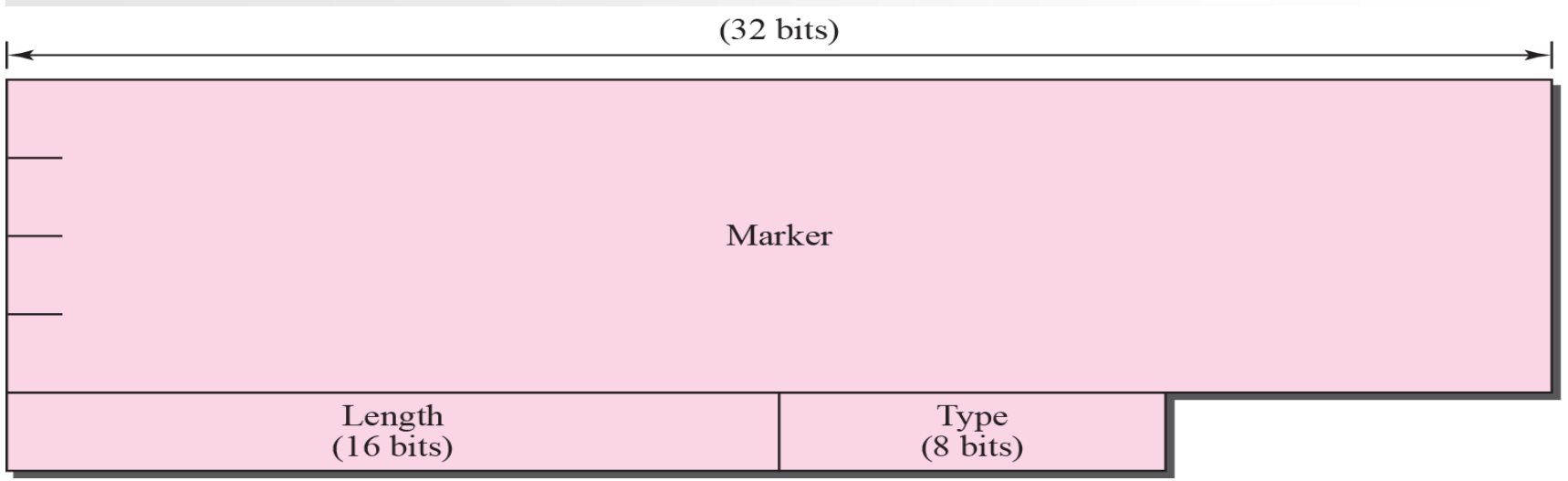
- AS1 gateway router 1c learns path *AS2,AS3,X* from 2a
- AS1 gateway router 1c learns path *AS3,X* from 3a
- based on *policy,* AS1 gateway router 1c chooses path *AS3,X* and advertises path within AS1 via iBGP

Types of BGP messages [RFC 4371]



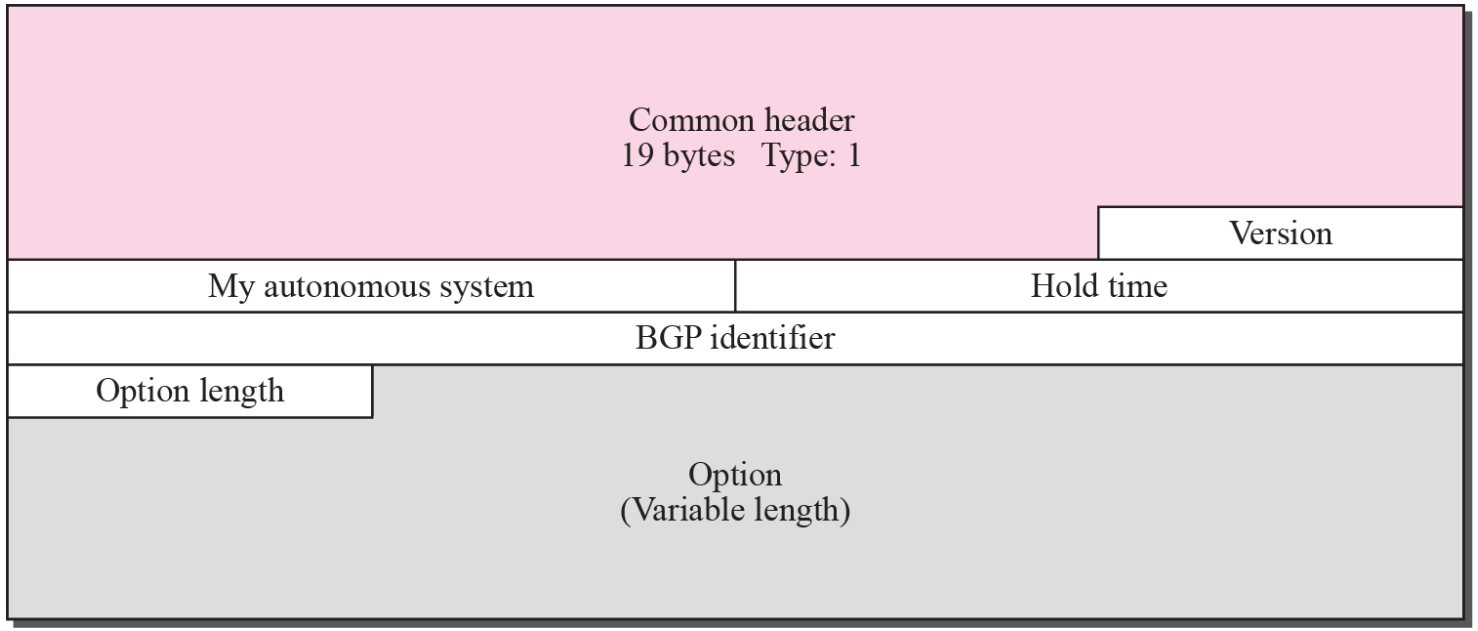
- BGP messages exchanged between peers over TCP connection
- BGP messages:
  - OPEN: opens TCP connection to remote BGP peer and authenticates sending BGP peer
  - UPDATE: advertises new path (or withdraws old)
  - KEEPALIVE: keeps connection alive in absence of UPDATES
  - NOTIFICATION: reports errors in previous msg; also used to close connection

BGP packet header



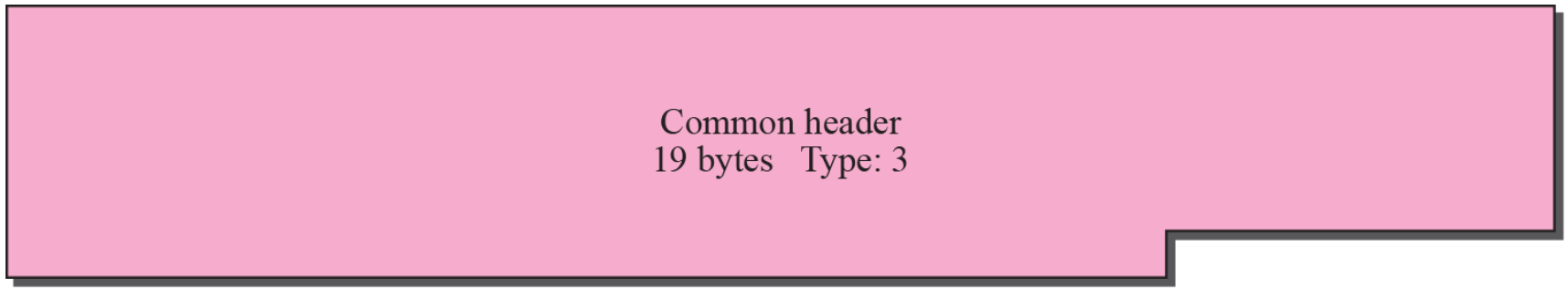
- **Type:** what type of packet this is and what type of information it contained in the packet.
- **Length:** This shows the total length of the BPG Message including the BGP headers. The value will always be between 19 and the maximum message allowed in any BGP message is 4096 byes.
- **Marker** (16-bytes): Authentication. If no authentication information is assigned or contained in this field then the marker is set to all ones or if the message is an open message.

Open message



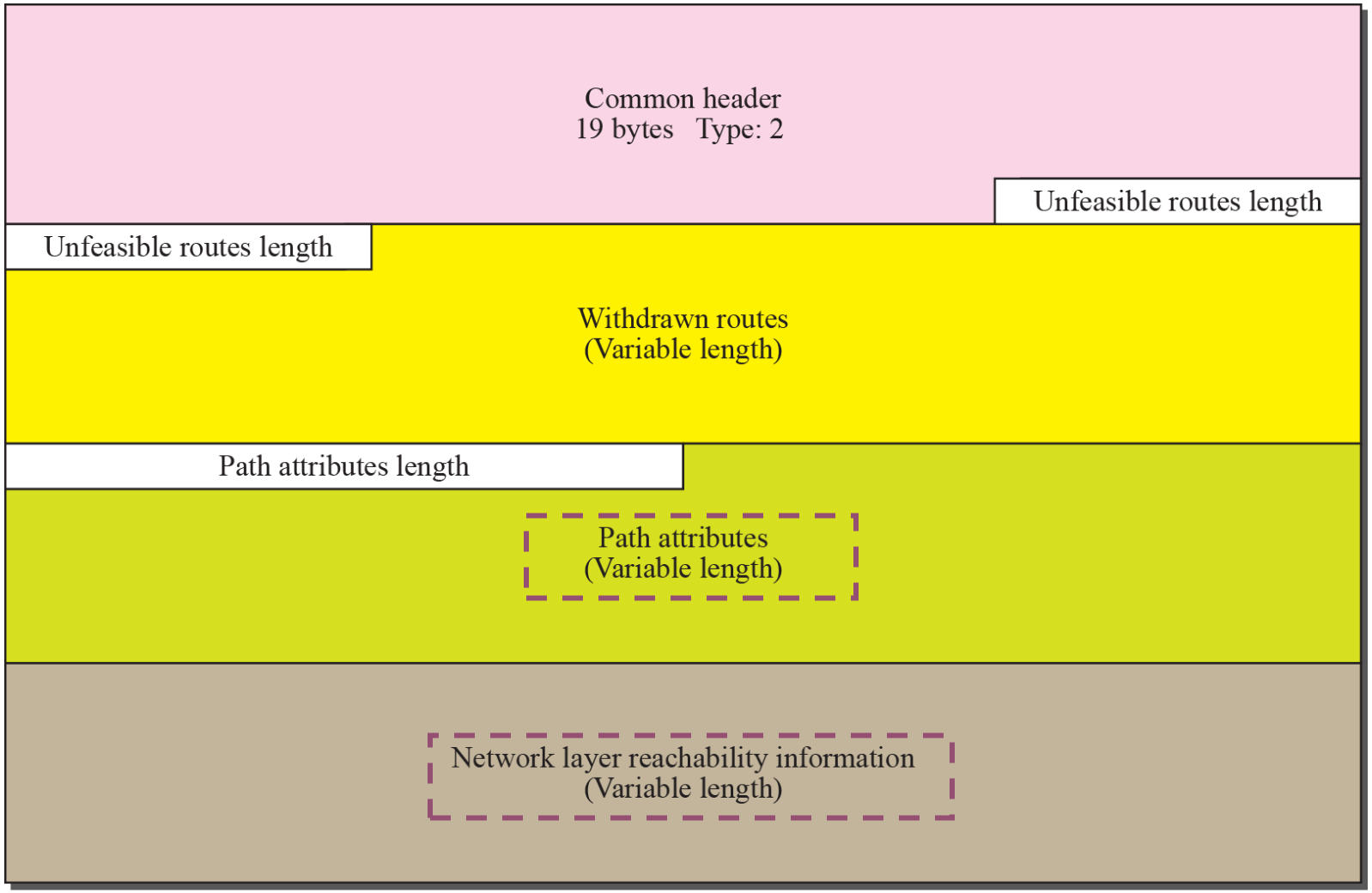
- Used to initiate a BGP session between two routers.
- **Version:** BGP version
- **AS:** Sender AS number
- **Hold time:** how long BGP session stays up if doesn't receive any message. (CISCO default is 180s)
- **BGP Identifier:** The BGP Router-ID (RID) is a 32-bit unique number that identifies the BGP router in the advertised prefixes as the BGP Identifier
- Option: BGP capabilities.

Keepalive message



- I am still alive. Let’s keep this BGP Session open.
- Send generally every one-third of the “hold time”. Cisco Default every 60 seconds.

Update message



- This message sends the Network Layer reachability information (NLRI).
  - Ex: 110.13.3.0/24
- Path attribute: Information about each route.
- Withdrawn routes:

(example: 11.12.1.128/24 is no longer reachable via this path.)

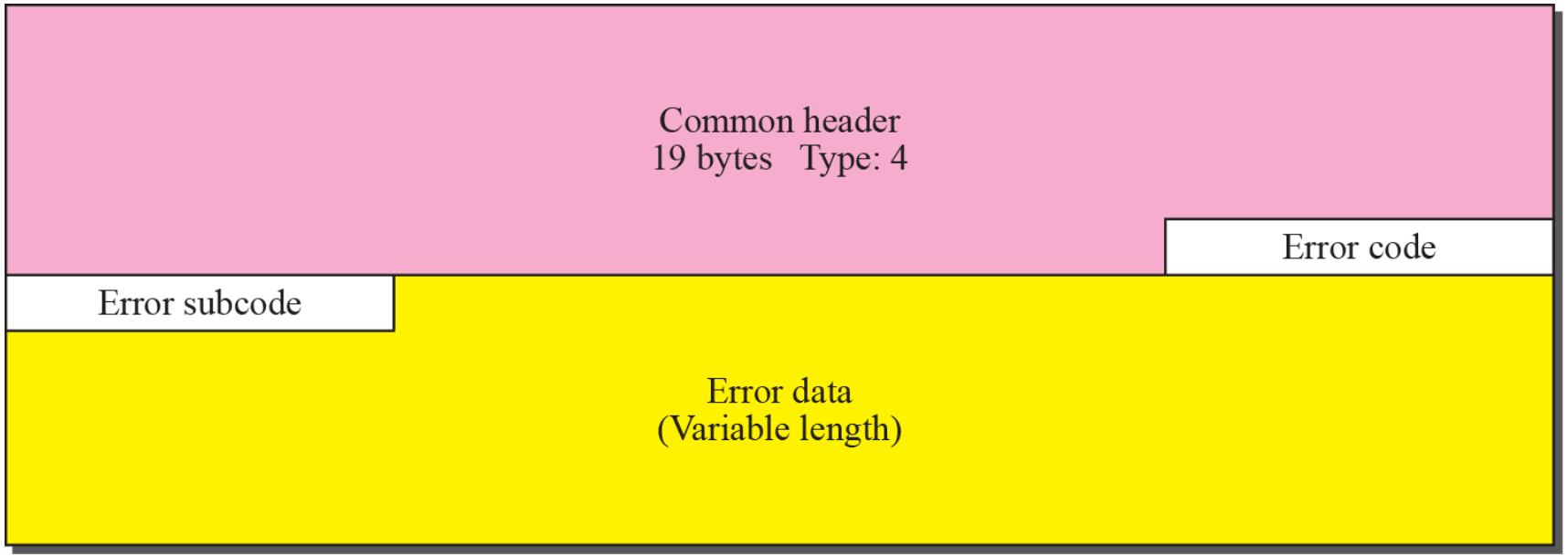
BGP Attribute Types

- Attributes are included in BGP updates.
- **Each NLRI (advertised route) has attributes.**
- The attributes are of two types: Well-known (understood by all BGP speaking routers) and optional (may not be understood by all BGP routers) attributes.
- Well known : Mandatory (include with every update) and optional (doesn't need to include in every update)

Path Attributes (Well known)

- ORIGIN
  - The source of the routing information (RIP, OSPF, etc)
  - IGP: Route originally learned from an IGP
  - EGP: Route originally learned from EGP, etc
- AS_PATH
  - The list of ASs through which the destination can be reached
  - List of AS's that the route has passed through.
    - Example: AS 300, AS 500, AS 700
- NEXT-HOP
  - The next router to which the data packet should be sent

Notification message



- Terminates or rejects a BGP connection.
- Can include error message. For ex: BGP version not supported.

Table 11.6 *Error Codes*

Error Code	*Error Code Description*	*Error Subcode Description*
1	Message header error	Three different subcodes are defined for this type of error: synchronization problem (1), bad message length (2), and bad message type (3).
2	Open message error	Six different subcodes are defined for this type of error: unsupported version number (1), bad peer AS (2), bad BGP identifier (3), unsupported optional parameter (4), authentication failure (5), and unacceptable hold time (6).
3	Update message error	Eleven different subcodes are defined for this type of error: malformed attribute list (1), unrecognized well-known attribute (2), missing well-known attribute (3), attribute flag error (4), attribute length error (5), invalid origin attribute (6), AS routing loop (7), invalid next hop attribute (8), optional attribute error (9), invalid network field (10), malformed AS_PATH (11).
4	Hold timer expired	No subcode defined.
5	Finite state machine error	This defines the procedural error. No subcode defined.
6	Cease	No subcode defined.

Note

BGP uses the services of TCP on port 179.

Why different Intra-, Inter-AS routing ?

policy:

- inter-AS: admin wants control over how its traffic routed, who routes through its network
- intra-AS: single admin, so policy less of an issue

scale:

- hierarchical routing saves table size, reduced update traffic

performance:

- intra-AS: can focus on performance
- inter-AS: policy dominates over performance

Check out IIT Bhilai AS Interconnections (BGP)

- What is my IP -> Get the public IP address
- Find out AS number (from IP address)
  - 103.147.138.252 - Indian Institute of Technology Bhilai - ASN Lookup
- Check the AS interconnections (BGP)
  - https://stats.apnic.net/vizas/
- BGP route update message
  - https://stat.ripe.net/widget/looking-glass
    - You can search with your public IP or AS number.

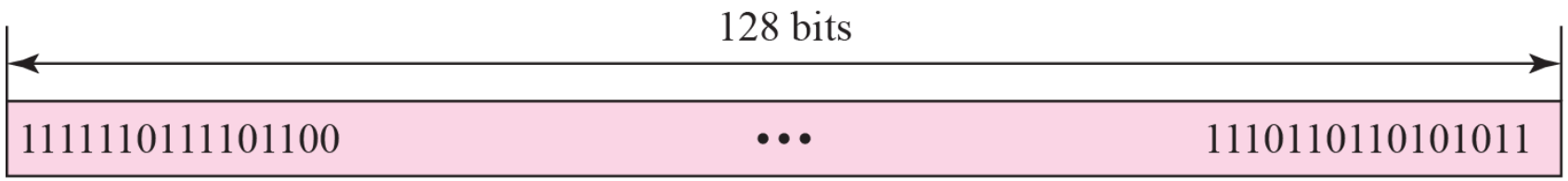
IPv6

IPv6: motivation

- initial motivation: 32-bit IPv4 address space would be completely allocated
- additional motivation:
  - speed processing/forwarding: 40-byte fixed length header
    - No checksum, No fragmentation/reassembly, no options (available as upper-layer, next-header protocol at router)
  - enable different network-layer treatment of “flows”

IPv6 address formats

- IPv6 addresses are 128 bits long. IPv4 addresses are only 32 bits long
- IPv6 addresses are written as eight hexadecimal groups. Each hexadecimal group, separated by a colon (:), consists of a 16-bit hexadecimal value. The following is an example of the IPv6 format:
  - xxxx:xxxx:xxxx:xxxx:xxxx:xxxx:xxxx:xxxx
- A group of xxxx represents the 16-bit hexadecimal value. Each individual x represents a 4-bit hexadecimal value. The following is an example of a possible IPv6 address:
  - 2001:0340:0000:0000:0000:F673:0029:0564


128 bits
1111110111101100
•••
1110110110101011

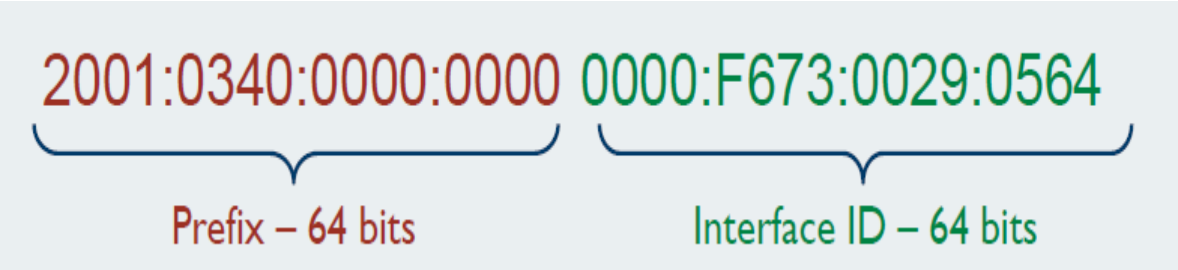
Colon hexadecimal notation

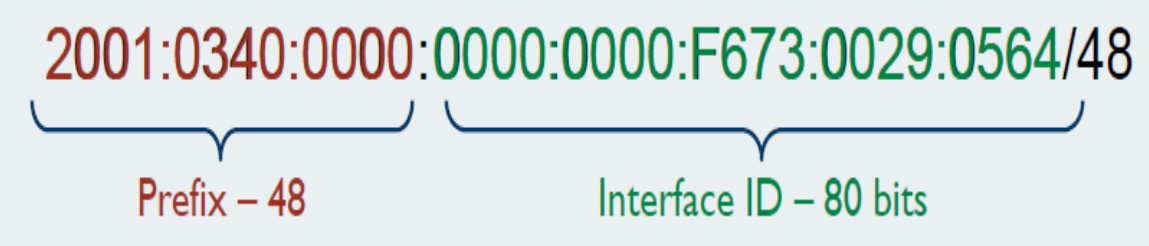
FDEC : BA98 : 7654 : 3210 : ADBF : BBFF : 2922 : FFFF

Prefix and interface ID

- IP addresses combine, in a single address, a network identifier (called the prefix) and a device identifier (the interface ID).
- The point at which to split the address into these two portions is given by the prefix length.
- The prefix length is written as /xx at the end of the address; e.g.

  - 2001:0340:0000:0000:0000:F673:0029:0564/64





Address optimization

- To make IPv6 addresses easier to write, some zeros can be removed. The leading zeros in a 4-digit block can be removed.
- Also, contiguous sets of 4 zeros, and their separating colons can be completely removed.

2001:0340:0000:0000:0000:F673:0029:0564

Address optimization Cont

- To avoid ambiguity, it is only possible to have one place in the address where a continuous set of 0s is replaced by ::

So 2001:0340:0000:0000:F673:0000:0000:0564

can be written as

2001:340::F673:0000:0000:564

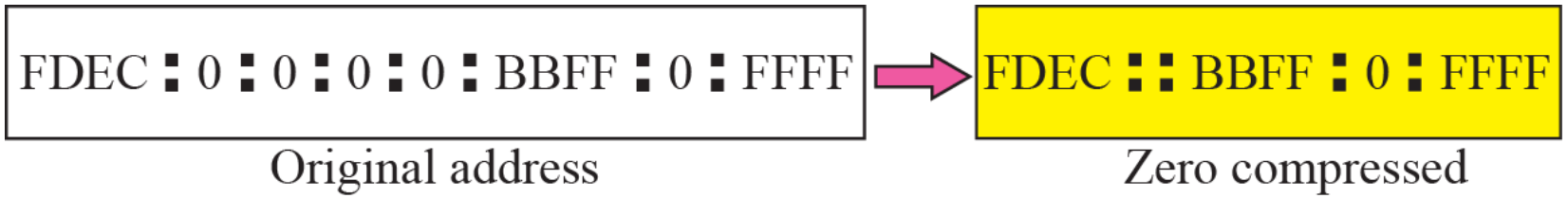
or

2001:340:0000:0000:F673::564

but NOT

2001:340::F673::564

Zero compression

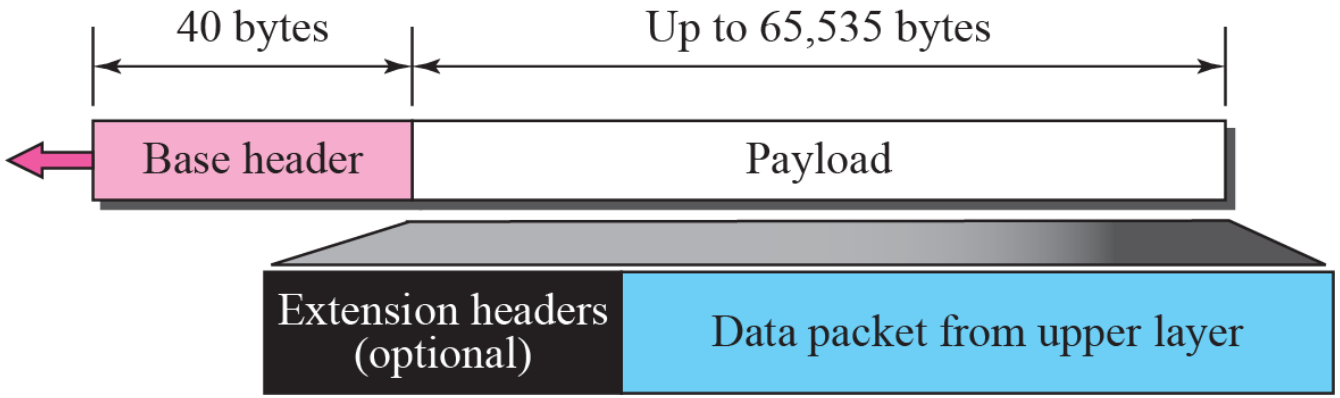


FDEC : : BBFF : 0 : FFFF/**60**

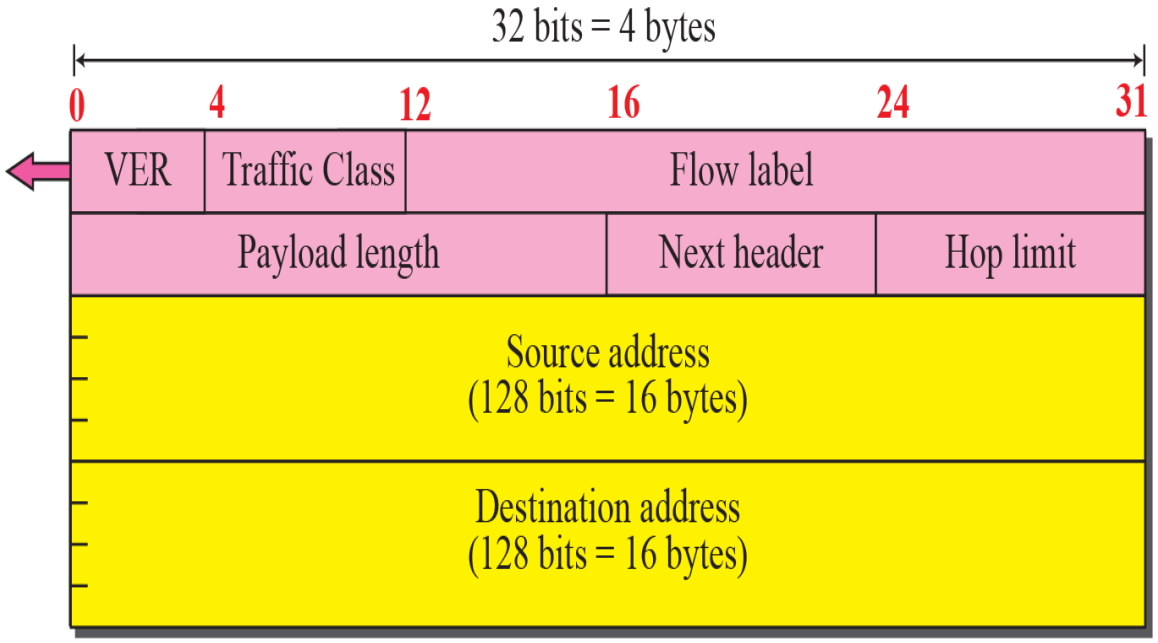
IPv6 Header Format

- Forty bytes long (IPv4 Header is 20 bytes)
- Each IPv6 address is four times the length of an IPv4 address
- The IPv6 header no longer contains the header length, identification, flags, fragment offset and header checksum fields
- Some of these options have been placed in extension headers
- The ‘Time To Live’ field has been replaced with a hop limit
- IPv4 Type of Service field is now replaced with a Traffic Class field

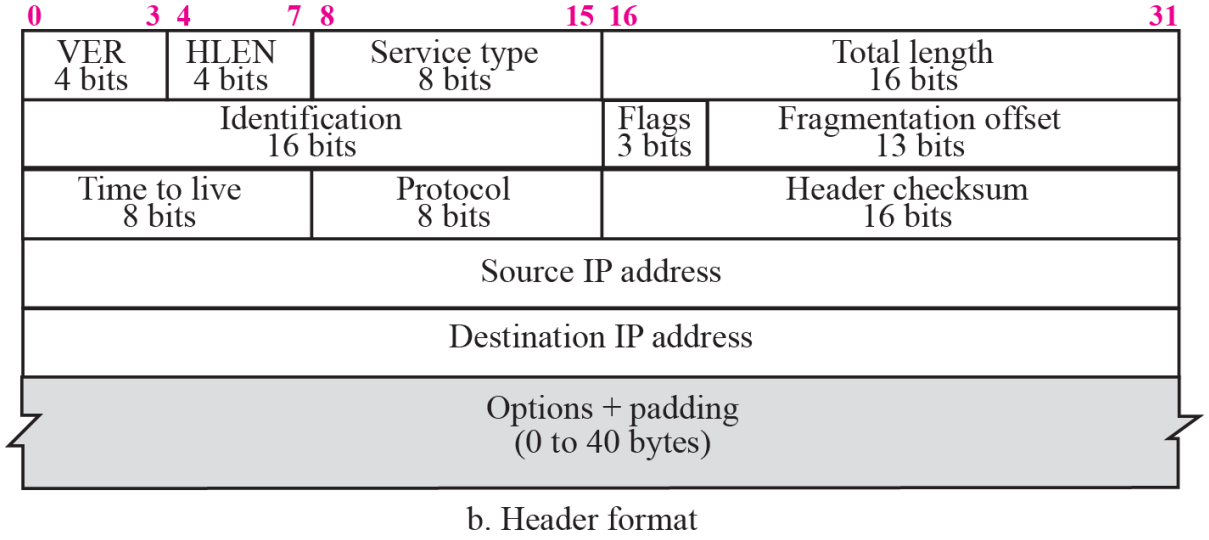
IPv6 datagram



Format of the base header



IPv6



b. Header format

IPv4

Table 27.1 *Next Header Codes*

Code	*Next Header*	*Code*	*Next Header*
0	Hop-by-hop option	44	Fragmentation
2	ICMP	50	Encrypted security payload
6	TCP	51	Authentication
17	UDP	59	Null (No next header)
43	Source routing	60	Destination option

Extension header format