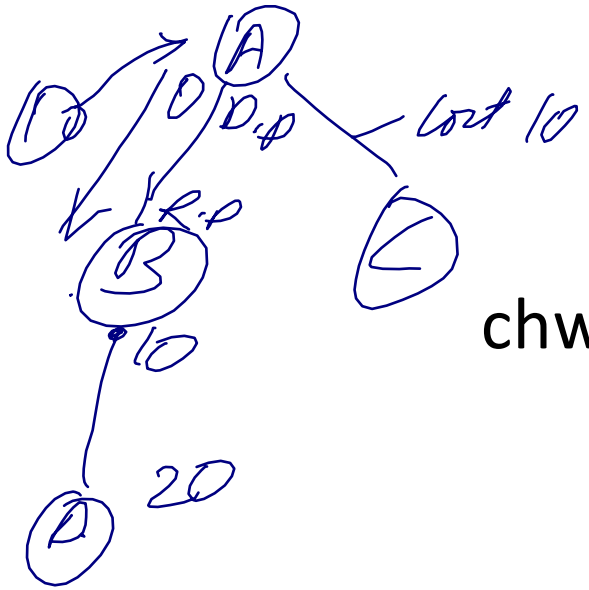
CS 540
Computer Networks II



Sandy Wang

chwang_98@yahoo.com

3. IPV4

Topics

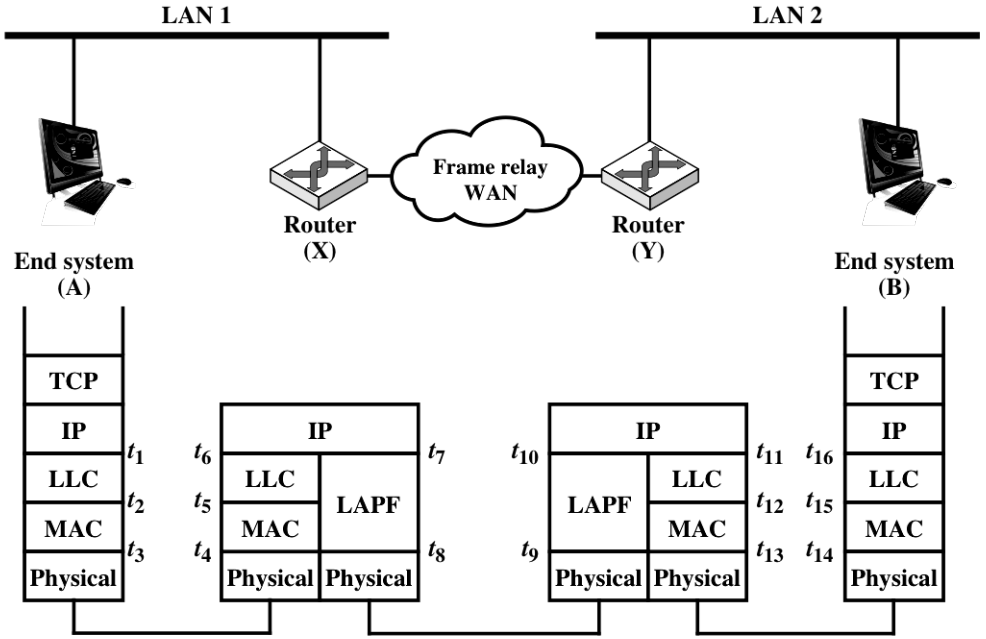
1. Overview
2. LAN Switching
3. IPv4
4. IPv6
5. Routing Protocols -- RIP, RIPng, OSPF
6. Routing Protocols -- ISIS, BGP
7. MPLS
8. Midterm Exam
9. Transport Layer -- TCP/UDP
10. Congestion Control & Quality of Service (QoS)
11. Access Control List (ACL)
12. Application Layer Protocols
13. Application Layer Protocols continue
14. Others – Multicast, SDN
15. Final Exam

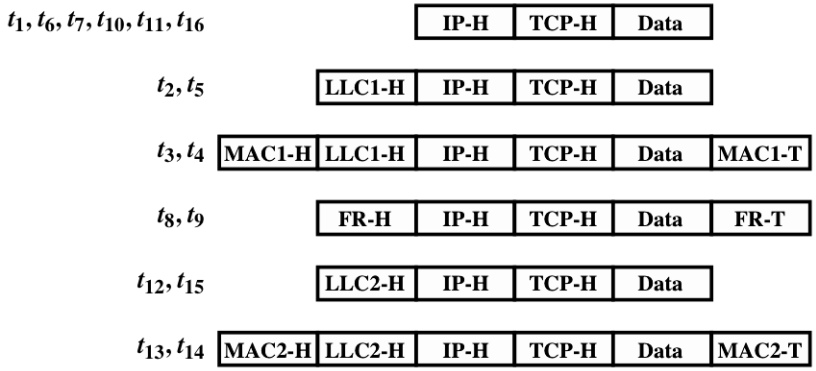
Reference Books

- **Cisco CCNA Routing and Switching ICND2 200-101 Official Cert Guide, Academic Edition** by Wendel Odom -- July 10, 2013**.** ISBN-13: 978-1587144882
- **The TCP/IP Guide: A Comprehensive, Illustrated Internet Protocols Reference** by Charles M. Kozierok – October 1, 2005. ISBN-13: 978-1593270476
- **Data and Computer Communications (10th Edition) (William Stallings Books on Computer and Data Communications)** by Williams Stallings – September 23, 2013. ISBN-13: 978-0133506488

Topics:

- IPv4 Forwarding
  - Life time of a IP datagram
- ARP/RARP/GARP
- ICMP
- First Hop Redundancy
- Policy Routing and VRF
- Tunnels





TCP-H = TCP header
IP-H = IP header
LLCi-H = LLC header
MACi-H = MAC header
MACi-T = MAC trailer
FR-H = Frame relay header
FR-T = Frame relay trailer

Figure 14.2 Example of Internet Protocol Operation

Connectionless Internetworking

- Connectionless internet facility is flexible
- IP provides a connectionless service between end systems
  - Advantages:
    - Is flexible
    - Can be made robust
    - Does not impose unnecessary overhead

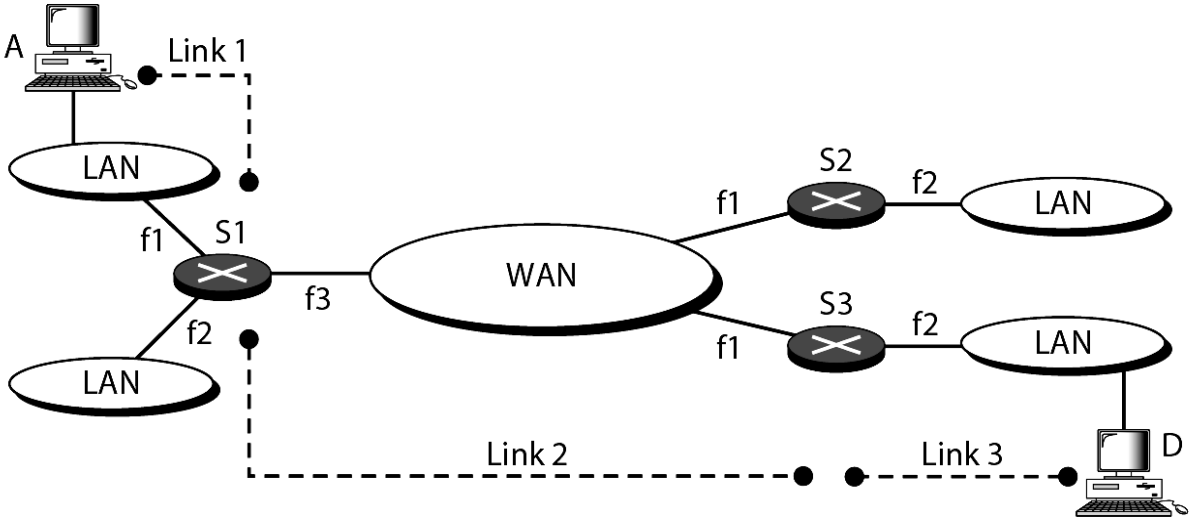
IP Design Issues

- Routing
- Datagram lifetime
- Fragmentation and reassembly
- Error control
- Flow control

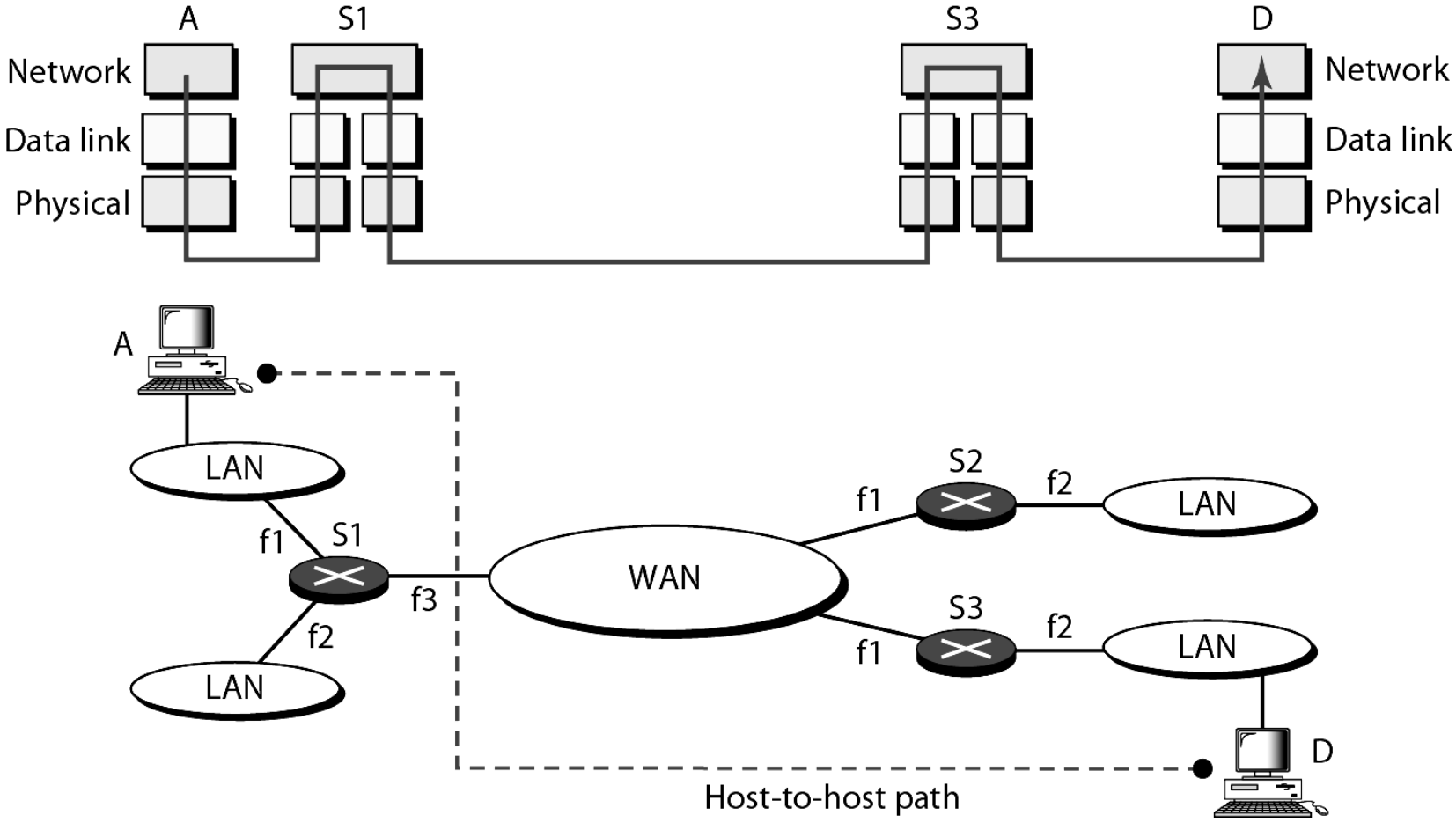
Network Layer

- Need
  - A frame has no routing info.
  - DL layer has no routing info.
  - For a router with 3+ NIC's,
    - how to deliver a packet through multiple links.
    - How to find a next hop router
- Responsibility
  - Host-to-host delivery
  - For routing packets through the router and switches.

Links between two hosts



Network layer in an internetwork

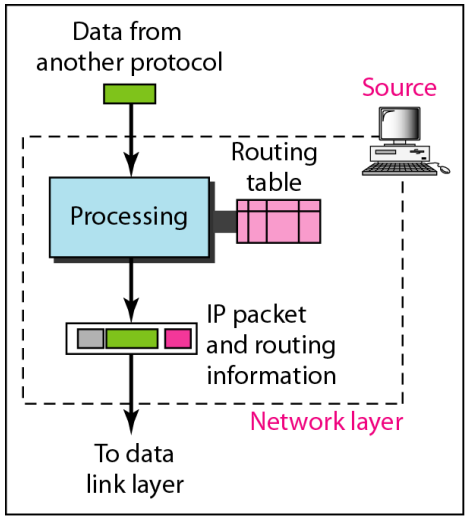


Network Layer

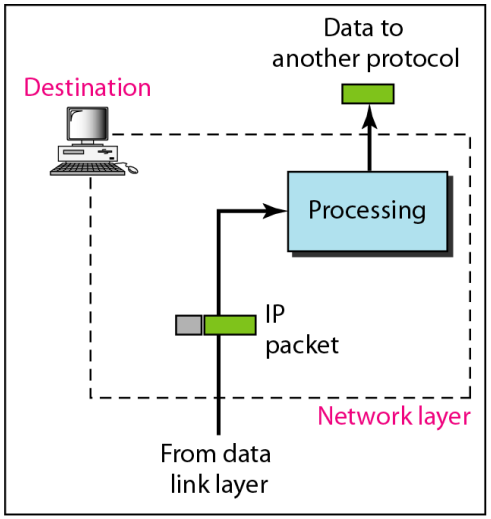
- Source
  - Creating a packet from the upper layer.
    - The header contains source and destination IP addresses.
  - Checking the routing table to find the routing info (eg. Outgoing interface, or machine address of the next hop)
  - If the packet is larger than MTU, fragment it.

- Router
  - Routing the packet by consulting the routing table for each incoming packet and find the interface that the packet must be sent to.
- Destination
  - Address verification.
  - For fragmented frames, wait for all fragmentations then reassemble them before delivering the packet to the upper layer.

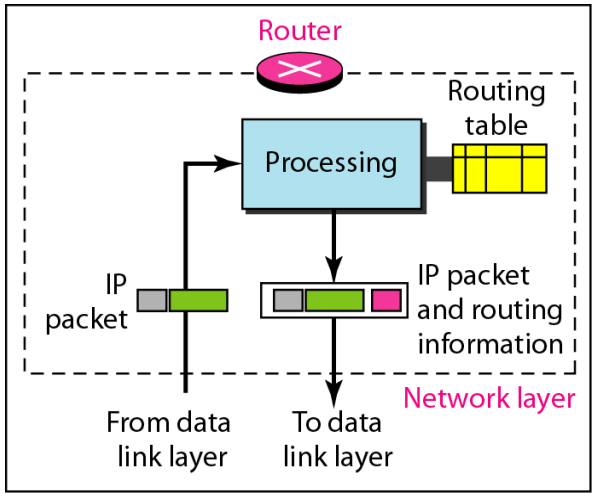
Network layer at the source, router, and destination



a. Network layer at source



b. Network layer at destination

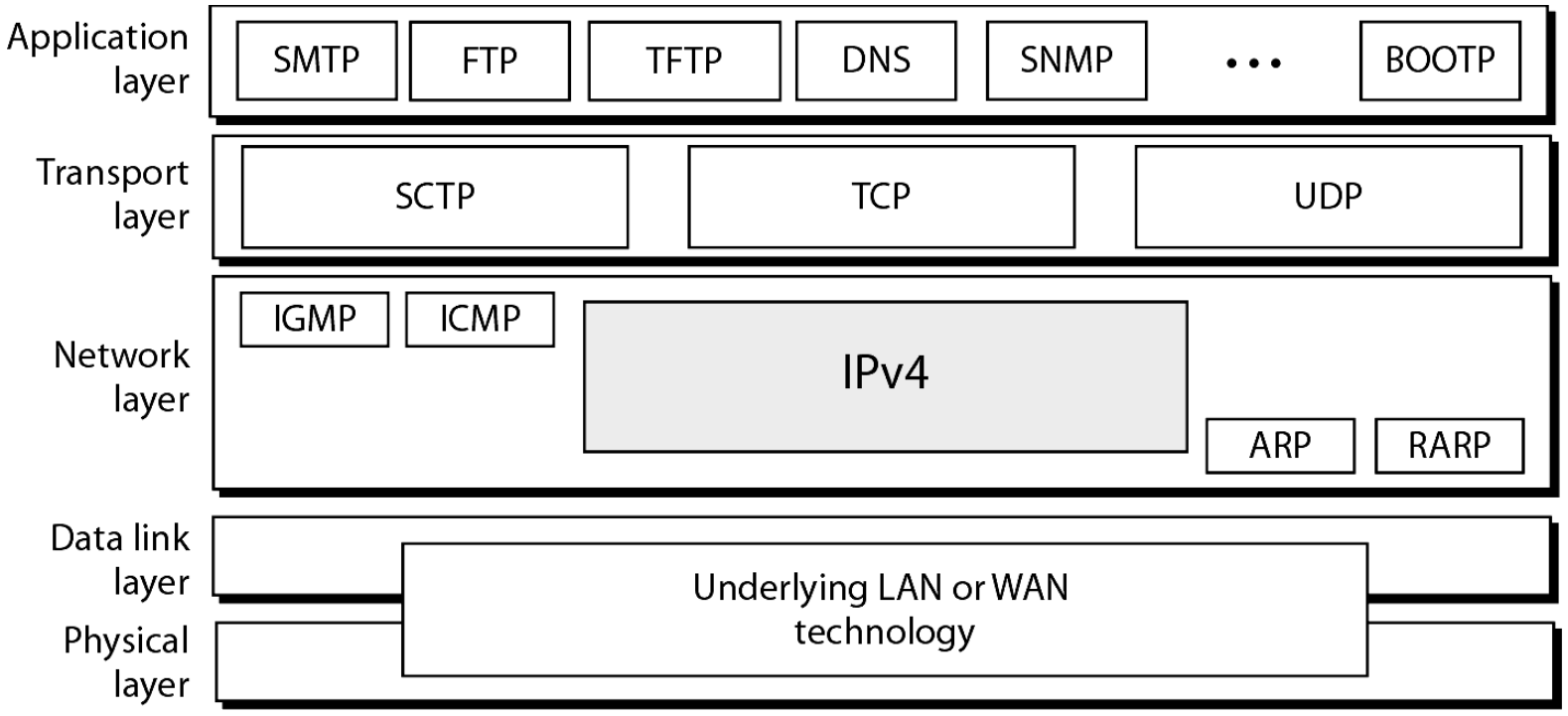


c. Network layer at a router

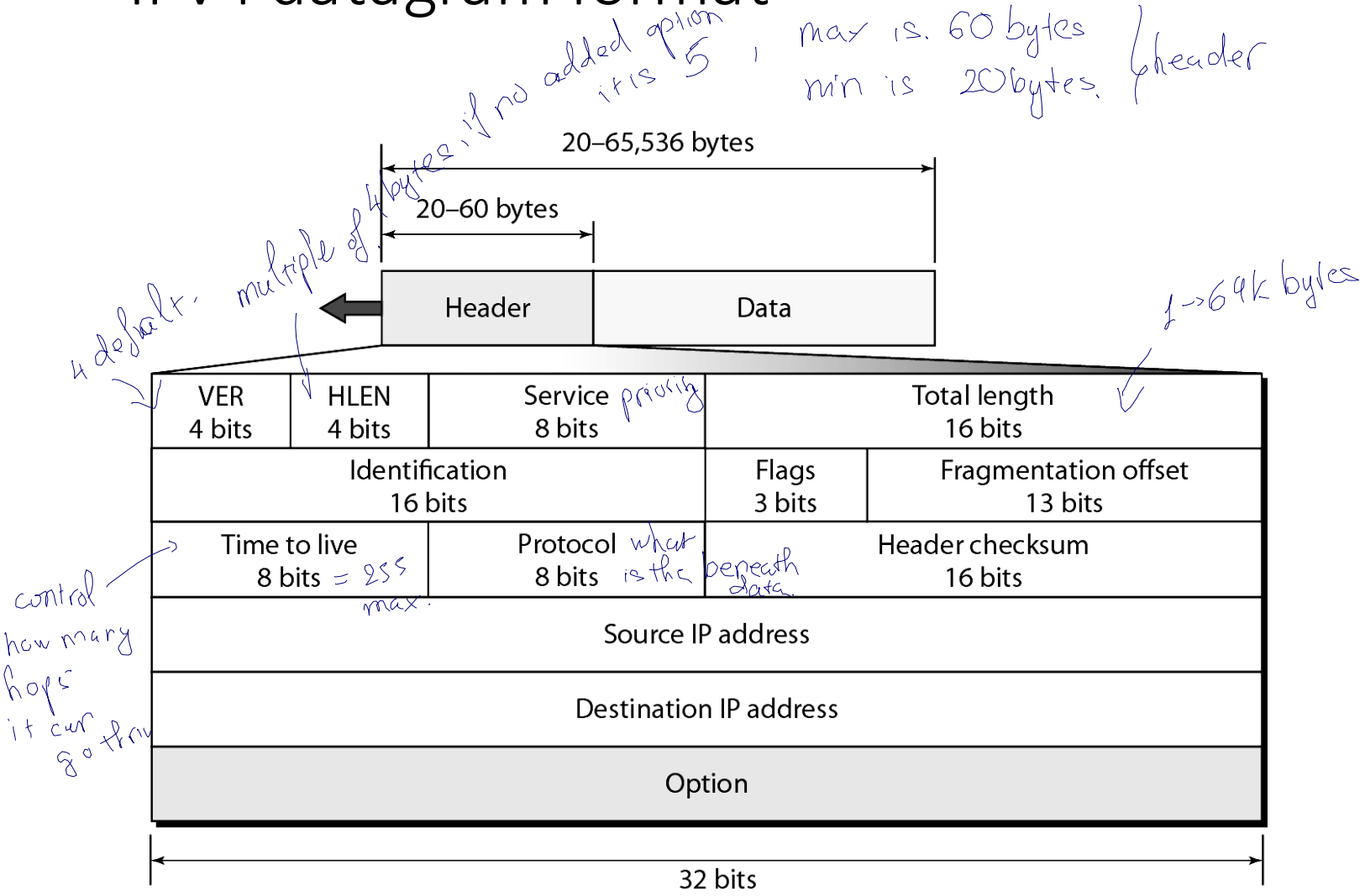
Internet as an Layer3

- Switching at the network layer in the Internet uses the datagram approach to packet switching.
  - Use of globally unique address for each packet
- Communication at the network layer in the Internet is connectionless.
  - Each packet is treated independently by the intermediate routers.
  - Packets in a message may travel through different paths.
- Why?

Position of IPv4 in TCP/IP protocol suite



IPv4 datagram format

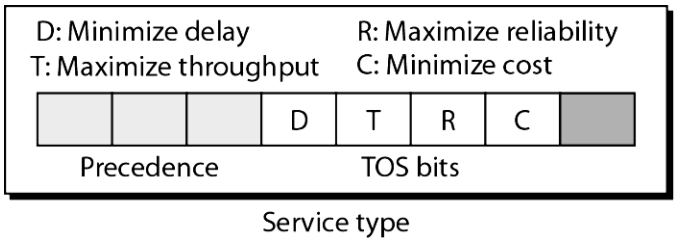


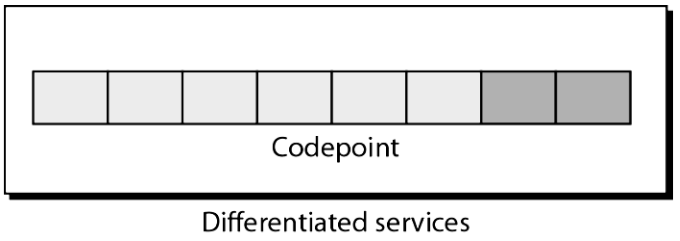
IPv4 Header

- Variable length: 20-60 byte (multiple of 4)
- Contains routing information

IPv4 Format

- Version (4-bit): currently 4.
- Header length (4-bit): the length of the IP header in 4-byte unit.
- Type of Services(TOS):
  - This field was not used earlier because of the lack of standard
  - DiffServ uses this field for differentiate packet types.





- Total length
  - to defines the total length of the datagram including the header in bytes.
  - 16-bit number, the maximum IP size is limited to 2^{16} bytes, or 64 Kbytes.

IPv4 Format

- Identification
  - A source node gives a unique ID to each packet.
  - Identification, Flags, Fragmentation offset fields are used for fragmentation (will be covered later)
- Time to Live (TTL)
  - A packet has a limited lifetime in the network to avoid zombie packets.
  - Hold the maximum number of hops the packet can travel thru the network. Each router decrements it by one. A packet is discarded by a router if TTL is zero.
- Protocol -- To define payload protocol type
  - 1 for ICMP
  - 2 for IGMP
  - 6 for TCP
  - 17 for UDP
  - 89 for OSPF

IPv4 Format

- Header checksum
  - Refer RFC 1071
  - An IP header is slightly modified by each router. At least TTL field.
  - The checksum must be re-calculated by routers
- Source IP address and Destination IP address
- Options
  - Variable length
  - For new protocols
- Padding
  - To make the header a multiple of 32-bit words

Example 20.1

An IPv4 packet has arrived with the first 8 bits as shown:

01000010

The receiver discards the packet. Why?

Solution

There is an error in this packet. The 4 leftmost bits (0100) show the version, which is correct. The next 4 bits (0010) show an invalid header length (2 × 4 = 8). The minimum number of bytes in the header must be 20. The packet has been corrupted in transmission.

Example 20.2

In an IPv4 packet, the value of HLEN is 1000 in binary. How many bytes of options are being carried by this packet?

Solution

The HLEN value is 8, which means the total number of bytes in the header is 8 × 4, or 32 bytes. The first 20 bytes are the base header, the next 12 bytes are the options.

Example 20.3

In an IPv4 packet, the value of HLEN is 5, and the value of the total length field is 0x0028. How many bytes of data are being carried by this packet?

Solution

The HLEN value is 5, which means the total number of bytes in the header is 5 × 4, or 20 bytes (no options). The total length is 40 bytes, which means the packet is carrying 20 bytes of data (40 – 20).

Example 20.4

An IPv4 packet has arrived with the first few hexadecimal digits as shown.

0x45000028000100000102 . . .

How many hops can this packet travel before being dropped? The data belong to what upper-layer protocol?

Solution

To find the time-to-live field, we skip 8 bytes. The time-to-live field is the ninth byte, which is 01. This means the packet can travel only one hop. The protocol field is the next byte (02), which means that the upper-layer protocol is IGMP.

Note

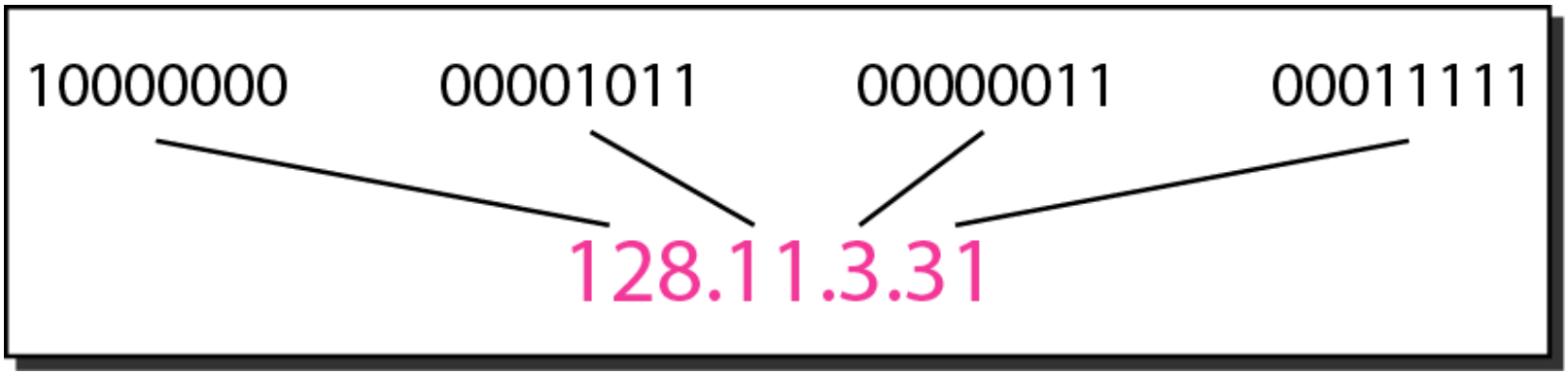
An IPv4 address is 32 bits long.

The IPv4 addresses are unique and universal (all nodes connecting Internet must have IP addresses).

The address space of IPv4 is 2^{32} or 4,294,967,296.

Figure 19.1 *Dotted-decimal notation and binary notation for an IPv4 address*

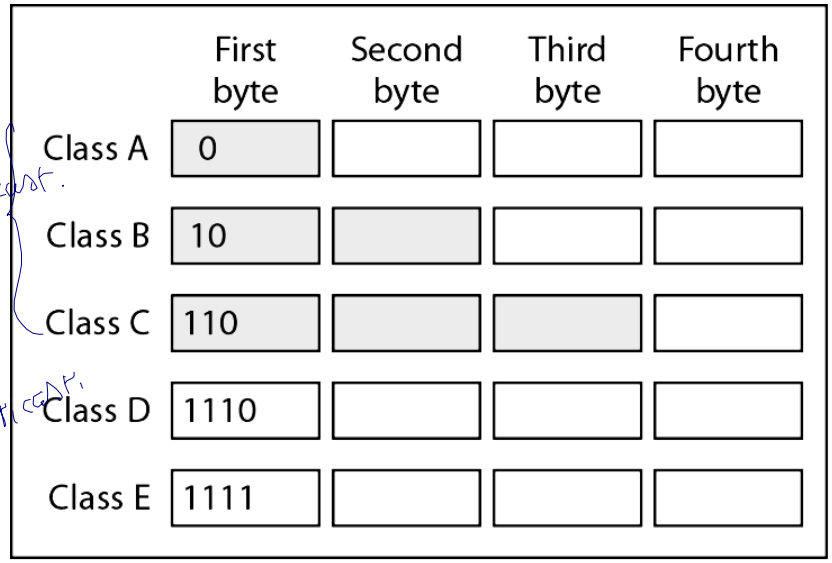
32bit ≅ 4 bytes



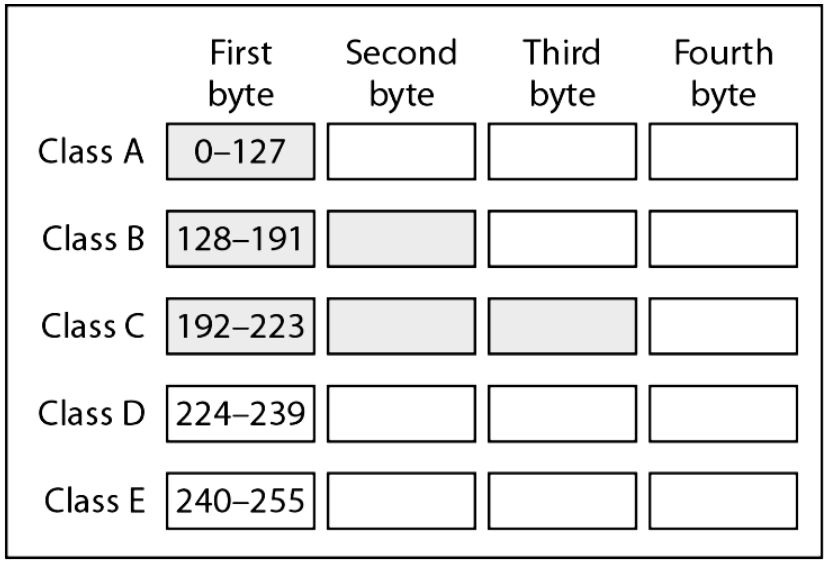
Note

In classful addressing, the address space is divided into five classes: A, B, C, D, and E.

Figure 19.2 *Finding the classes in binary and dotted-decimal notation*



a. Binary notation



b. Dotted-decimal notation

Class D: multicast
Class E: reserved

Table 19.1 *Number of blocks and block size in classful IPv4 addressing*

Class	*Number of Blocks*	*Block Size*	*Application*
A	128	16,777,216	Unicast
B	16,384	65,536	Unicast
C	2,097,152	256	Unicast
D	1	268,435,456	Multicast
E	1	268,435,456	Reserved

Note

In classful addressing, a large part of the available addresses were wasted.

Note

Classful addressing, which is almost obsolete, is replaced with classless addressing.

Note

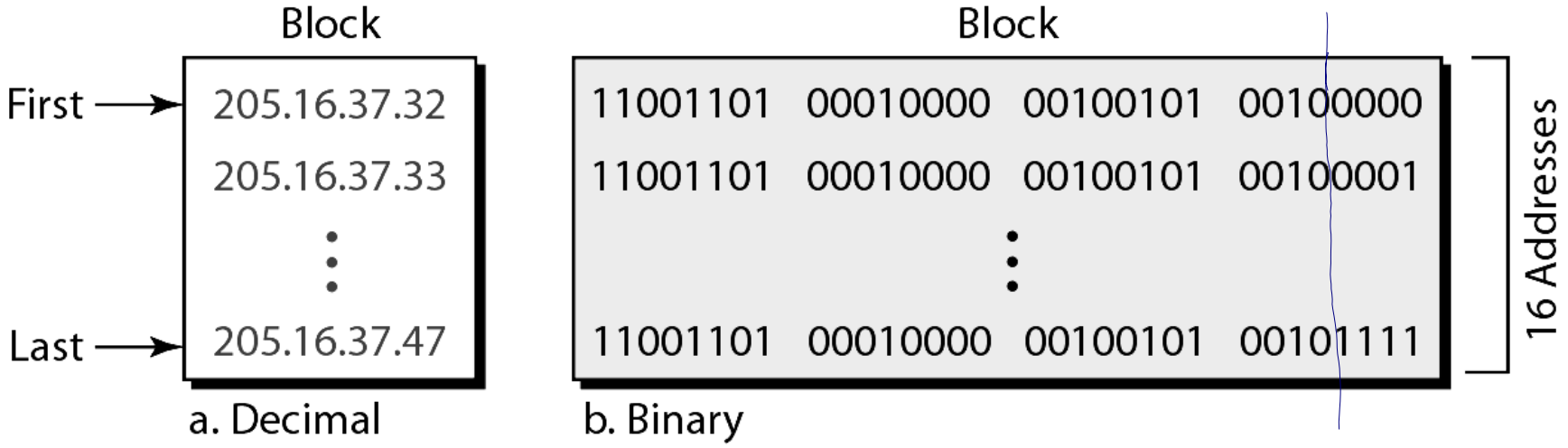
In IPv4 addressing, a block of addresses can be defined as x.y.z.t /*n* in which x.y.z.t defines one of the addresses and the /*n* defines the mask.

Usually, x.y.z.t is the first address in the address block

Note

The first address in the block can be found by setting the rightmost $32 - n$ bits to 0s.

Figure 19.3 *A block of 16 addresses granted to a small organization*



We can see that the restrictions are applied to this block. The addresses are contiguous. The number of addresses is a power of 2 ($16 = 2^4$). This block of IP addresses is represented by:

205.16.37.32/28

Example 19.6

A /28 block of addresses is granted to a small organization. We know that one of the addresses is 205.16.37.39. What is the first address in the block? What is its x.y.z.t/n representation?

128 64 32 16 | 8 4 2 1
1 0 | 0 1 1 1

Solution

The binary representation of the given address is

11001101 00010000 00100101 00100111

If we set 32−28 rightmost bits to 0, we get

11001101 00010000 00100101 0010000 0

or

205.16.37.32

The block representation is 205.16.37.32/28

Note

The last address in the block can be found by setting the rightmost 32 − n bits to 1s.

Example 19.7

Find the last address for the block in Example 19.6.

Solution

The binary representation of the given address is

11001101 00010000 00100101 00100111

If we set 32 − 28 rightmost bits to 1, we get

11001101 00010000 00100101 00101111

or

205.16.37.47

This is actually the block shown in Figure 19.3.

Note

The number of addresses in the block can be found by using the formula 2^{32-n}.

Example 19.9

Another way to find the first address, the last address, and the number of addresses is to represent the mask as a 32-bit binary (or 8-digit hexadecimal) number. This is particularly useful when we are writing a program to find these pieces of information. In Example 19.5 the /28 can be represented as

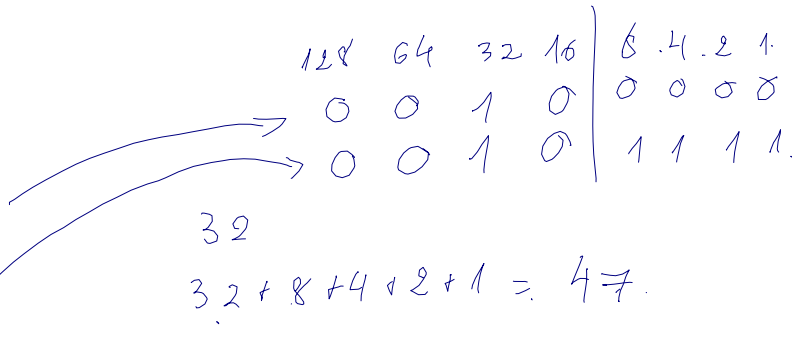
11111111 11111111 11111111 11110000

(twenty-eight 1s and four 0s).

Find

a. The first address

b. The last address



Example 19.9 (continued)

Solution

a. ***The first address can be found by ANDing the given addresses with the mask. ANDing here is done bit by bit. The result of ANDing 2 bits is 1 if both bits are 1s; the result is 0 otherwise.***

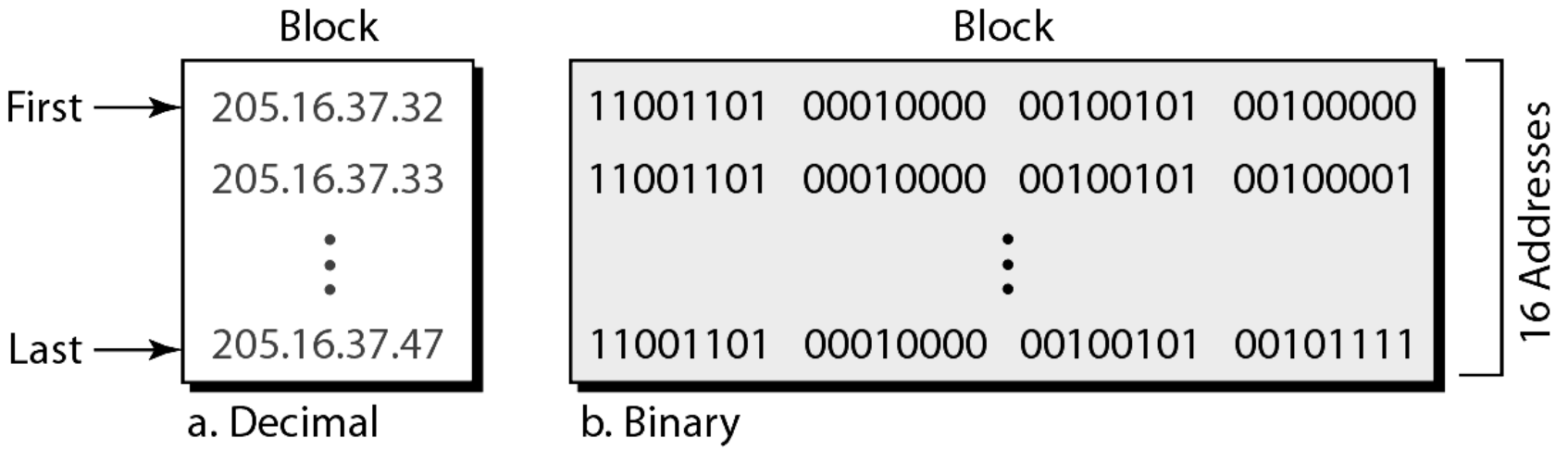
Address:	11001101	00010000	00100101	00100111
Mask:	**11111111**	**11111111**	**11111111**	**11110000**
First address:	11001101	00010000	00100101	00100000

Example 19.9 (continued)

b. ***The last address can be found by ORing the given addresses with the complement of the mask. ORing here is done bit by bit. The result of ORing 2 bits is 0 if both bits are 0s; the result is 1 otherwise. The complement of a number is found by changing each 1 to 0 and each 0 to 1.***

Address:	11001101	00010000	00100101	00100111
Mask complement:	**00000000**	**00000000**	**00000000**	**00001111**
Last address:	11001101	00010000	00100101	00101111

Figure 19.4 *A network configuration for the block 205.16.37.32/28*



Note

The first address in a block is normally not assigned to any device; it is used as the network address that represents the organization to the rest of the world.

Figure 19.6 *hierarchy in IP addressing*

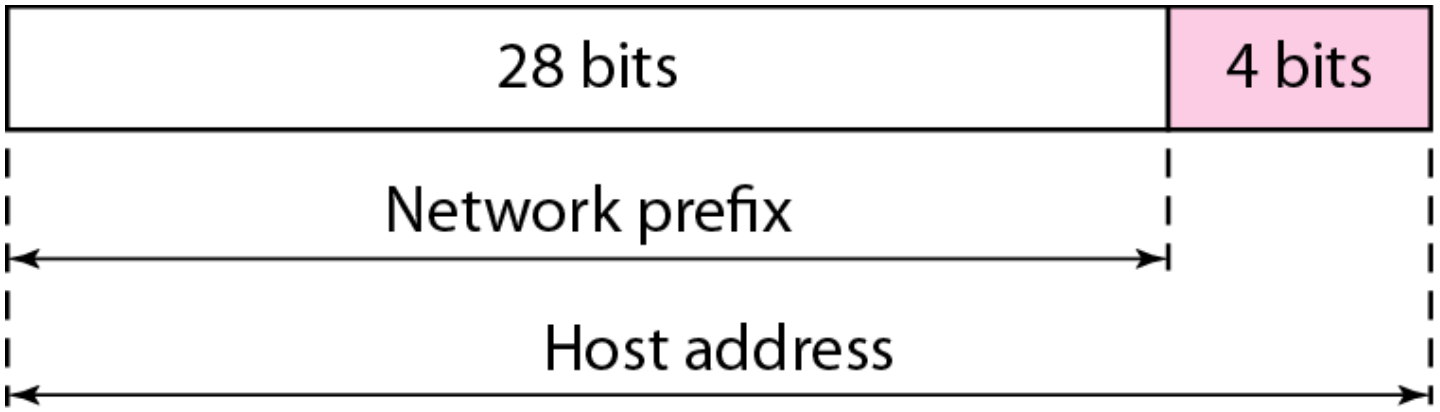


Figure 19.7 *Configuration and addresses in a subnetted network*

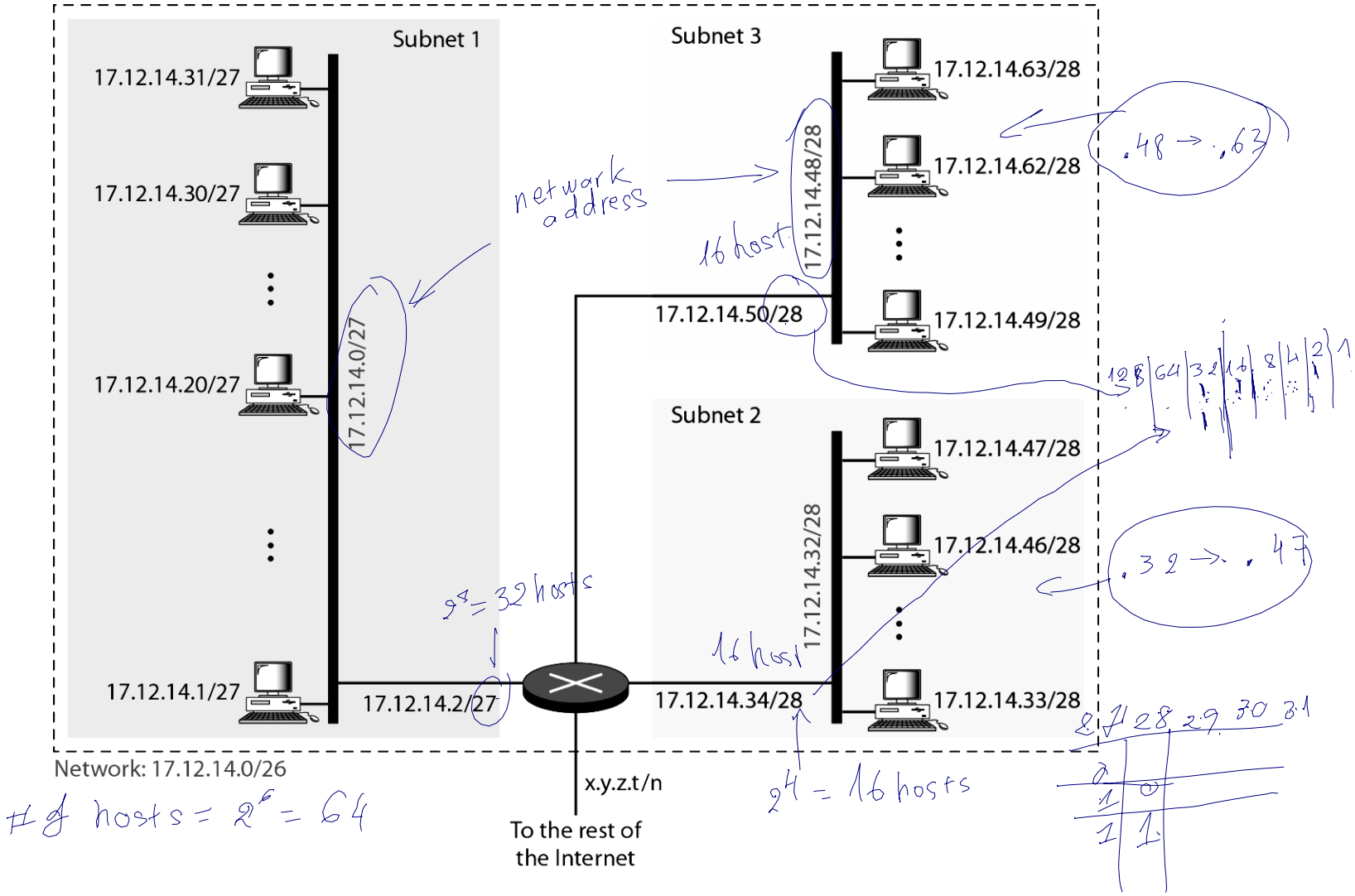
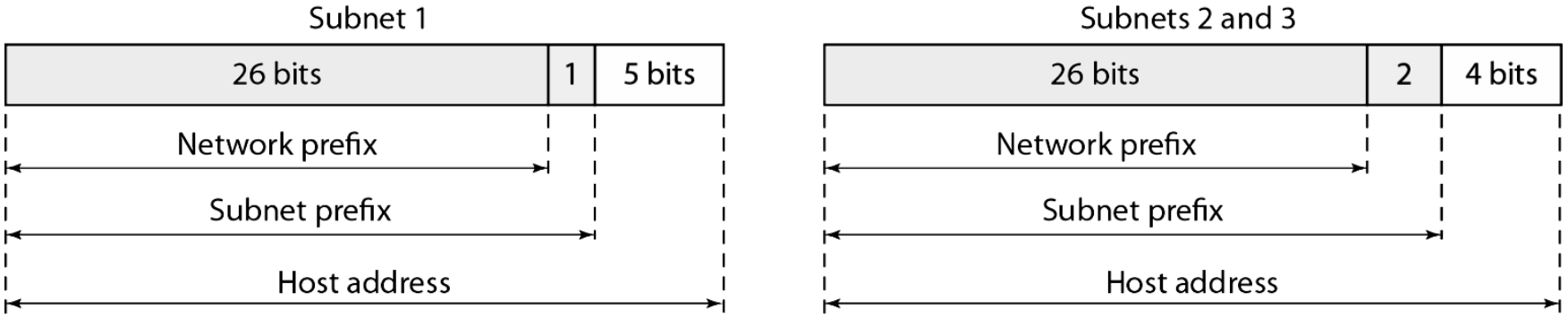


Figure 19.8 *Three-level hierarchy in an IPv4 address*



Example 19.10

An ISP is granted a block of addresses starting with 190.100.0.0/16 (65,536 addresses). The ISP needs to distribute these addresses to three groups of customers as follows:

a. The first group has 64 customers; each needs 256 addresses.

b. The second group has 128 customers; each needs 128 addresses.

c. The third group has 128 customers; each needs 64 addresses.

Assume the blocks of IPs are sequentially assigned. Design the subblocks and find out how many addresses are still available after these allocations.

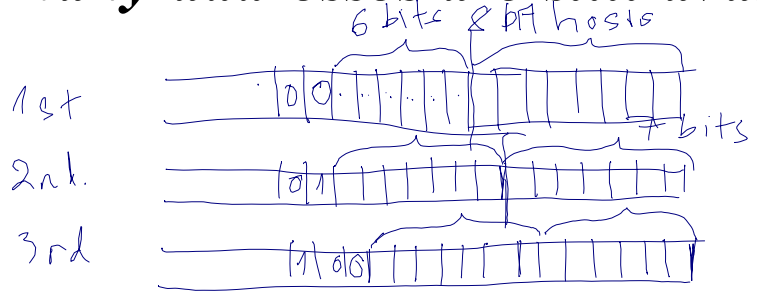
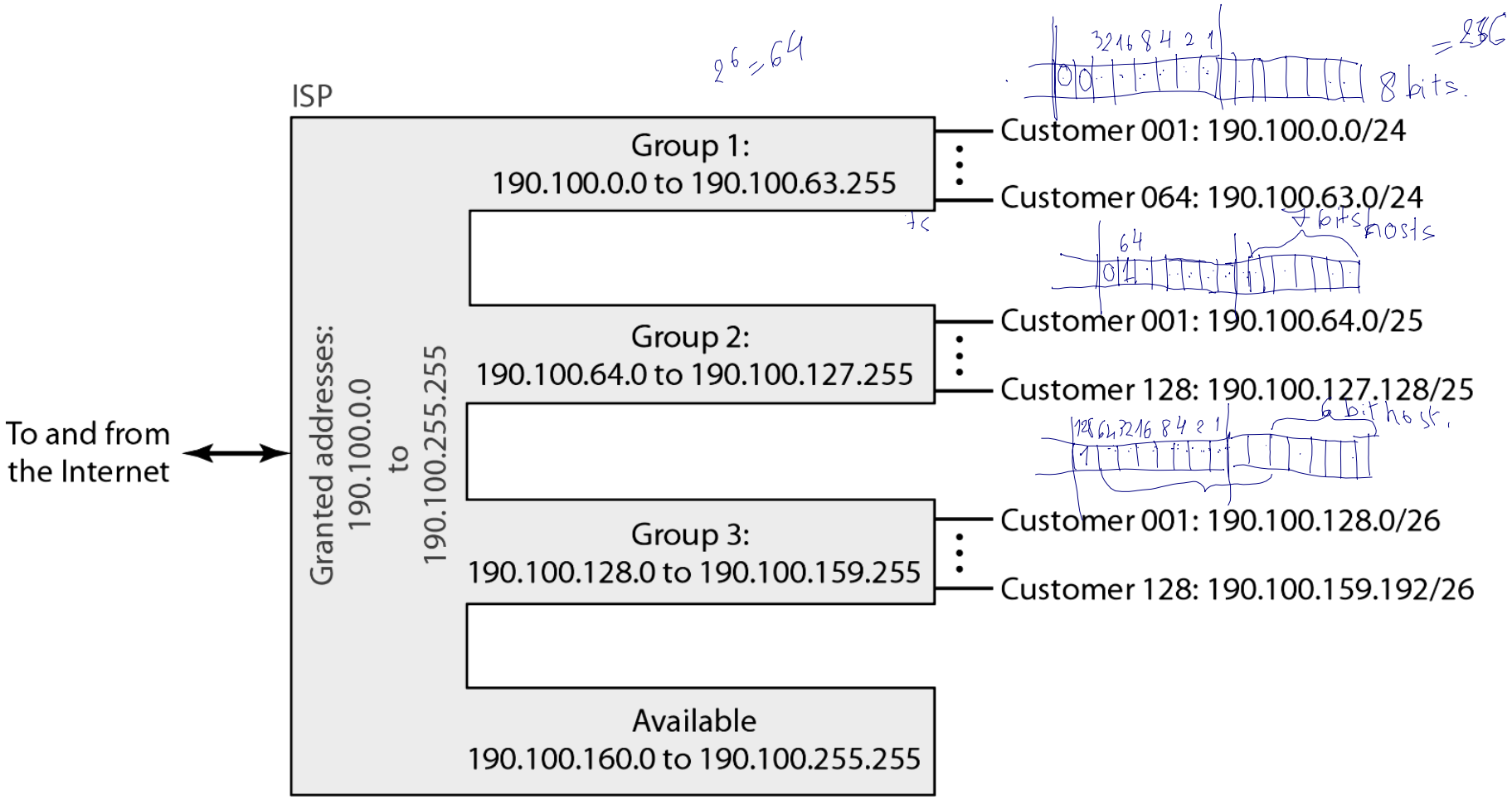
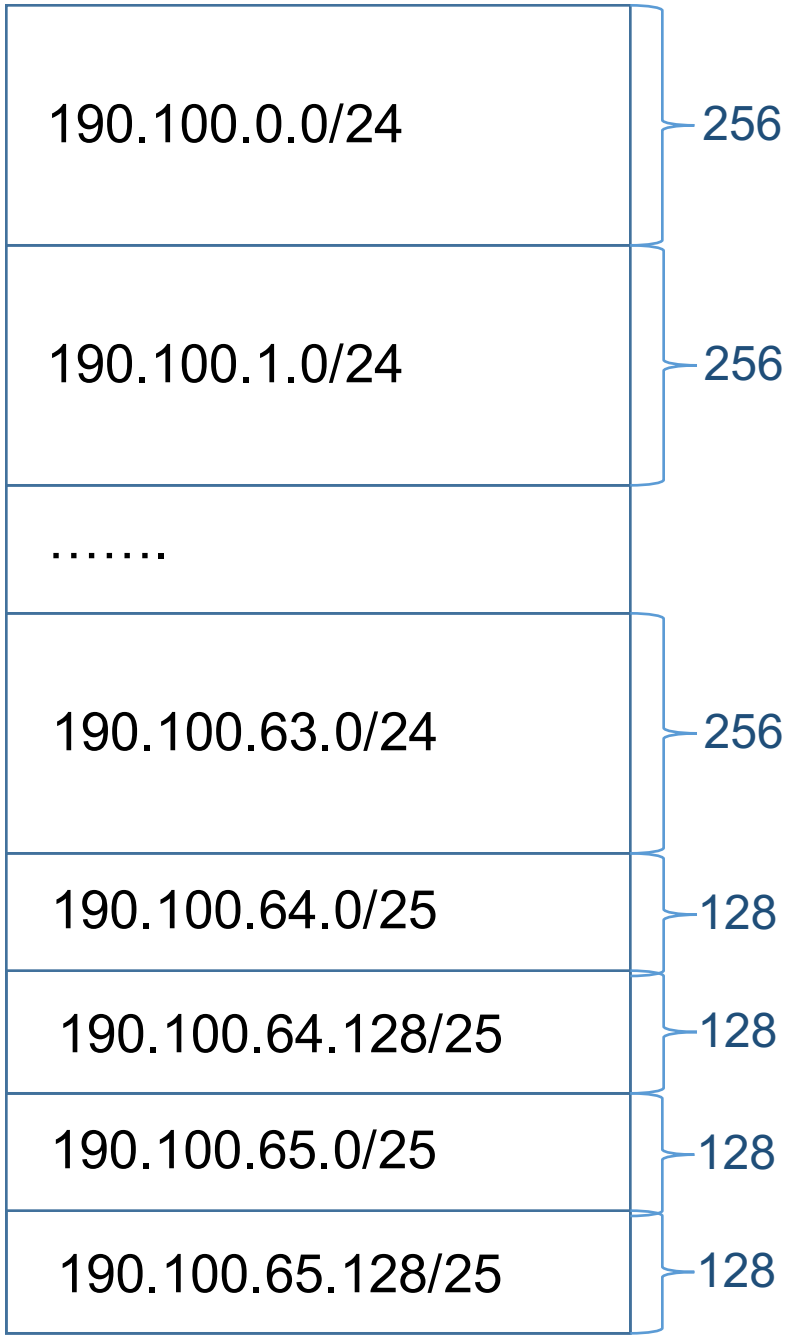
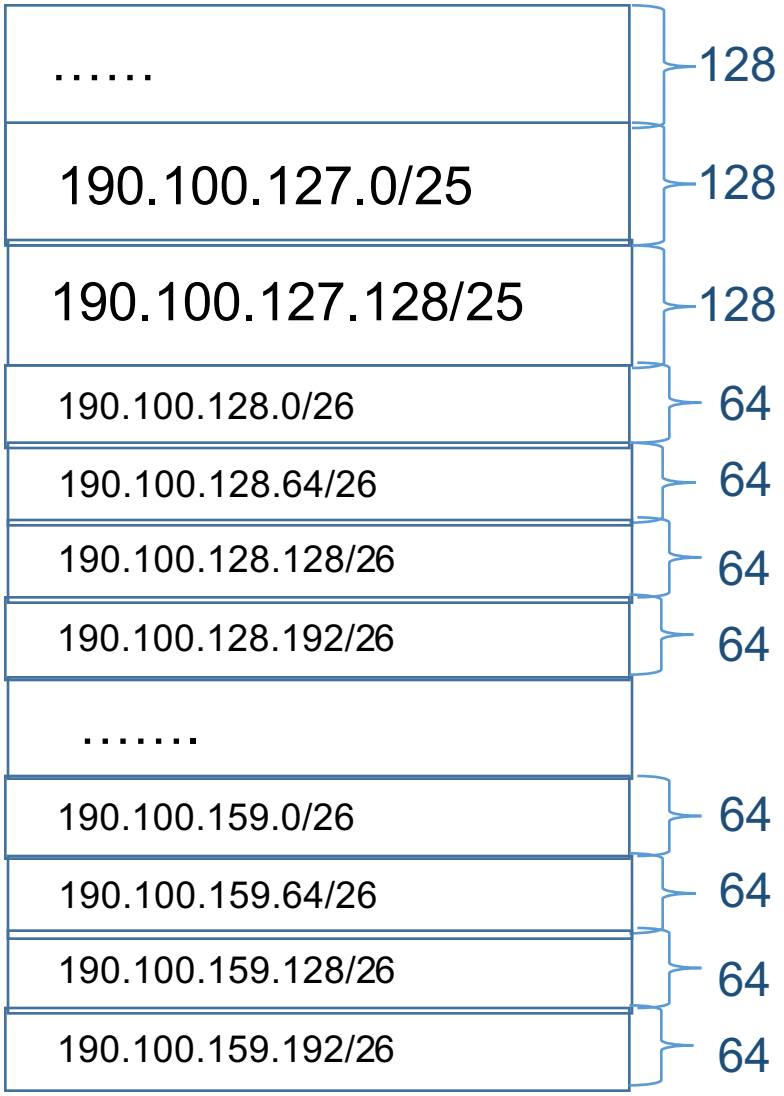


Figure 19.9 *An example of address allocation and distribution by an ISP*







190.100.160.0 to 190.100.255.255
Are available

0.0.0.0 : unspecified. ← can only be used as src (when comp boots up), not dest addr

127.0.0.0 /8 : loopback → not for public.

x x 255.255 : broadcast addr

Table 19.3 *Addresses for private networks*

8.

	Range			Total
A	10.0.0.0 /8	to	10.255.255.255	2^{24}
B	172.16.0.0 /12	to	172.31.255.255	2^{20}
C	192.168.0.0 /16	to	192.168.255.255	2^{16}

Home used wireless router usually uses 192.168.1.0/24 or 192.168.0.0/24 IP block

Figure 20.1 *Links between two hosts*

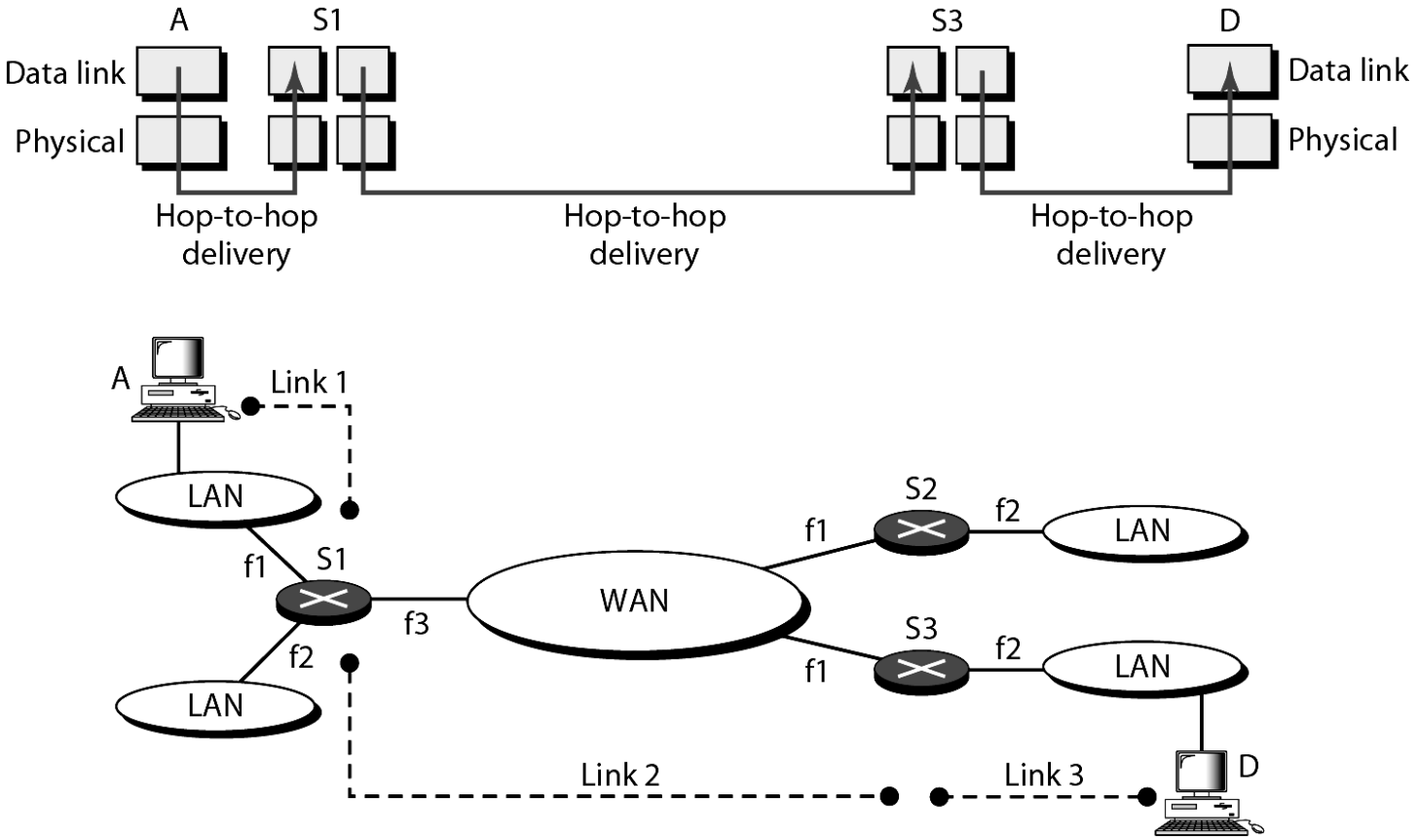


Figure 20.2 *Network layer in an internetwork*

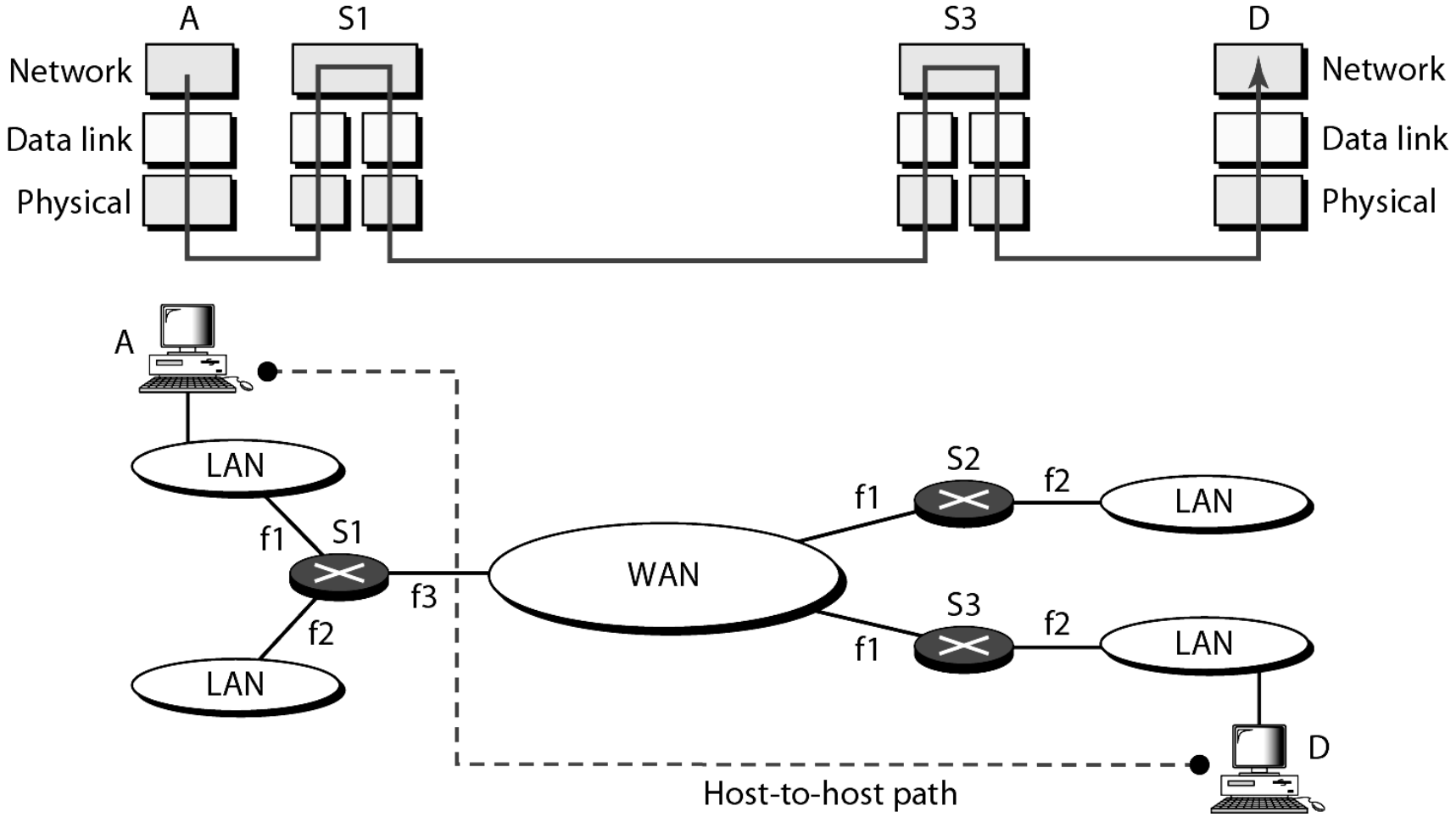
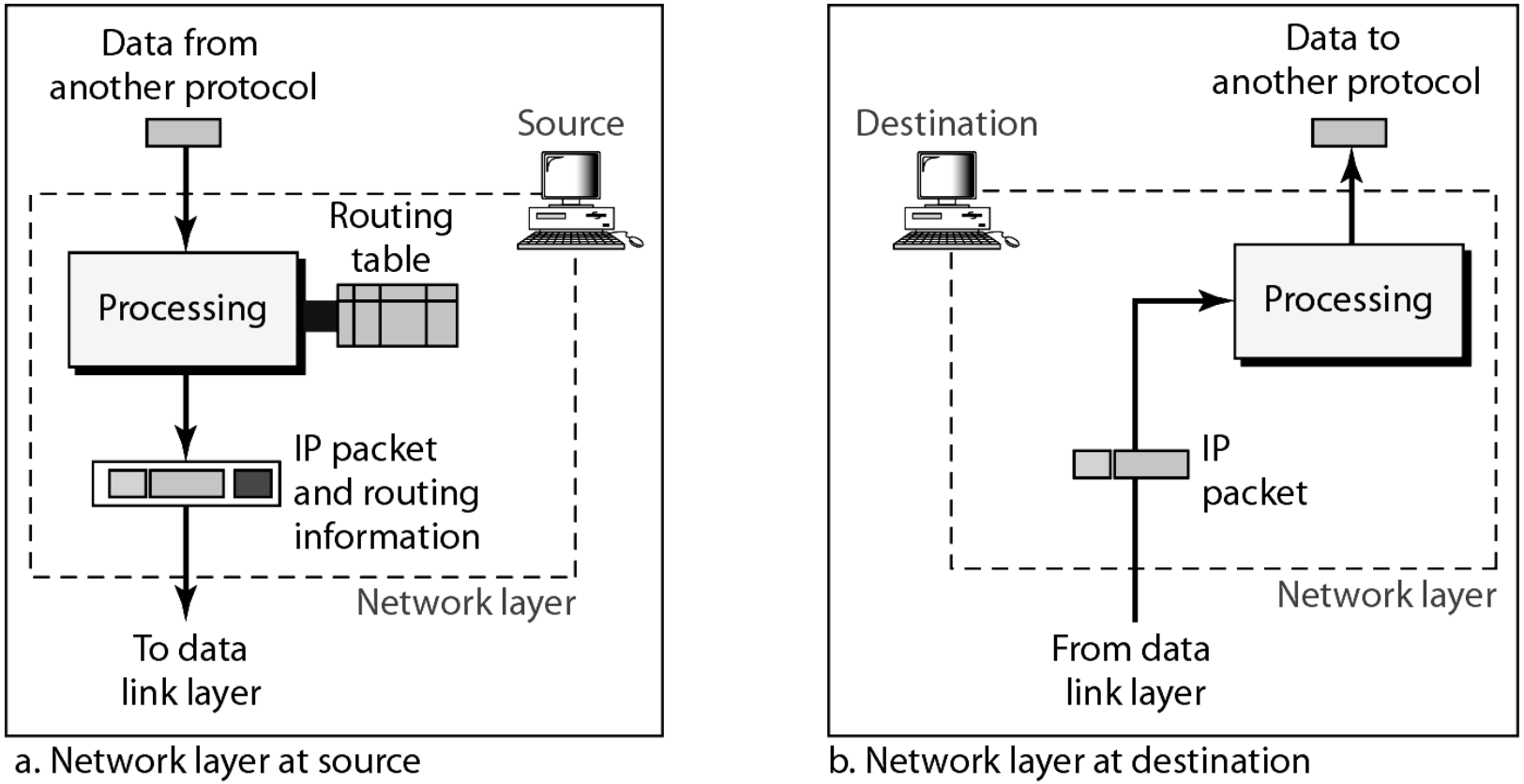


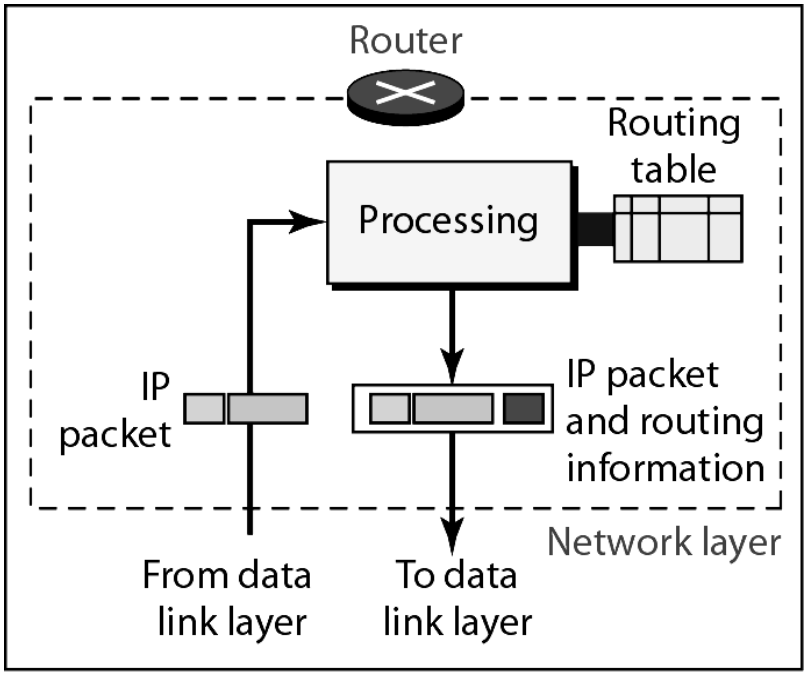
Figure 20.3 *Network layer at the source, router, and destination*



a. Network layer at source

b. Network layer at destination

Figure 20.3 *Network layer at the source, router, and destination (continued)*

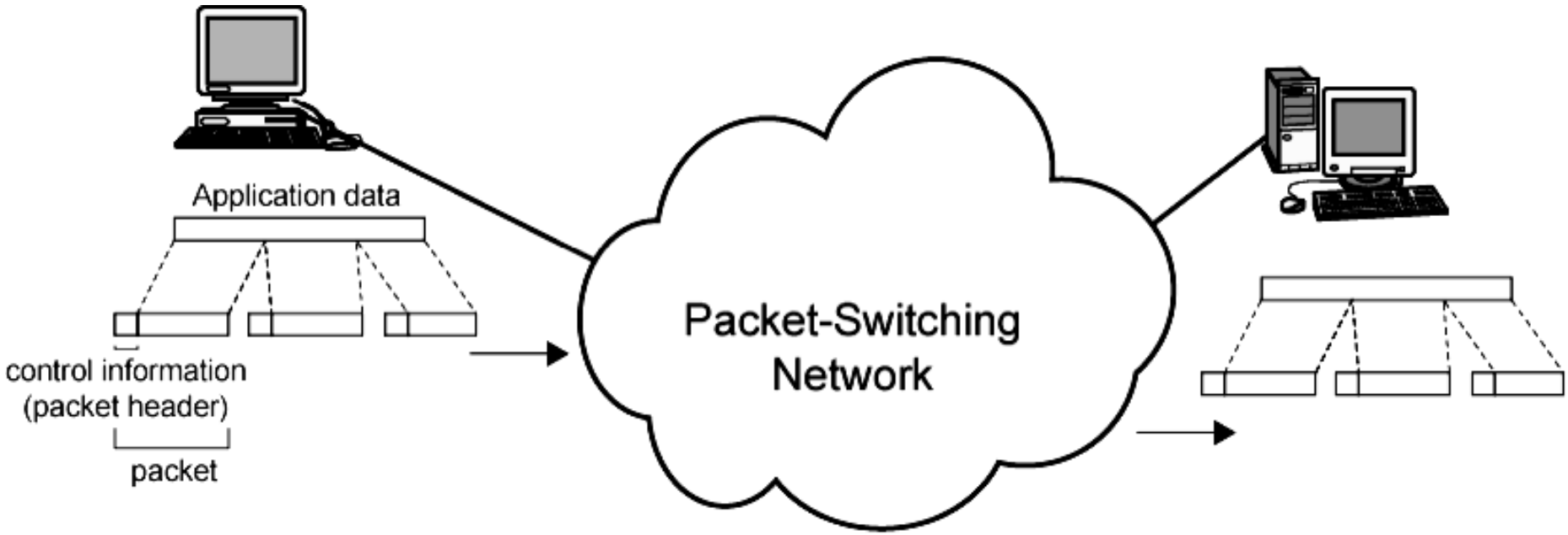


c. Network layer at a router

Packet Switching

- Data transmitted in small packets
  - Typically less than 1500 bytes (why?)
  - Longer messages split into series of packets
  - Each packet contains a portion of user data plus some control info
- Control info
  - Routing (addressing) info
- Packets are received, stored briefly (buffered) and past on to the next node
  - Store and forward

Use of Packets



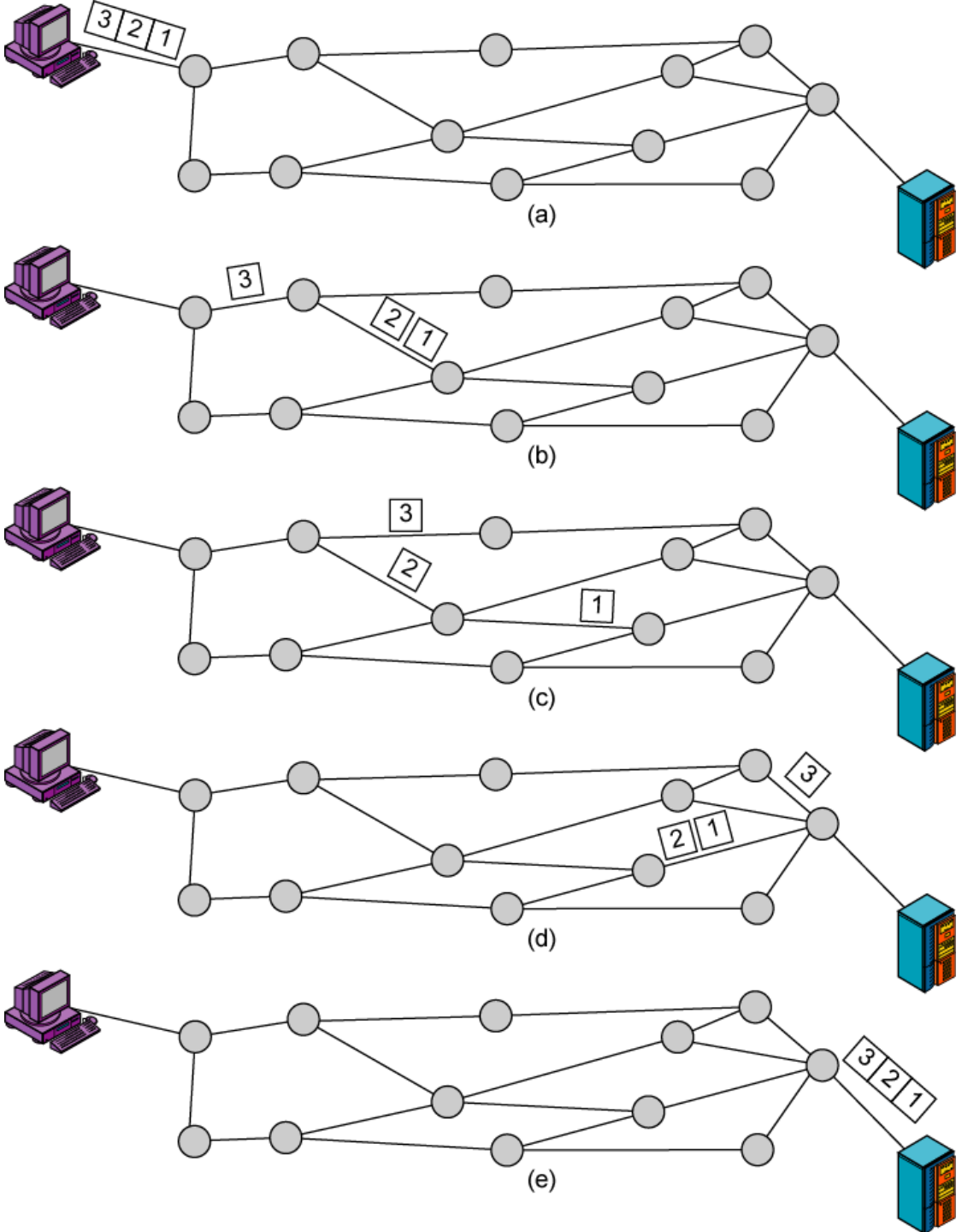
Switching Technique

- Station breaks long message into packets
- Packets sent one at a time to the network
- Packets handled in two ways
  - Datagram
  - Virtual circuit

Datagram

- Each packet treated independently
- Packets can take any practical route
- Packets may arrive out of order
- Packets may go missing
- Up to receiver to re-order packets and recover from missing packets

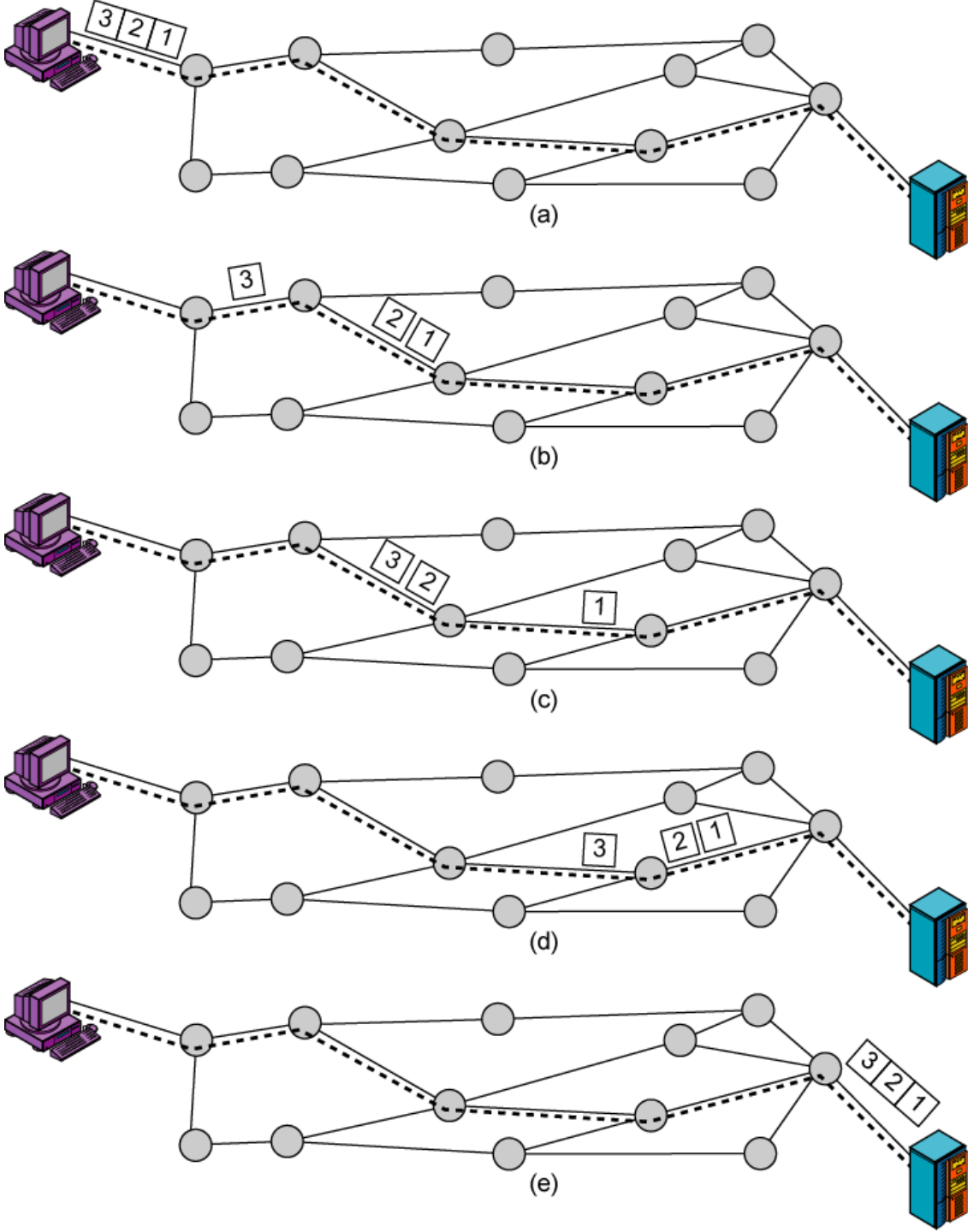
Datagram Diagram



Virtual Circuit

- Preplanned route established before any packets sent
- Call request and call accept packets establish connection (handshake)
- Each packet contains a virtual circuit identifier instead of destination address
- No routing decisions required for each packet
- Clear request to drop circuit
- Not a dedicated path

Virtual Circuit Diagram



Virtual Circuits v Datagram

- Virtual circuits
  - Network can provide sequencing and error control
  - Packets are forwarded more quickly
    - No routing decisions to make
  - Less reliable
    - Loss of a node looses all circuits through that node
- Datagram
  - No call setup phase
    - Better if few packets
  - More flexible
    - Routing can be used to avoid congested parts of the network

Note

Switching at the network layer in the Internet uses the datagram approach to packet switching.

Note

Communication at the network layer in the Internet is connectionless.

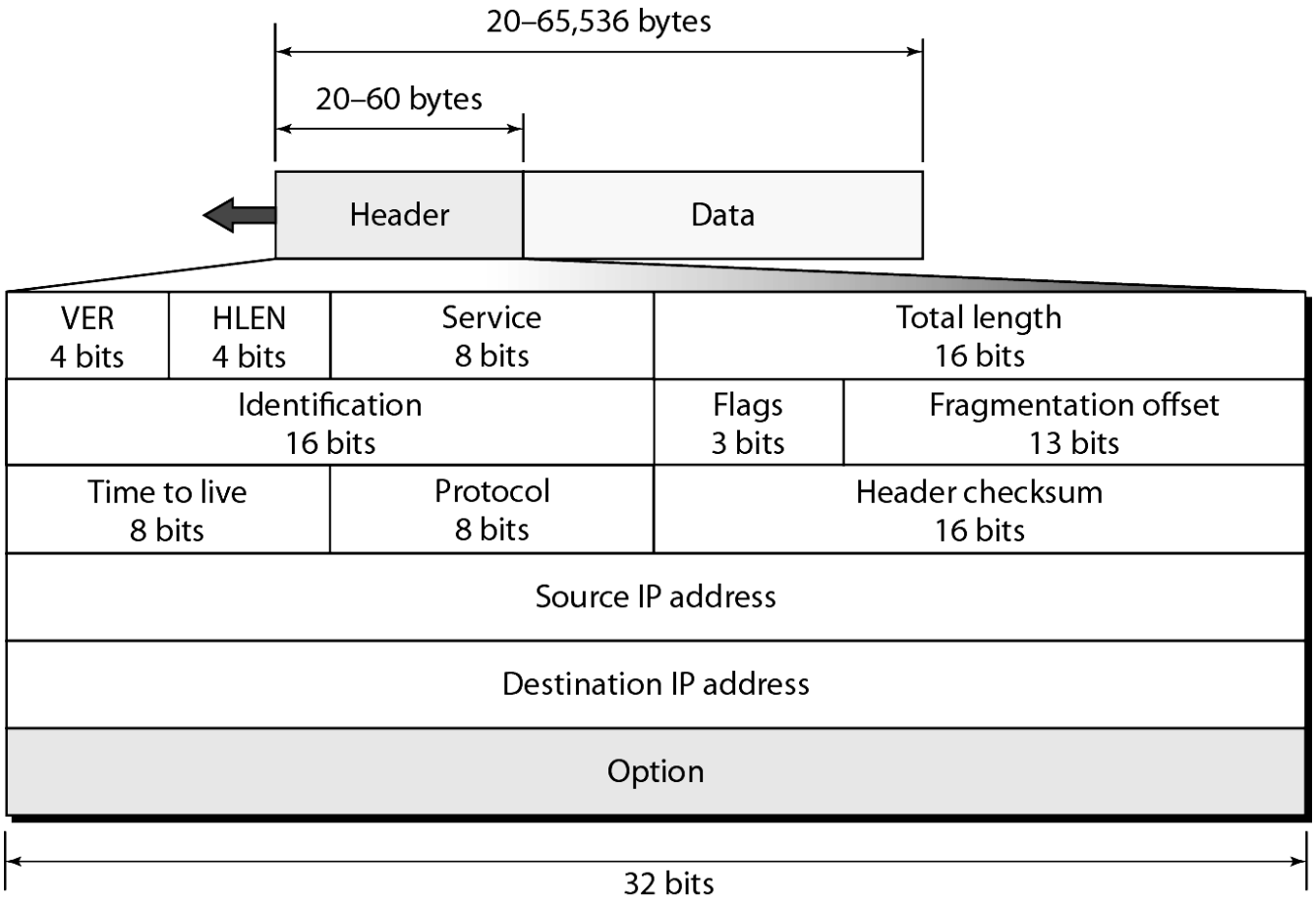
Note

IPv4 is an unreliable and connectionless datagram protocol – a best effort delivery

Best effort means that IPv4 provides no error control (except for error detection on the header) or flow control

IPv4 does its best to get a transmission through to its destination, but with no guarantees

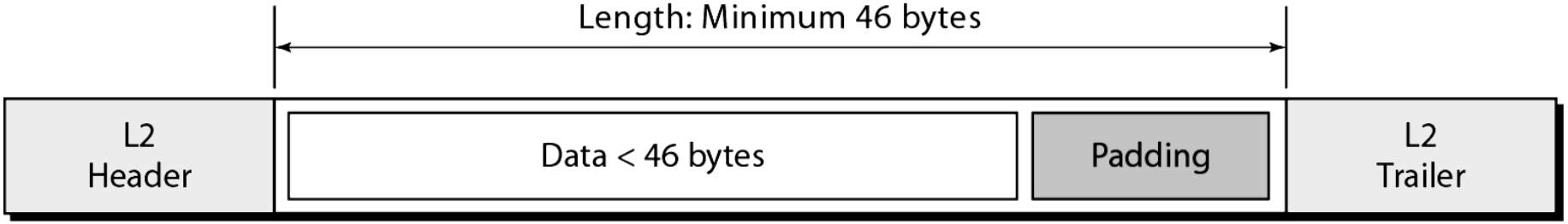
Figure 20.5 *IPv4 datagram format*



Note

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

Figure 20.7 *Encapsulation of a small datagram in an Ethernet frame*



One of the reason why "total length" field is required.

Figure 20.8 *Protocol field and encapsulated data*

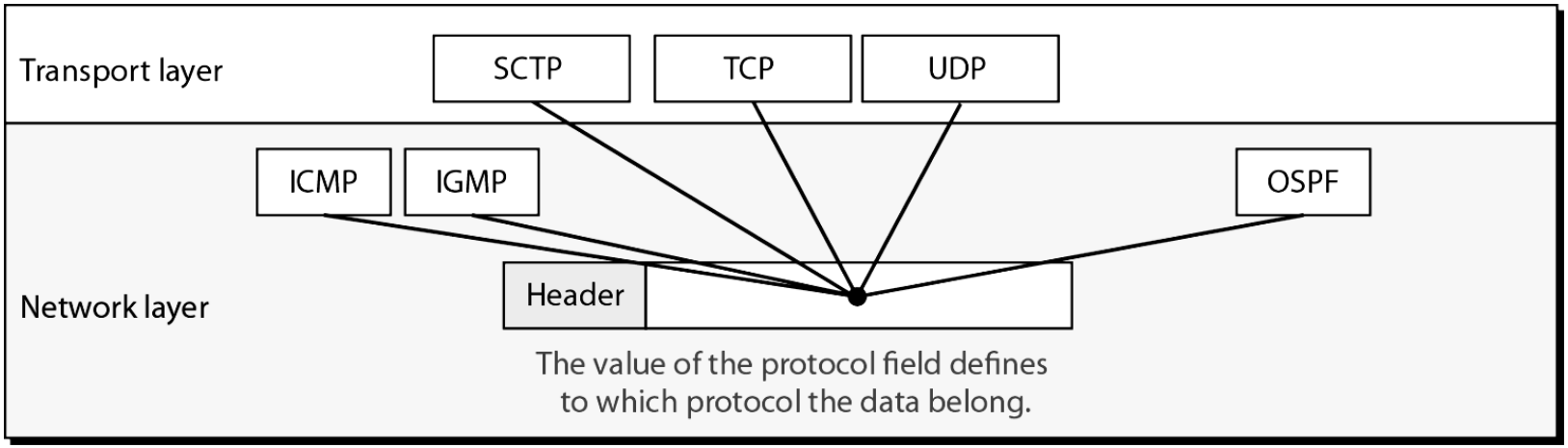
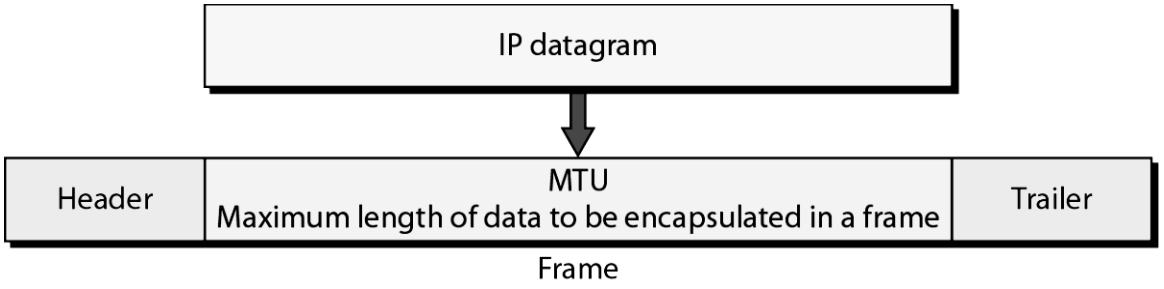


Table 20.4 ***Protocol values***

Value	*Protocol*
1	ICMP
2	IGMP
6	TCP
17	UDP
89	OSPF

Fragmentation

- A IP packet can travel through many different networks using different L2 (Data Link layers).
- The source node has no idea of the path and data link layer its packets will travel.
- MTU
  - Each DL has its own frame format and limitation.
  - One of such limitation is the maximum size of the frame, which is imposed by software, hardware, performance, and standards.



MTUs for some networks

Protocol	*MTU*
Hyperchannel	65,535
Token Ring (16 Mbps)	17,914
Token Ring (4 Mbps)	4,464
FDDI	4,352
Ethernet	1,500
X.25	576
PPP	296

Fields Related to Fragmentation

- **Identification:** identifies a datagram originating from the source host. A combination of the identification and source address must uniquely define a datagram as it leaves the source node.
- **Flags:** see next slide.
- **Fragmentation offset:** is the offset of the data in the original datagram measured in units of 8 bytes.

Figure 20.10 *Flags (3 bits) used in fragmentation*

- first bit: reserved (not used)
- second bit: = 1 requires the packet not to be fragmented
  drops the packet if it is > MTU

- third bit: =1 more fragmented packets later
  =0 the last fragmented packet

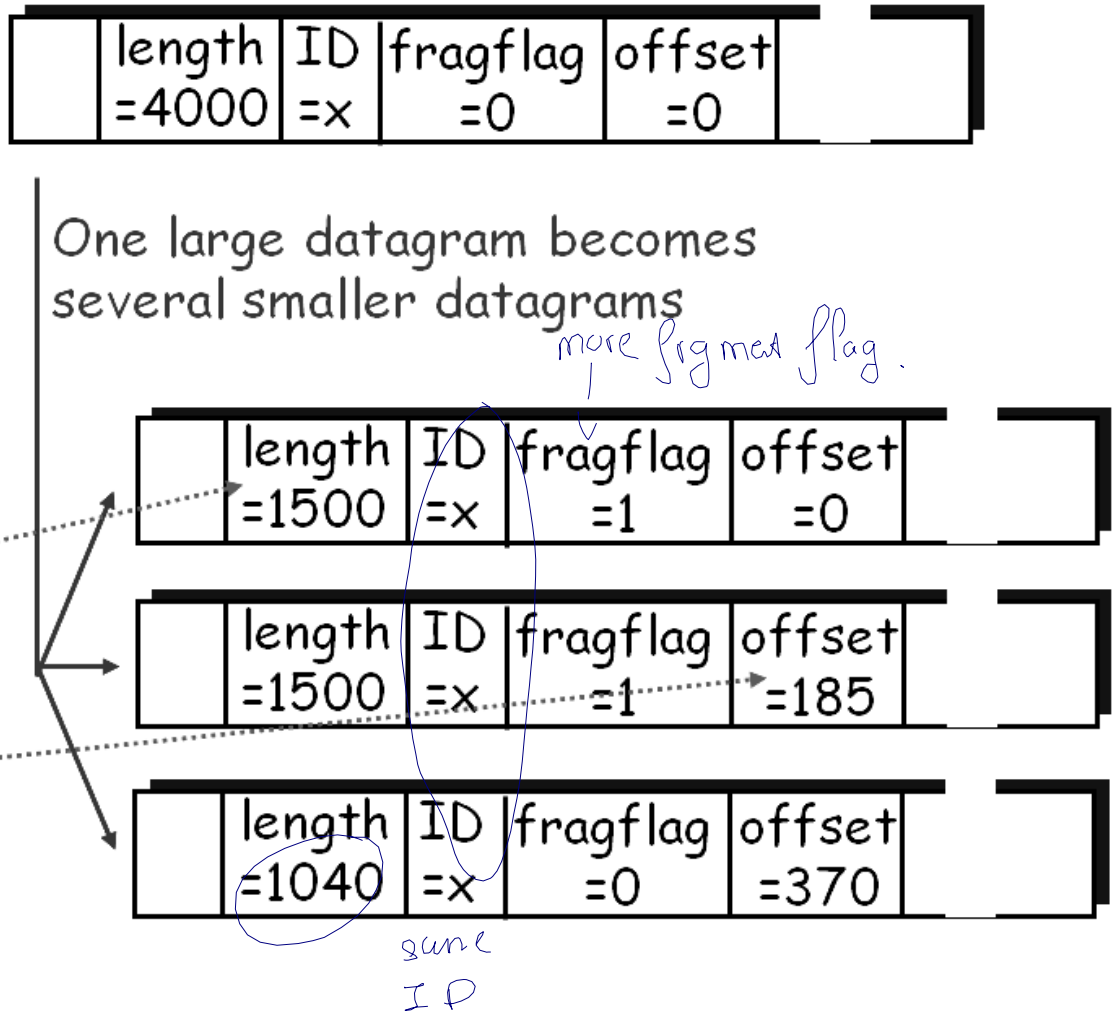
Fragmentation of IP

- The source node usually does not fragment the packet. Instead, L4 will segment the data into a size that can fit into L3 and L2 of the source.
- But, there is a possibility that a packet travel thru a link whose MTU is smaller than one of the source node.
  - Then, the packet must be fragmented to go forward the next hop.
  - Each fragment has its own header mostly repeated from the original packet.
  - A fragmented packet can be further fragmented into even smaller packet.
  - Fragmented packets will be re-assembled only by the final destination.

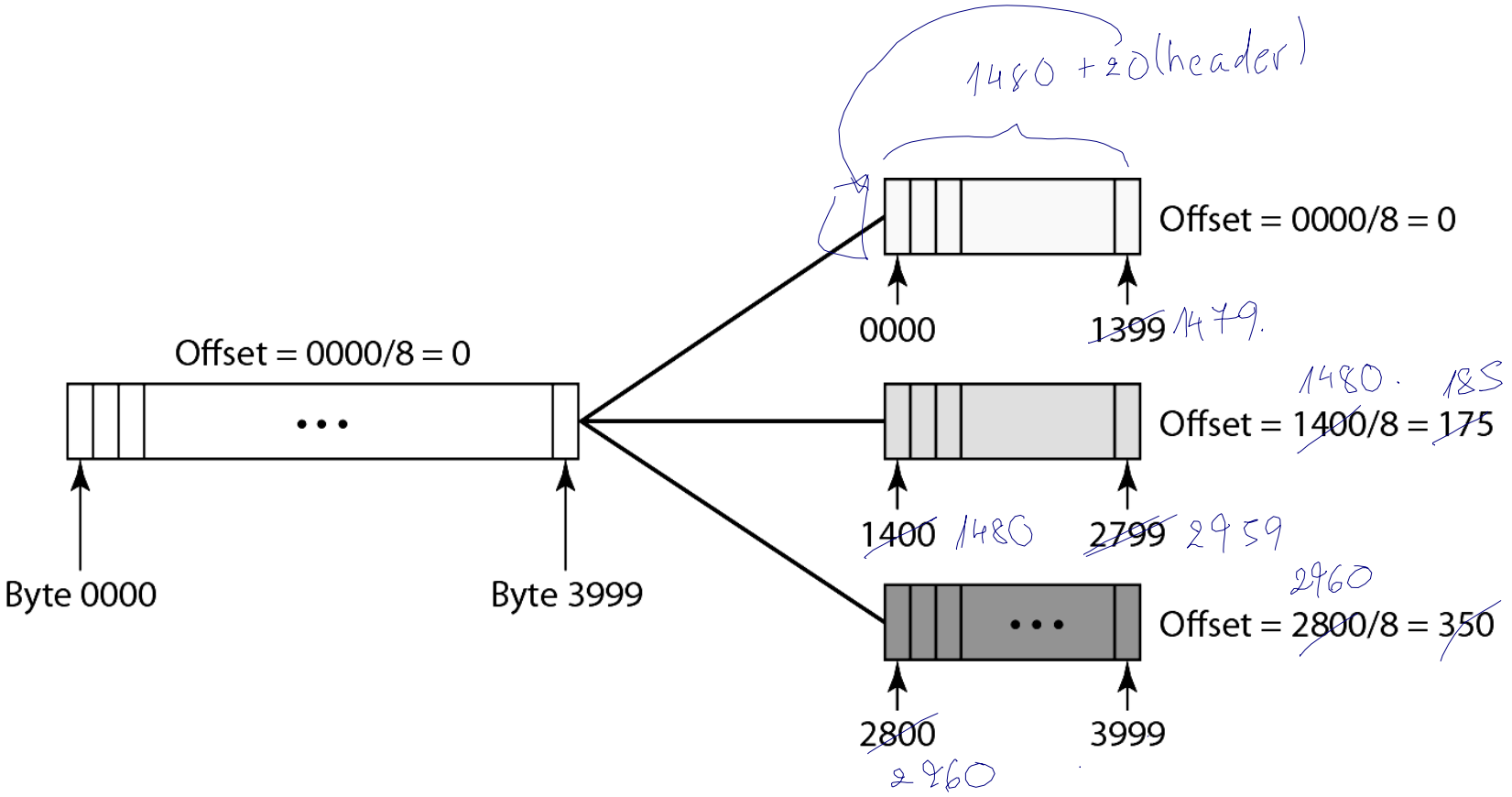
IP Fragmentation and Reassembly

Example

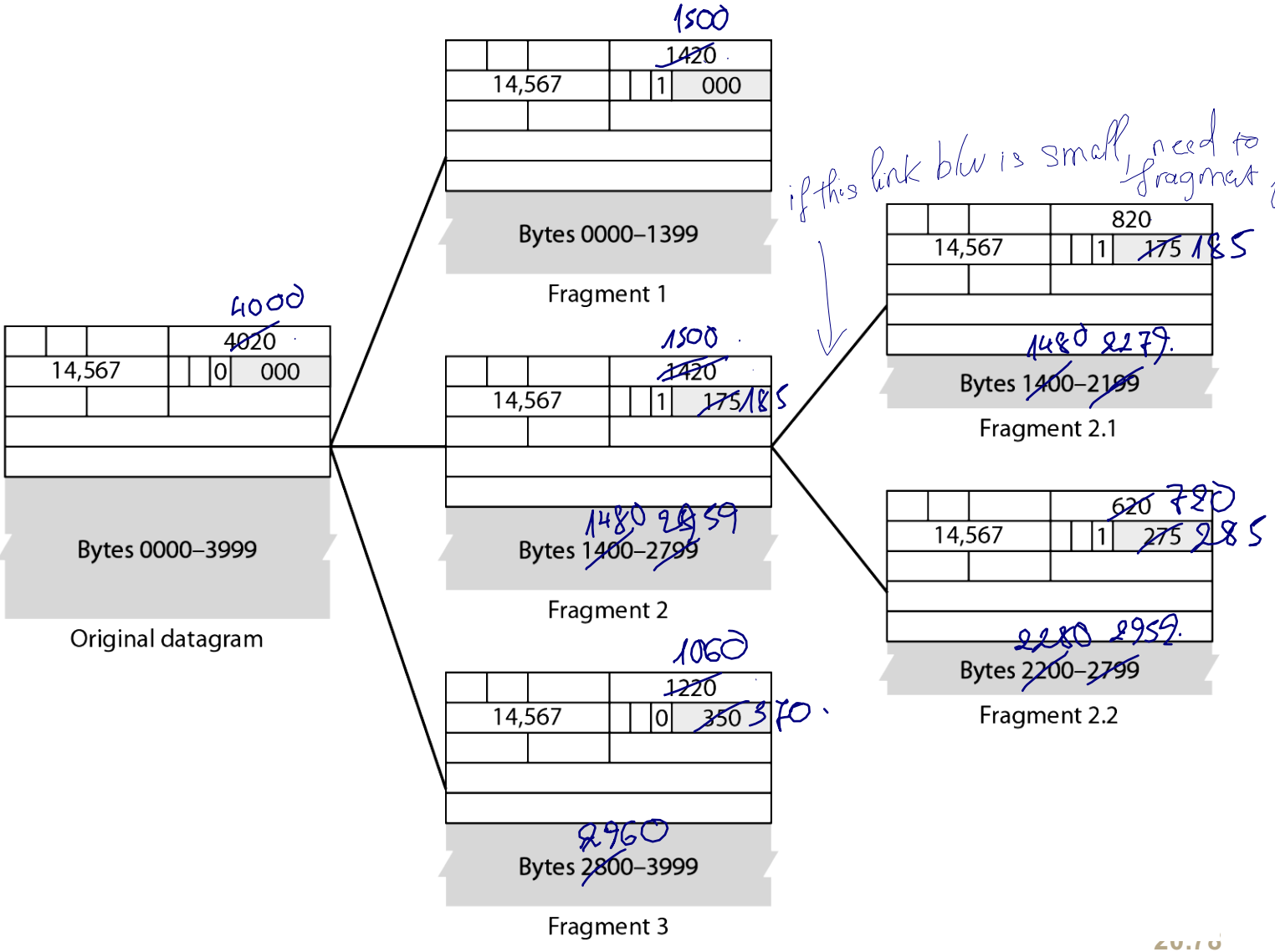
- 4000 byte datagram
- MTU = 1500 bytes



Fragmentation example



Detailed fragmentation example



Example 20.5

A packet has arrived with an M bit value of 0. Is this the first fragment, the last fragment, or a middle fragment? Do we know if the packet was fragmented?

Solution

If the M bit is 0, it means that there are no more fragments; the fragment is the last one. However, we cannot say if the original packet was fragmented or not. A non-fragmented packet is considered the last fragment.

Example 20.6

A packet has arrived with an M bit value of 1. Is this the first fragment, the last fragment, or a middle fragment? Do we know if the packet was fragmented?

Solution

If the M bit is 1, it means that there is at least one more fragment. This fragment can be the first one or a middle one, but not the last one. We don't know if it is the first one or a middle one; we need more information (the value of the fragmentation offset).

Example 20.7

A packet has arrived with an M bit value of 1 and a fragmentation offset value of 0. Is this the first fragment, the last fragment, or a middle fragment?

Solution

Because the M bit is 1, it is either the first fragment or a middle one. Because the offset value is 0, it is the first fragment.

Example 20.8

A packet has arrived in which the offset value is 100. What is the number of the first byte? Do we know the number of the last byte?

Solution

To find the number of the first byte, we multiply the offset value by 8. This means that the first byte number is 800. We cannot determine the number of the last byte unless we know the length.

Example 20.9

A packet has arrived in which the offset value is 100, the value of HLEN is 5, and the value of the total length field is 100. What are the numbers of the first byte and the last byte?

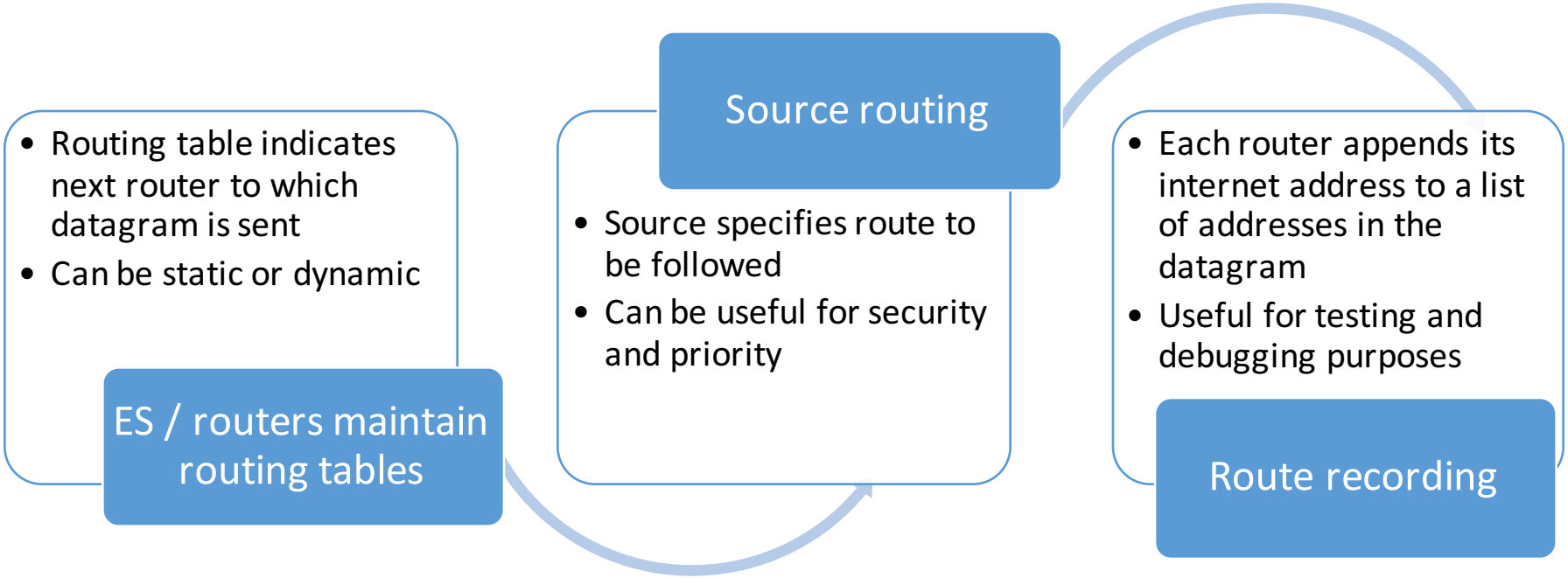
Solution

The first byte number is 100 × 8 = 800. The total length is 100 bytes, and the header length is 20 bytes (5 × 4), which means that there are 80 bytes in this datagram. If the first byte number is 800, the last byte number must be 879.

Internet Checksum

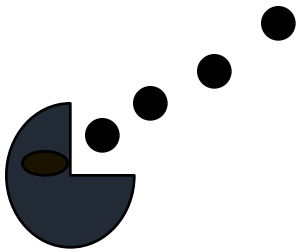
- Only for the header, but not for the payload
  - Each router modifies the IP header, but not the payload.
  - No special hardware can be used.
    - Computationally efficient.
  - The upper layers will check the integrity of the payload by their own schemes.

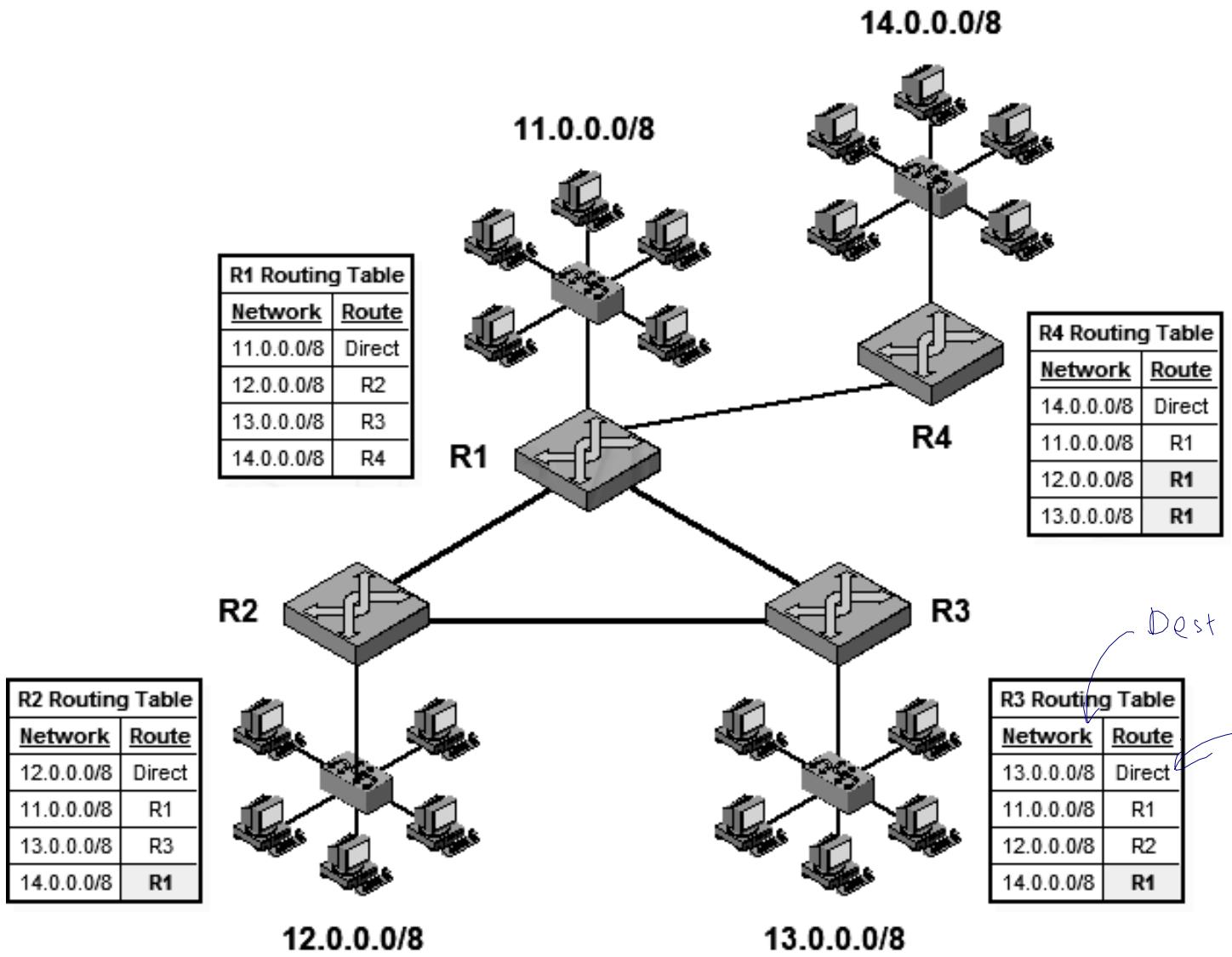
Routing


• Routing table indicates next router to which datagram is sent
• Can be static or dynamic
ES / routers maintain routing tables
Source routing
• Source specifies route to be followed
• Can be useful for security and priority
• Each router appends its internet address to a list of addresses in the datagram
• Useful for testing and debugging purposes
Route recording

Datagram Lifetime

- If dynamic or alternate routing is used the potential exists for a datagram to loop indefinitely
  - Consumes resources
  - Transport protocol may need upper bound on lifetime of a datagram
    - Can mark datagram with lifetime
    - When lifetime expires, datagram is discarded





R1 Routing Table	
Network	**Route**
11.0.0.0/8	Direct
12.0.0.0/8	R2
13.0.0.0/8	R3
14.0.0.0/8	R4

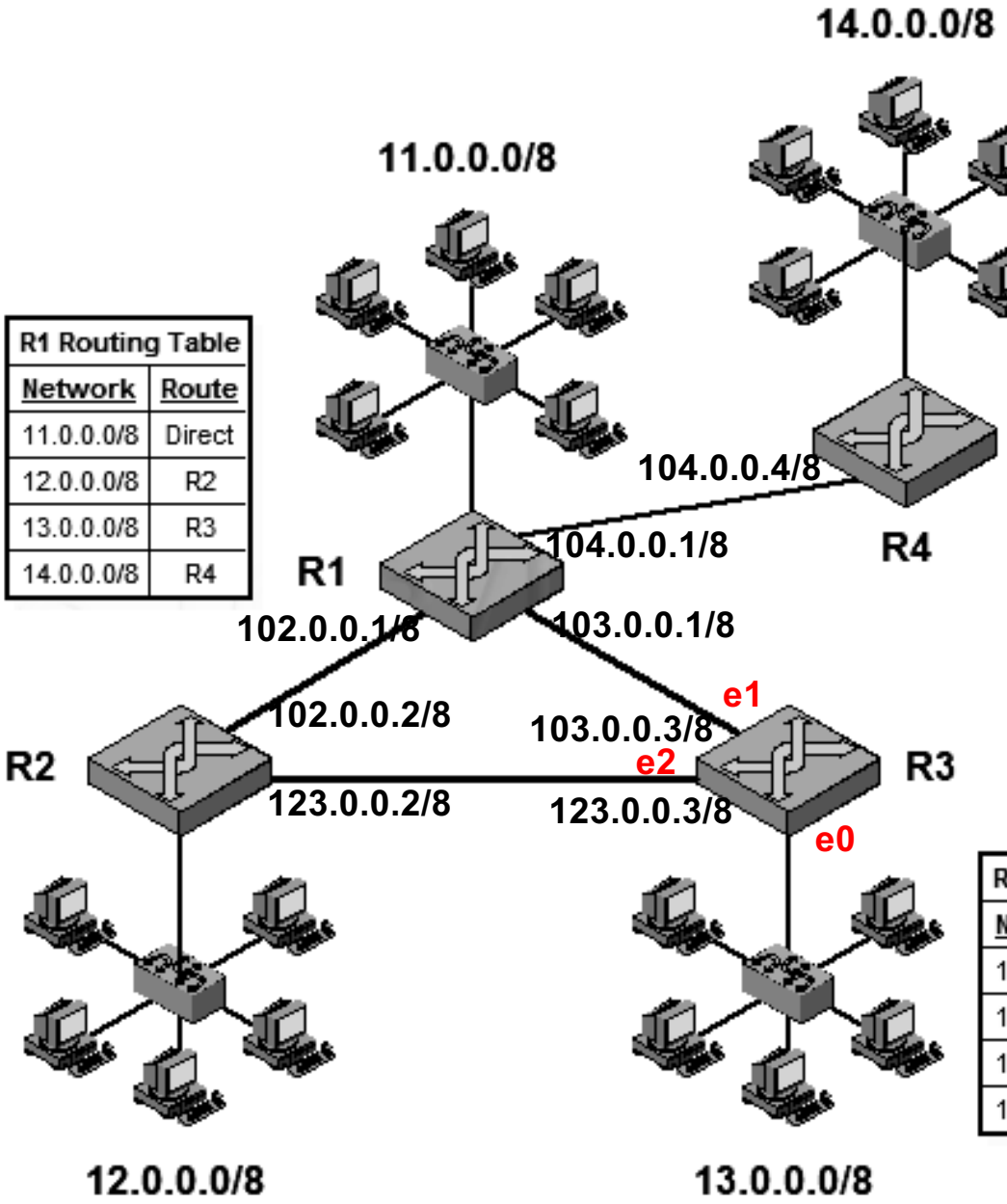
R4 Routing Table	
Network	**Route**
14.0.0.0/8	Direct
11.0.0.0/8	R1
12.0.0.0/8	**R1**
13.0.0.0/8	**R1**

R2 Routing Table	
Network	**Route**
12.0.0.0/8	Direct
11.0.0.0/8	R1
13.0.0.0/8	R3
14.0.0.0/8	**R1**

R3 Routing Table	
Network	**Route**
13.0.0.0/8	Direct
11.0.0.0/8	R1
12.0.0.0/8	R2
14.0.0.0/8	**R1**

Dest address prefix.

where should I send it to



R1 Routing Table	
Network	**Route**
11.0.0.0/8	Direct
12.0.0.0/8	R2
13.0.0.0/8	R3
14.0.0.0/8	R4

R4 Routing Table	
Network	**Route**
14.0.0.0/8	Direct
11.0.0.0/8	R1
12.0.0.0/8	**R1**
13.0.0.0/8	**R1**

R2 Routing Table	
Network	**Route**
12.0.0.0/8	Direct
11.0.0.0/8	R1
13.0.0.0/8	R3
14.0.0.0/8	**R1**

R3 Routing Table	
Network	**Route**
13.0.0.0/8	Direct
11.0.0.0/8	R1
12.0.0.0/8	R2
14.0.0.0/8	**R1**

next hop
e0 IPadd

103.0.0.1
123.0.02
103.0.0.1

Example of IP Routing Table

Destination	Gateway	Genmask	Flags	Interface
217.136.39.1	0.0.0.0	255.255.255.255	UH	ppp0
192.168.0.0	0.0.0.0	255.255.0.0	U	eth0
169.254.0.0	0.0.0.0	255.255.0.0	U	eth0
172.16.0.0	0.0.0.0	255.240.0.0	U	eth0
10.0.0.0	0.0.0.0	255.0.0.0	U	eth0
0.0.0.0	217.136.39.1	0.0.0.0	UG	ppp0

exact match

Destination	Gateway (next hop IP addr)	Flags	Interface
217.136.39.1/32	0.0.0.0	UH	ppp0
192.168.0.0 /16	0.0.0.0	U	eth0
169.254.0.0/16	0.0.0.0	U	eth0
172.16.0.0 /12	0.0.0.0	U	eth0
10.0.0.0/8	0.0.0.0	U	eth0
0.0.0.0/0	217.136.39.1	UG	ppp0

first 16 bits only need to match

no gateway → direct connect

no match, anything or default route

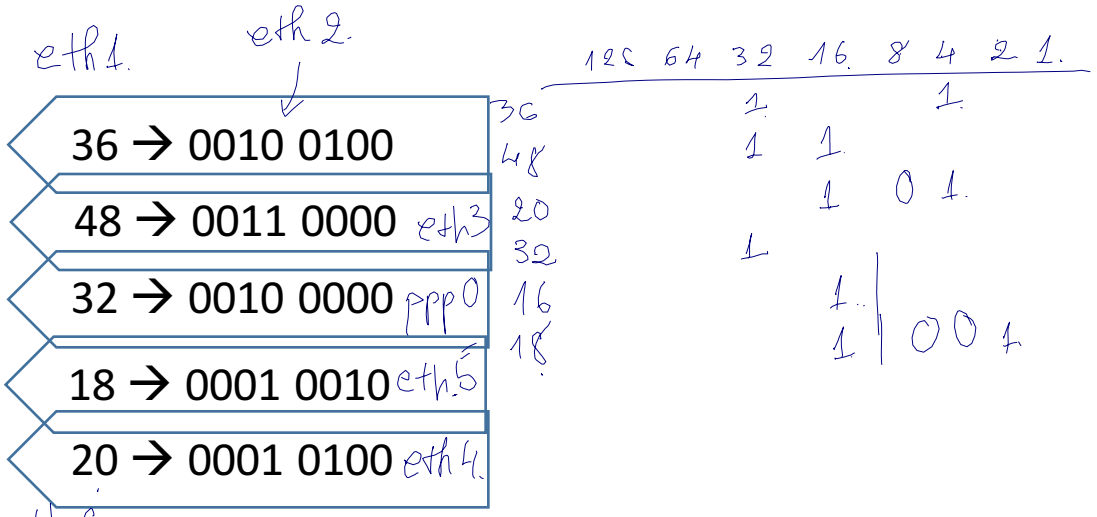
(if 2 routes match, choose the more specific one)

quiz

Longest Prefix Match

Destination	Interface	Binary Number
192.168.**1**.0/24	eth1	1 → 0000 0001
192.168.**32**.0 /20	eth2	32 → 0010 0000
192.168.0.0/16	eth3	
172.**20**.0.0/16	eth4	20 → 0001 0100
172.**16**.0.0 /12	eth5	16 → 0001 0000
10.0.0.0/8	eth6	
0.0.0.0/0	ppp0	

a. 192.168.1.10
b. 192.168.36.10
c. 192.168.48.10
d. 172.32.0.10
e. 172.18.0.10
f. 172.20.1.10
g. 10.1.0.10 → eth6
h. 20.1.0.10 → ppp0



Address Resolution Protocol (ARP)

IP addr to MAC

Need MAC address to send to LAN host

- Manual
- Included in network address
- Use central directory
- Use address resolution protocol

ARP (RFC 826) provides dynamic IP to Ethernet address mapping

- Source broadcasts ARP request
- Destination replies with ARP response

21-1 ADDRESS MAPPING

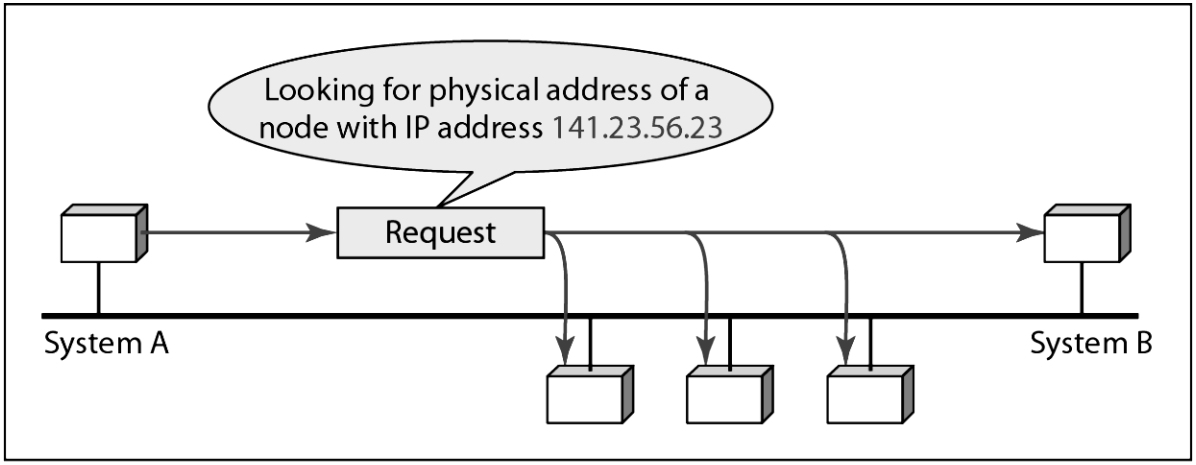
The delivery of a packet to a host or a router requires two levels of addressing: logical and physical. We need to be able to map a logical address to its corresponding physical address and vice versa. This can be done by using either static or dynamic mapping.

Topics discussed in this section:

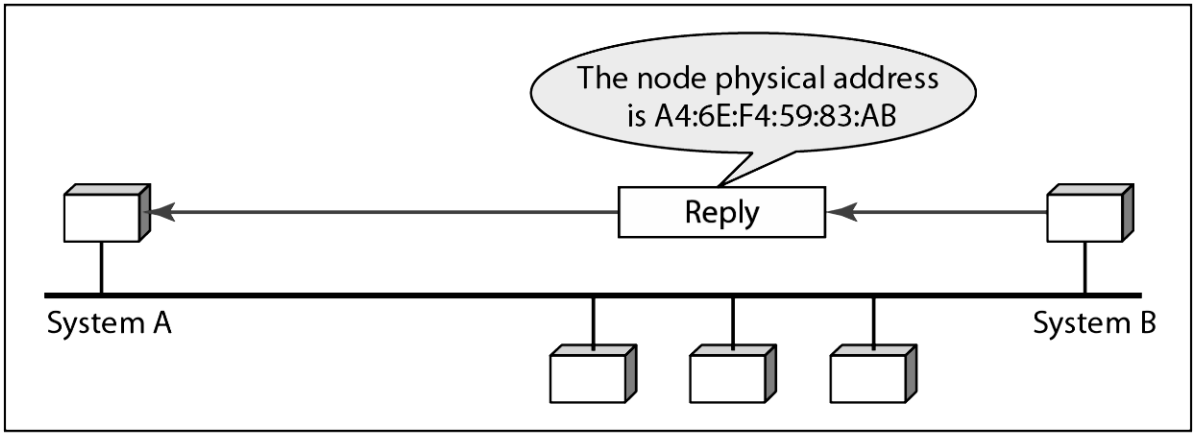
Mapping Logical to Physical Address
Mapping Physical to Logical Address

Figure 21.1 Mapping Logical to Physical Address

ARP (address resolution protocol)



a. ARP request is broadcast



b. ARP reply is unicast

ARP can be useful if the ARP reply is cached (kept in cache memory for a while).

Figure 21.2 *ARP packet*

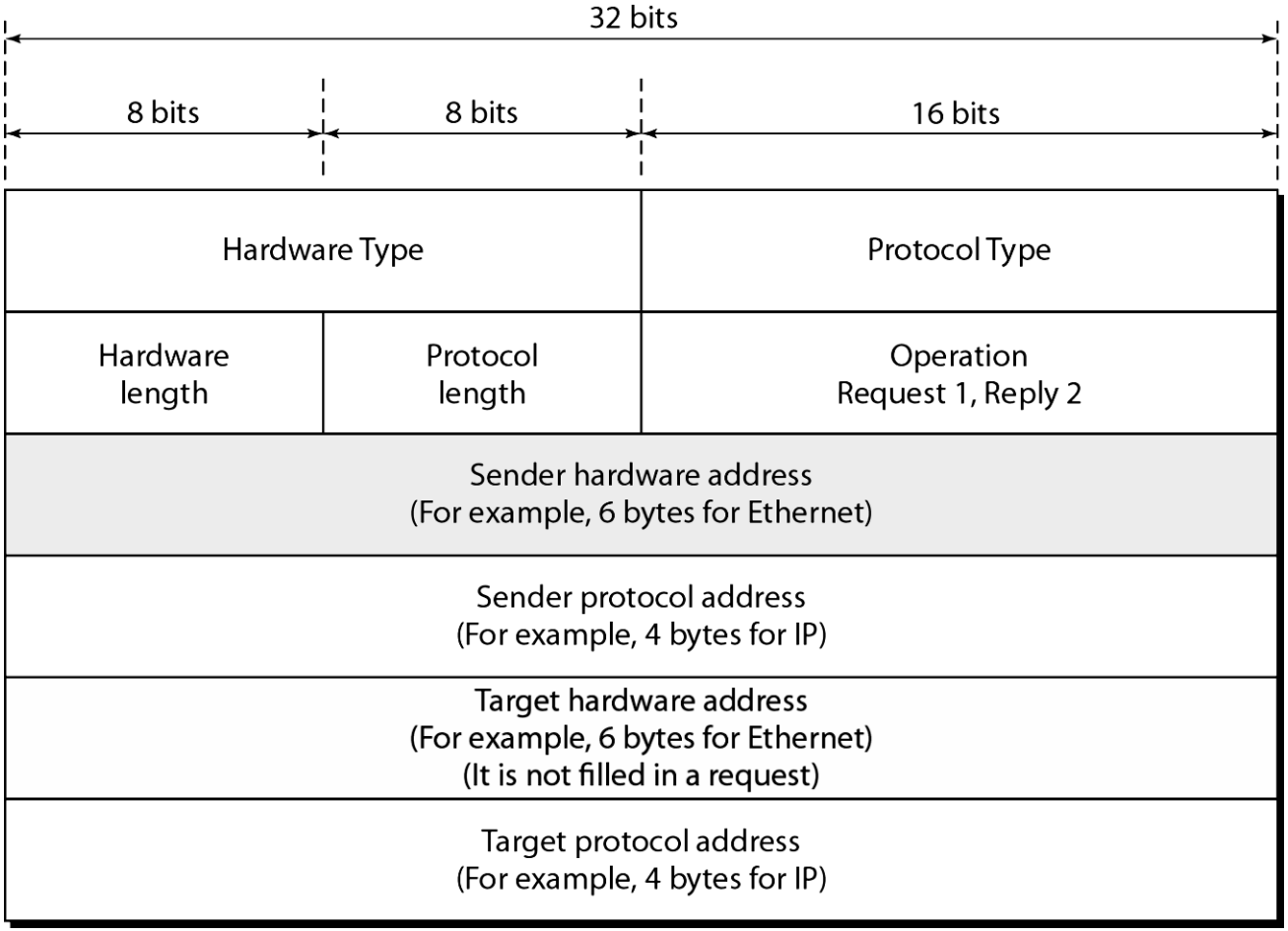


Figure 21.3 *Encapsulation of ARP packet*

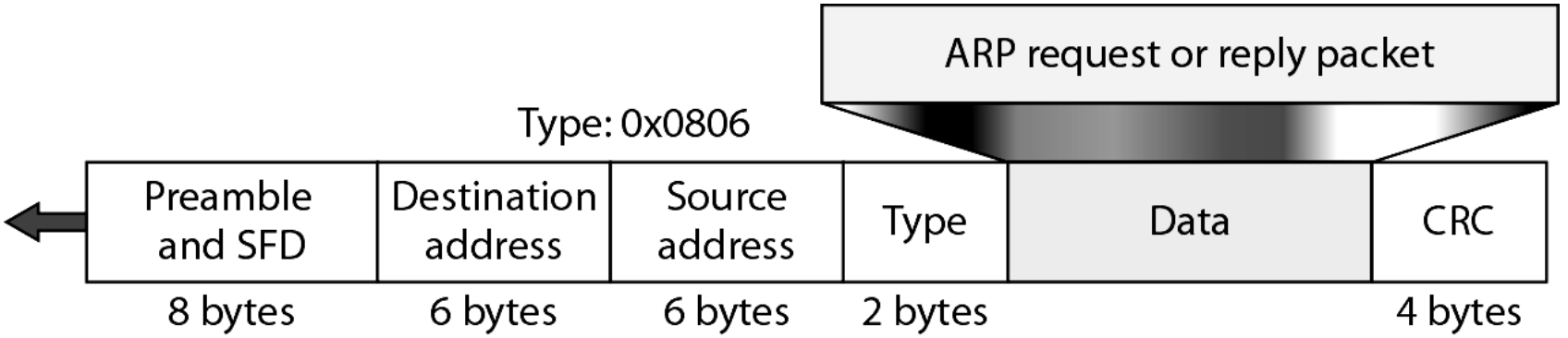
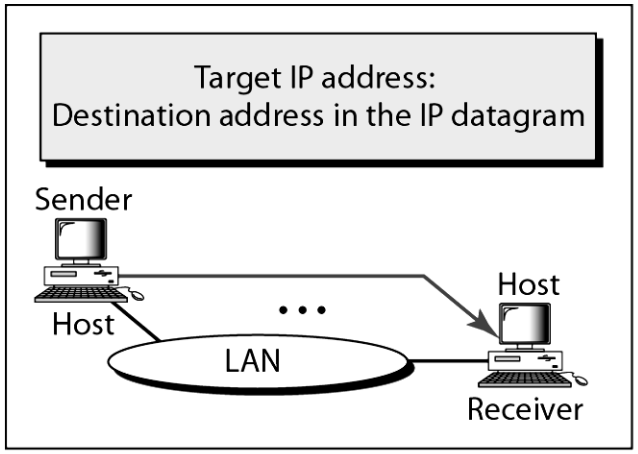
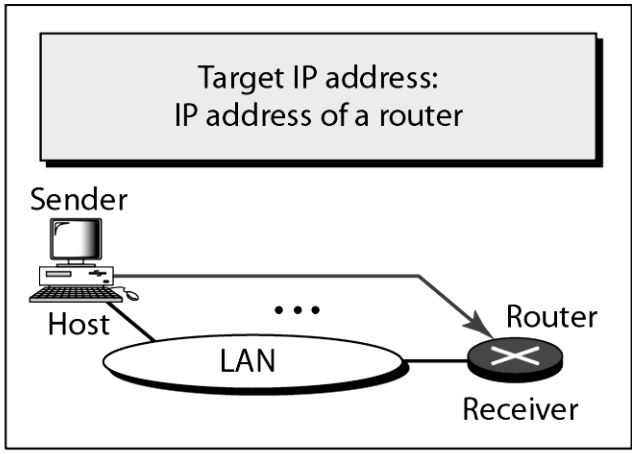


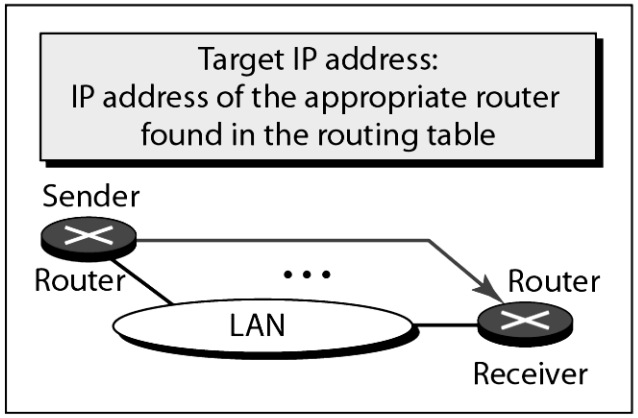
Figure 21.4 *Four cases using ARP*



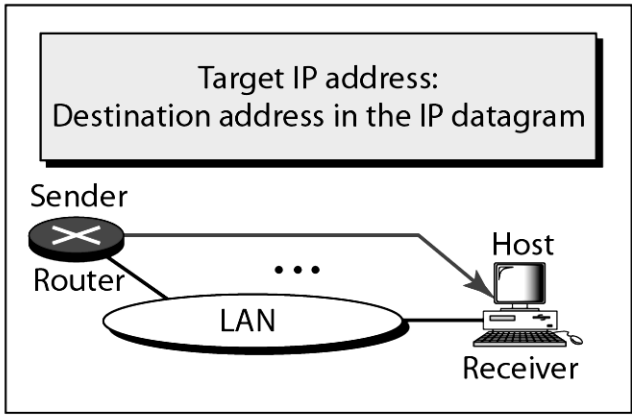
Case 1. A host has a packet to send to another host on the same network.



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

Note

An ARP request is broadcast; an ARP reply is unicast.

Example 21.1

A host with IP address 130.23.43.20 and physical address B2:34:55:10:22:10 has a packet to send to another host with IP address 130.23.43.25 and physical address A4:6E:F4:59:83:AB. The two hosts are on the same Ethernet network. Show the ARP request and reply packets encapsulated in Ethernet frames.

Solution

Figure 21.5 shows the ARP request and reply packets. Note that the ARP data field in this case is 28 bytes, and that the individual addresses do not fit in the 4-byte boundary. That is why we do not show the regular 4-byte boundaries for these addresses.

Figure 21.5 *Example 21.1, an ARP request and reply*

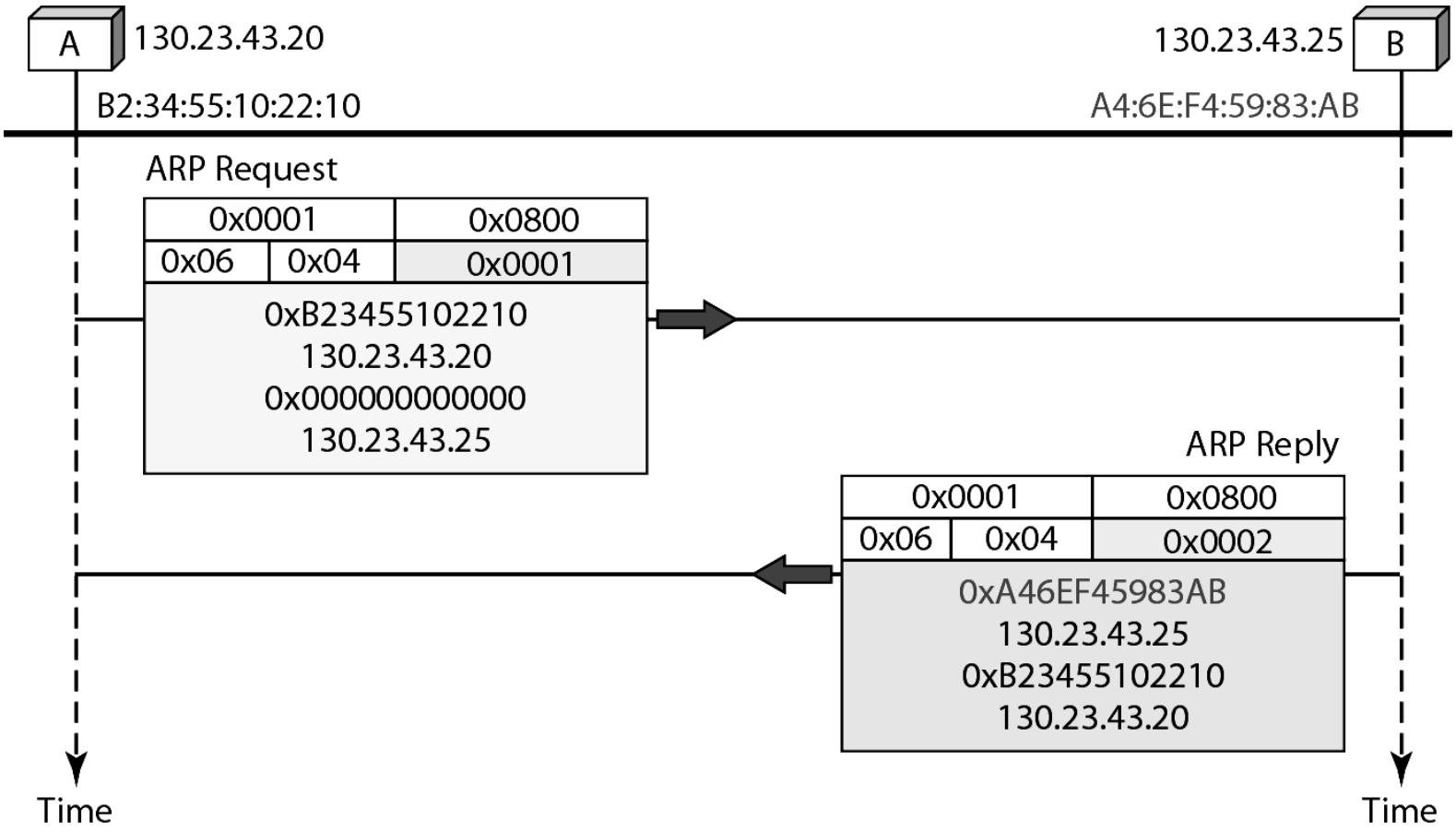
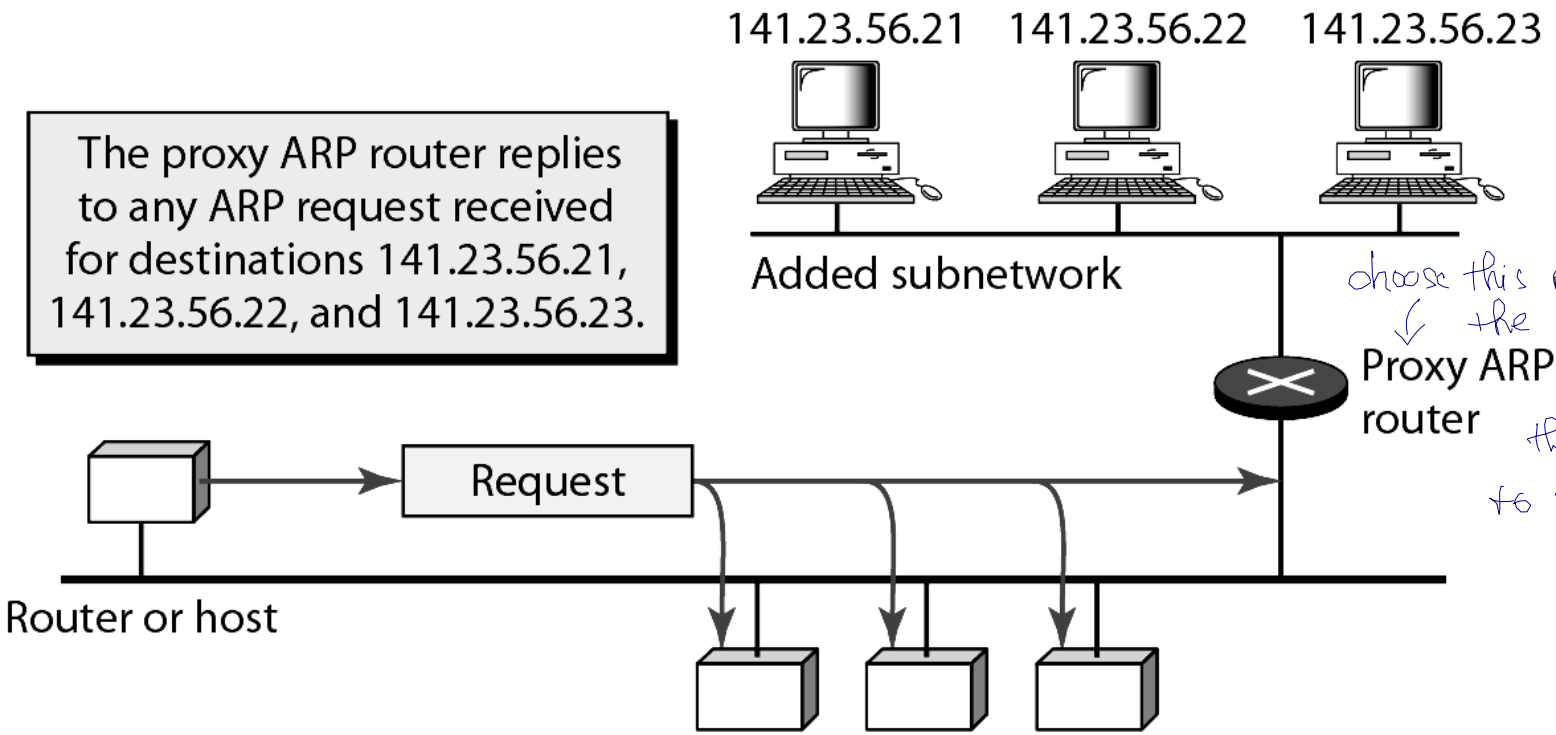


Figure 21.6 *Proxy ARP*



choose this router as the representative
Instead the host has to reply.

Reverse Address Resolution Protocol (RARP)

Stop here.

- A machine can use the phy address to get the logical address using RARP.
- A RARP messages is created and brodcast on the local network.
- The machine on the local network that knows the logical address will respond with a RARP reply.
- Broadcasting is done at data link layer.
- Broadcast requests does not pass the boundaries of a network.

GARP (gracious ARP) : advertise the MAC before ppl ask

benefits - when you move, can inform new add
- reduce traffic of broadcast

21-2 ICMP

The IP protocol has no error-reporting or error-correcting mechanism. The IP protocol also lacks a mechanism for host and management queries. The Internet Control Message Protocol (ICMP) has been designed to compensate for the above two deficiencies. It is a companion to the IP protocol.

Topics discussed in this section:

Types of Messages
Message Format
Error Reporting and Query
Debugging Tools

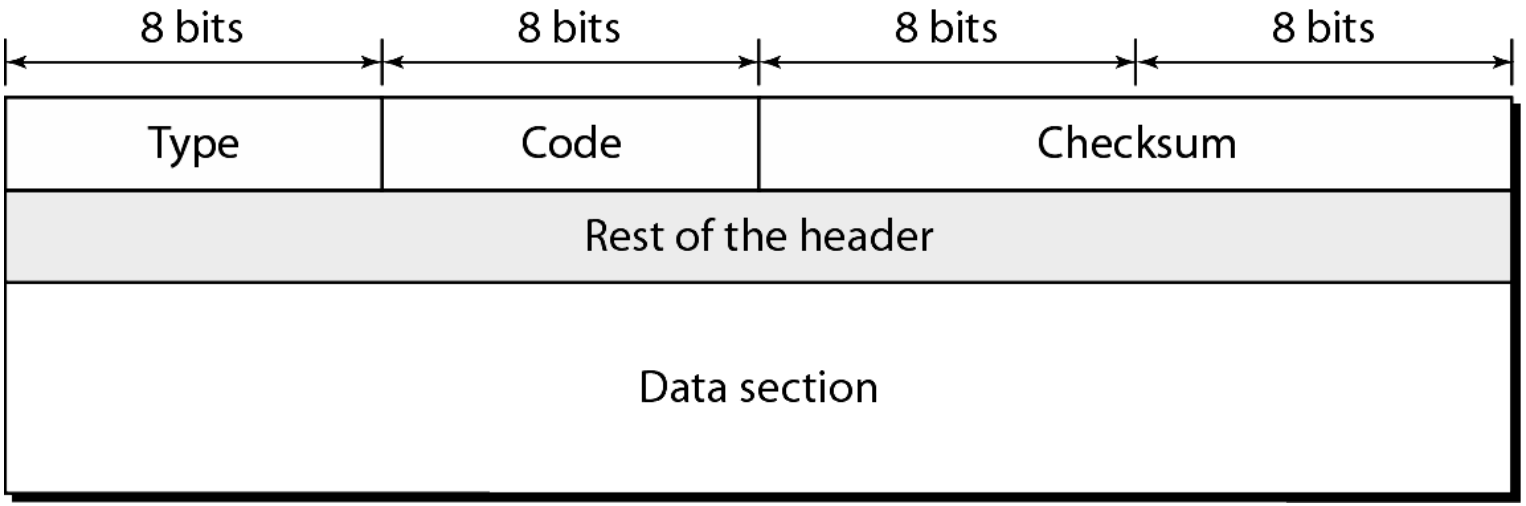
Internet Control Message Protocol (ICMP)

- RFC 792
- Provides a means for transferring messages from routers and other hosts to a host
- Provides feedback about problems
    - Datagram cannot reach its destination
    - Router does not have buffer capacity to forward
    - Router can send traffic on a shorter route
- Encapsulated in IP datagram
    - Hence not reliable

Common ICMP Messages

- Destination unreachable
- Time exceeded
- Parameter problem something wrong in header
- Source quench deprecated, no one use
- Redirect
- Echo and echo reply for pinging.
- Timestamp and timestamp reply not common deprecated
- Address mask request and reply deprecated.

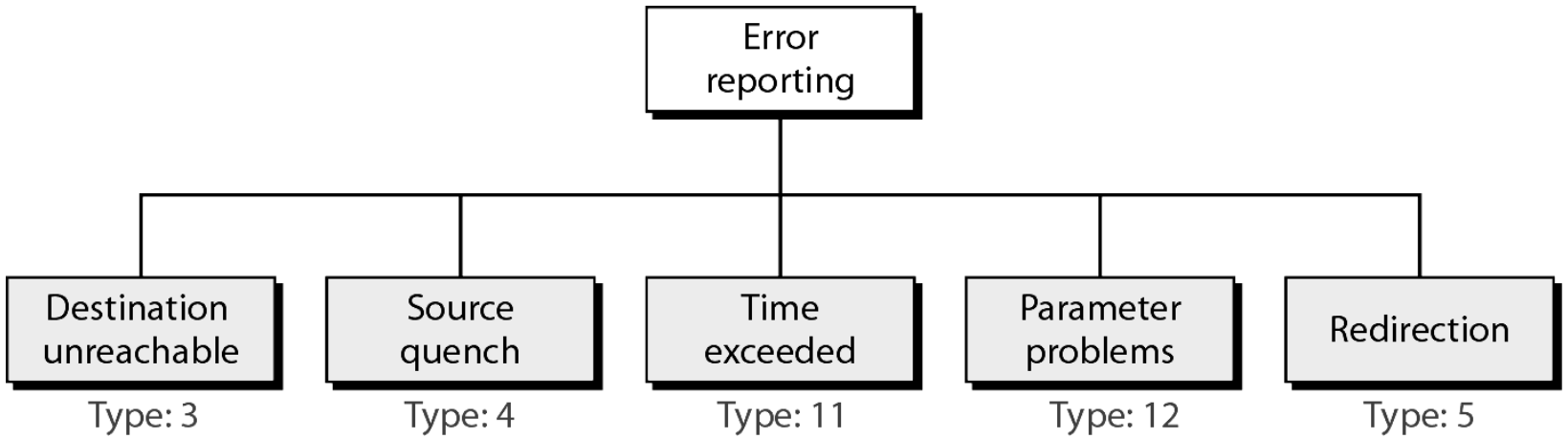
Figure 21.8 *General format of ICMP messages*



Note

ICMP always reports error messages to the original source.

Figure 21.9 *Error-reporting messages*



Note

Important points about ICMP error messages:

- No ICMP error message will be generated in response to a datagram carrying an ICMP error message.
- No ICMP error message will be generated for a fragmented datagram that is not the first fragment.
- No ICMP error message will be generated for a datagram having a multicast address.
- No ICMP error message will be generated for a datagram having a special address such as 127.0.0.0 or 0.0.0.0.

Figure 21.10 *Contents of data field for the error messages*

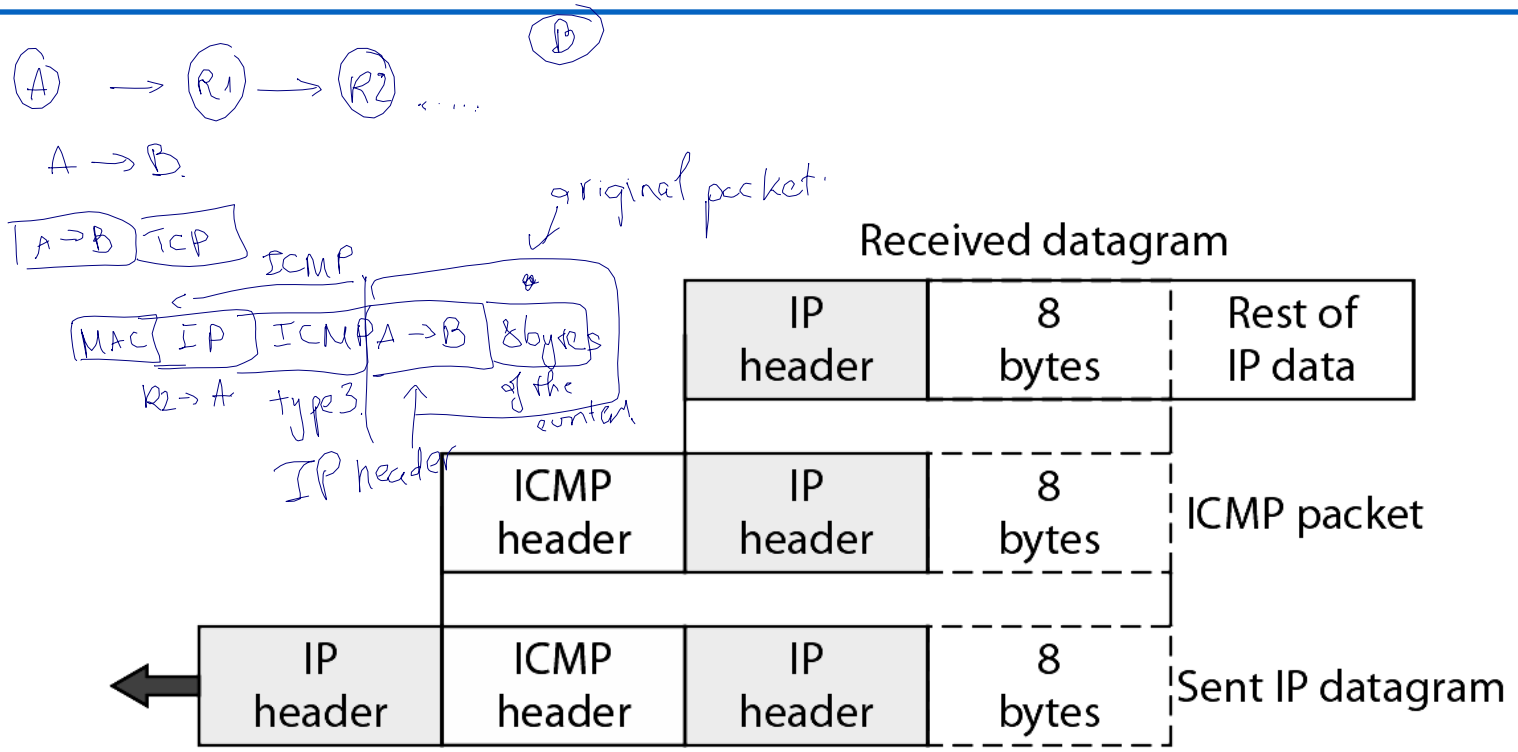


Figure 21.12 *Query messages*

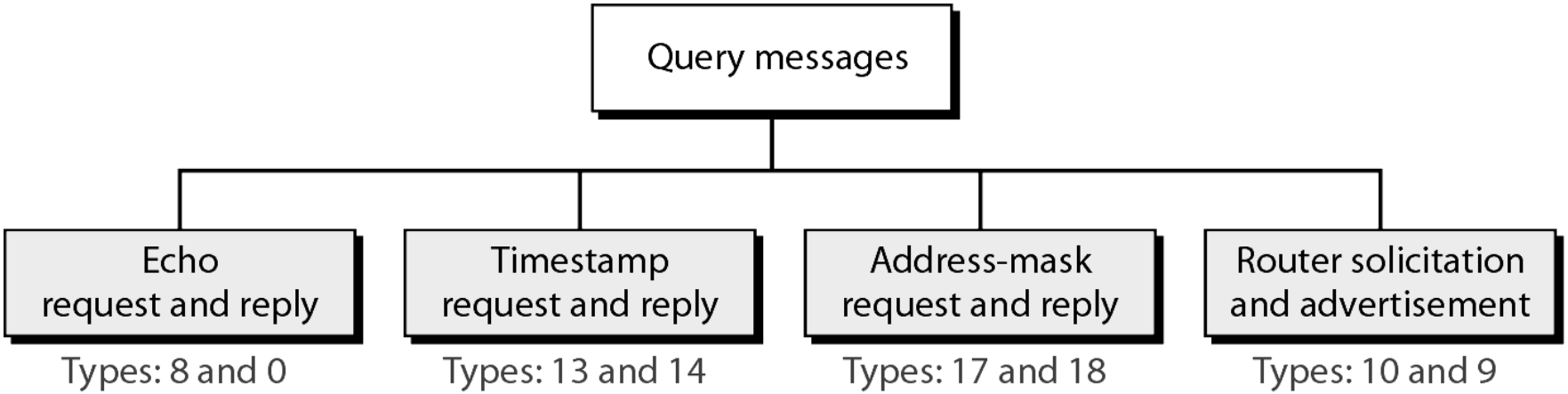
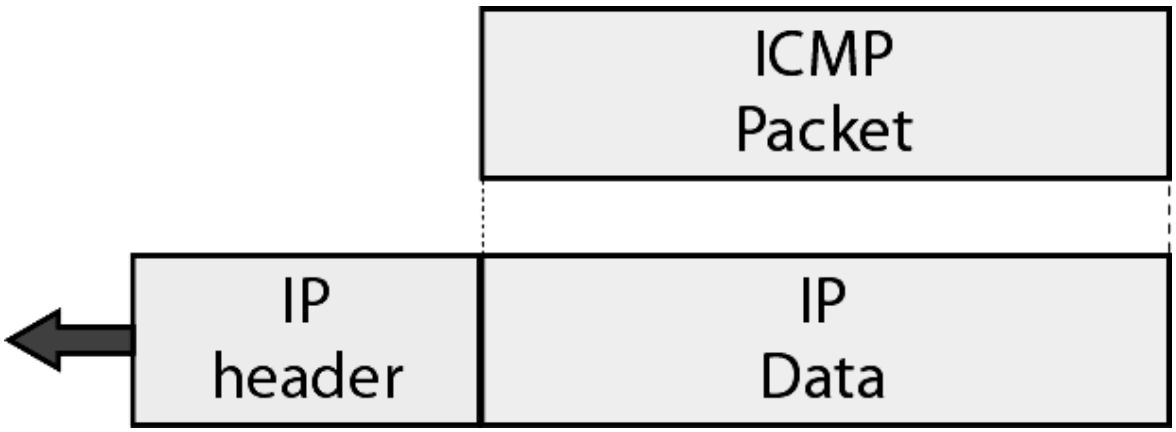


Figure 21.13 *Encapsulation of ICMP query messages*



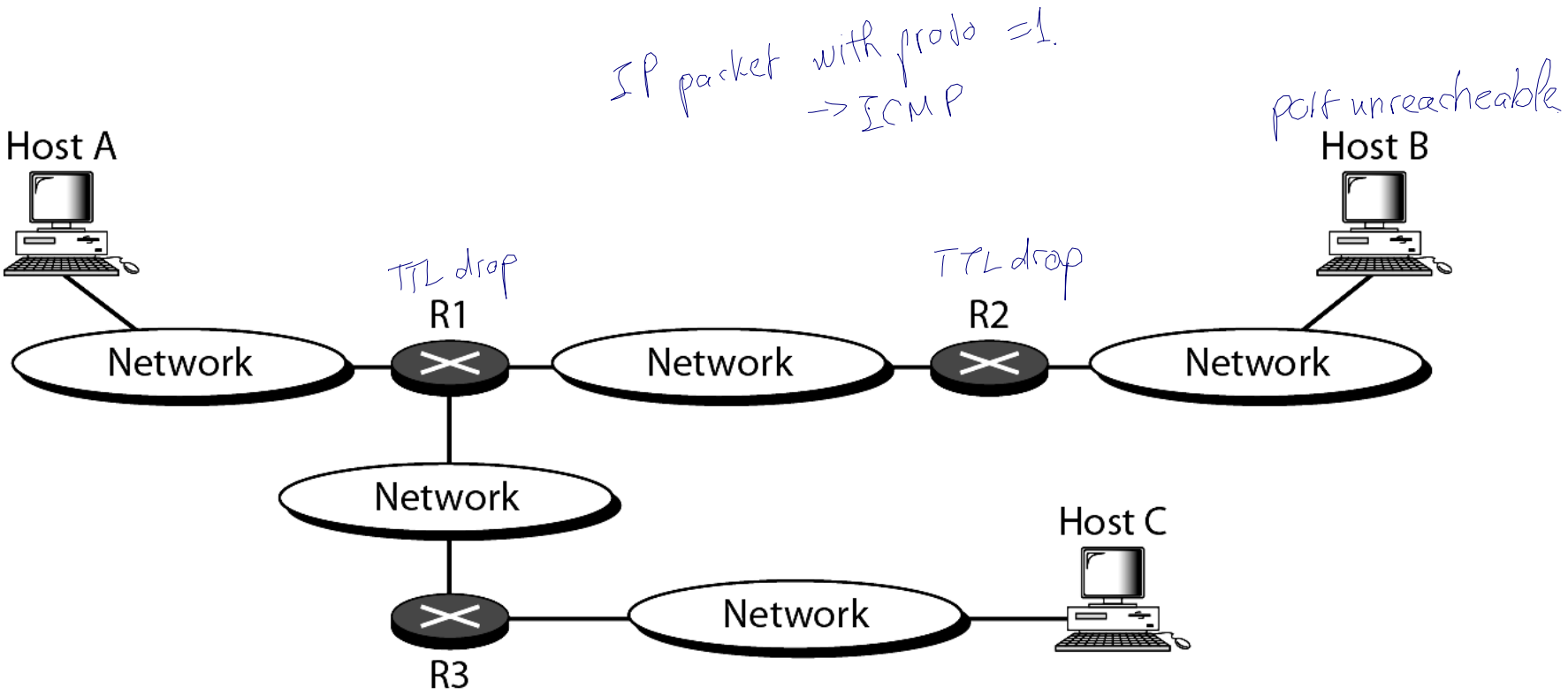
Example 21.3

We use the ping program to test the server fhda.edu. The result is shown on the next slide. The ping program sends messages with sequence numbers starting from 0. For each probe it gives us the RTT time. The TTL (time to live) field in the IP datagram that encapsulates an ICMP
message has been set to 62. At the beginning, ping defines the number of data bytes as 56 and the total number of bytes as 84. It is obvious that if we add 8 bytes of ICMP header and 20 bytes of IP header to 56, the result is 84. However, note that in each probe ping defines the number of bytes as 64. This is the total number of bytes in the ICMP packet (56 + 8).

Example 21.3 (continued)

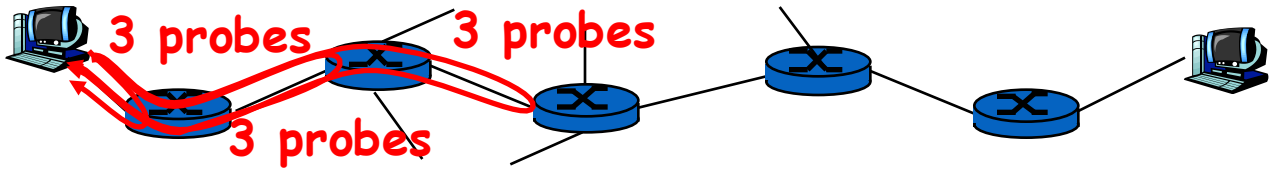
```
$ ping fhda.edu
PING fhda.edu (153.18.8.1)    56 (84)  bytes of data.
64 bytes from tiptoe.fhda.edu (153.18.8.1): icmp_seq=0     ttl=62     time=1.91 ms
64 bytes from tiptoe.fhda.edu (153.18.8.1): icmp_seq=1     ttl=62     time=2.04 ms
64 bytes from tiptoe.fhda.edu (153.18.8.1): icmp_seq=2     ttl=62     time=1.90 ms
64 bytes from tiptoe.fhda.edu (153.18.8.1): icmp_seq=3     ttl=62     time=1.97 ms
64 bytes from tiptoe.fhda.edu (153.18.8.1): icmp_seq=4     ttl=62     time=1.93 ms
64 bytes from tiptoe.fhda.edu (153.18.8.1): icmp_seq=5     ttl=62     time=2.00 ms
64 bytes from tiptoe.fhda.edu (153.18.8.1): icmp_seq=6     ttl=62     time=1.94 ms
64 bytes from tiptoe.fhda.edu (153.18.8.1): icmp_seq=7     ttl=62     time=1.94 ms
64 bytes from tiptoe.fhda.edu (153.18.8.1): icmp_seq=8     ttl=62     time=1.97 ms
64 bytes from tiptoe.fhda.edu (153.18.8.1): icmp_seq=9     ttl=62     time=1.89 ms
64 bytes from tiptoe.fhda.edu (153.18.8.1): icmp_seq=10    ttl=62     time=1.98 ms

--- fhda.edu ping statistics ---
11 packets transmitted, 11 received, 0% packet loss, time 10103ms
      rtt min/avg/max = 1.899/1.955/2.041 ms
```

Figure 21.15 *The traceroute program operation*



“Real” Internet delays and routes

- What do “real” Internet delay & loss look like?
- **`Traceroute`** program: provides delay measurement from source to router along end-end Internet path towards destination. For all *i:*
  - sends three packets that will reach router *i* on path towards destination
  - router *i* will return packets to sender
  - sender times interval between transmission and reply.



First Hop Redundancy

- Protect the default gateway used on a subnetwork by allowing two or more routers to provide backup for that address; in the event of failure of the/an active router, the backup router will take over the address

Introduction

- Virtual Router Redundancy Protocol (VRRP) [RFC 5798]
  - is designed to eliminate the single point of failure inherent in the static default routed environment.
  - specifies an election protocol that dynamically assigns responsibility for a virtual router to one of the VRRP routers on a LAN.
- Cisco Proprietary Protocols
  - Hot Standby Router Protocol (HSRP).
  - Gateway Load Balancing Protocol (GLBP).

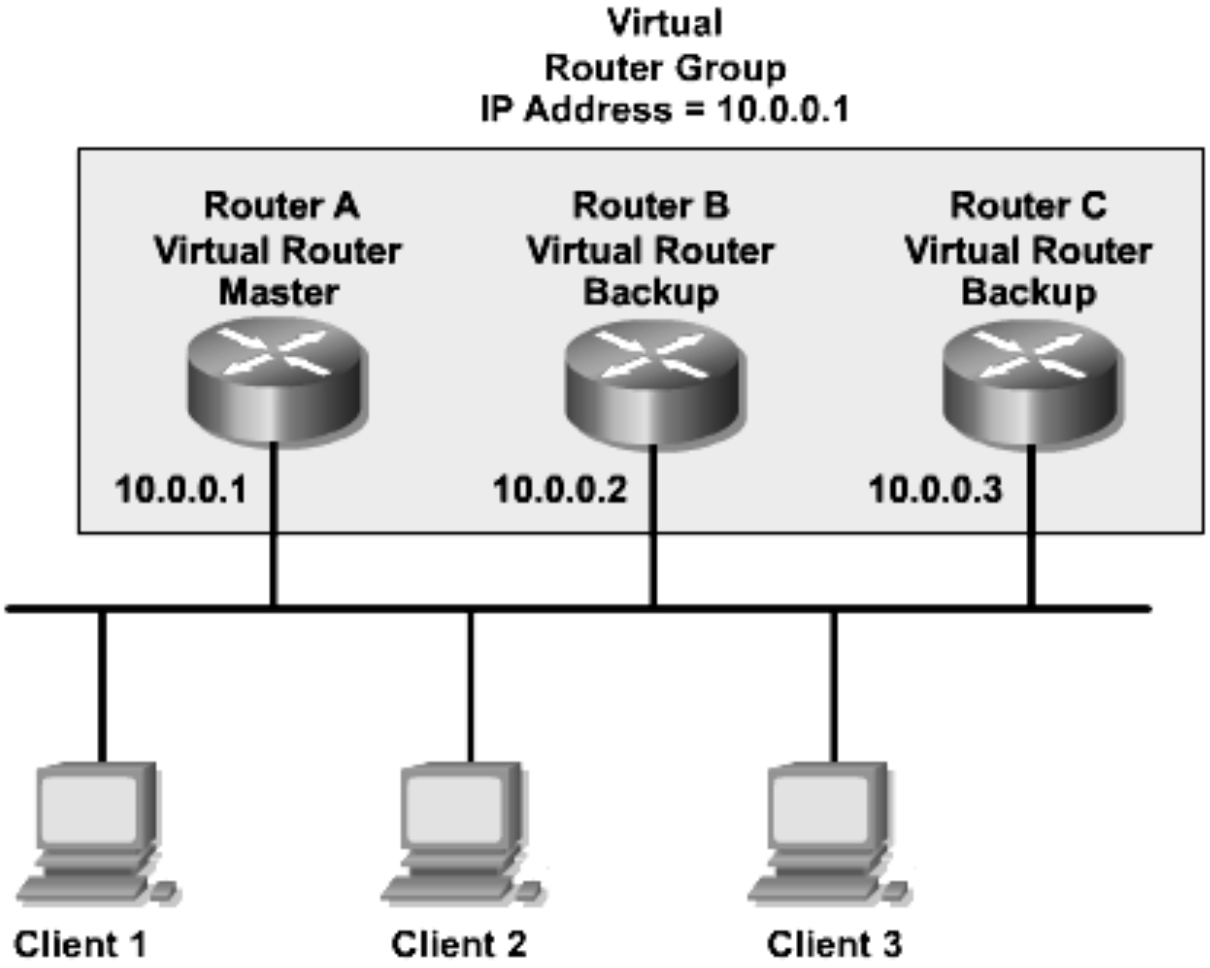
VRRP

- A group of routers function as one virtual router by sharing ONE virtual IP address and ONE virtual MAC address
- One (master) router
  - performs packet forwarding for local hosts
  - answers ARP requests for these IP address(es) associated with a virtual router
  - only one master router doing the actual routing
- The rest of the routers act as “back up” in case the master router fails
- Backup routers stay idle as far as packet forwarding from the client side is concerned
- consists of a Virtual Router Identifier (VRID) and a set of associated IP address(es) across a common LAN.

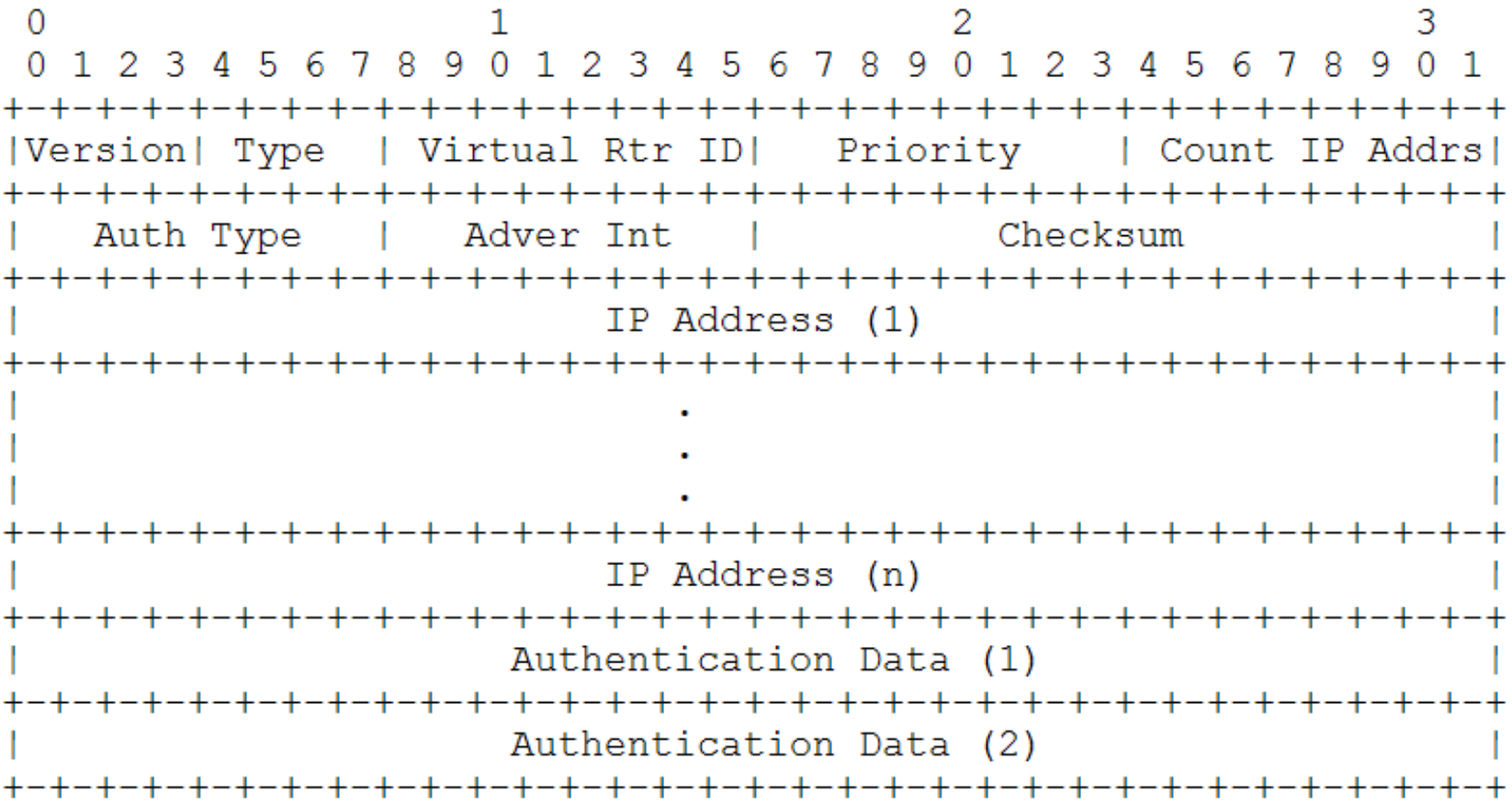
VRRP Redundancy Features

- VRRP provides redundancy for the real IP address of a router, or for a virtual IP address shared among the VRRP group members.
  - If a real IP address is used, the owning router becomes the master.
  - If a virtual IP address is used, the master is the router with the highest priority.
- A VRRP group has one master router and one or more backup routers.
- The master router uses VRRP messages to inform group members of the IP addresses of the backup routers.

VRRP Example



Relevant fields in the VRRP header



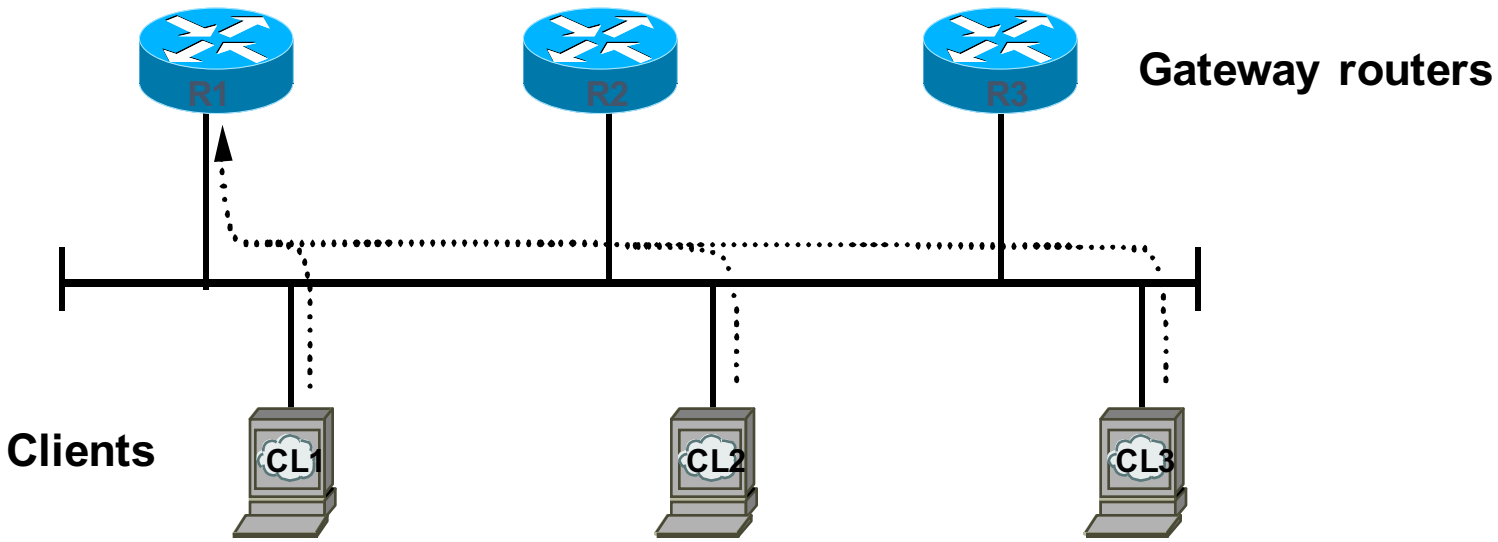
Relevant fields in the VRRP header

- Priority
  - A value between 0-255.
  - **0**: indicate the current Master has stopped participating in VRRP
  - **255**: for the VRRP router that owns the IP address(es) associated with the virtual router
    - Note that if the IP address owner is available, then it will always become the Master.
  - **1-254**: for the VRRP routers backing up a virtual router
- VRID (Virtual Router IDentifier)
  - different for each virtual router in the network
  - used by only one physical router at a time
  - in the range 1-255

First Hop Redundancy with VRRP

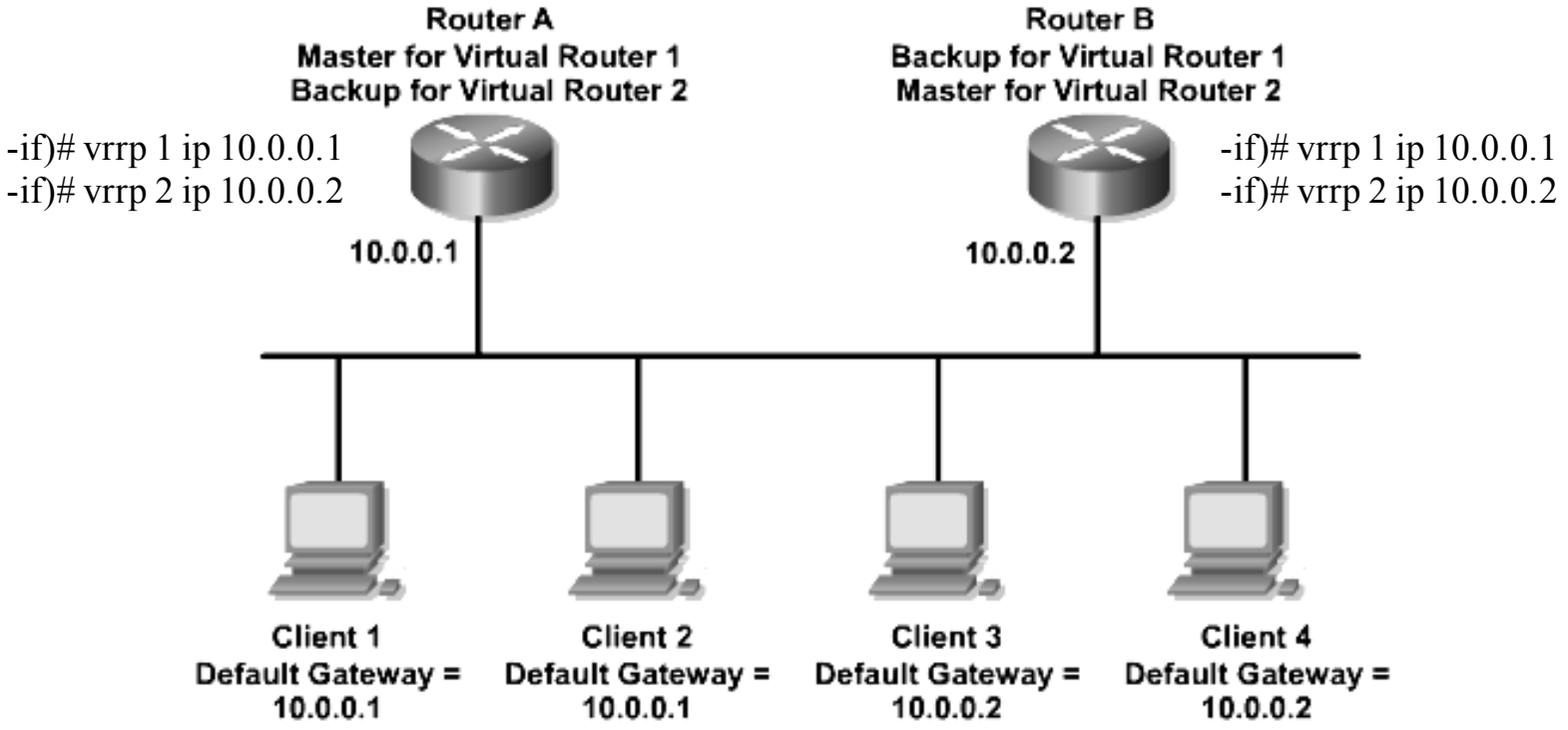
R1- Master, forwarding traffic; R2, R3 - backup

VRRP ACTIVE	VRRP BACKUP	VRRP BACKUP
IP: 10.0.0.254 MAC: 0000.0c12.3456 vIP: 10.0.0.10 vMAC: 0000.5e00.0100	IP: 10.0.0.253 MAC: 0000.0C78.9abc vIP: vMAC:	IP: 10.0.0.252 MAC: 0000.0cde.f123 vIP: vMAC:



IP: 10.0.0.1 MAC: aaaa.aaaa.aa01 GW: 10.0.0.10 ARP: 0000.5e00.0100	IP: 10.0.0.2 MAC: aaaa.aaaa.aa02 GW: 10.0.0.10 ARP: 0000.5e00.0100	IP: 10.0.0.3 MAC: aaaa.aaaa.aa03 GW: 10.0.0.10 ARP: 0000.5e00.0100

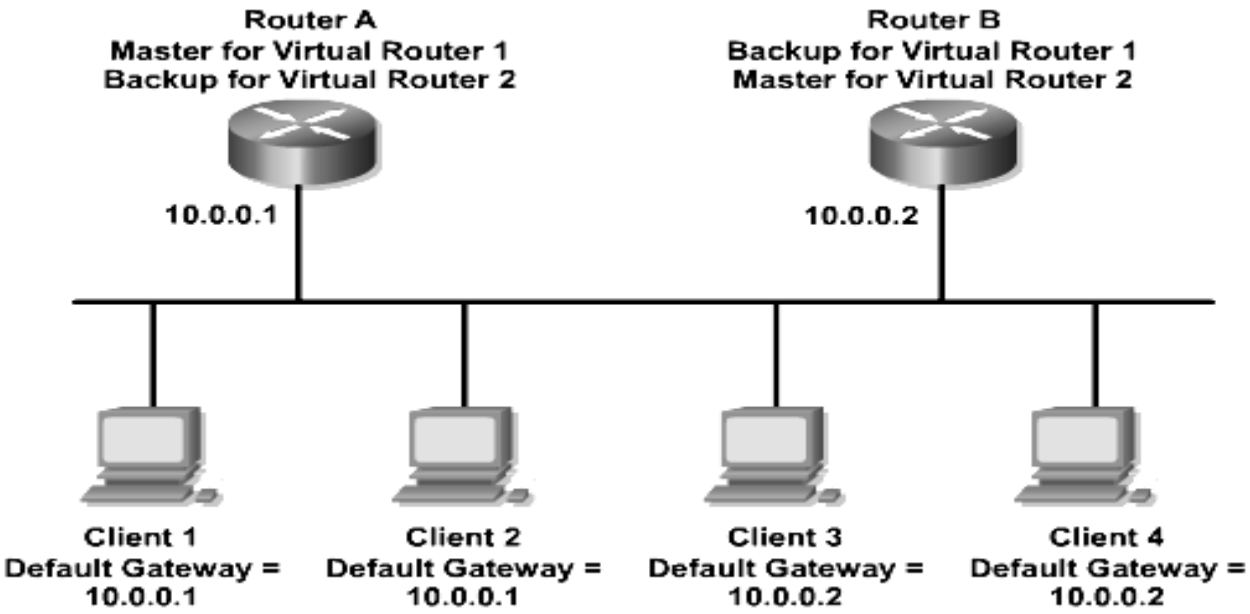
The VRRP Operation Process



Two virtual routers groups are configured: Virtual Router 1 and 2.
Virtual Router 1 is 10.0.0.1 and Virtual Router 2 is 10.0.0.2.

Since each router owns one of these IPs it will be the Master Router for that group and the other router set with the same IP will be the backup.

The VRRP Operation Process



The priority of the Master Router is set to 255.
Backup router priority values can range from 1 to 254; the default value is 100. The VRRP MAC address is 0000.5e00.01xx.

The master sends the advertisement on multicast 224.0.0.18 on a default interval of 1 second (advertisement interval).
The master-down interval is the time interval for backup to declare the master down (seconds).

VRRP Configuration

Master Router:

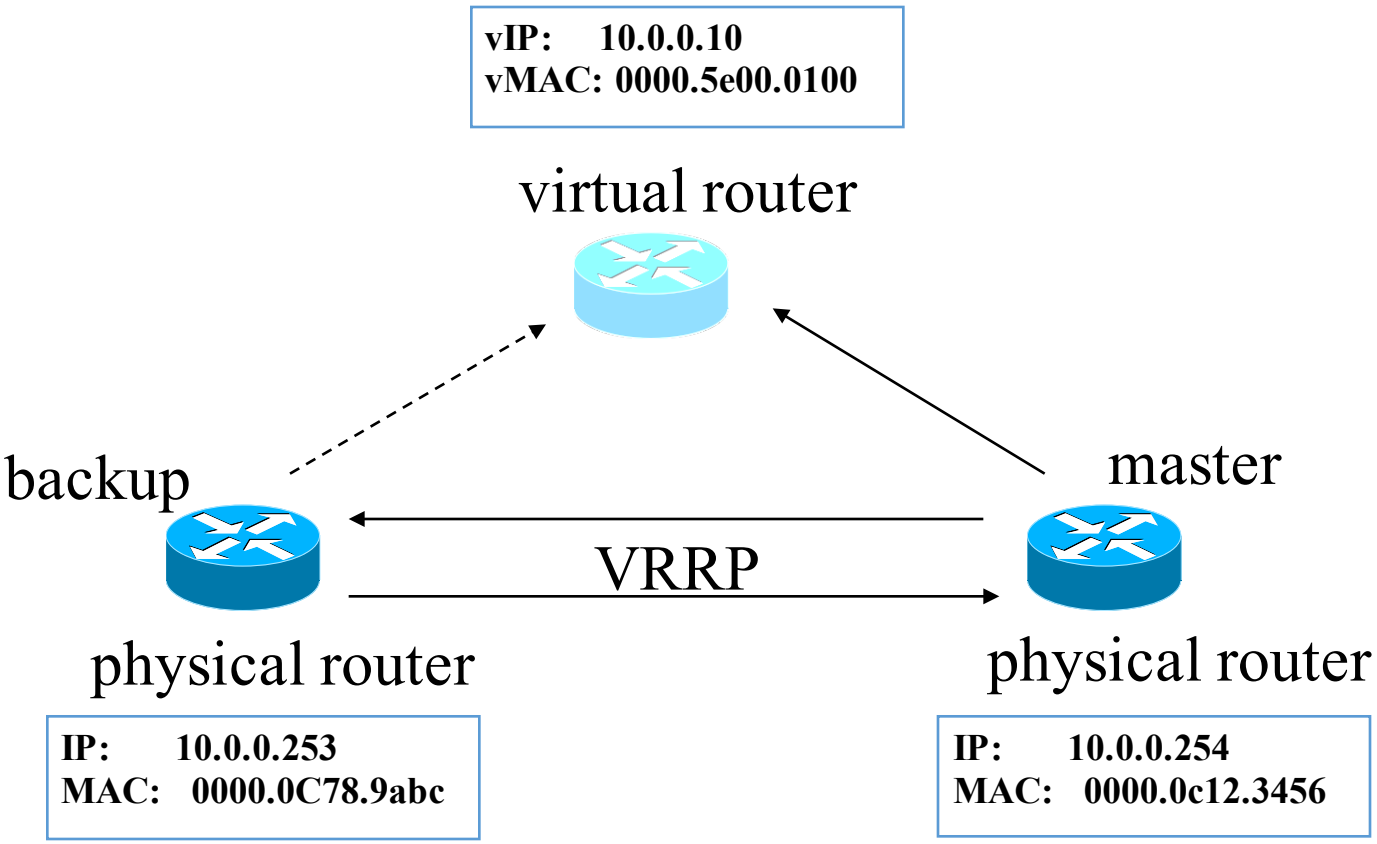
```
interface FastEthernet0/1
 ip address 10.1.2.3 255.255.255.0
 duplex auto
 speed auto
 vrrp 10 ip 10.1.2.3
```

Even with the higher priority, the router below is still the backup because the route above is using its own IP address.

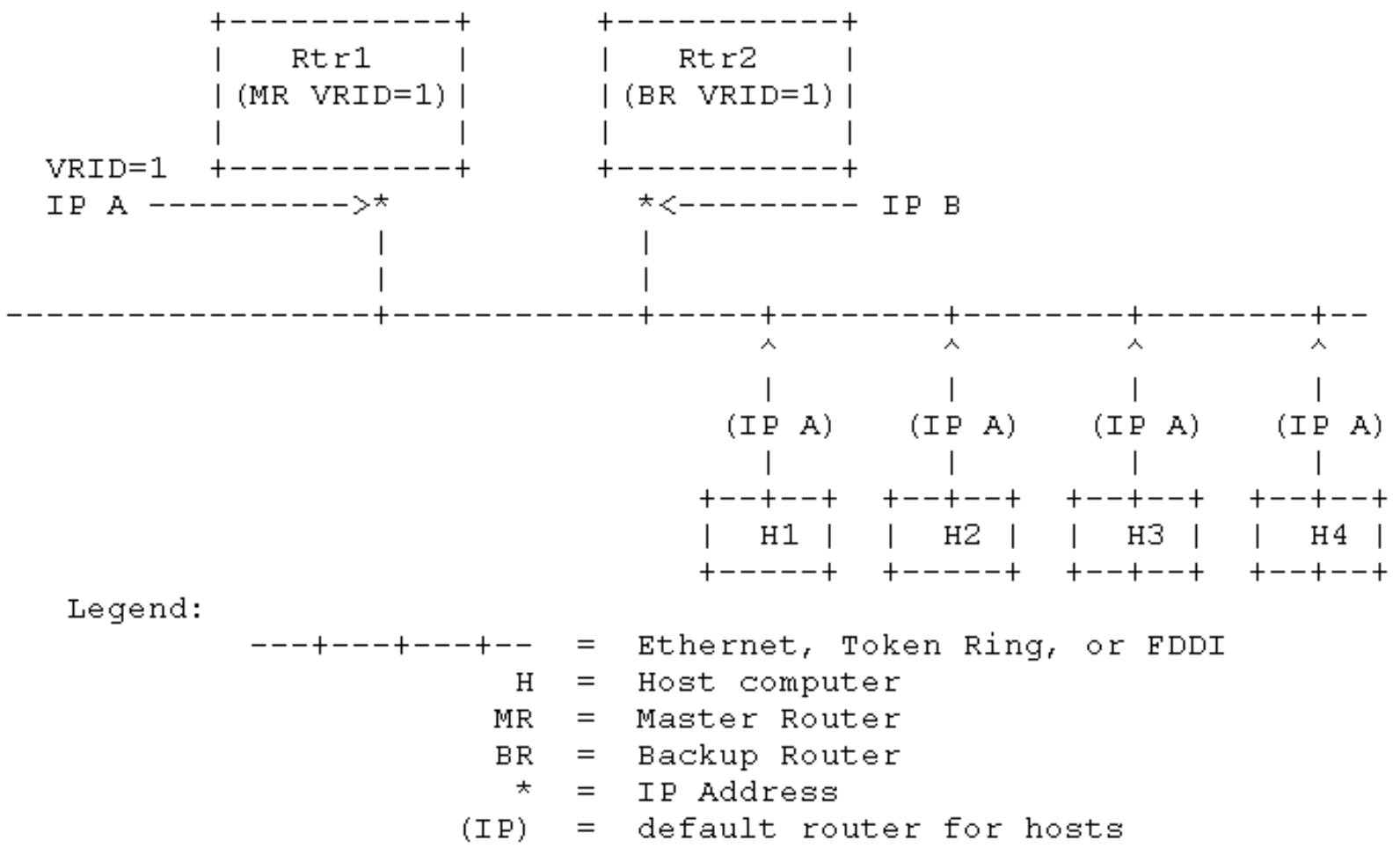
Backup Router:

```
interface FastEthernet0/1
 ip address 10.1.2.2 255.255.255.0
 duplex auto
 speed auto
 vrrp 10 ip 10.1.2.3
 vrrp 10 priority 150
```

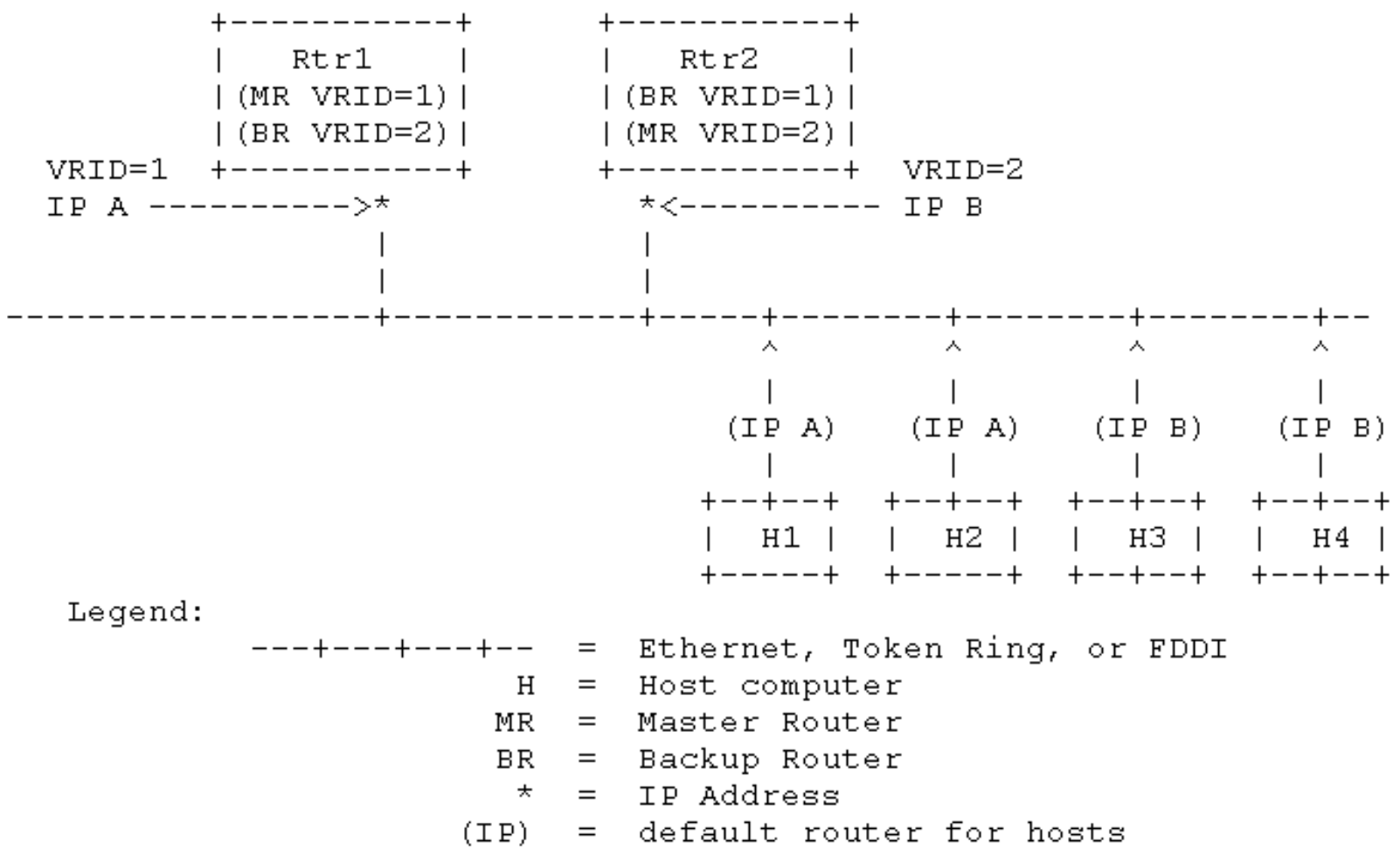
- MAC address: 00-00-5E-00-01-[VRID]
  - This address is used by only one physical router at a time, and it will reply with this MAC address when an ARP request is sent for the virtual router's IP address.
- Physical routers within the virtual router must communicate within themselves using packets with
  - multicast IP address 224.0.0.18
  - and IP protocol number 112.


vIP: 10.0.0.10
vMAC: 0000.5e00.0100
virtual router
backup
master
VRRP
physical router
physical router
IP: 10.0.0.253
MAC: 0000.0C78.9abc
IP: 10.0.0.254
MAC: 0000.0c12.3456

Sample Configuration 1



Sample Configuration 2



Elections of master routers

- Master router sends an advertisement to the backups.
  - Advertisement intervals can be set by the user; the VRRP default is 1 second.
- If the advertisements suddenly stop, the backups set interval timers, typically for three times the advertisement frequency.
- If no further advertisements appear, the backups assume the master is down and the failover routine is activated.
  - From that point, the election of the next-in-line master typically takes less than a second.

Policy-Based Routing

- Make routing decisions based on policies set by the network administrator
- Forwarding decision not based on destination address
- Select next-hop based on attributes of user packet
  - Source/destination address
  - Application ports
  - Packet length
  - Or other information available in a packet header or payload

Policy-Based Routing – Action Example

- set ip next-hop ip-address1 […]
- set ip default next-hop ip-address1 […]
- set interface interface1 […]
- set ip precedence value
- set ip tos value
- set vrf value

Policy-Based Routing Example – Cisco

```
interface Ethernet0
   ip address 192.168.93.10 255.255.255.0
   ip policy route-map foo
interface Serial1
   ip address 11.0.0.2 255.0.0.0
interface BRI0
   ip address 10.0.0.2 255.0.0.0
route-map foopermit 10
   match ip address 101
   set ip next-hop 11.0.0.1
route-map foopermit 11
   match ip address 103
   set ip next-hop 10.0.0.1
route-map foopermit 12
   set default interface Null0
access-list 101 permit tcp 192.168.93.0 0.0.0.255 any eq telnet
access-list 101 permit icmp any any
access-list 103 permit tcp 192.168.93.0 0.0.0.255 any eq ftp
```

Policy-Based Routing Example -- Linux

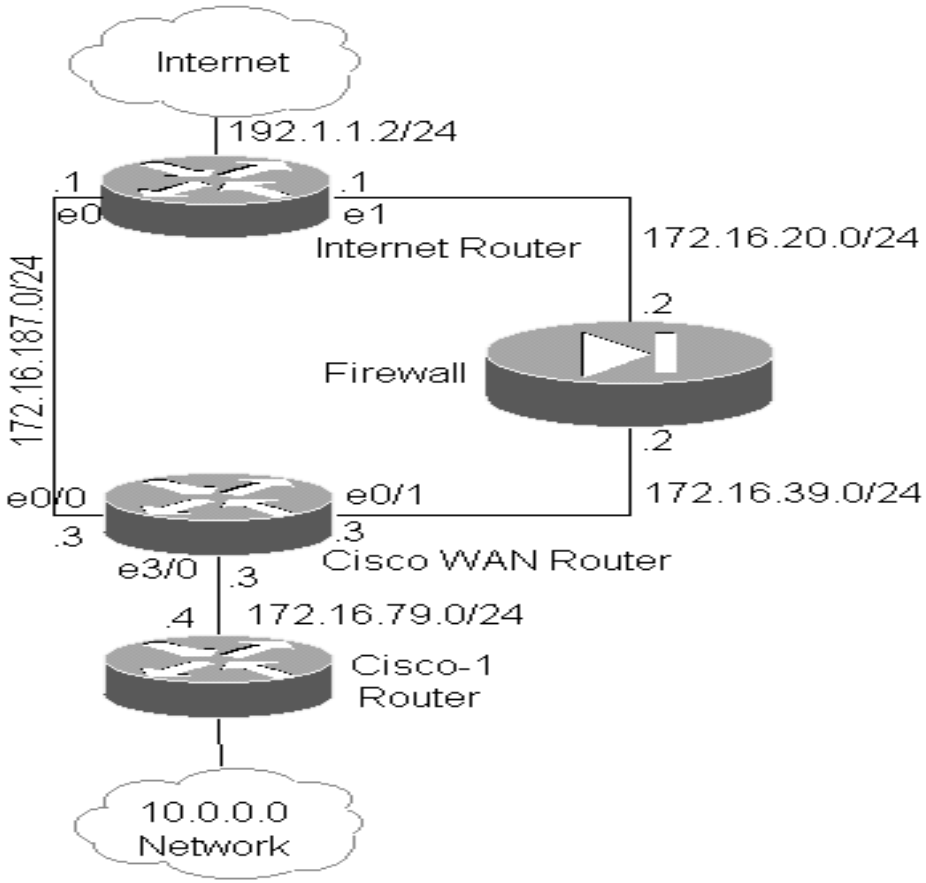
$ ip route list table main

```
195.96.98.253 dev ppp2  proto kernel  scope link  src 212.64.78.148
212.64.94.1 dev ppp0  proto kernel  scope link  src 212.64.94.251
10.0.0.0/8 dev eth0  proto kernel  scope link  src 10.0.0.1
127.0.0.0/8 dev lo  scope link
default via 212.64.94.1 dev ppp0
```

- echo 200 John >> /etc/iproute2/rt_tables
- ip rule add from 10.0.0.10 table John
- ip rule ls

```
0:   from all lookup  local
32765:       from 10.0.0.10 lookup John
32766:       from all lookup  main
32767:       from all lookup  default
```

- ip route add default via 195.96.98.253 dev ppp2 table John


Internet
192.1.1.2/24
.1
e0
.1
e1
Internet Router
172.16.20.0/24
.2
Firewall
.2
172.16.187.0/24
e0/0
.3
e0/1
172.16.39.0/24
.3
Cisco WAN Router
e3/0
.3
.4
172.16.79.0/24
Cisco-1
Router
10.0.0.0
Network

Virtual Routing and Forwarding (VRF)

- VRFs provide a valuable routing tool to provide isolation between different networks that are sharing the same network infrastructure.

- Virtual routing and forwarding (VRF) is a technology included in IP (Internet Protocol) network routers that allows multiple instances of a routing table to exist in a router and work simultaneously.