Flow Control and Error Control

Flow and Error Control

- The most important responsibilities of the data link layer are:
  - Flow control and
  - Error control
- Collectively, these functions are known as data link control

Flow Control

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

- Receiver has a limited speed at which it can process incoming data and a limited amount of memory in which to store incoming data.

- Receiver must inform the sender before the limits are reached and request the transmitter to send fewer frames or stop temporarily

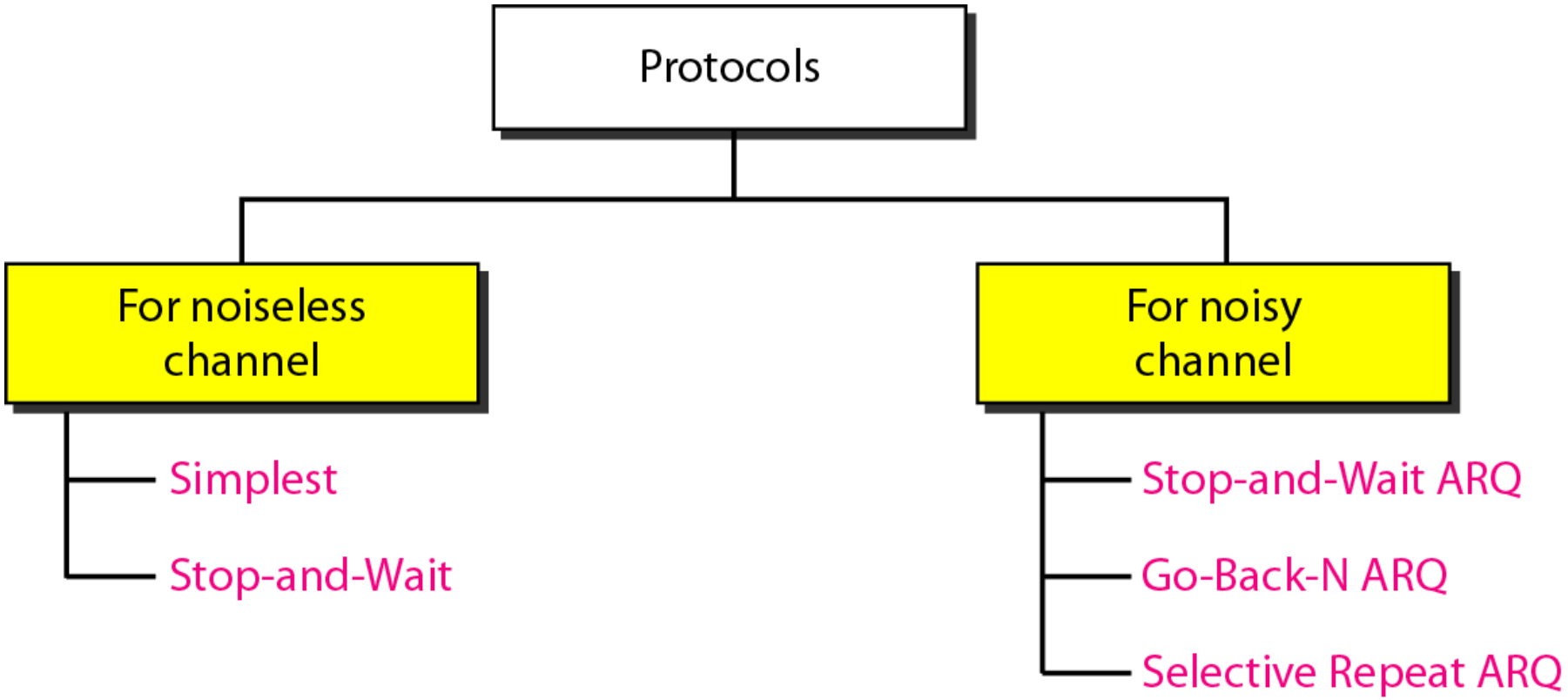
- Since the rate of processing is often slower than the rate of transmission, receiver has a block of memory (buffer) for storing incoming data until they are processed.

Error Control

- Error control includes both error detection and error correction.
- It allows the receiver to inform the sender if a frame is lost or damaged during transmission and coordinates the retransmission of those frames by the sender.
- Error control in the data link layer is based on automatic repeat request (ARQ). Whenever an error is detected, specified frames are retransmitted.

Flow and Error Control Mechanisms

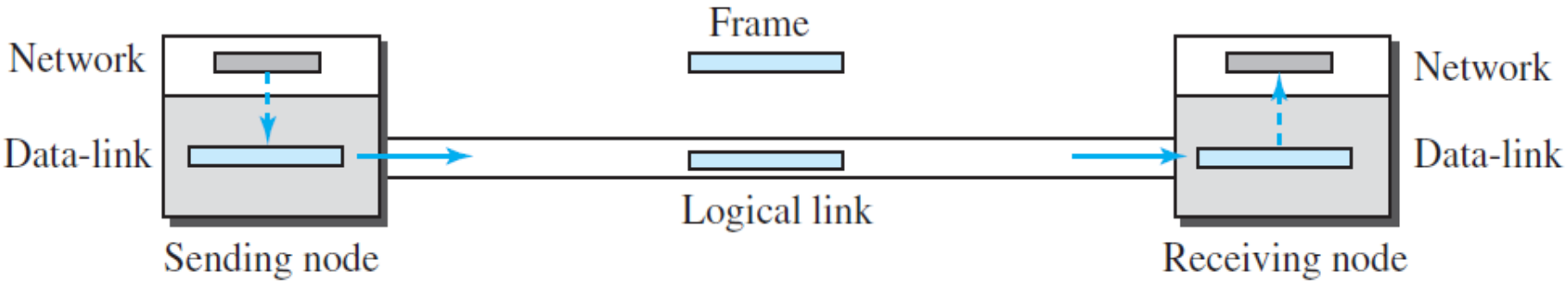
- Data link layer can combine framing, flow control, and error control to achieve the delivery of data from one node to another
- The protocols are normally implemented in software



Simple Protocol

- Simple (or Simplest or Simplex) protocol require neither flow nor error control
- Here, we assume that the receiver can immediately handle any frame it receives
- In other words, the receiver can never be overwhelmed with incoming frames

Simple protocol with no flow or error control

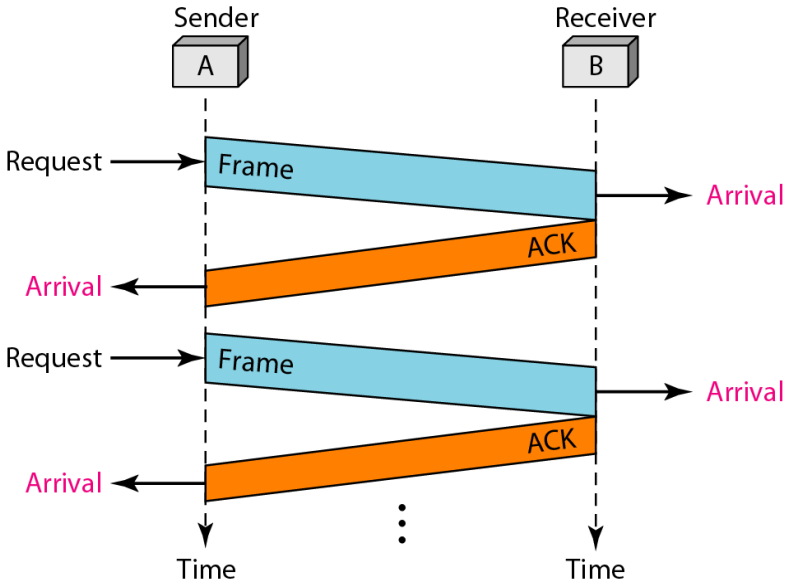


Noiseless Channels

- Let us first assume that we have an ideal channel in which no frames are lost, duplicated, or corrupted.
- Two protocols have been designed for this type of channel
  - Stop-and-Wait
  - Sliding-Window

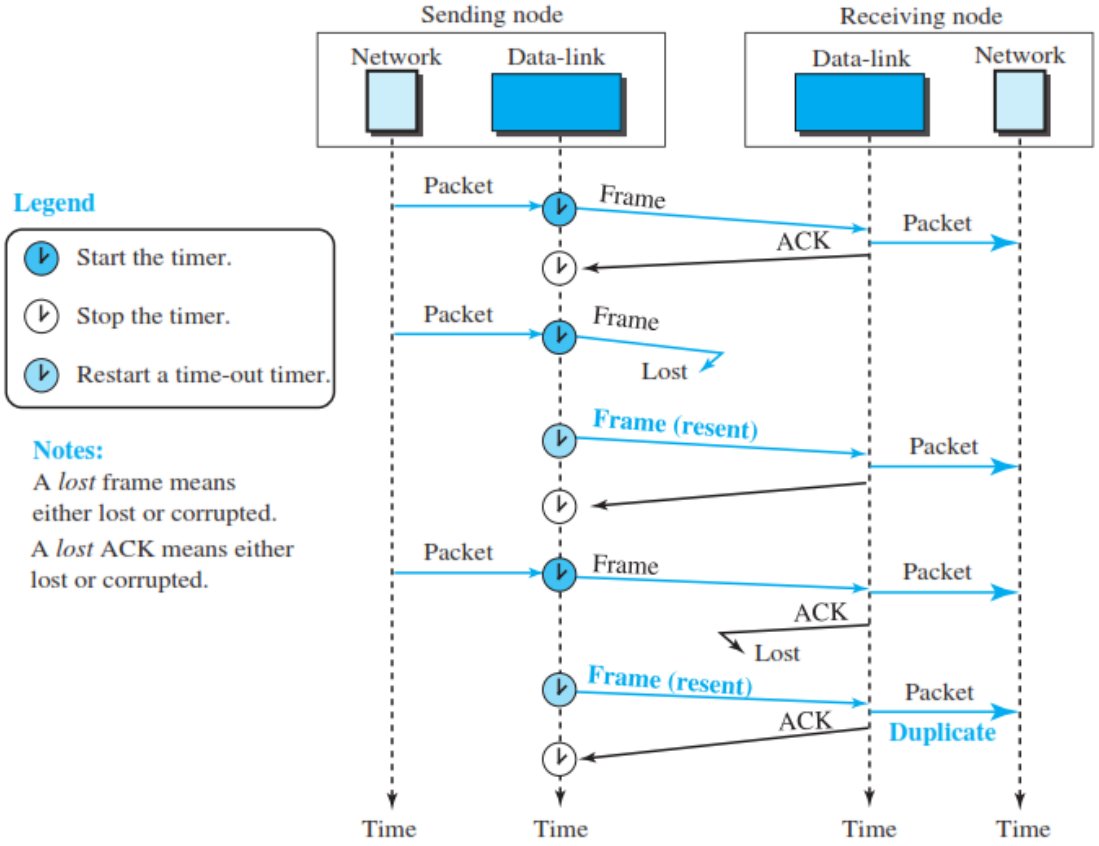
Stop-and-Wait Protocol

- The simplest form of flow control, known as stop-and-wait flow control
- It is also known as request/reply sometimes
- Here, the sender sends one frame, stops until it receives confirmation from the receiver, and then sends the next frame
- The destination can stop the flow of data simply by withholding acknowledgment



NOISY CHANNELS

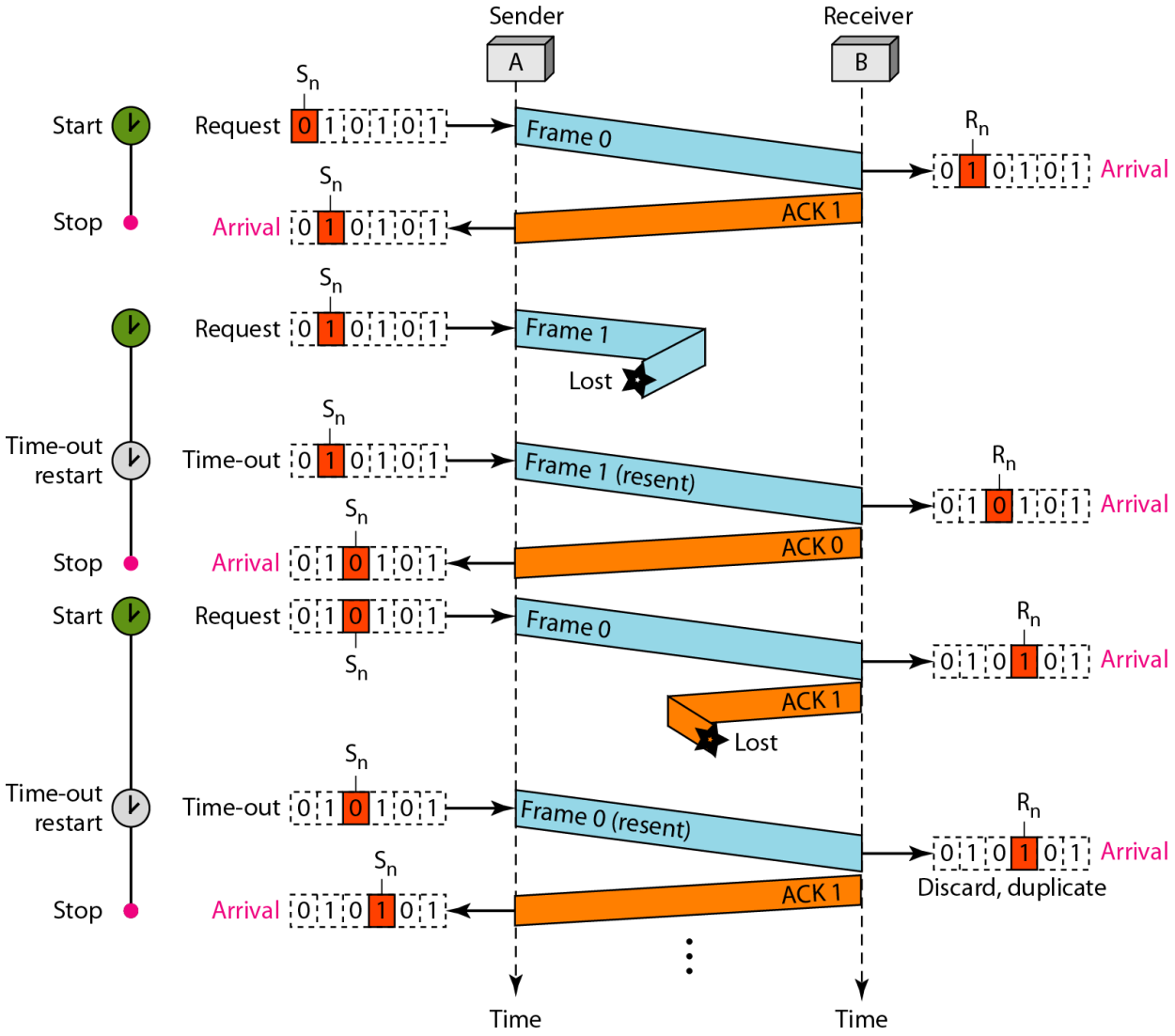
- Although the Stop-and-Wait Protocol gives us an idea of how to add flow control
- noiseless channels are nonexistent.



Stop-and-Wait ARQ

- Error correction in Stop-and-Wait ARQ is done by keeping a copy of the sent frame and retransmitting of the frame when required
- These types of errors could occur in Stop-and-Wait:
  - The frame that arrives at the destination could be corrupted.
  - The frame is lost.
  - Damaged acknowledgment.
  - Delayed acknowledgment.
- Sender uses a timer after transmitting each frame.
- Error correction is done by keeping a copy of the sent frame and retransmitting of the frame when the timer expires.
- Uses sequence numbers based on modulo-2 arithmetic to number the frames.
- The acknowledgment number always announces the sequence number of the next frame expected.

Stop-and-Wait ARQ



Stop-and-Wait Protocol

- **Link Utilization in Stop-and-Wait:**

  The link utilization U = 1/(1+2a),

  where a = Propagation time / Transmission time

- It is evident from the above equation that:
  - When the propagation time is small, as in case of LAN environment, the link utilization is good
  - But, in case of long propagation delays, as in case of satellite communication, the utilization can be very poor
- To improve the link utilization, we can use the sliding-window protocol instead of using stop-and-wait protocol

Example

Assume that, in a Stop-and-Wait ARQ system, the bandwidth of the line is 1 Mbps, and 1 bit takes 20 ms to make a round trip. What is the bandwidth-delay product? If the system data frames are 1000 bits in length, what is the utilization percentage of the link?

Solution

The bandwidth-delay product is

$$(1 \times 10^6) \times (20 \times 10^{-3}) = 20{,}000 \text{ bits}$$

Example

The system can send 20,000 bits during the time it takes for the data to go from the sender to the receiver and then back again.

However, the system sends only 1000 bits.

We can say that the link utilization is only 1000/20,000, or 5 percent.

Stop-and-Wait Protocol

Advantages:

- It is simple
- Each frame is checked and acknowledged well

Disadvantages:

- Only one frame can be in transmission at a time
- It is inefficient, if the distance between devices is long

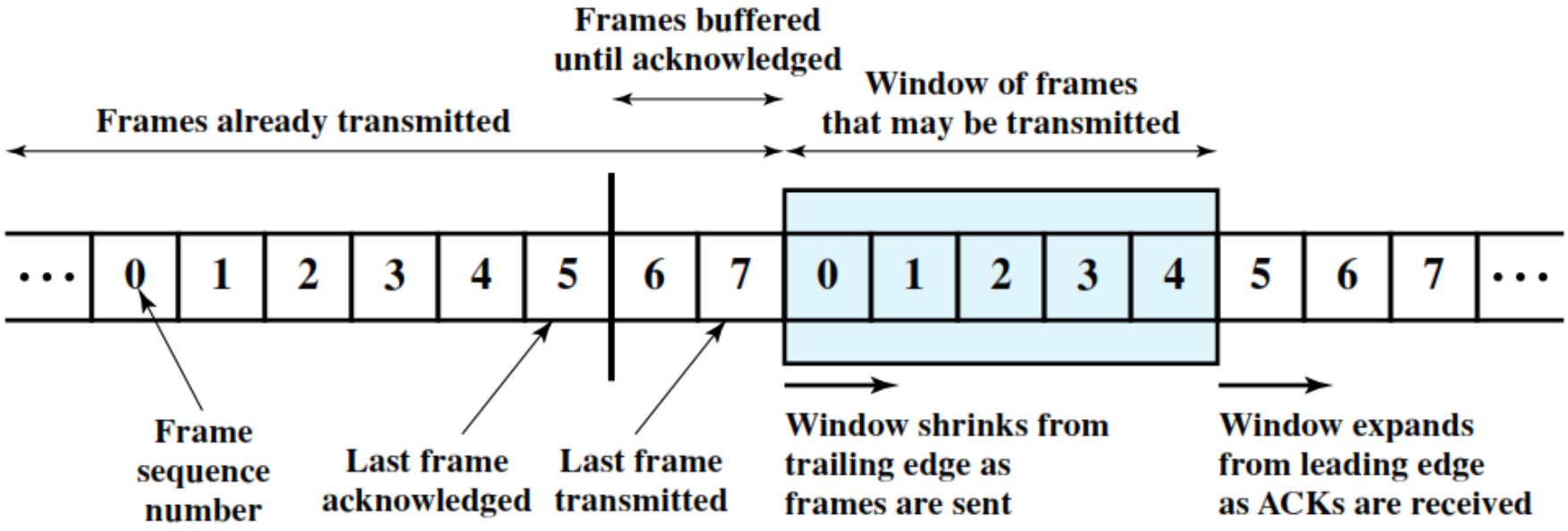
Sliding Window

- In sliding window protocol multiple frames are allowed to sent without waiting for any acknowledgments.
- Receiver have a buffer to store multiple frames.
- To keep track of which frames have been acknowledged, each is labeled with a sequence number.
- Receiver acknowledges a frame by sending an acknowledgment that includes the sequence number of the next frame expected.
- This scheme can also be used to acknowledge multiple frames.
- It uses two type of acknowledgement
  - Receive Ready (RR)
    - I have received all frames up to frame number **n** and **am ready** to receive frame number **n+1**
  - Receive Not Ready (RNR)
    - I have received all frames up to frame number **n** and **am not ready** to receive frame number **n+1**

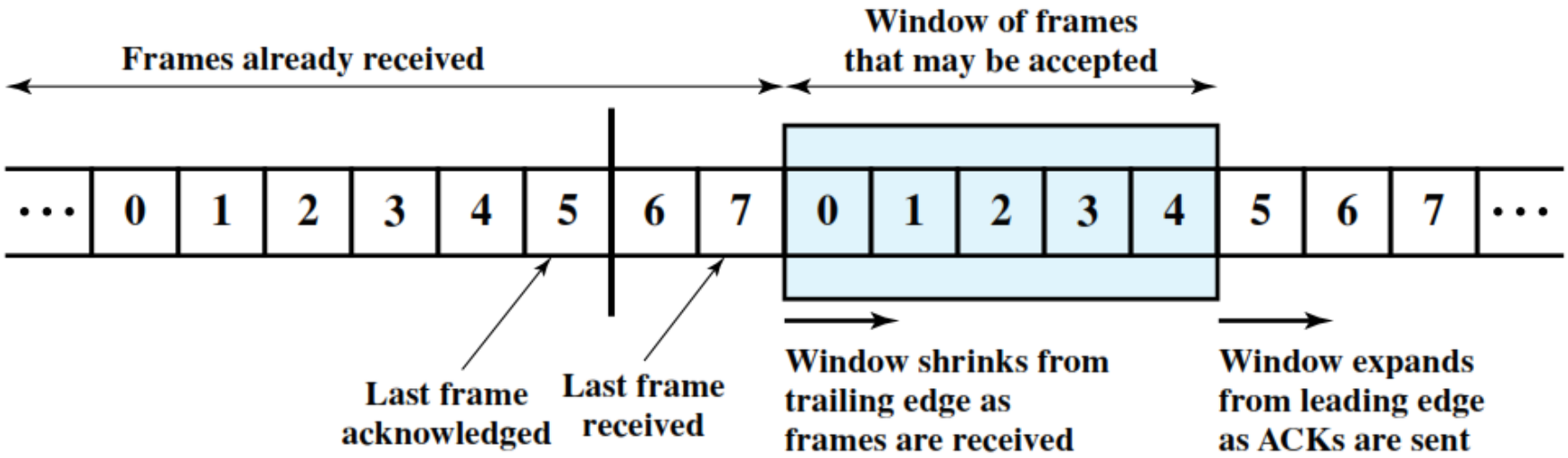
Sliding Window

- Sender maintains a list of sequence numbers that it is allowed to send.
- Receiver maintains a list of sequence numbers that it is prepared to receive.
- Each of these lists can be thought of as a window of frames. The operation is referred to as **sliding-window flow control**
- Since the sequence number to be used occupies a field in the frame, it should be of limited size
- For a *k*-bit field the range of sequence numbers is **0 through $2^k - 1$**, and frames are numbered **modulo 2^k**
- The window size need not be the maximum possible size for a given sequence number length

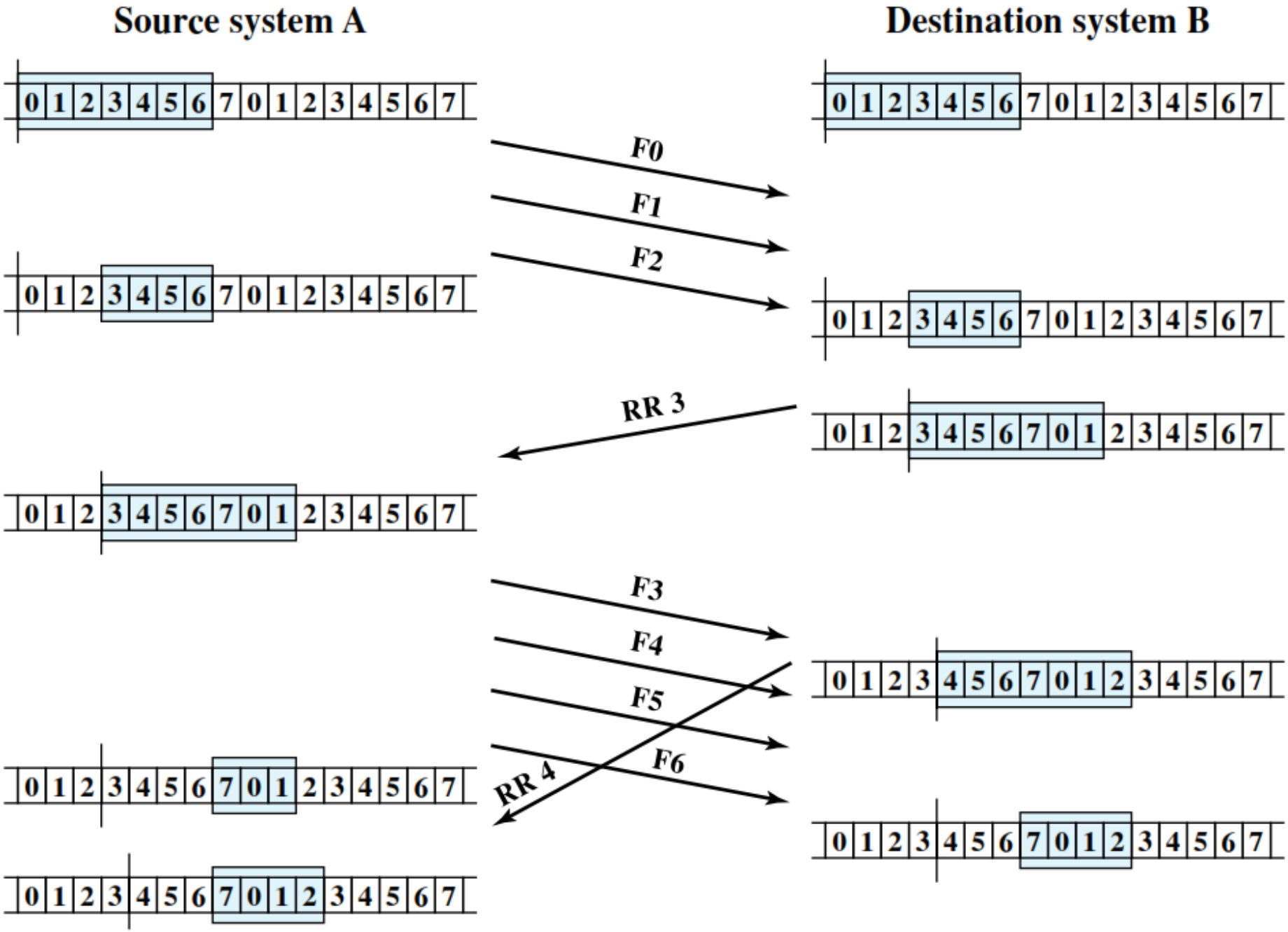
Sliding Window



(a) Sender's perspective



(b) Receiver's perspective


Source system A
Destination system B
0 1 2 3 4 5 6 7 0 1 2 3 4 5 6 7
F0
F1
F2
0 1 2 3 4 5 6 7 0 1 2 3 4 5 6 7
0 1 2 3 4 5 6 7 0 1 2 3 4 5 6 7
RR 3
0 1 2 3 4 5 6 7 0 1 2 3 4 5 6 7
F3
F4
F5
F6
RR 4
0 1 2 3 4 5 6 7 0 1 2 3 4 5 6 7
0 1 2 3 4 5 6 7 0 1 2 3 4 5 6 7
0 1 2 3 4 5 6 7 0 1 2 3 4 5 6 7

Sliding Window

- If two stations exchange data, each needs to maintain two windows, one for transmit and one for receive, and each side needs to send the data and acknowledgments to the other
- To provide efficient support for this requirement, a feature known as **piggybacking** is typically provided

Piggybacking:

- In two-way communication, when a frame is received, the receiver waits and does not send the ACK back to the sender immediately
- The receiver waits until its network layer passes the next data packet to it. The delayed acknowledgement is then attached to this outgoing data frame
- This technique of temporarily delaying the acknowledgement so that it can be hooked with next outgoing data frame is known as piggybacking.

Sliding Window

- The link utilization in case of Sliding Window Protocol:

$$U = \begin{cases} 1 & W \geq 2a + 1 \\ \dfrac{W}{2a + 1} & W < 2a + 1 \end{cases}$$

where W = the window size,
and a = Propagation time / Transmission time

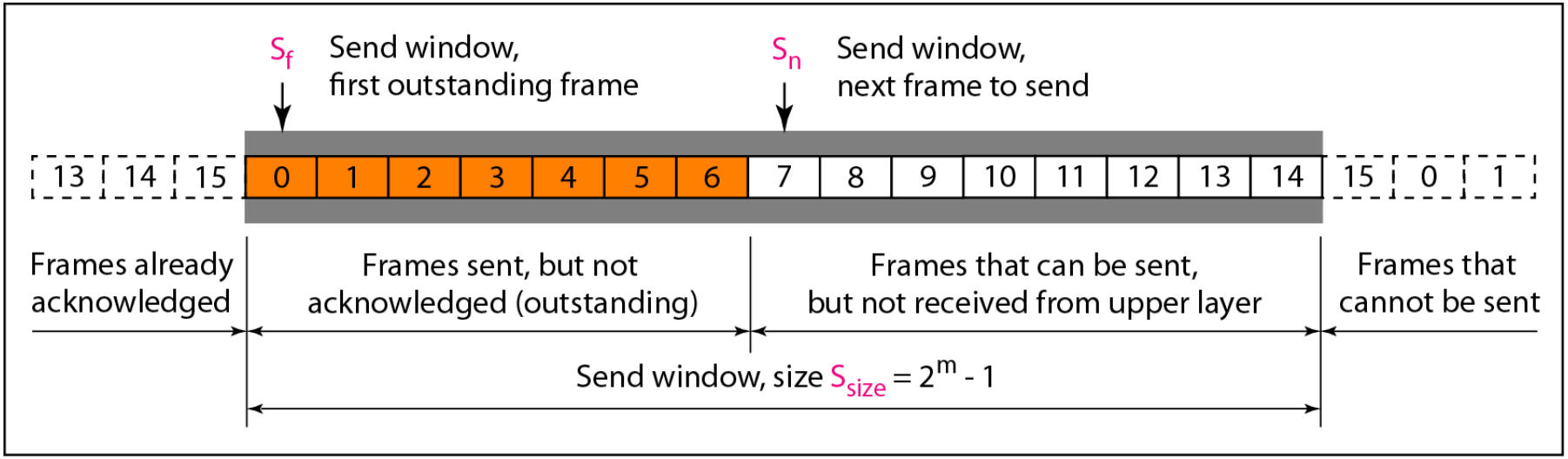
Go-back-N ARQ

- The form of error control based on sliding-window flow control.
- The size of the send window must be less than 2^k
- The size of the receiver window is always 1.
- The send window can slide one or more slots when a valid acknowledgment arrives.
- The receiver window slides when a correct frame has arrived; sliding occurs one slot at a time.
- While no errors occur, the destination will acknowledge incoming frames as usual.
- Receiver sends a negative acknowledge (NACK (REJ = reject)), in case a frame is incorrectly received.

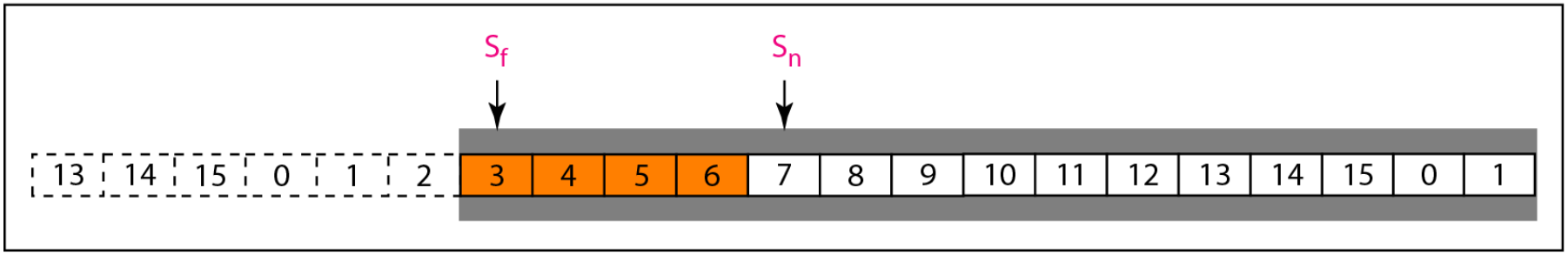
Go-back-N ARQ

- When the sender receives a NACK (NAK or REJ = reject), it retransmits the frame in error plus all the succeeding frames.
- Damaged (or lost) frame: If a frame is lost, the receiver sends NAK after receiving the next frame.
- In case there is long delay before sending the NAK, the sender will resend the lost frame after its timer times out
- Damaged (or lost) ACK (i.e. RR): If the ACK frame sent by the receiver is lost, the sender resends the frames after its timer times out.
- Damaged NACK (NAK or REJ): If a REJ is lost, the sender resends the frames after its timer times out.

Go-back-N ARQ

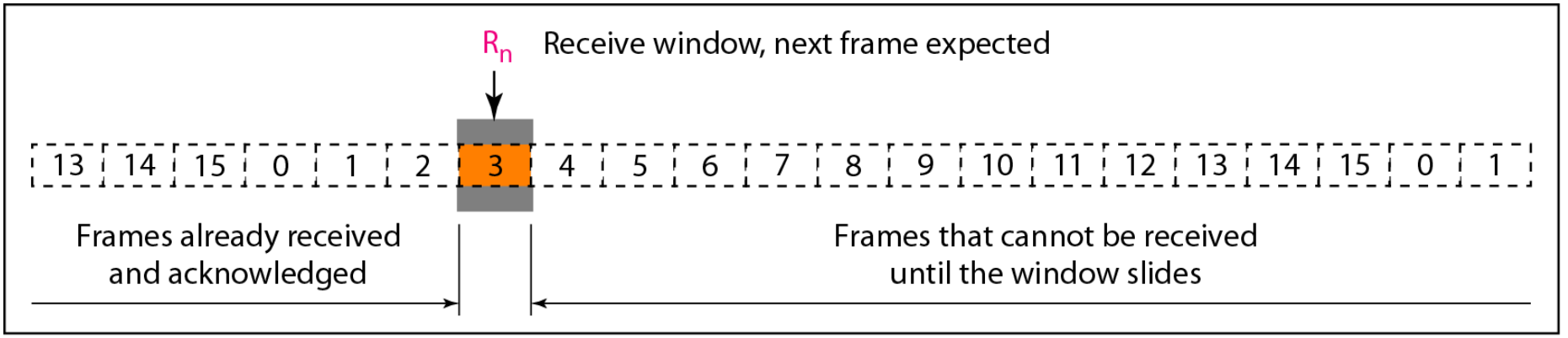


a. Send window before sliding

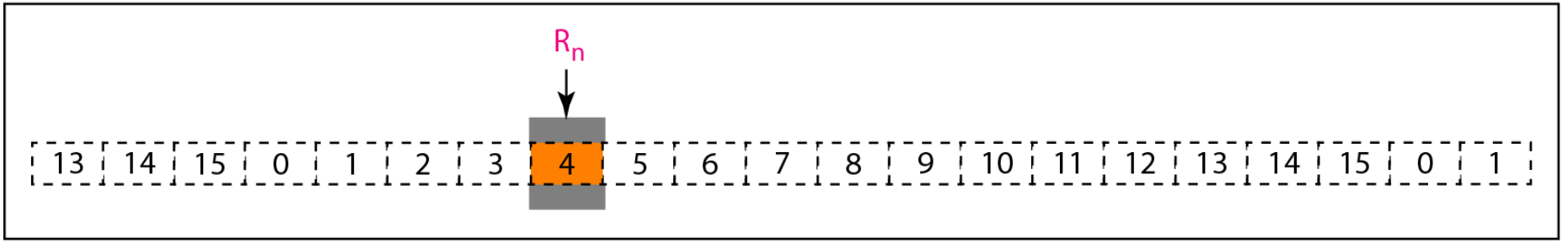


b. Send window after sliding

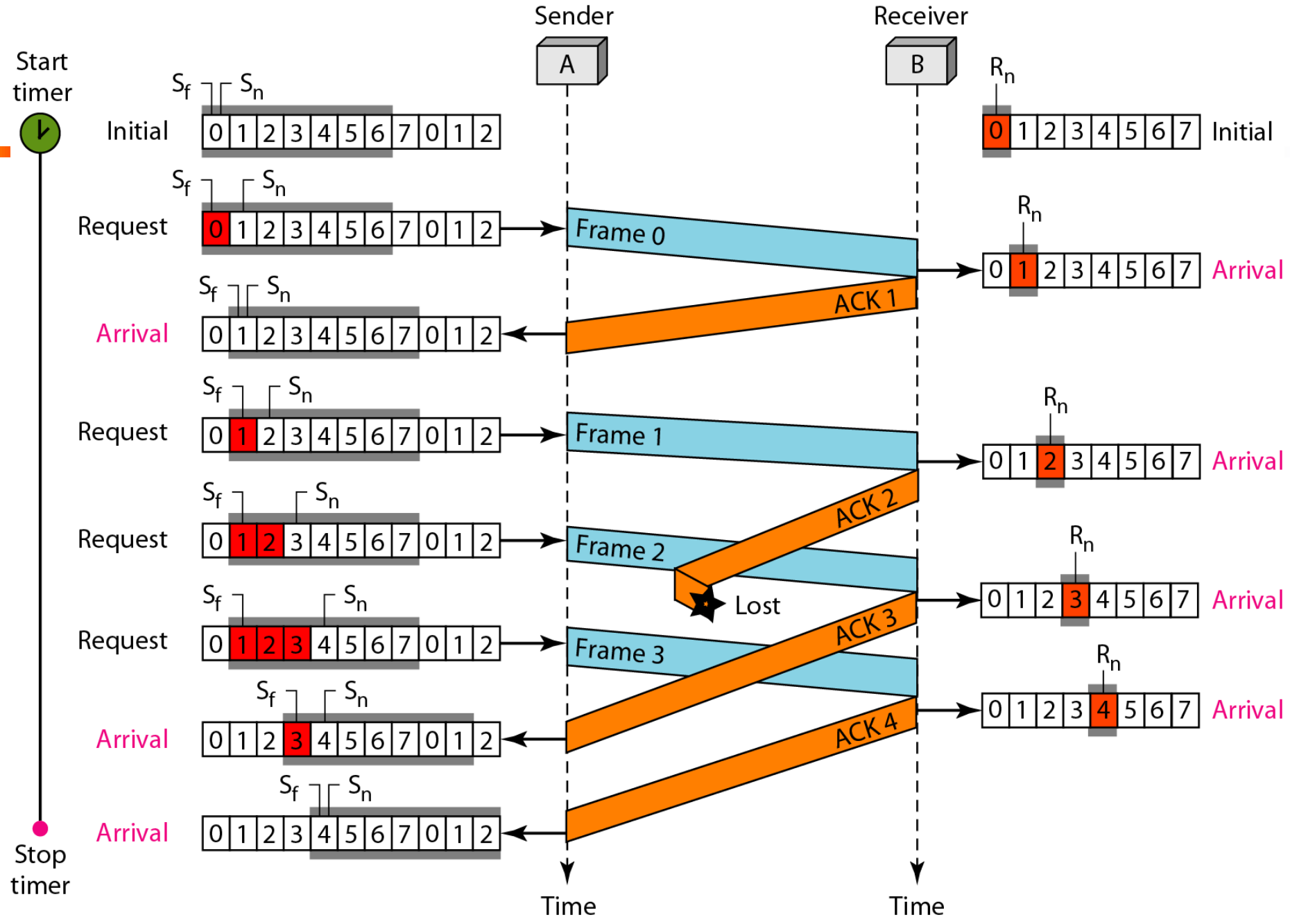
Go-back-N ARQ

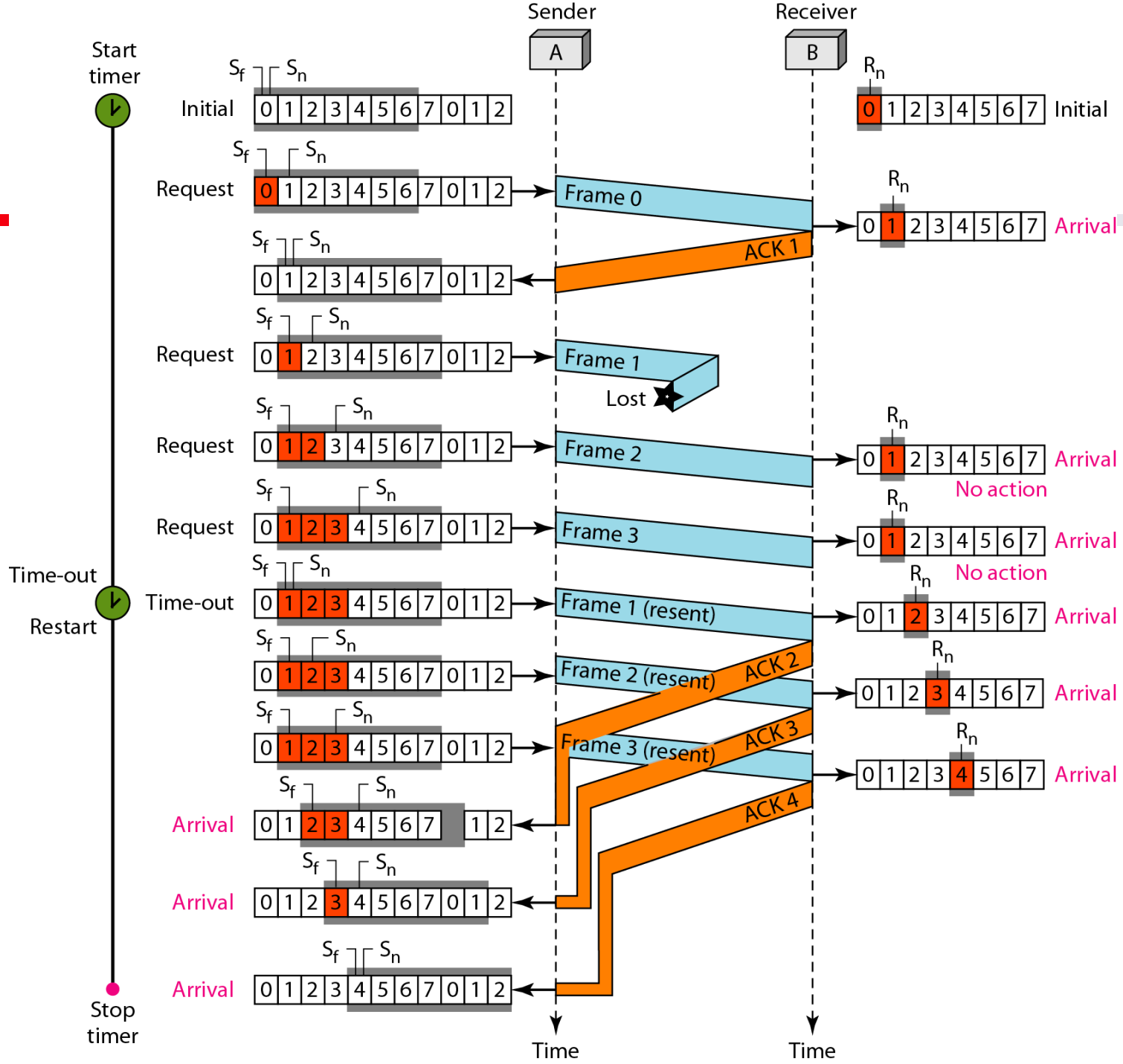


a. Receive window

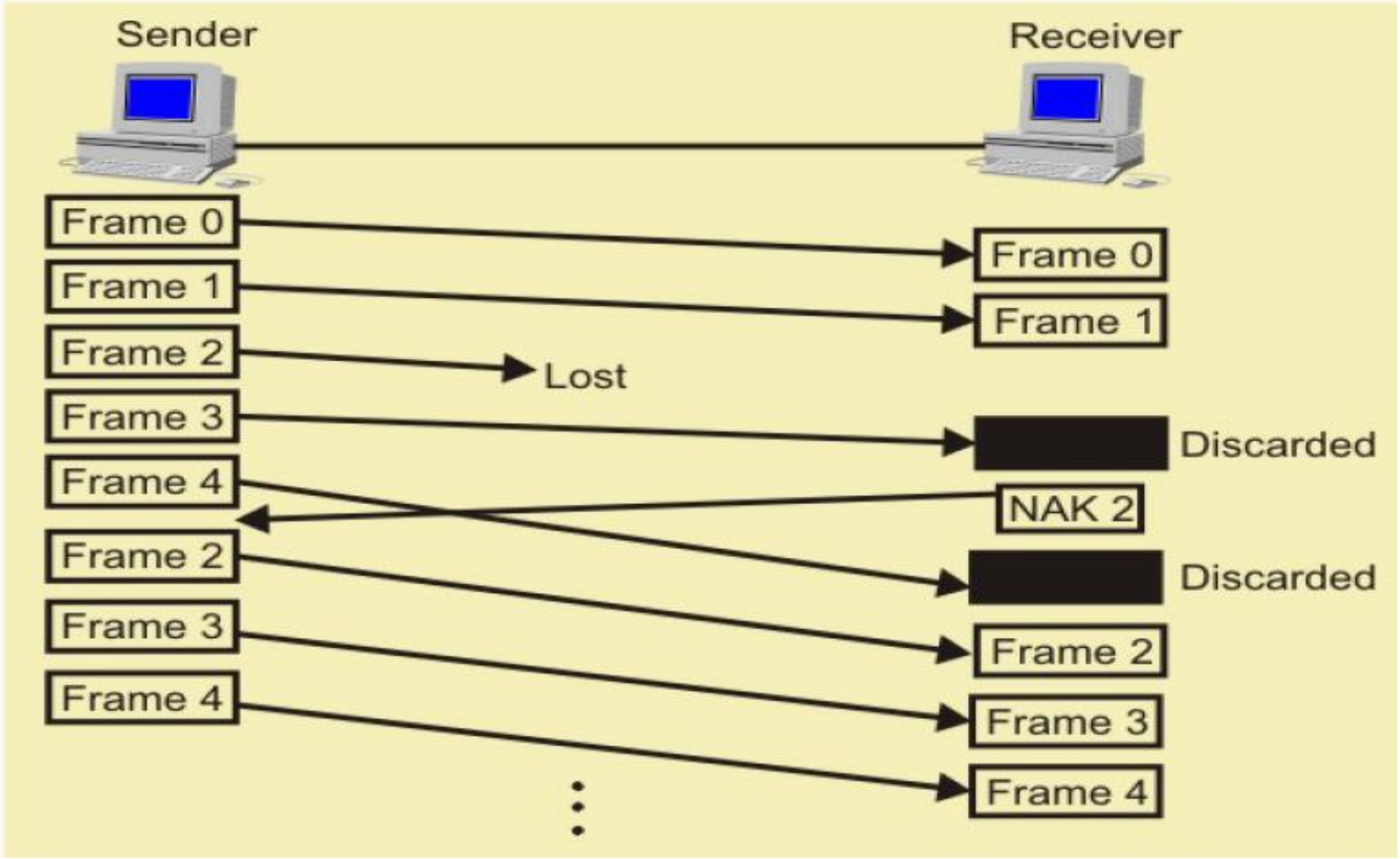


b. Window after sliding

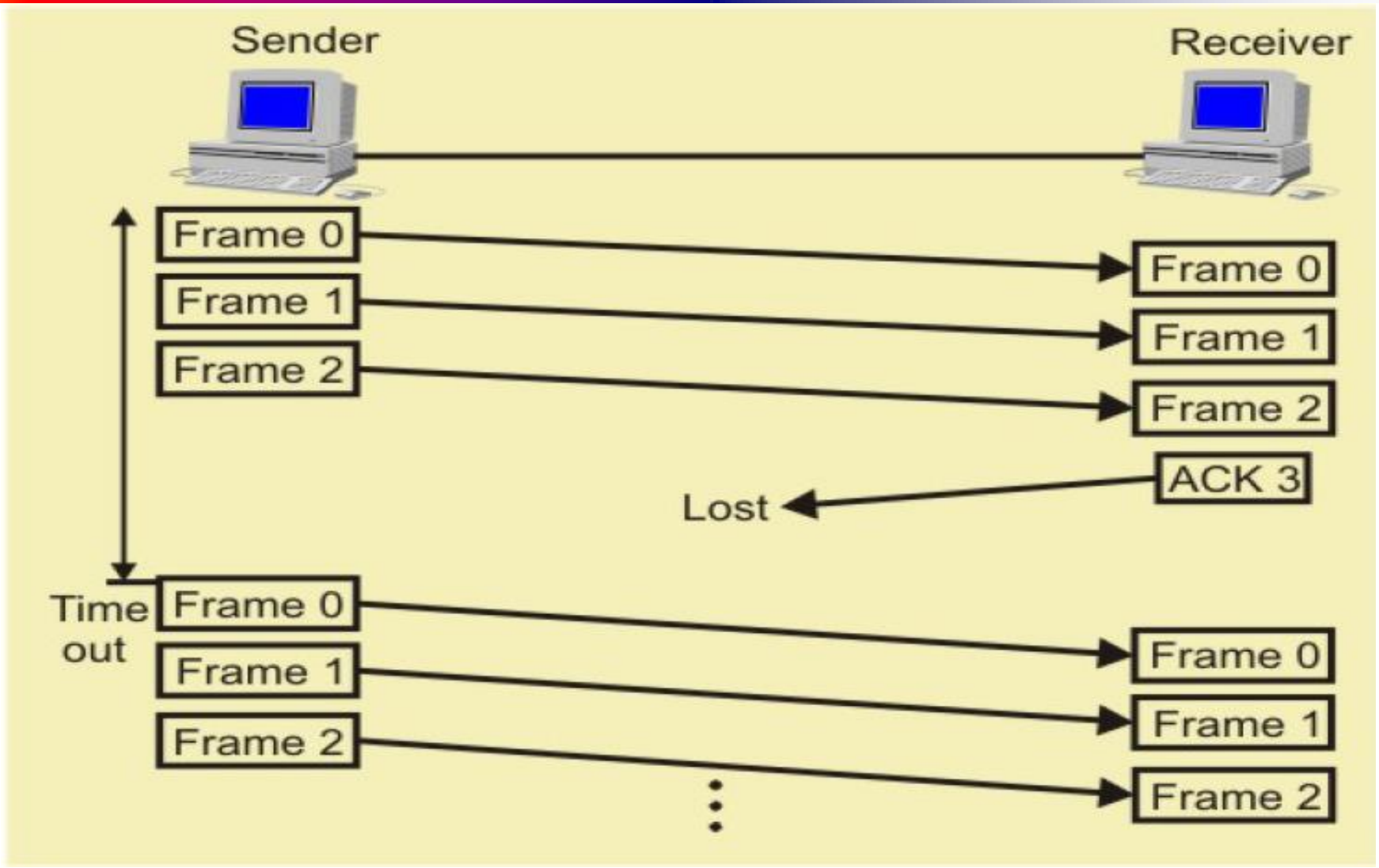

Sender
Receiver
A
B
Start timer
Stop timer
Initial
Request
Arrival
Frame 0
Frame 1
Frame 2
Frame 3
ACK 1
ACK 2
ACK 3
ACK 4
Lost
Time
Time


Sender
Receiver
A
B
Start timer
Initial
Request
Frame 0
ACK 1
Arrival
Request
Frame 1
Lost
Request
Frame 2
Arrival
No action
Request
Frame 3
Arrival
No action
Time-out
Restart
Time-out
Frame 1 (resent)
Arrival
Frame 2 (resent)
ACK 2
Arrival
Frame 3 (resent)
ACK 3
Arrival
ACK 4
Arrival
Stop timer
Time
Time

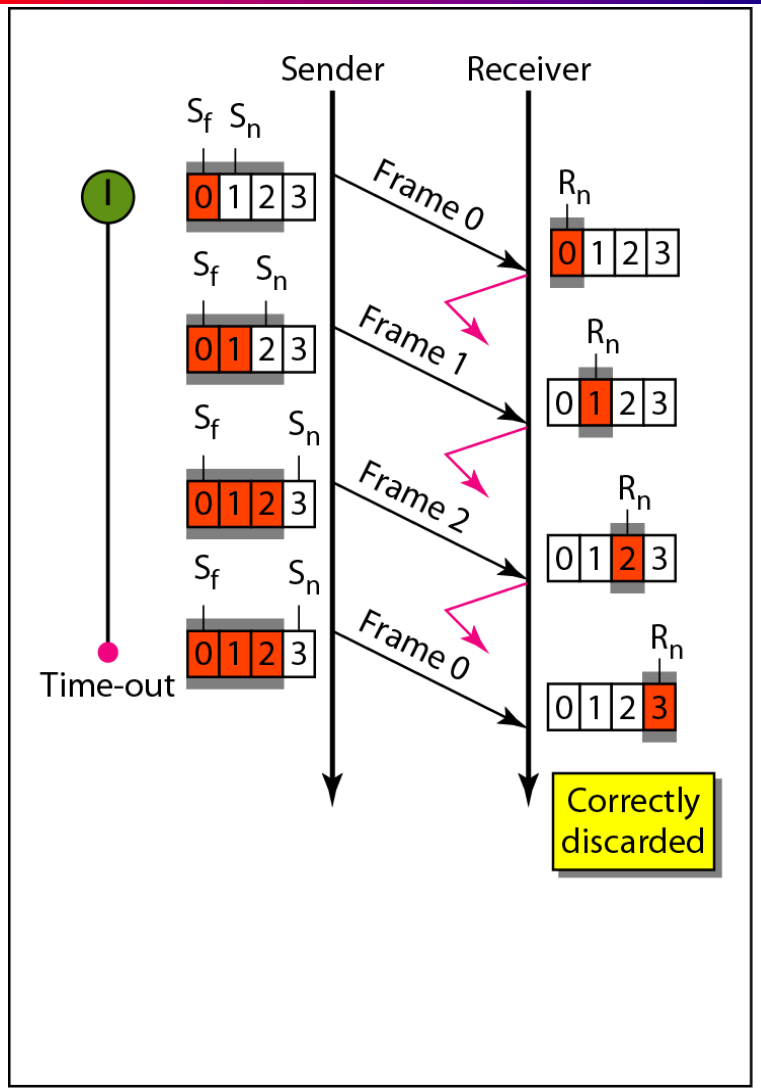
Go-back-N ARQ


Sender
Receiver
Frame 0
Frame 1
Frame 2
Lost
Frame 3
Frame 4
Frame 2
Frame 3
Frame 4
Frame 0
Frame 1
Discarded
NAK 2
Discarded
Frame 2
Frame 3
Frame 4

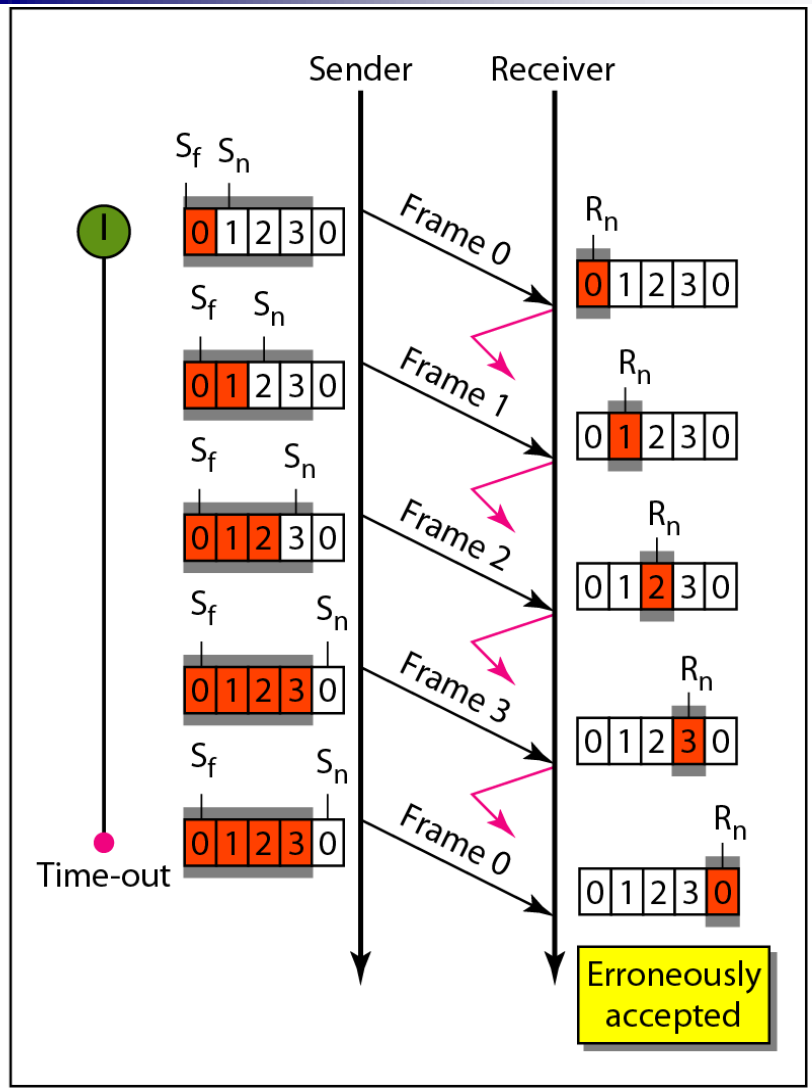
Go-back-N ARQ


Sender
Receiver
Frame 0
Frame 1
Frame 2
Frame 0
Frame 1
Frame 2
ACK 3
Lost
Time out
Frame 0
Frame 1
Frame 2
Frame 0
Frame 1
Frame 2

Go-back-N ARQ (Window Size)



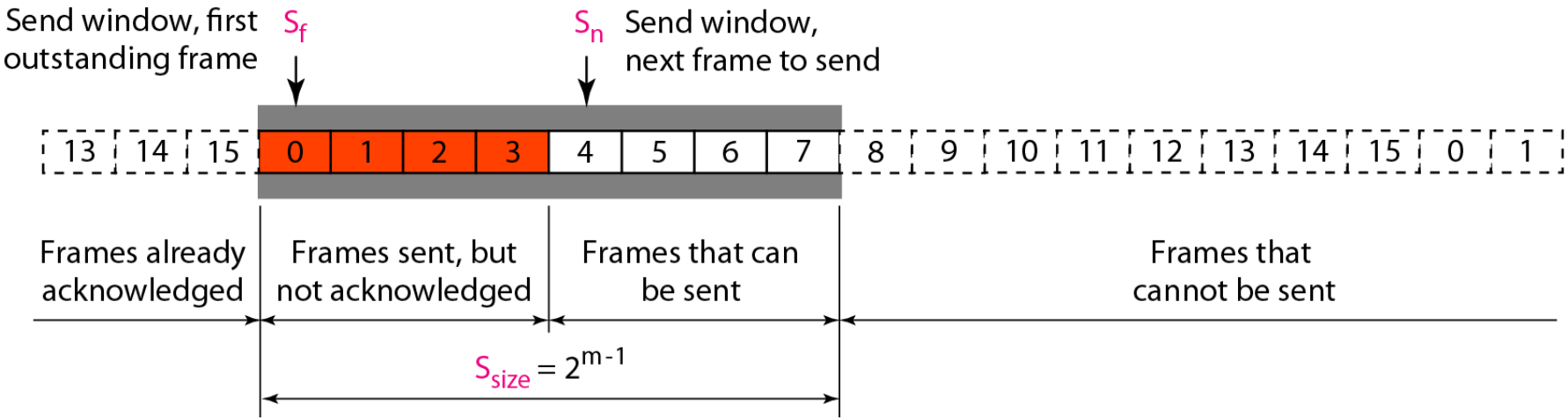
a. Window size < 2^m



b. Window size = 2^m

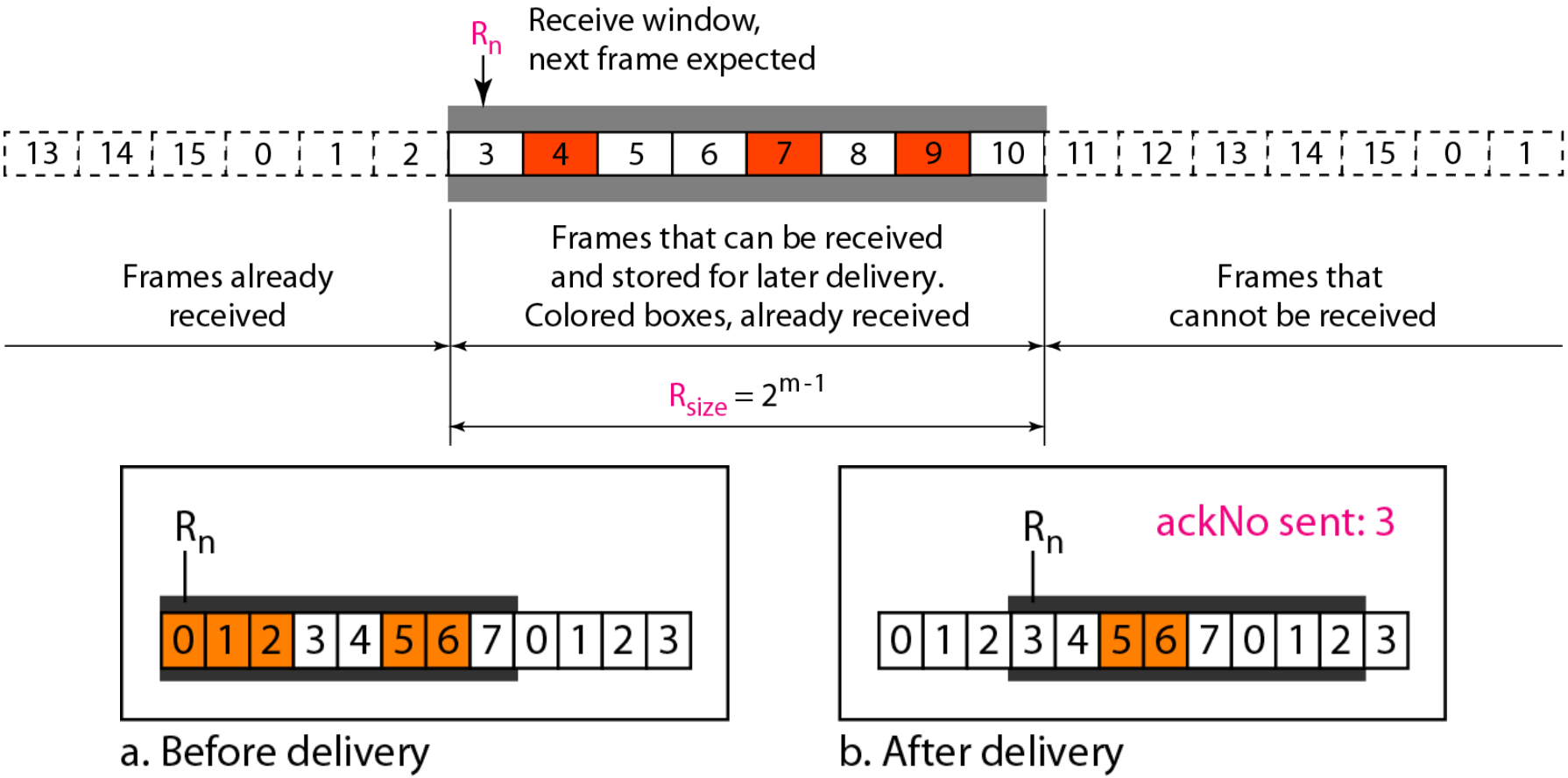
Selective Repeat ARQ

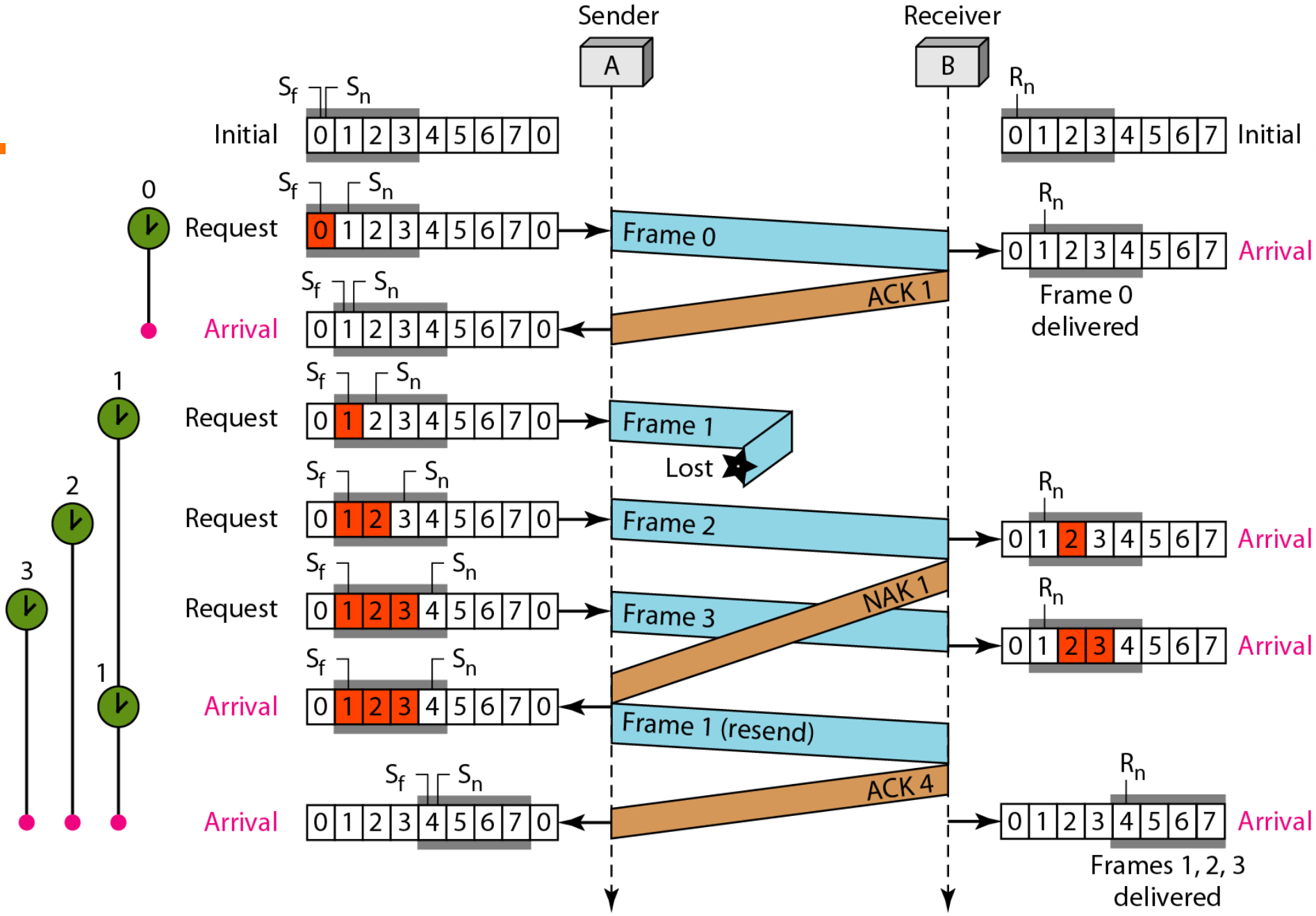
- The form of error control based on sliding-window flow control.
- The size of the sender and receiver window must be at most 2^{m-1}.



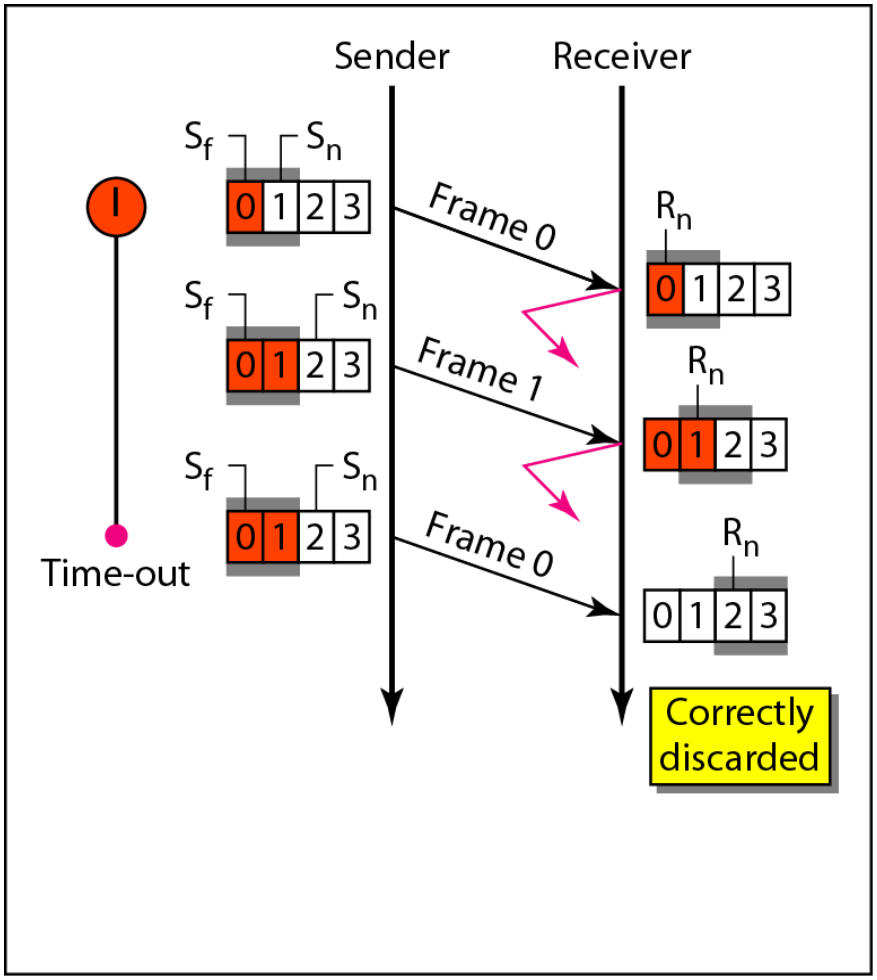
Selective Repeat ARQ

- The form of error control based on sliding-window flow control.
- The size of the sender and receiver window must be at most 2^{m-1}.

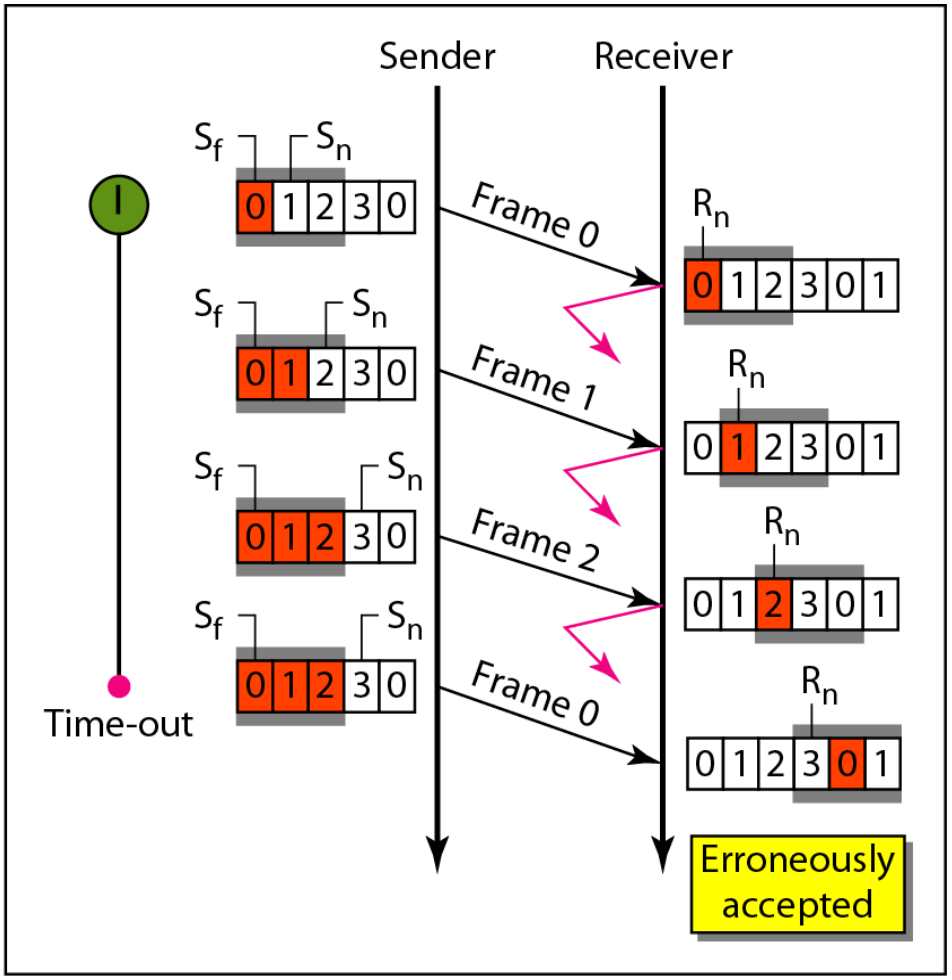



Sender
Receiver
A
B
Initial
Request
Arrival
Frame 0
ACK 1
Frame 0 delivered
Frame 1
Lost
Frame 2
Frame 3
NAK 1
Frame 1 (resend)
ACK 4
Frames 1, 2, 3 delivered

Selective Repeat ARQ (Window Size)



a. Window size = 2^{m-1}



b. Window size > 2^{m-1}

Exercise-1

- A sender sends a series of packets to the same destination using 5-bit sequence numbers. If the sequence numbers start with 0, what is the sequence number of the 100th packet?

Exercise-2

- If transmission delay and propagation delay in a sliding window protocol are 1 msec. and 49.5 msec. respectively, then:
  - What should be the sender window size to get the maximum efficiency?
  - What is the minimum number of bits required in the sequence number field?
  - f only 6 bits are reserved for sequence numbers, then what will be the efficiency?

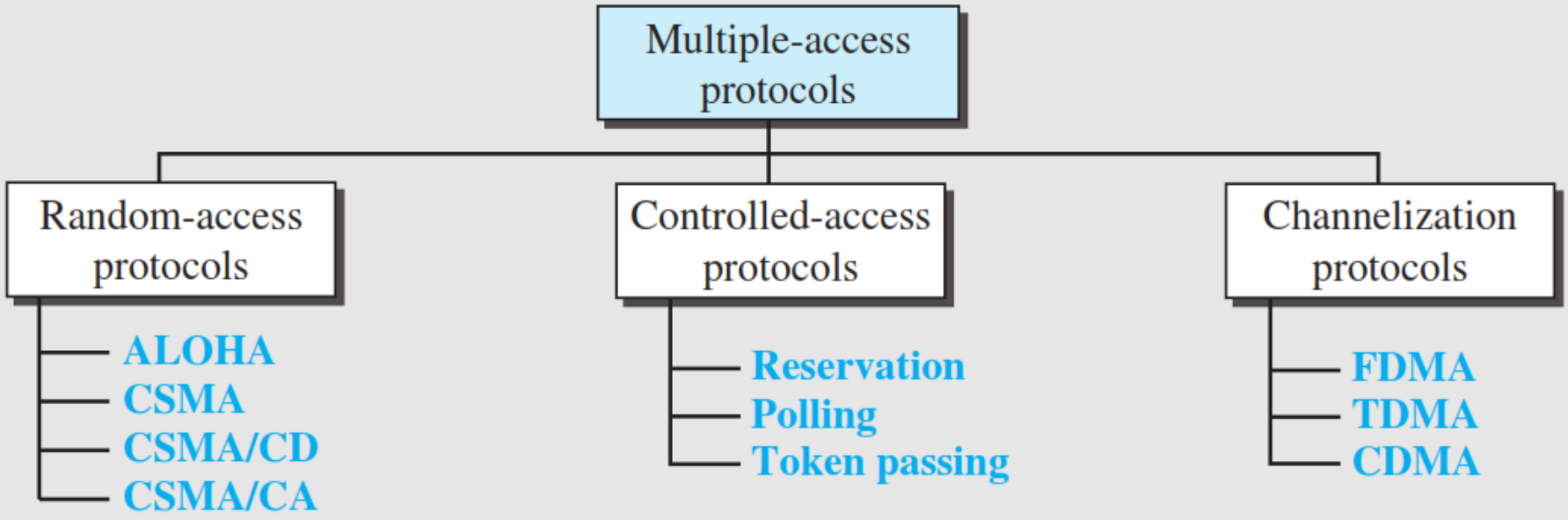
Exercise-3

- A 20 Kbps satellite link has a propagation delay of 400 msec., the transmitter employs the “Go back N” ARQ” scheme with N set to 10. Assuming that each frame is 100 bytes long, what is the maximum data rate possible?

Media Access Control (MAC)

Multiple Access Protocols

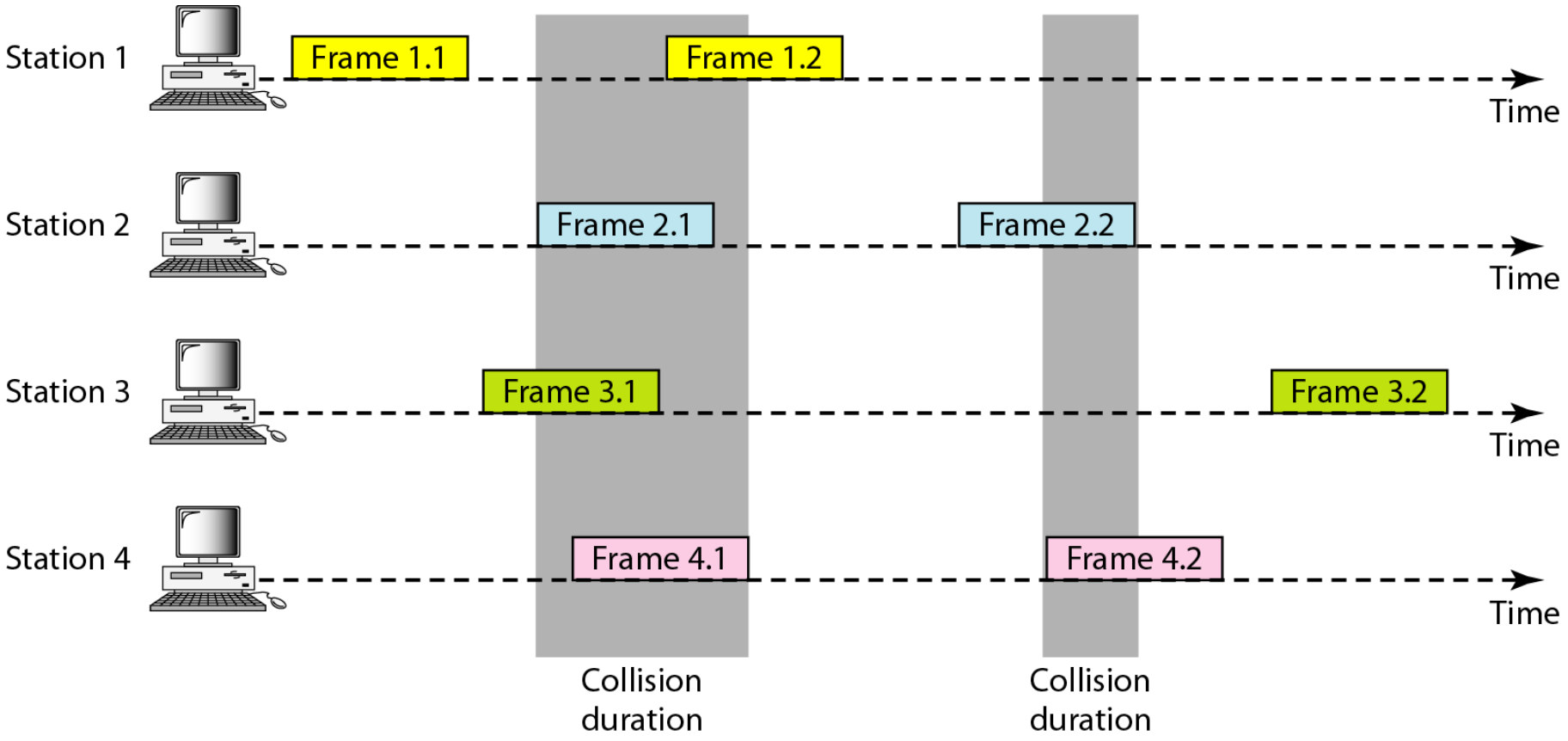
- When nodes or stations are connected and use a common link, called a multipoint or broadcast link.
- A multiple-access protocol is required to coordinate access to the link.
- All of these protocols belong to a sublayer in the data-link layer called media access control (MAC).



Random Access Methods

- No station is superior to another station.
- No one is assigned control over another.
- Each station can transmit when it desires by following protocol.
- There is no scheduled time for a station to transmit.
- No rules specify which station should send next.
- Stations **compete with one another** to access the medium.
- Also called **contention** methods.
- If more than one station tries to send, there is an access conflict—**collision**.
- The frames will be either destroyed or modified after collision.

Collision


Station 1
Frame 1.1
Frame 1.2
Time
Station 2
Frame 2.1
Frame 2.2
Time
Station 3
Frame 3.1
Frame 3.2
Time
Station 4
Frame 4.1
Frame 4.2
Time
Collision duration
Collision duration

ALOHA

- The original ALOHA protocol is called **pure ALOHA**.
- The pure ALOHA protocol relies on acknowledgments from the receiver.
- Resend the frame after time-out period.

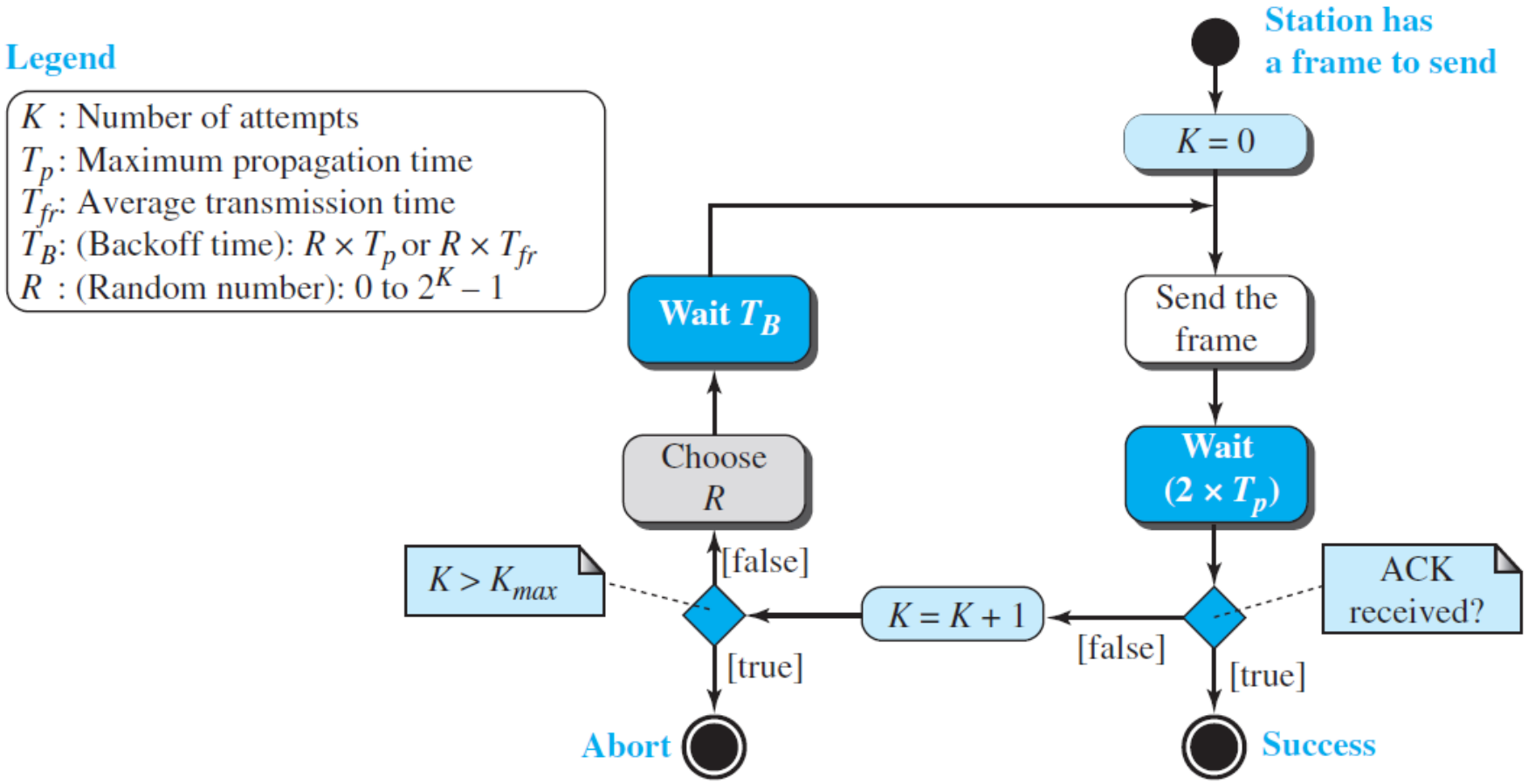
time out period = 2* T_p (T_p *is maximum propagation time)*

- Each station waits a random amount of time before resending its frame.
- This random waiting time in known as backoff time T_B .

$T_B = R * Tp$ $\qquad R = 0 \text{ to } 2^K - 1$

- After a maximum number of retransmission attempts K_{max}, a station must give up and try later.

ALOHA


Legend
K : Number of attempts
T_p: Maximum propagation time
T_{fr}: Average transmission time
T_B: (Backoff time): $R \times T_p$ or $R \times T_{fr}$
R : (Random number): 0 to $2^K - 1$
Station has a frame to send
K = 0
Send the frame
Wait ($2 \times T_p$)
ACK received?
[false]
K = K + 1
[true]
Success
$K > K_{max}$
[false]
Choose R
Wait T_B
[true]
Abort

Slotted ALOHA

- Vulnerable time : length of time in which there is a possibility of collision.

$$\text{Pure ALOHA vulnerable time} = 2 \times T_{fr}$$

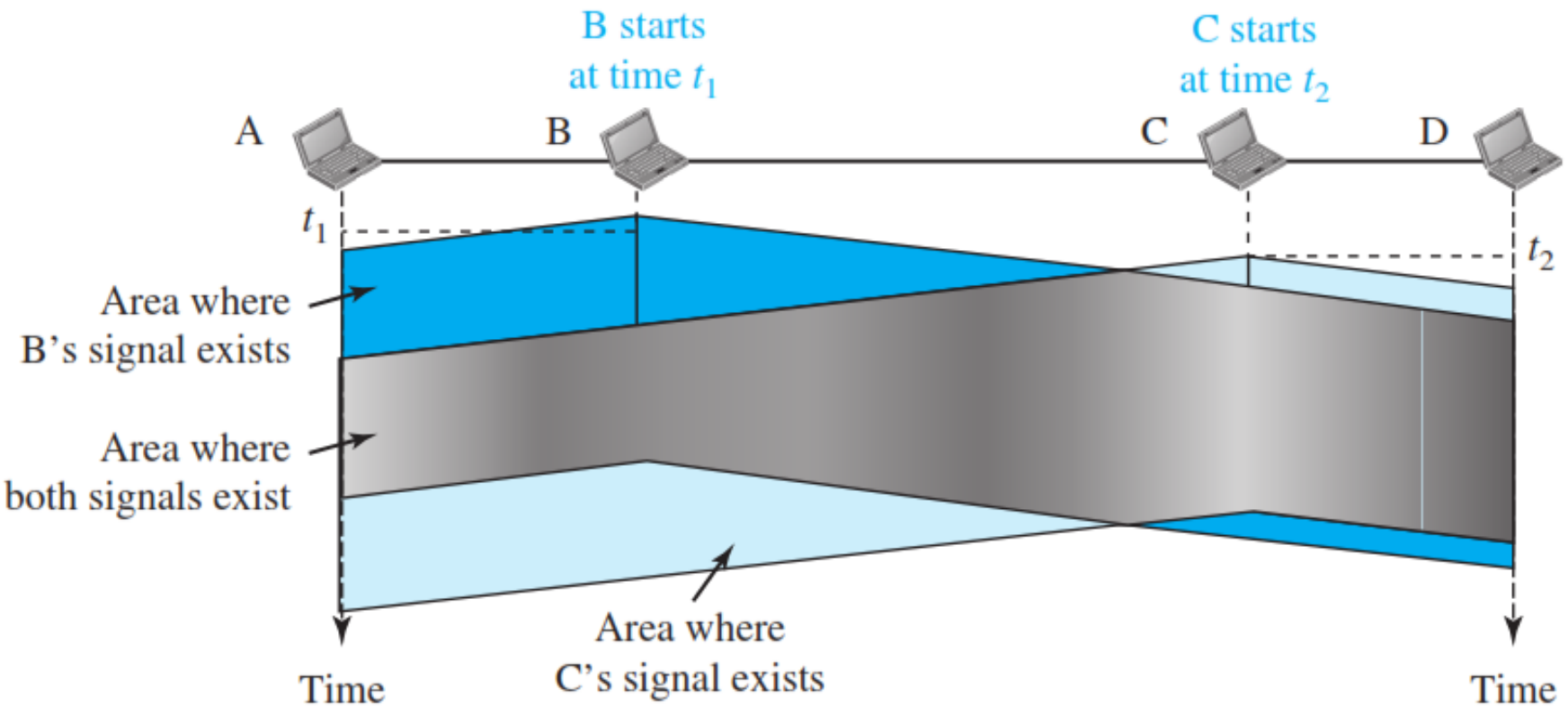
- Time is divided into slots of T_{fr}
- *Each station force to send frame at the beginning of the time slot.*

$$\text{Slotted ALOHA vulnerable time} = T_{fr}$$

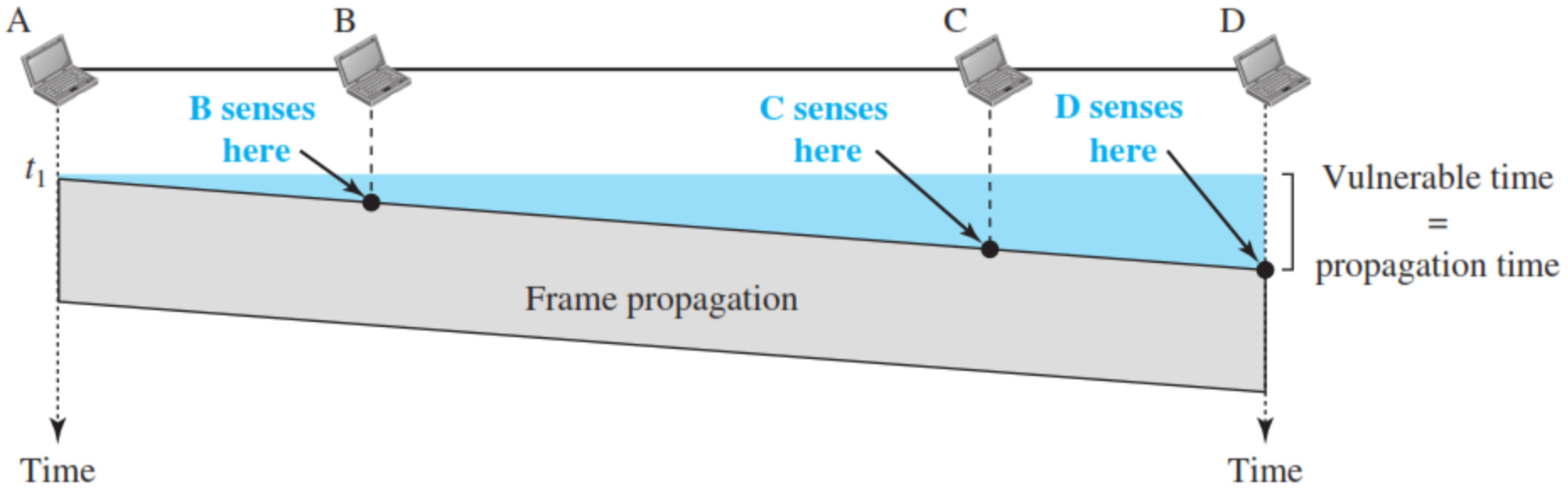
Carrier Sense Multiple Access (CSMA)

- To minimize the chance of collision CSMA method was developed
- **CSMA** requires that each station first **sense\listen** to the medium before sending
- CSMA is based on the principle "sense before transmit" or "listen before talk".
- CSMA can reduce the possibility of collision, but it cannot eliminate it.
- The possibility of collision still exists because of propagation delay
  - A station may sense the medium and find it idle, only because the first bit sent by another station has not yet been received.

Carrier Sense Multiple Access (CSMA)



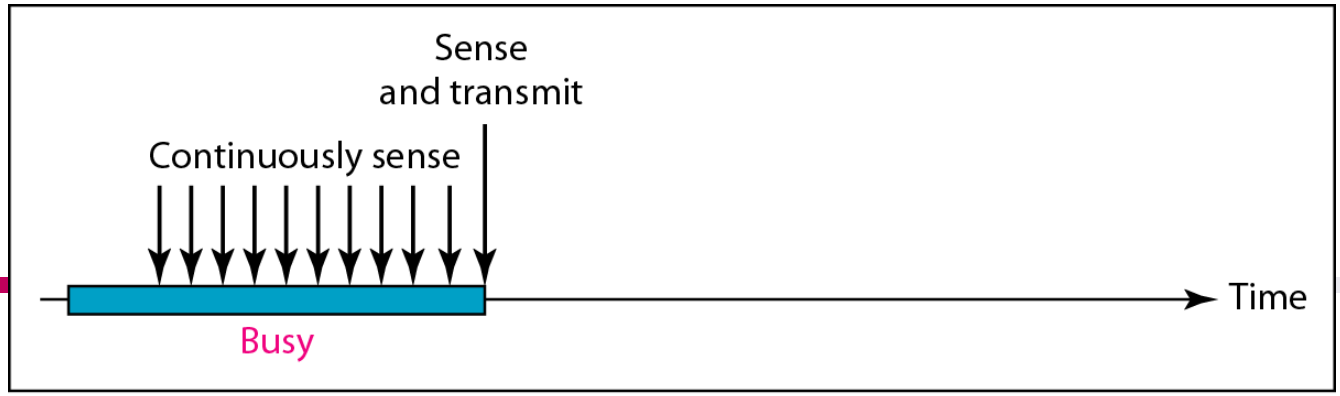
Vulnerable time in CSMA



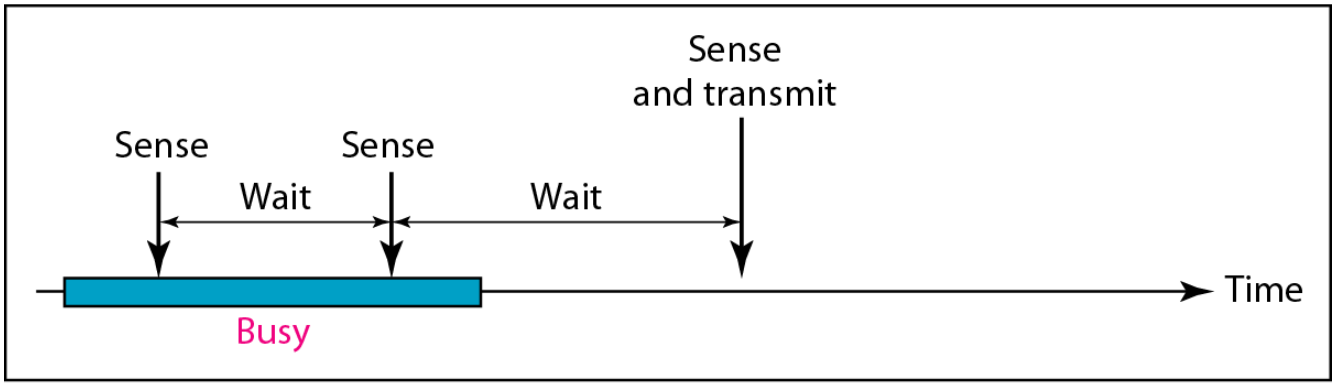
- The vulnerable time for CSMA is the propagation time **Tp**

Carrier Sense Multiple Access (CSMA)

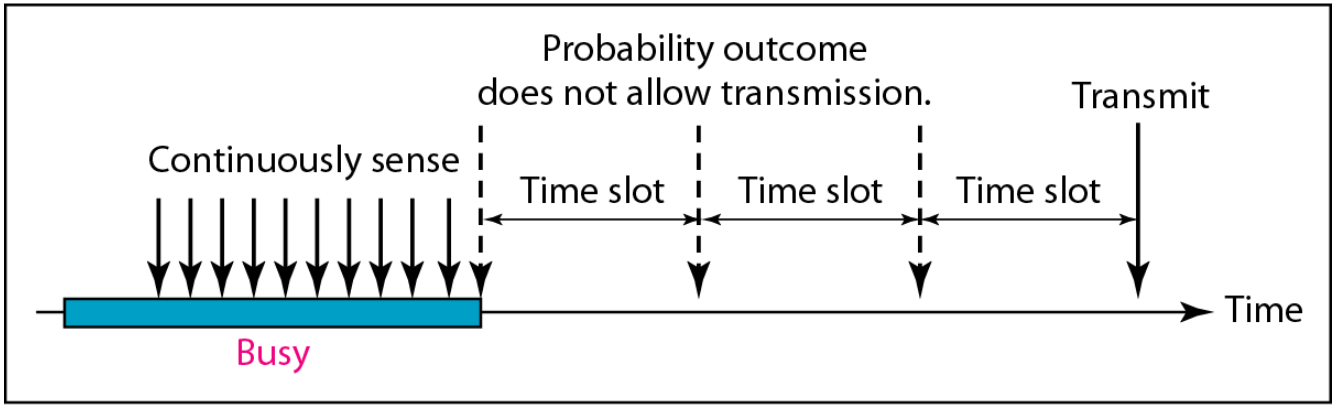
- **Persistence Methods:**
  - What should a station do if the channel is busy/ideal?
  - Three methods have been devised to answer these questions:
    - 1-persistent
    - Nonpersistent
    - *p-persistent*



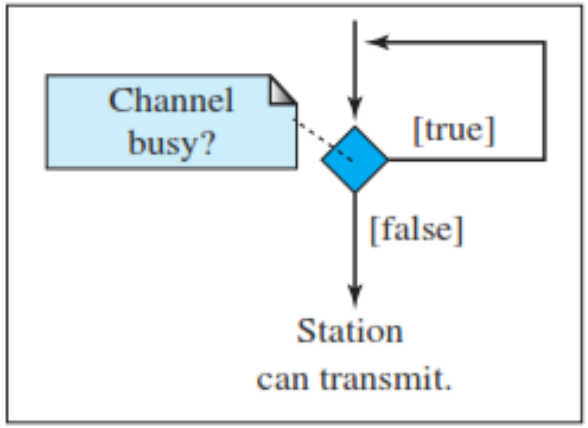
a. 1-persistent



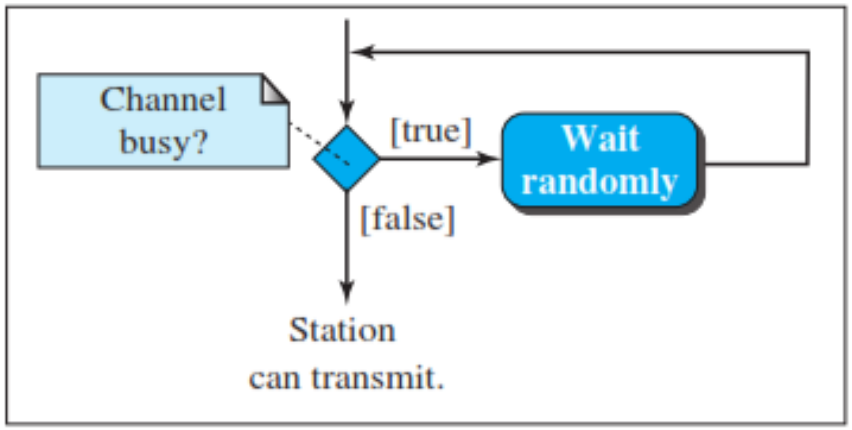
b. Nonpersistent



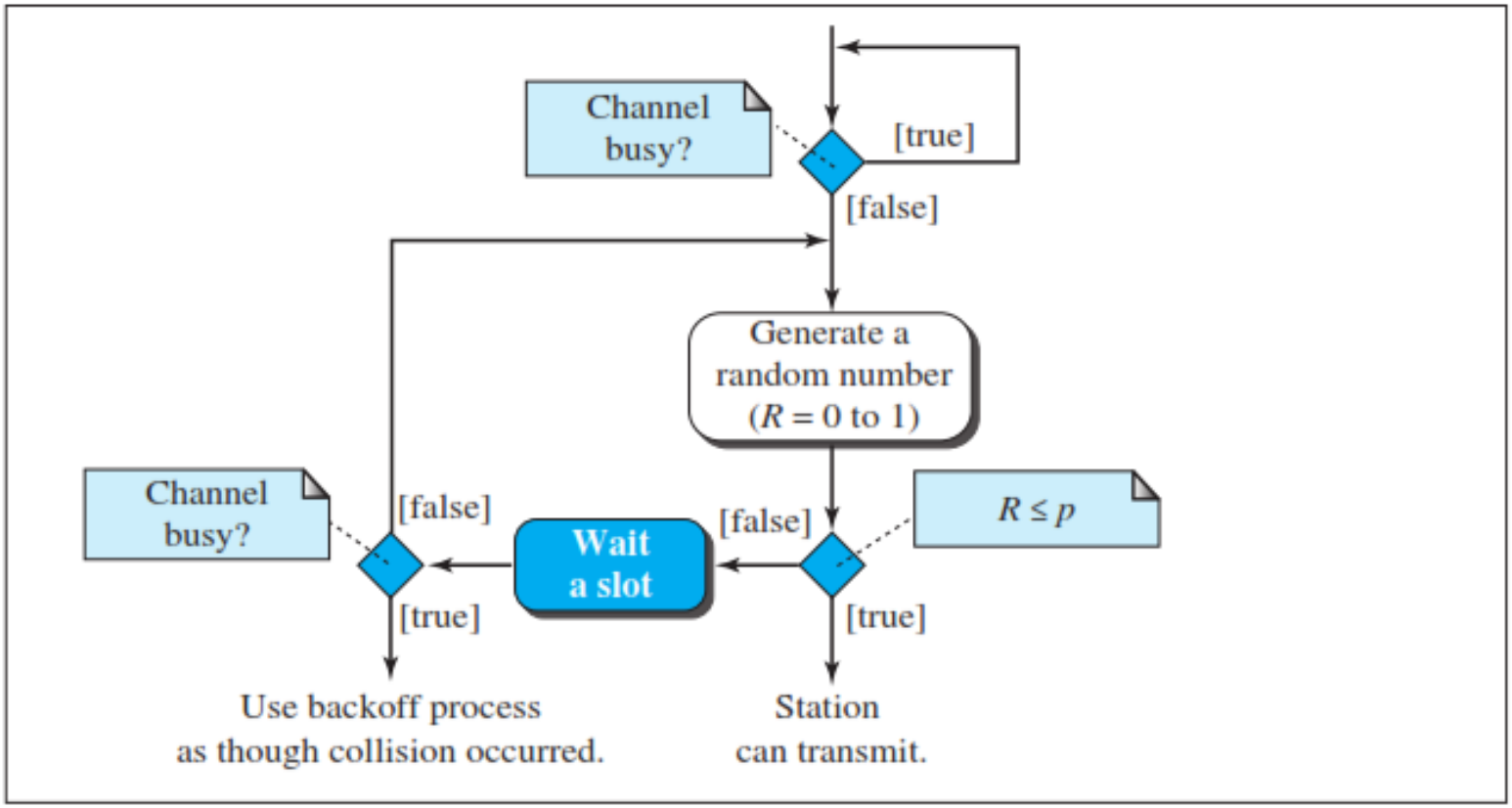
c. p-persistent



a. 1-Persistent



b. Nonpersistent

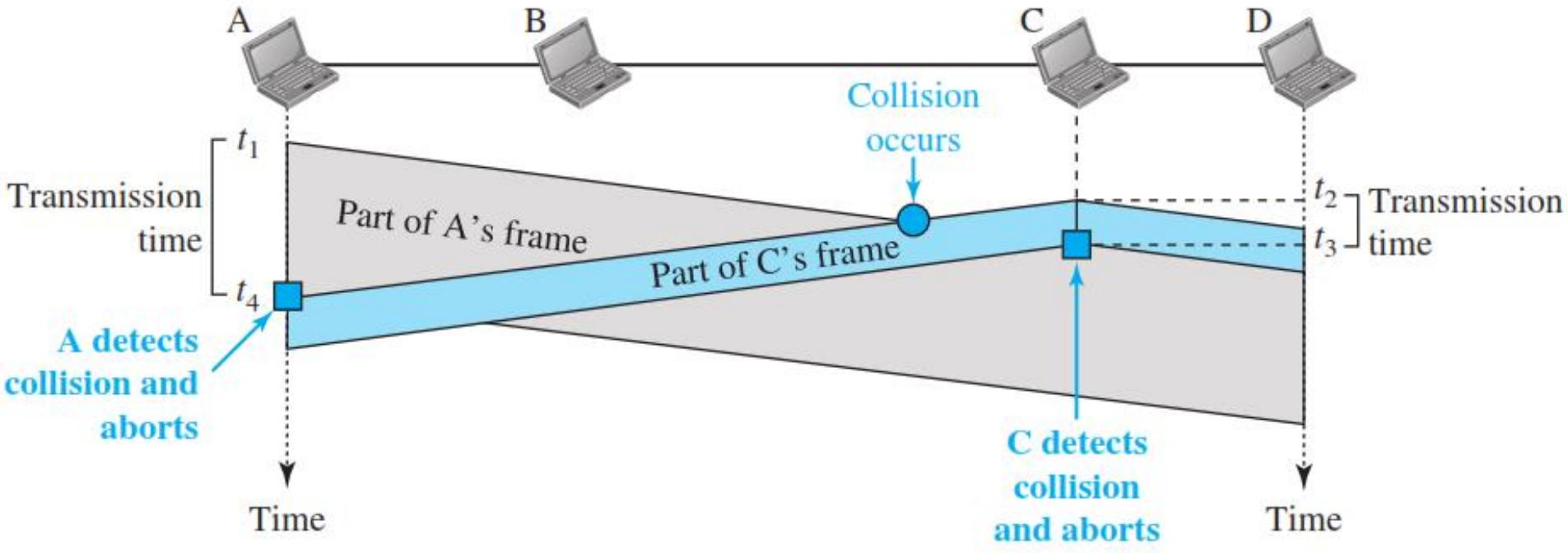


c. *p*-Persistent

CSMA/CD

- The CSMA method does not specify the procedure following a collision.
- CSMA/CD (Carrier Sense Multiple Access with Collision Detection) is designed to handle the collision.
- In CSMA/CD, a station monitors the medium after it sends a frame to see if the transmission was successful
  - If so, the station is finished
  - If, however, there is a collision, the frame is sent again

CSMA/CD


A
B
C
D
Collision
occurs
t_1
Transmission
time
t_4
Part of A's frame
Part of C's frame
t_2
t_3
Transmission
time
A detects
collision and
aborts
C detects
collision
and aborts
Time
Time

CSMA/CD

Minimum Frame Size:

- For CSMA/CD to work, we need a restriction on the frame size.
- Before sending the last bit of the frame, the sending station must detect a collision, if any, and abort the transmission
- This is so because the station, once the entire frame is sent, does not monitor the line for collision detection
- Therefore, the frame transmission time must be at least two times the maximum propagation time

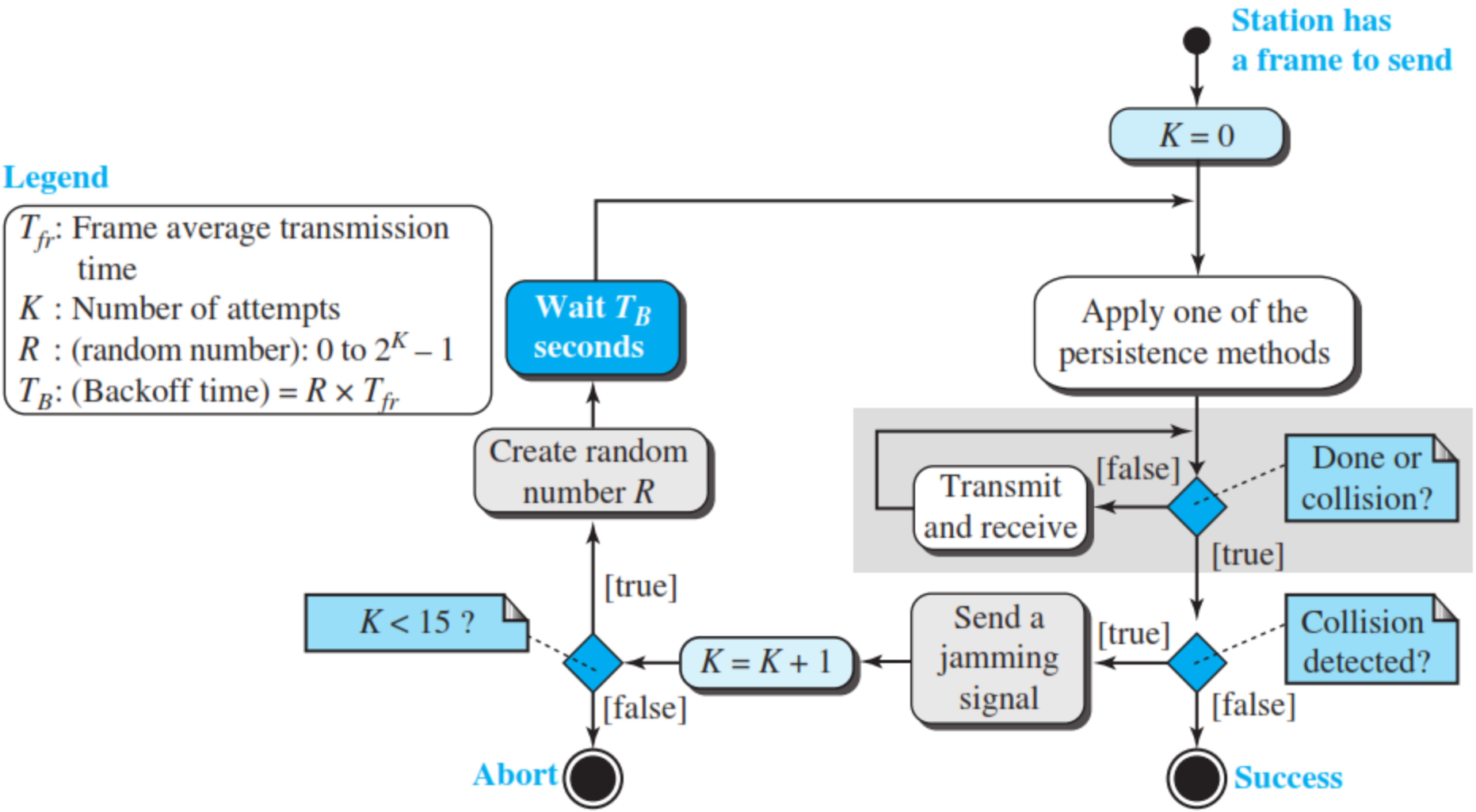
$$T_{fr} \geq 2 * T_p$$

Minimum Frame Size = T_{fr} * Data Rate

CSMA/CD

A network using CSMA/CD has a bandwidth of 10 Mbps. If the maximum propagation time (including the delays in the devices and ignoring the time needed to send a jamming signal, as we see later) is 25.6 μs, what is the minimum size of the frame?

Flow diagram for the CSMA/CD



CSMA/CA

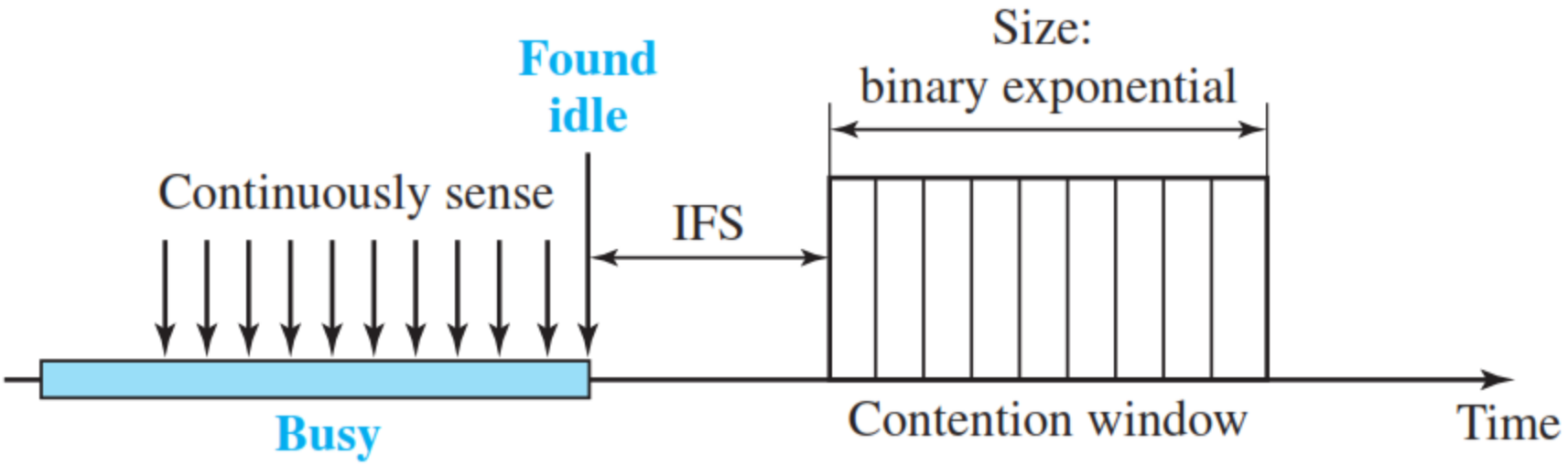
- Carrier sense multiple access with collision avoidance (CSMA/CA)
- Collisions are avoided through three strategies:
  - The interframe space,
  - The contention window,
  - Acknowledgments

Interframe Space (IFS)

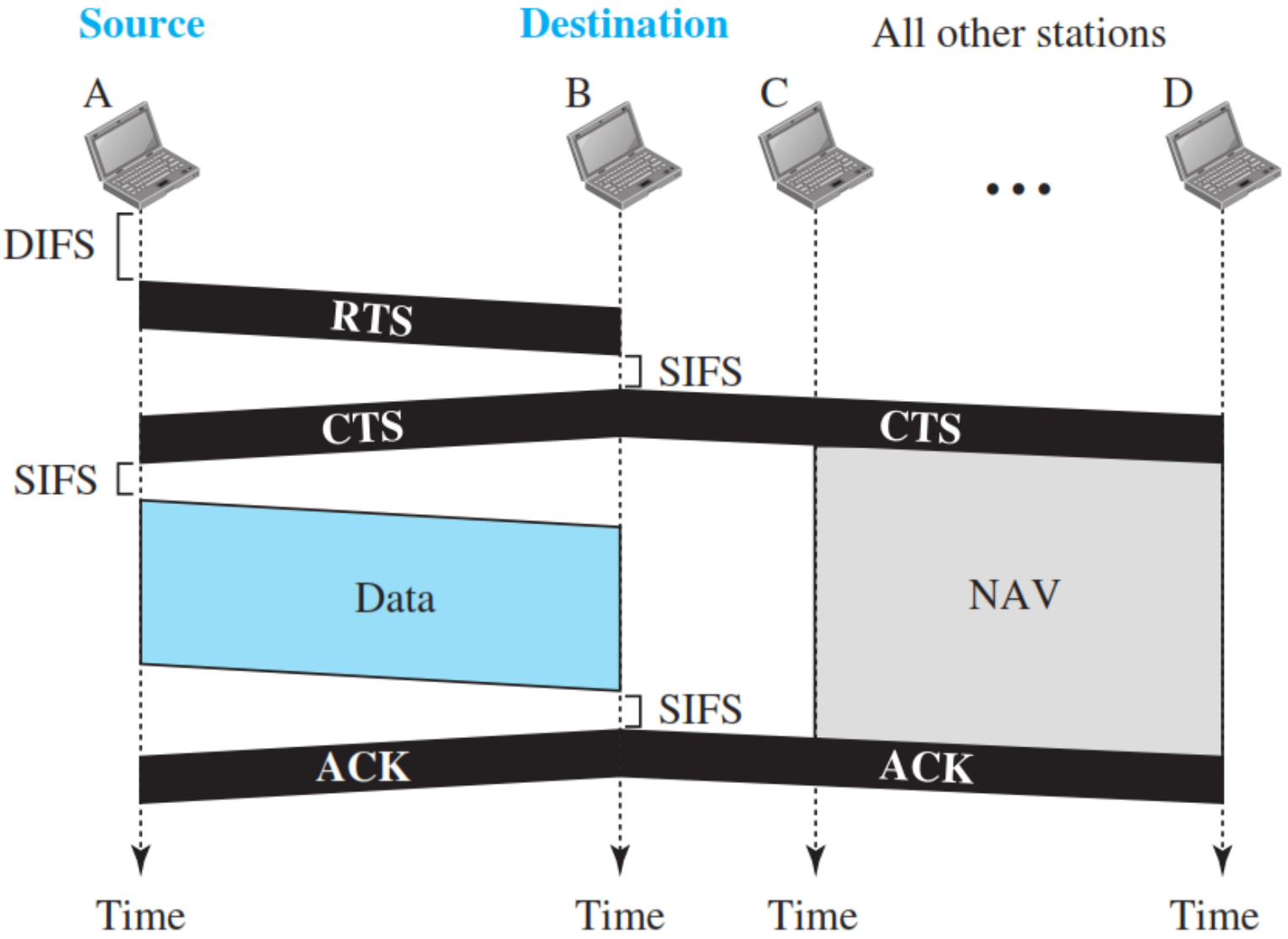
- When an idle channel is found, the station does not send immediately.
- It waits for a period of time called the interframe space or IFS.
- After waiting an IFS time, if the channel is still idle, the station can send after contention window.

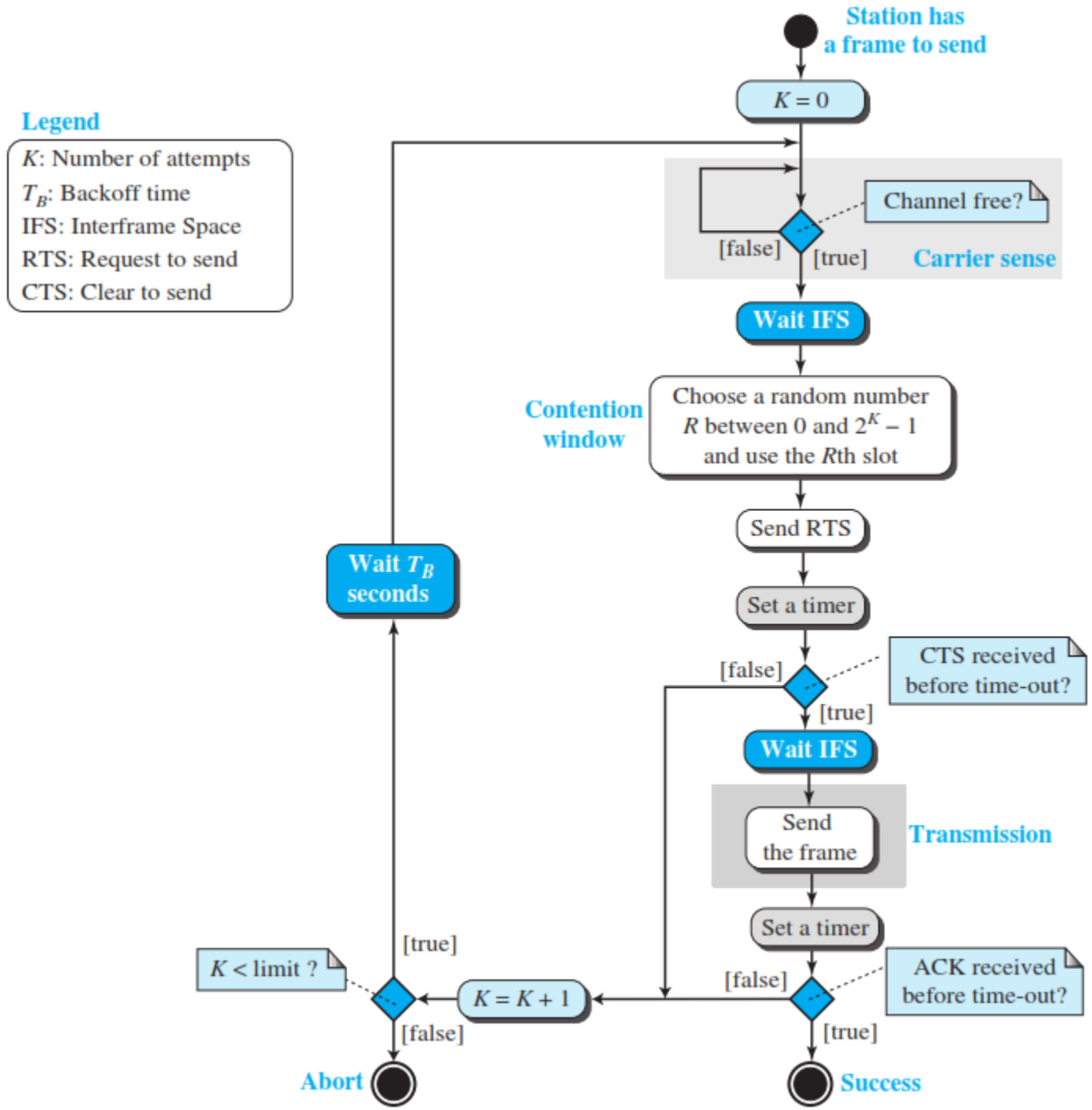
Contention Window

- The contention window is an amount of time divided into slots.
- A station that is ready to send chooses a random number of slots as its wait time.



Acknowledgment

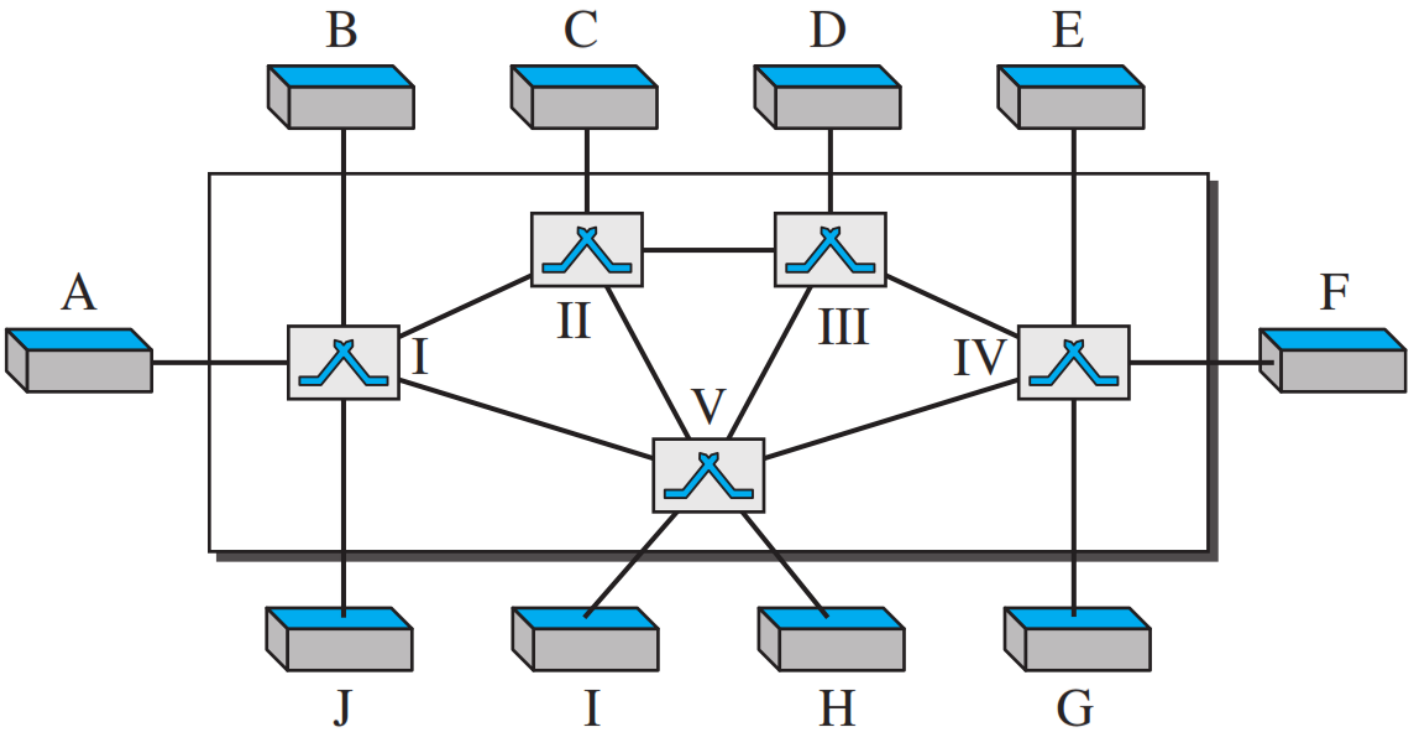



Legend
K: Number of attempts
T_B: Backoff time
IFS: Interframe Space
RTS: Request to send
CTS: Clear to send
Station has a frame to send
K = 0
Channel free?
[false]
[true]
Carrier sense
Wait IFS
Contention window
Choose a random number R between 0 and $2^K - 1$ and use the Rth slot
Send RTS
Set a timer
Wait T_B seconds
[false]
CTS received before time-out?
[true]
Wait IFS
Send the frame
Transmission
Set a timer
[true]
K < limit ?
[false]
K = K + 1
[false]
ACK received before time-out?
[true]
Abort
Success

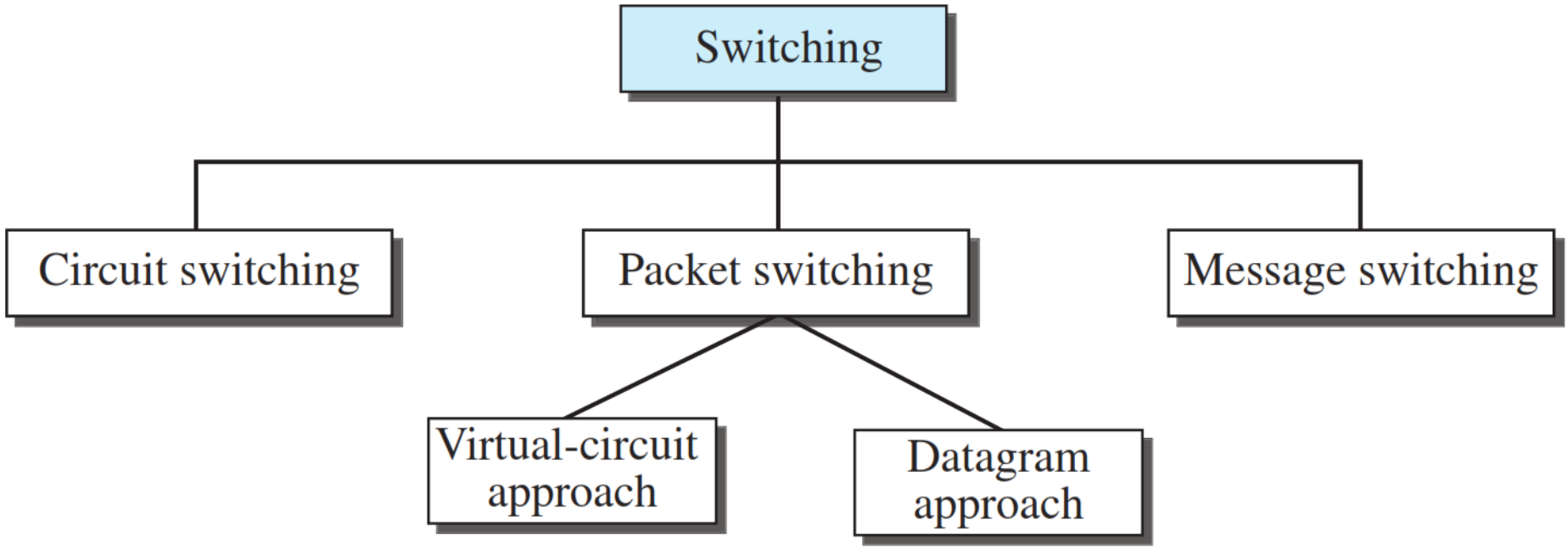
Switching

Switching

- A switched network consists of a series of interlinked nodes, called switches.
- Switches are devices capable of creating temporary connections between two or more devices.



Methods of Switching

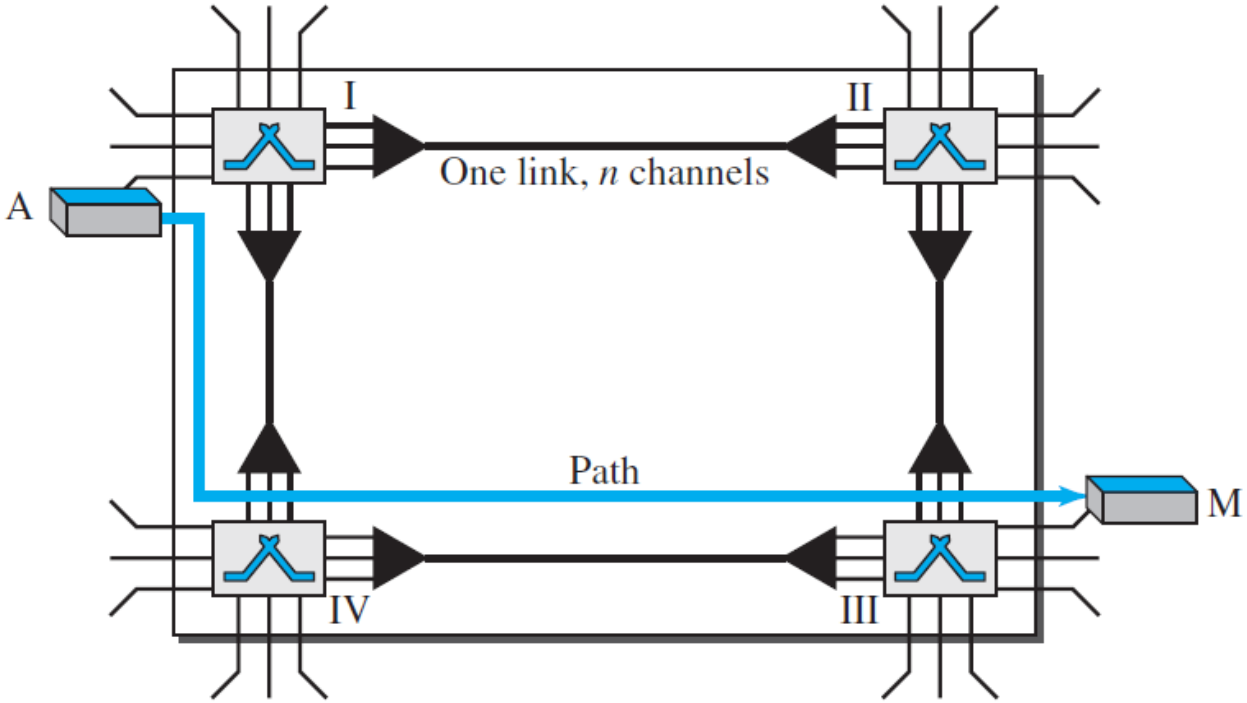


Switching and Network Layers

- Switching can happen at several layers
- Switching at Physical Layer
  - only circuit switching
- Switching at Data-Link Layer
  - packet switching using virtual circuit
- Switching at Network Layer
  - packet switching using virtual circuit
  - packet switching using a datagram
- Switching at Application Layer
  - only message switching

CIRCUIT-SWITCHING

- A circuit-switched network is made of a set of switches connected by physical links.
- Each link is divided into n channels.
- Each connection uses only one dedicated channel on each link.



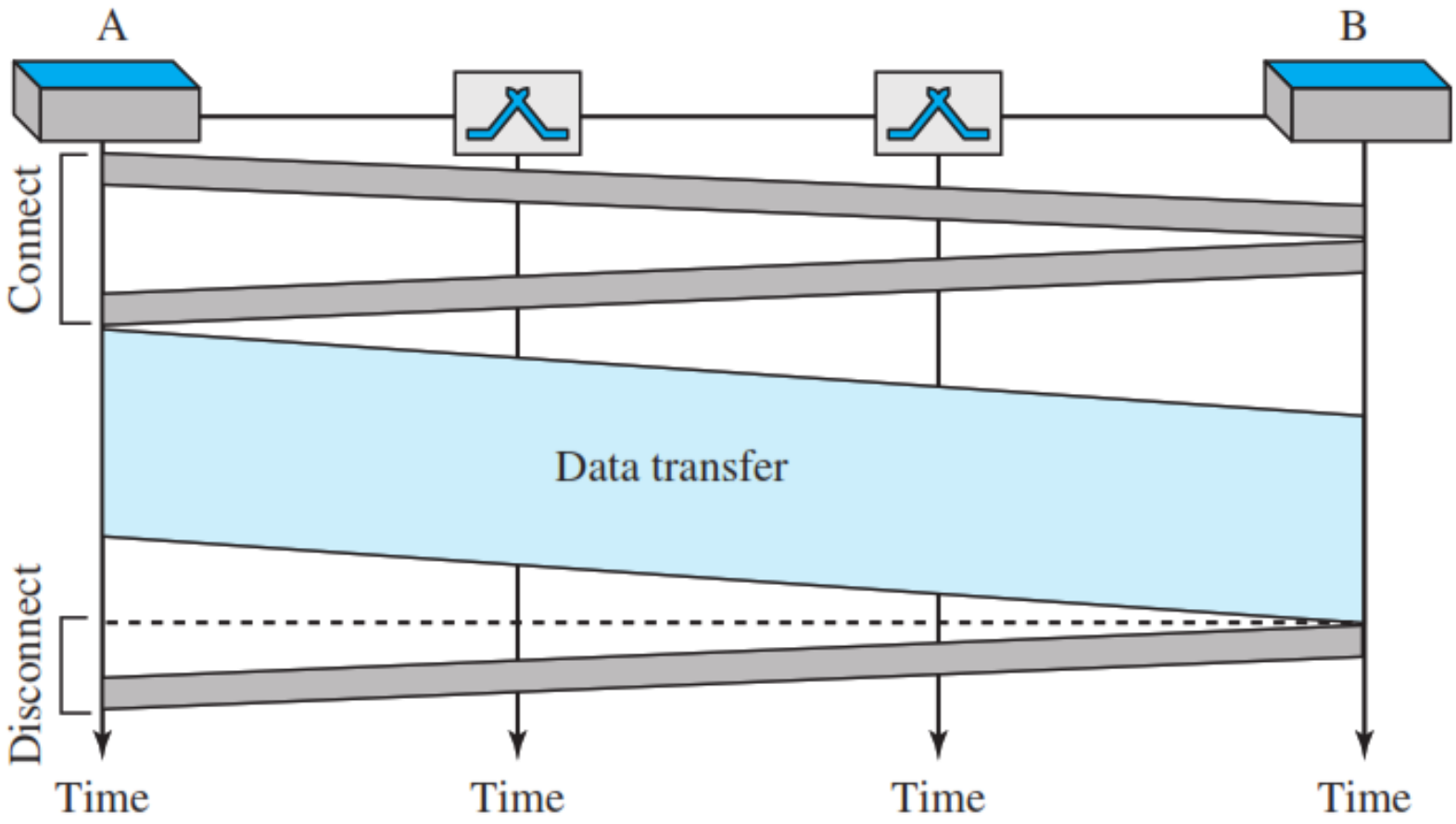
CIRCUIT-SWITCHING

- Actual communication have three phases:
  - Setup Phase
  - Data Transfer Phase
  - Teardown Phase
- A dedicated circuit needs to be established.
- Circuit is a combination of channels in links between two end systems
- The resources need to be reserved during the setup phase.
- These resources, such as channels (bandwidth in FDM and time slots in TDM), switch buffers, switch processing time, and switch input/output ports
- Must remain dedicated during the entire duration of data transfer until the teardown phase

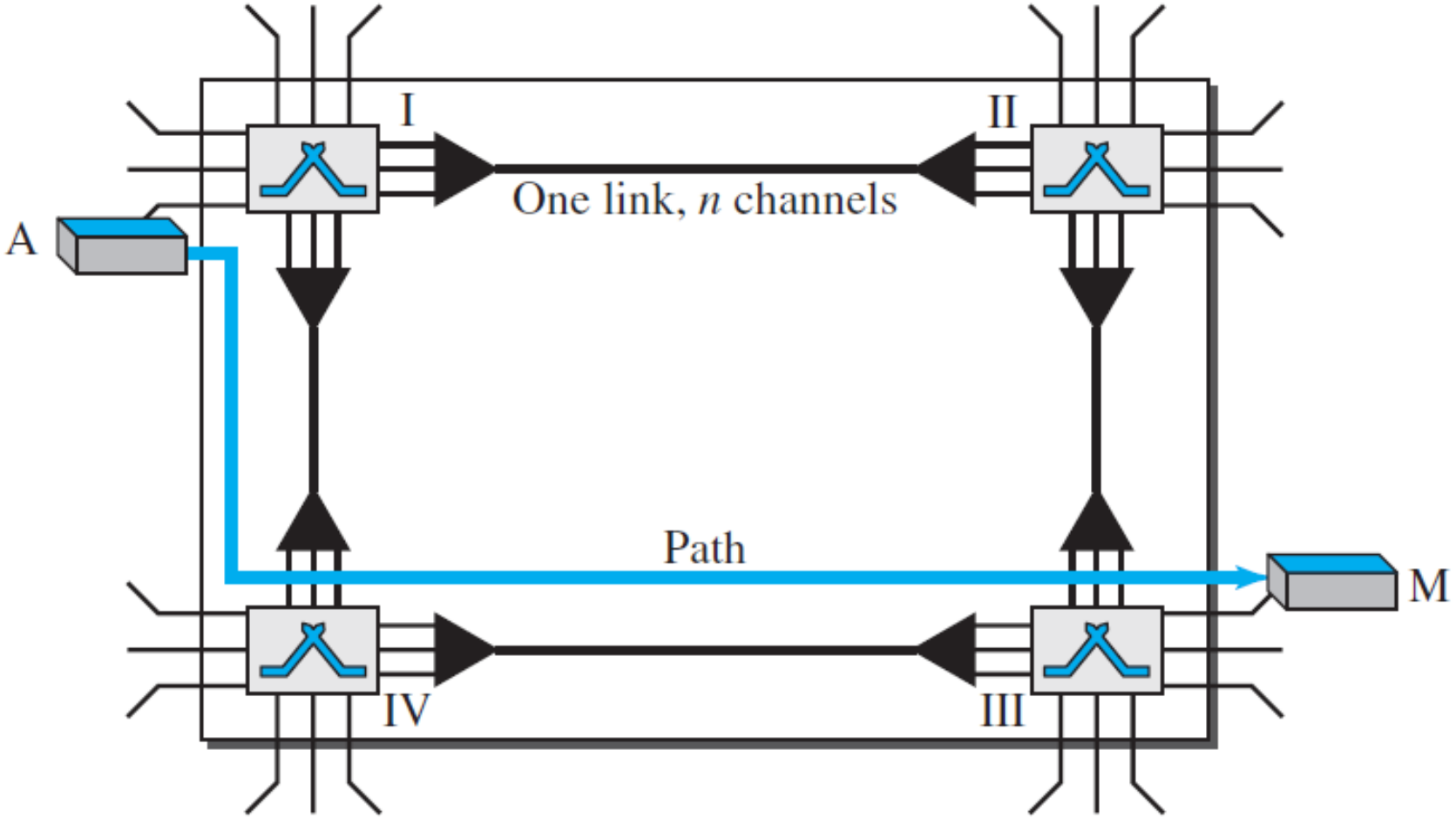
CIRCUIT-SWITCHING

Setup Phase

Teardown Phase



CIRCUIT-SWITCHING



CIRCUIT-SWITCHING

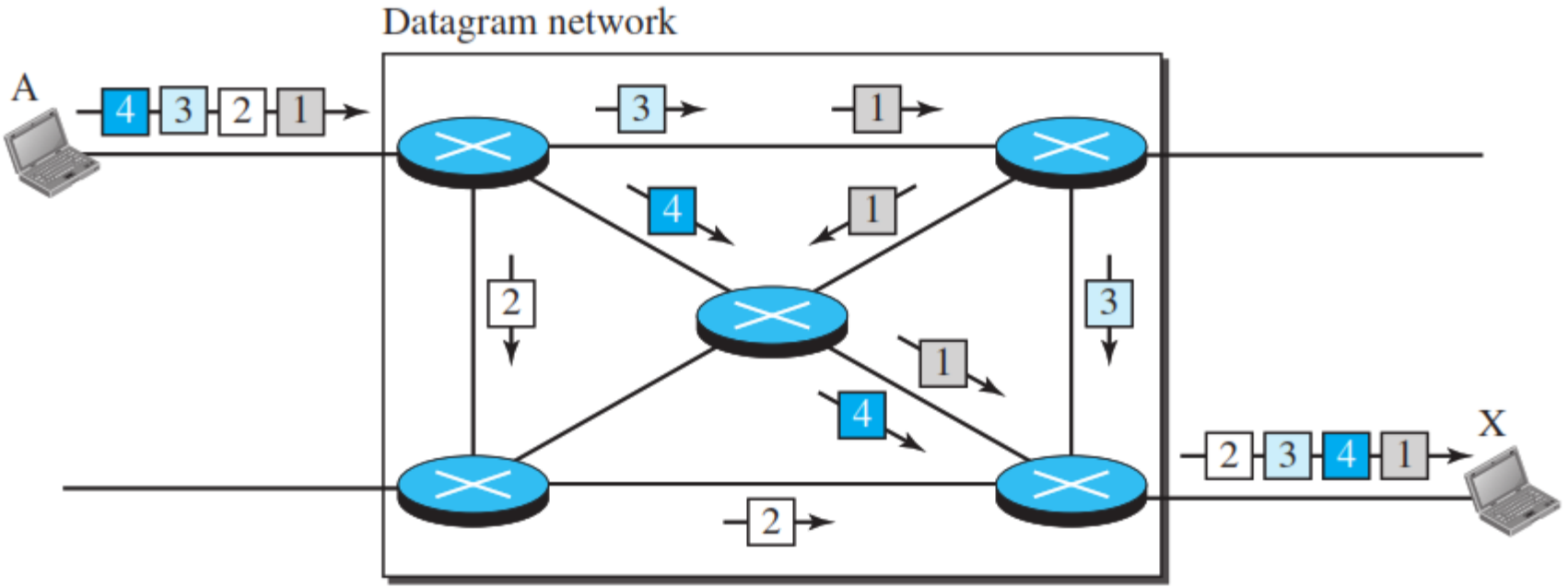
- It requires reservation of resources during the entire duration of the connection.
- It is not an efficient switching techniques for computer networks.
- The delay in this type of network is minimal.
- End-to-End addressing is required for creating a connection between the two end systems.
- Data transferred between the two stations are not packetized.

PACKET-SWITCHING

- A message needs to be divided into packets of fixed or variable size.
- There is no resource reservation.
- Resources are allocated on demand.
- The allocation is done on a FCFS basis.
- Two types of packet-switched networks:
  - Datagram networks
  - Virtual circuit networks

PACKET-SWITCHING : Datagram Approach

- Each packet is treated independently of all others.
- Packets in this approach are referred to as datagrams.
- The switches in a datagram network are traditionally referred to as routers.



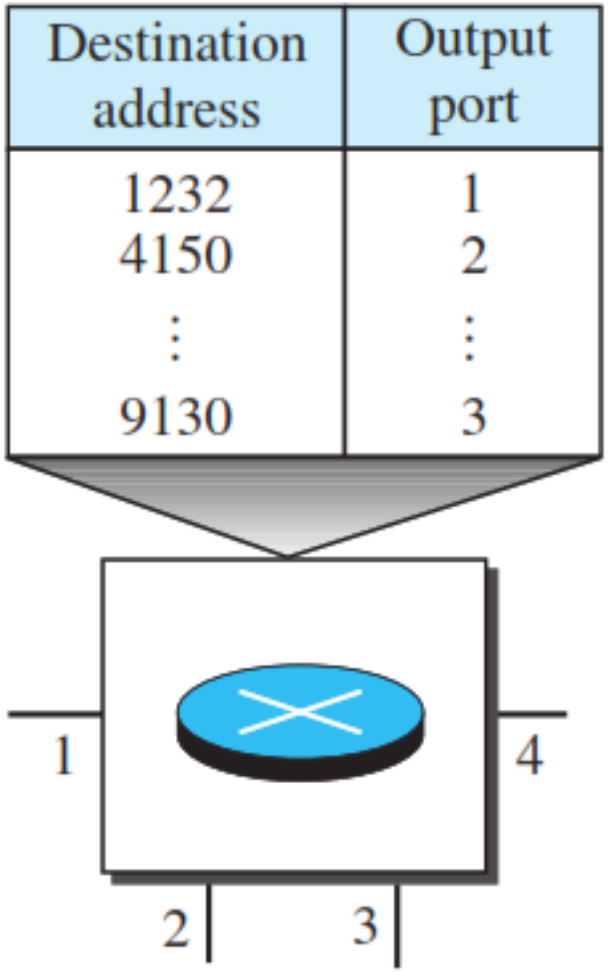
PACKET-SWITCHING : Datagram Approach

- Datagrams may arrive out of order with different delays.
- Packets may be lost or dropped.
- It is the responsibility of an upper-layer protocol to reorder the datagrams.
- Each switch (or packet switch) has a routing table.
- The routing tables are dynamic and updated periodically.
- The destination addresses and the corresponding forwarding output ports are recorded in the tables.

PACKET-SWITCHING : Datagram Approach

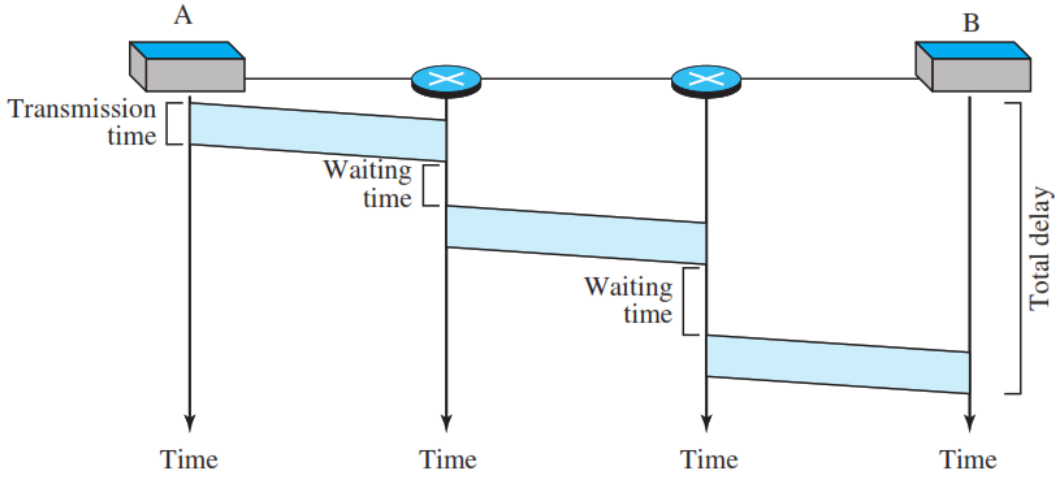
If there are no setup or teardown phases, how are the packets routed to their destinations in a datagram network?

PACKET-SWITCHING : Datagram Approach



PACKET-SWITCHING : Datagram Approach

- Every packet carries a header that contains the destination address of the packet.
- Remains the same during the entire journey of the packet.
- The efficiency of a datagram network is better than that of a circuit-switched network.
- There may be greater delay in a datagram network than in a circuit switched network.



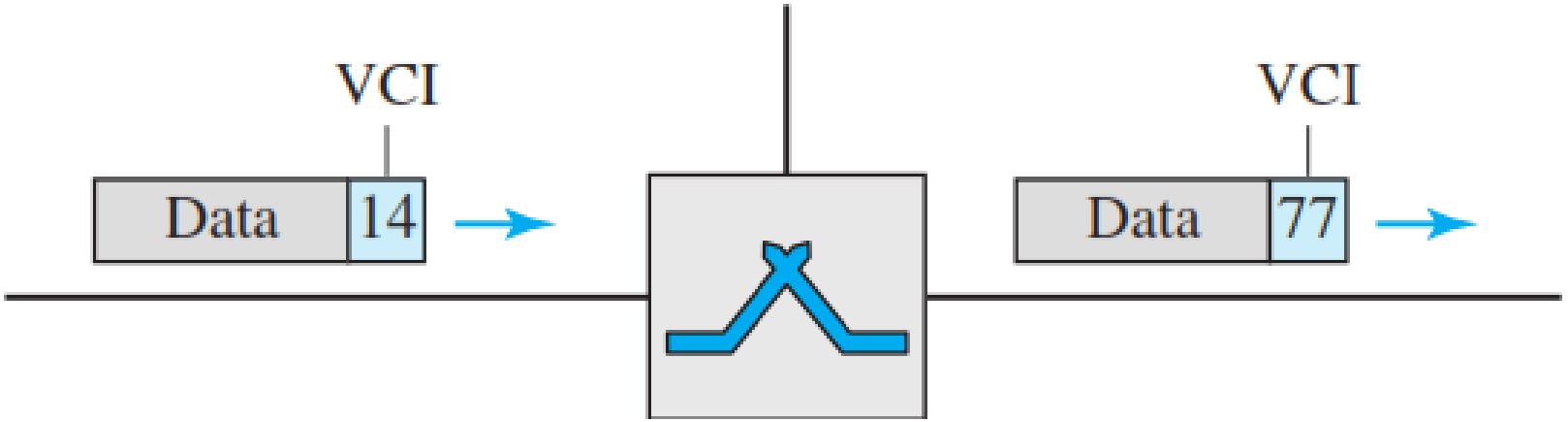
PACKET-SWITCHING : Virtual-Circuit

- A virtual-circuit switching is a cross between a circuit-switching and a datagram.
- There are setup and teardown phases in addition to the data transfer phase.
- Resources can be allocated during the setup phase or on demand.
- Data are packetized and each packet carries an address in the header.
- All packets follow the same path

PACKET-SWITCHING : Virtual-Circuit

Addressing

- Two types of addressing are involved:
  - Global
    - an address that can be unique in the scope of the network
    - global address is used only to create a virtual-circuit identifier
  - Local (virtual-circuit identifier)



PACKET-SWITCHING : Virtual-Circuit

Three Phases

- Setup
  - The source and destination use their global addresses.
  - Switches make table entries for the connection.
- Data transfer
  - Transfers the data in form of packets with VCI.
- Teardown
  - The source and destination inform the switches to delete the corresponding entry.

PACKET-SWITCHING : Virtual-Circuit

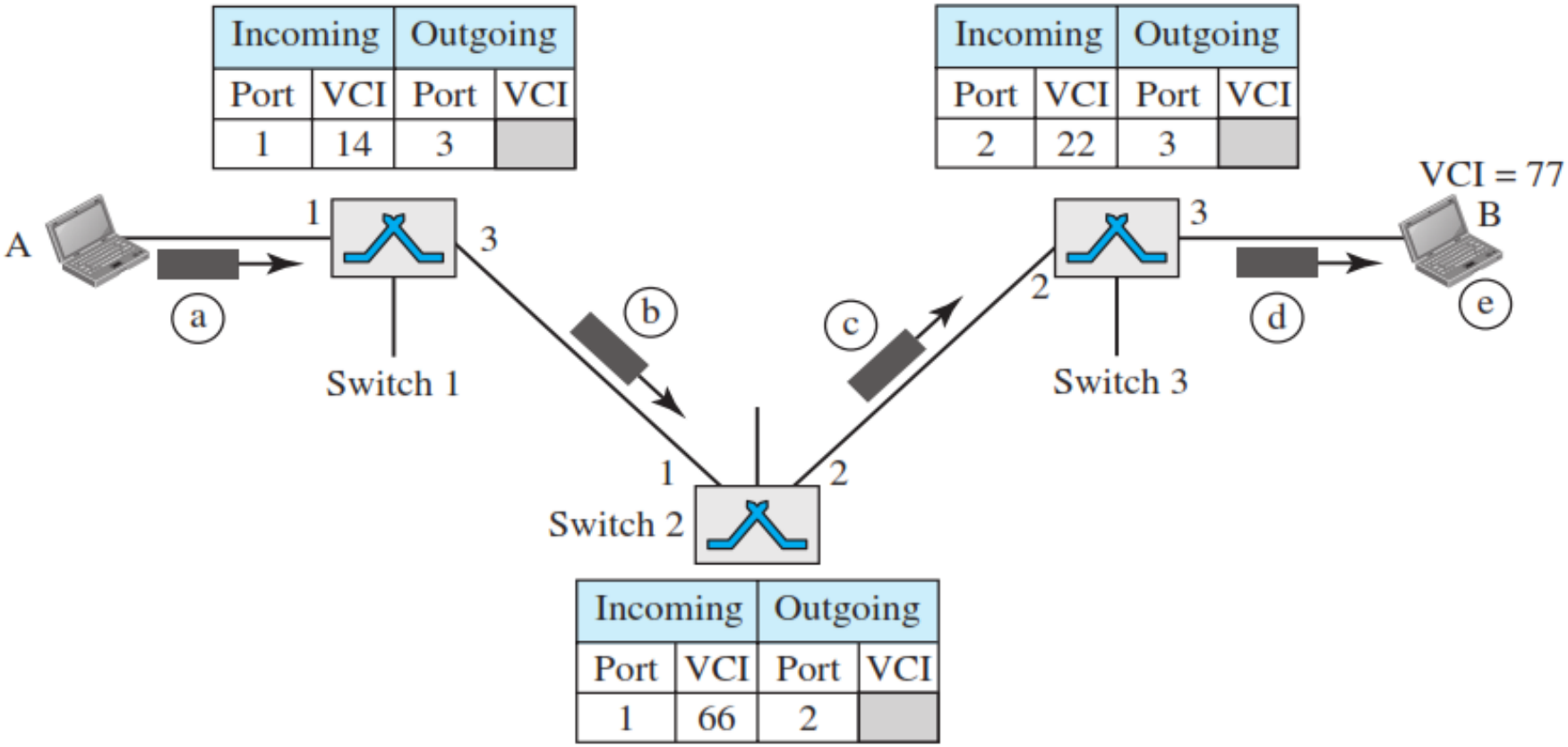
Setup Phase

- A switch creates an entry for a virtual circuit.
- Two steps are required:
  - The setup request
    - A setup request frame is sent from the source to the destination.
  - The acknowledgment
    - A acknowledgment frame is sent from the destination to the source.

PACKET-SWITCHING : Virtual-Circuit

Setup Phase

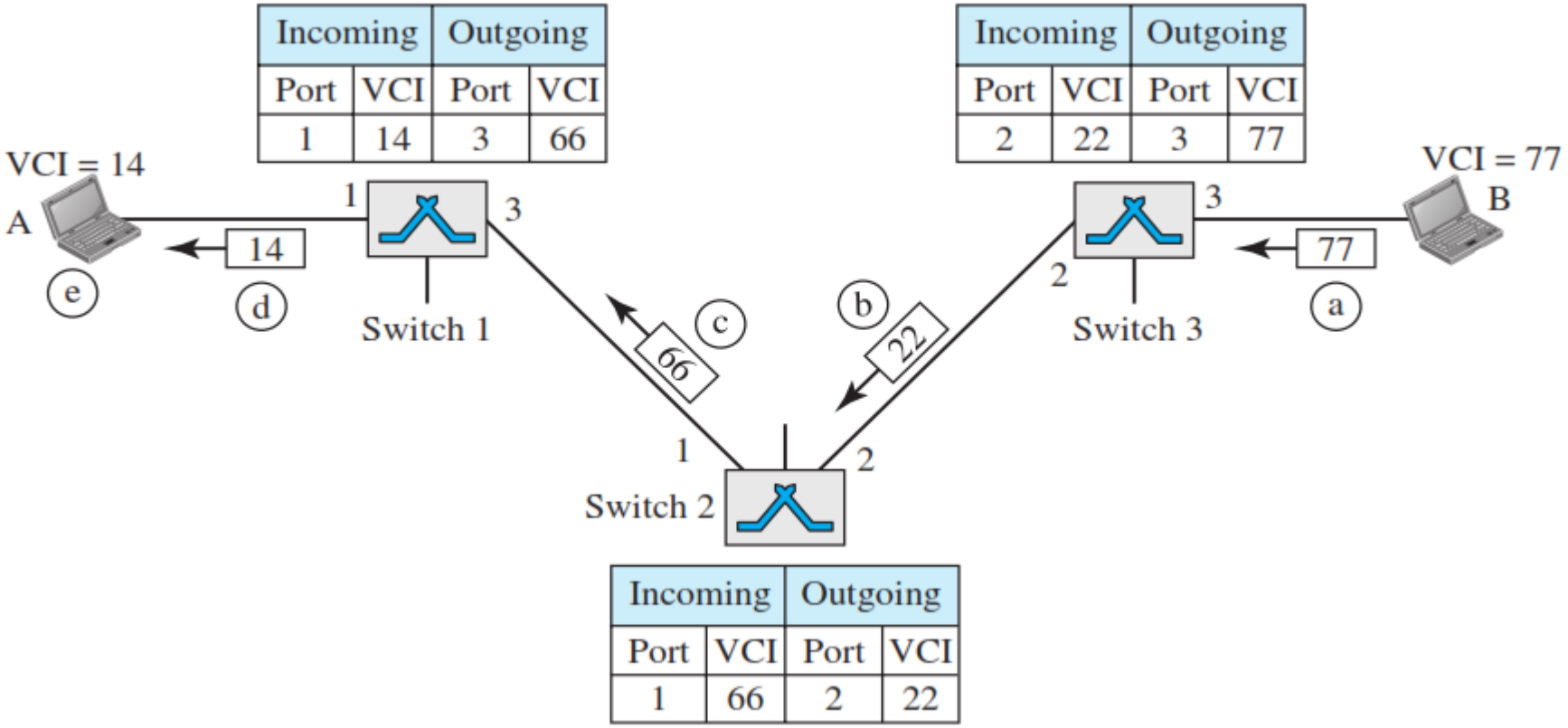
- The setup request



PACKET-SWITCHING : Virtual-Circuit

Setup Phase

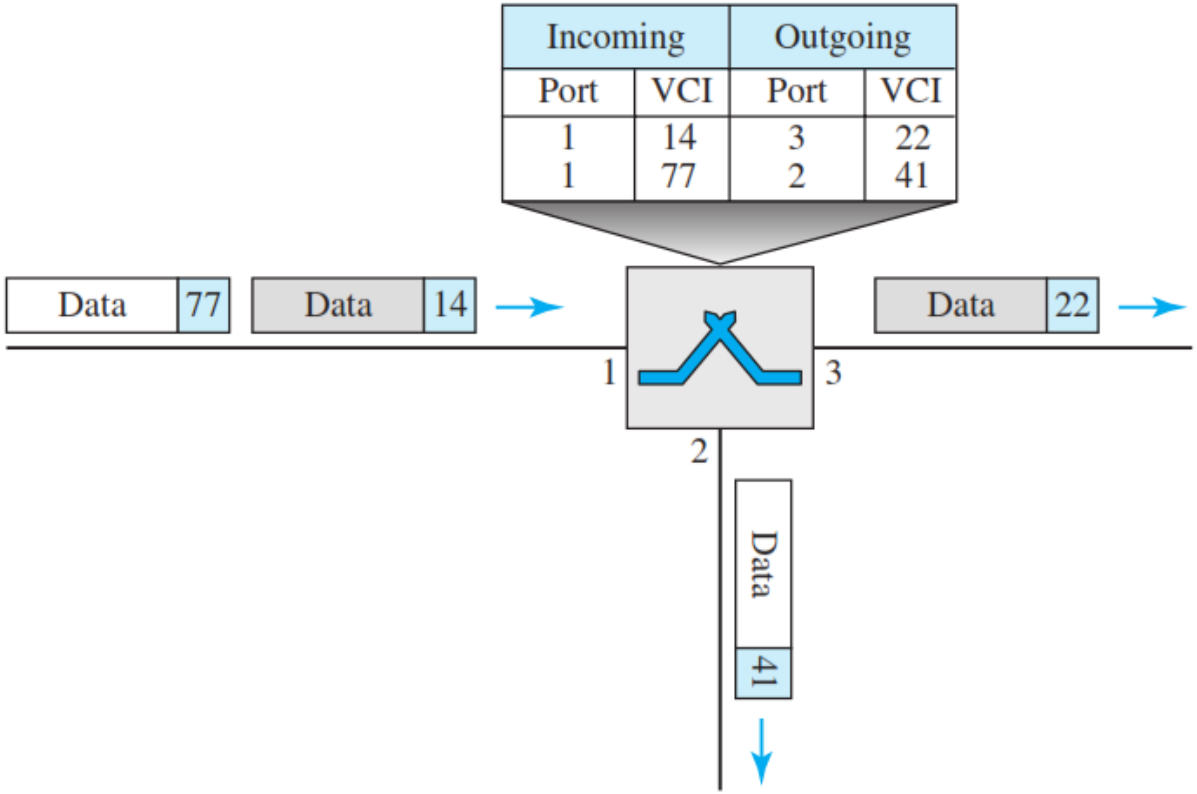
- The acknowledgment



PACKET-SWITCHING : Virtual-Circuit

Data Transfer Phase

- To transfer a frame from a source to its destination, all switches need to have a table entry for this virtual circuit.



PACKET-SWITCHING : Virtual-Circuit

Teardown Phase

- Source sends a special frame called a teardown request, after sending all frames to Destination.
- Destination responds with a teardown confirmation frame.
- All switches delete the corresponding entry from their tables.

Unicast Routing

Routing

- In an internet, the goal of the network layer is to deliver a datagram from its source to its destination or destinations.
- The routers are responsible for receiving and forwarding packets through.
- A router has a routing\forwarding table to forward a packet.
- To make the forwarding tables of the router, the Internet needs routing protocols.
- Routing Protocols executes routing algorithms.

Routing

- Routing is a process that is performed by router in order to deliver the packet by choosing an optimal path from source to destination.
- Routing metrics and costs are used for determining the best route.
- The most common metrics are
  - Hop count
  - Delay
  - Bandwidth
  - Throughput
  - Reliability

Type of Routing

- There are 3 types of routing:
  - **Static routing**
    - Network admin manually add routes to the routing table.
  - **Default Routing**
    - send all packets towards a single router (next hop).
  - **Dynamic Routing**
    - Route is updated in the routing table in response to the changes in the condition or topology of the network.

Routing Algorithms

- The routing protocol uses a routing algorithm that provides the best path from the source to the destination.
- The best path is the path that has the "least-cost path"
- The Routing algorithm is divided into two categories:
  - Adaptive Routing algorithm (Dynamic)
    - Centralized algorithm
    - Distributed algorithm
    - Isolation algorithm
  - Non-adaptive Routing algorithm (Static)
    - Flooding
    - Random walks

Routing Table

- A routing table is a set of rules, often viewed in table format.
- Routers use *Routing Tables* to determine out which interface the packet will be sent.
- A routing table lists all networks for which routes are known.
- Each router's routing table is unique and stored in the RAM of the device.

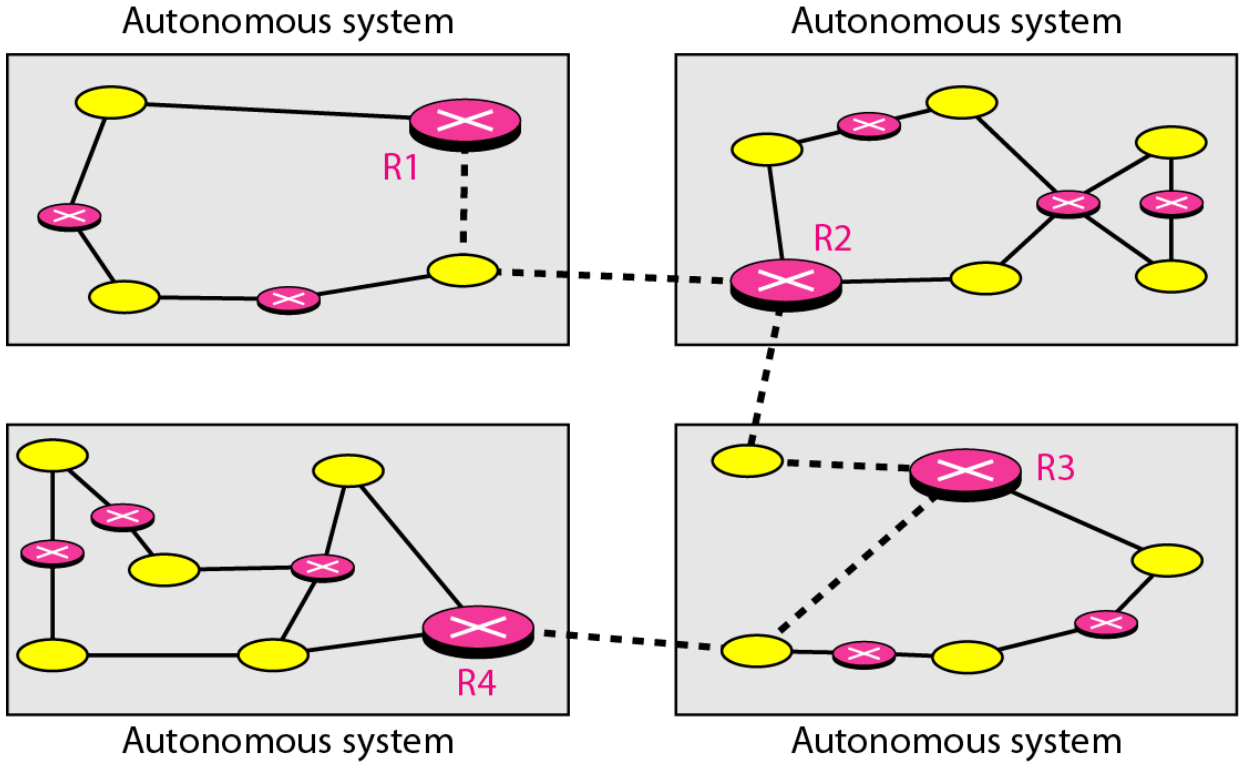
Routing Table

- Each entry in the routing table consists of
  - Network ID
  - Subnet Mask
  - Next Hop
  - Outgoing Interface
  - Metric

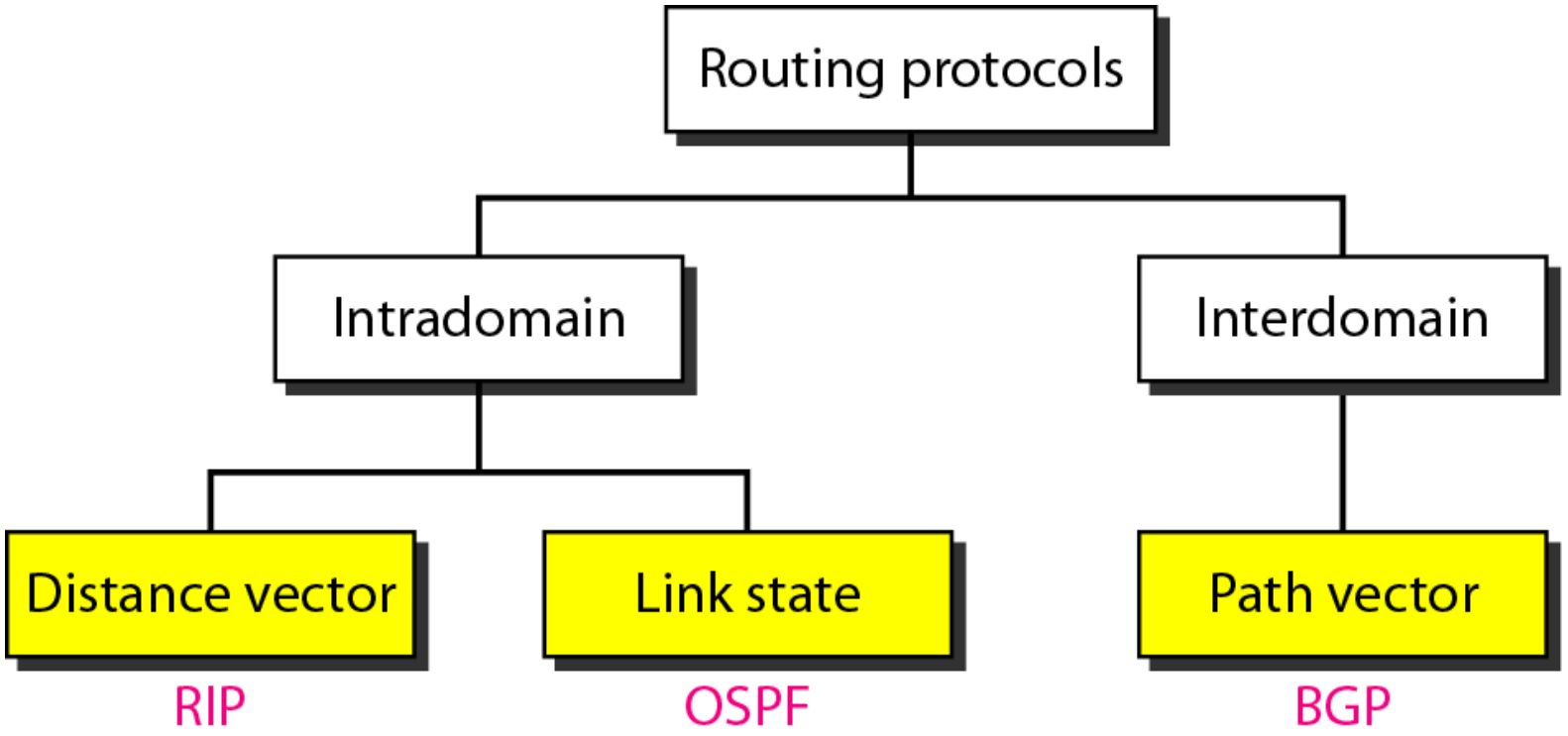
Destination	Subnet Mask	Interface
200.1.2.0	255.255.255.192	a
200.1.2.64	255.255.255.192	b
200.1.2.128	255.255.255.192	c
200.1.2.192	255.255.255.192	d
Default	0.0.0.0	e

Unicast Routing Protocols in Internet

- A routing protocol is a combination of rules and procedures that lets routers in the Internet inform each other of changes.



Unicast Routing Protocols in Internet



Distance vector routing

- Each node shares its routing table with its immediate neighbors periodically and when there is a change.
- A router continuously tells all of its neighbors what it knows about the whole internet.
- It uses Bellman-Ford equation to update distance vector.

$$D_{xy} = \min\{D_{xy}, (c_{xz} + D_{zy})\}$$

Distance-Vector Routing Algorithm

Step - 1

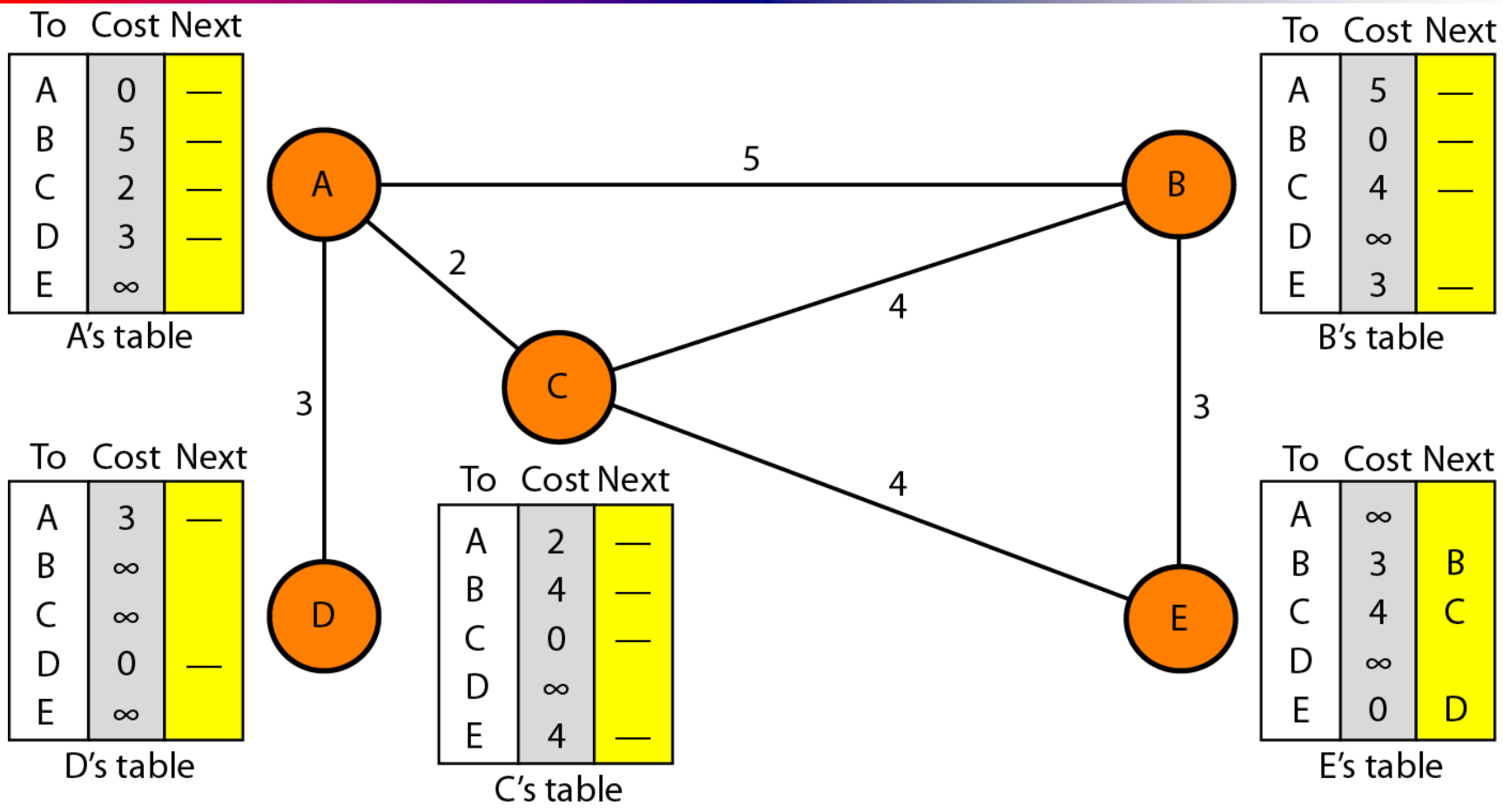
- Each router prepares its routing table by their local knowledge.
- Each router knows about-
  - All the routers present in the network
  - Distance to its neighboring routers.
- Put 0 distance for itself
- Put infinite distance for directly non reachable routers.

Distance-Vector Routing Algorithm

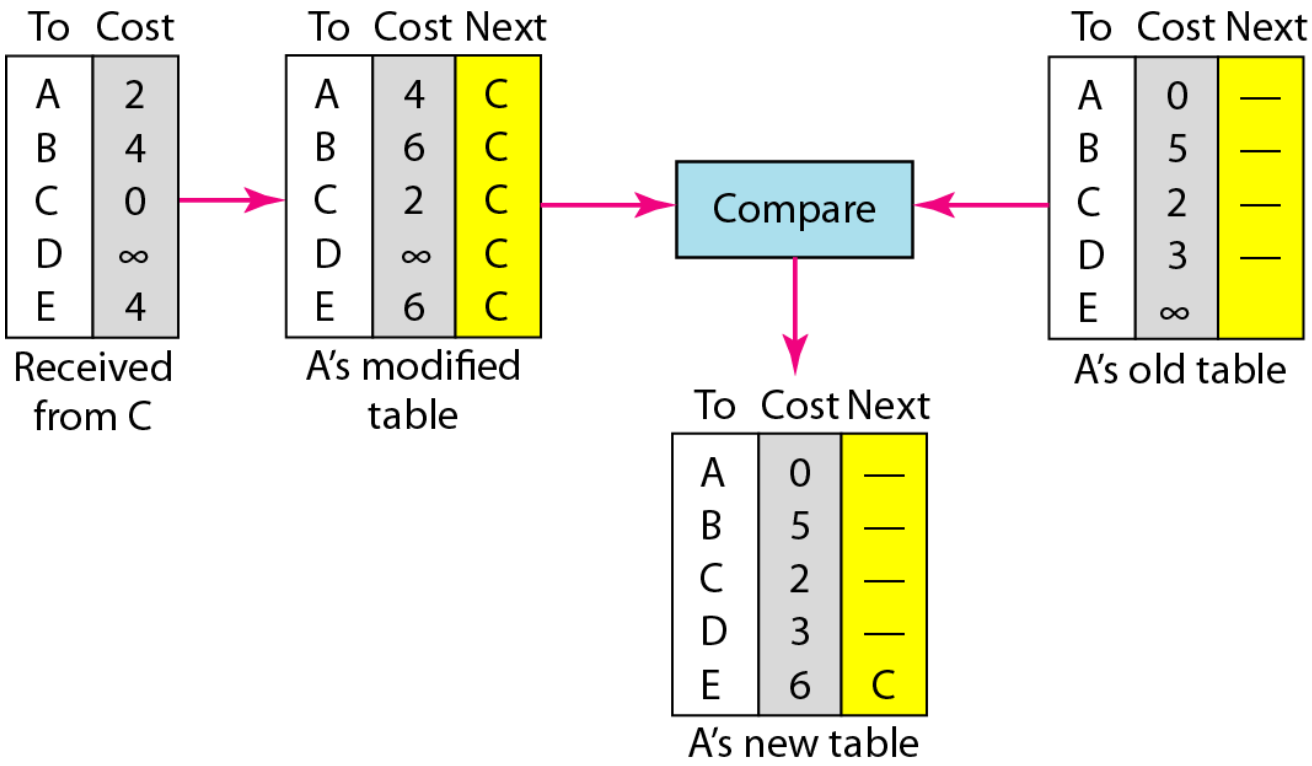
Step - 2

- Each router exchanges its distance vector with its neighboring routers.
- Each router prepares a new routing table using the distance vectors it has obtained from its neighbors.
- Send the updates to neighbors, if any change occurred.

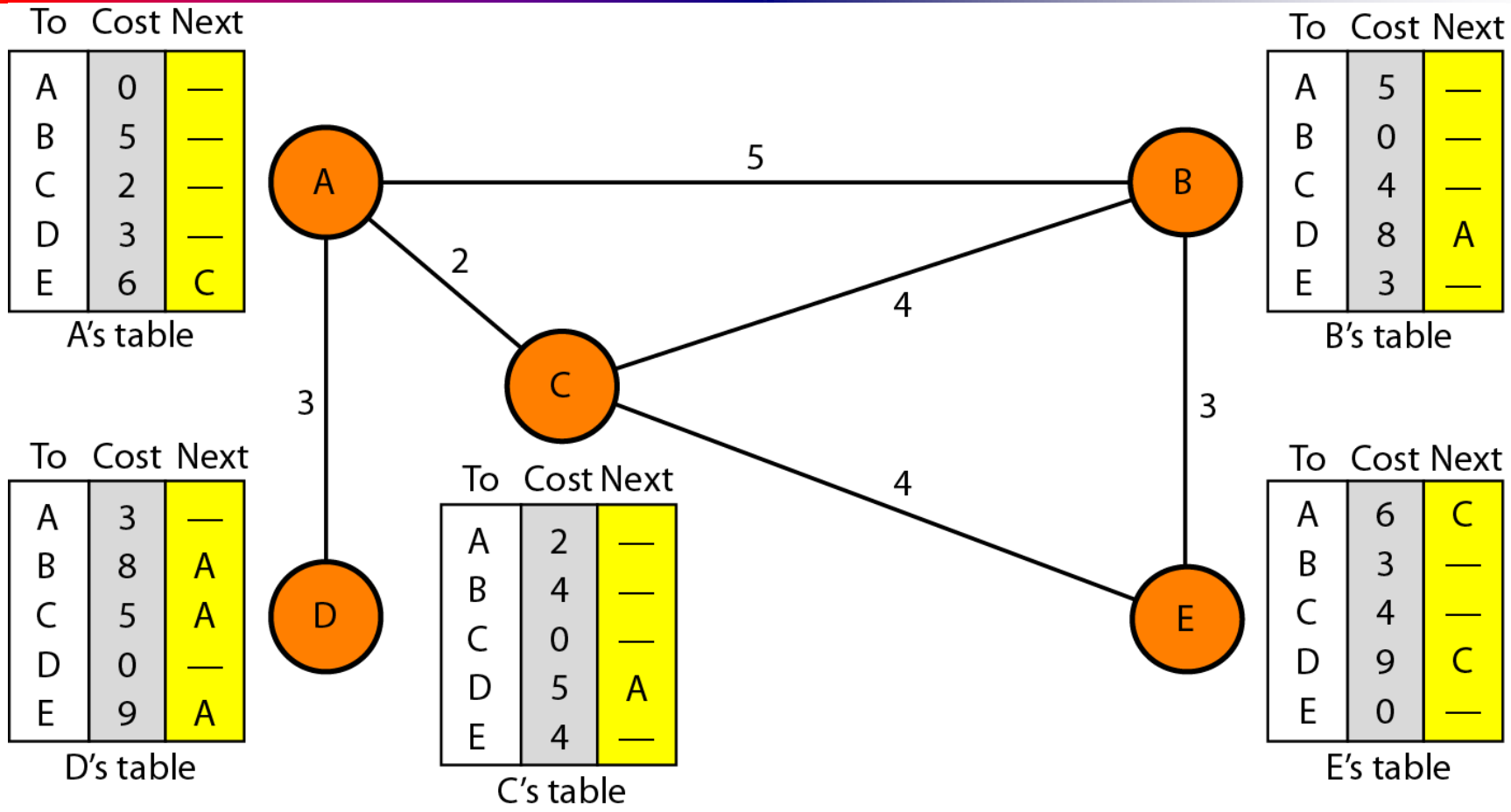
Initialization of tables in distance vector routing



Updating in distance vector routing



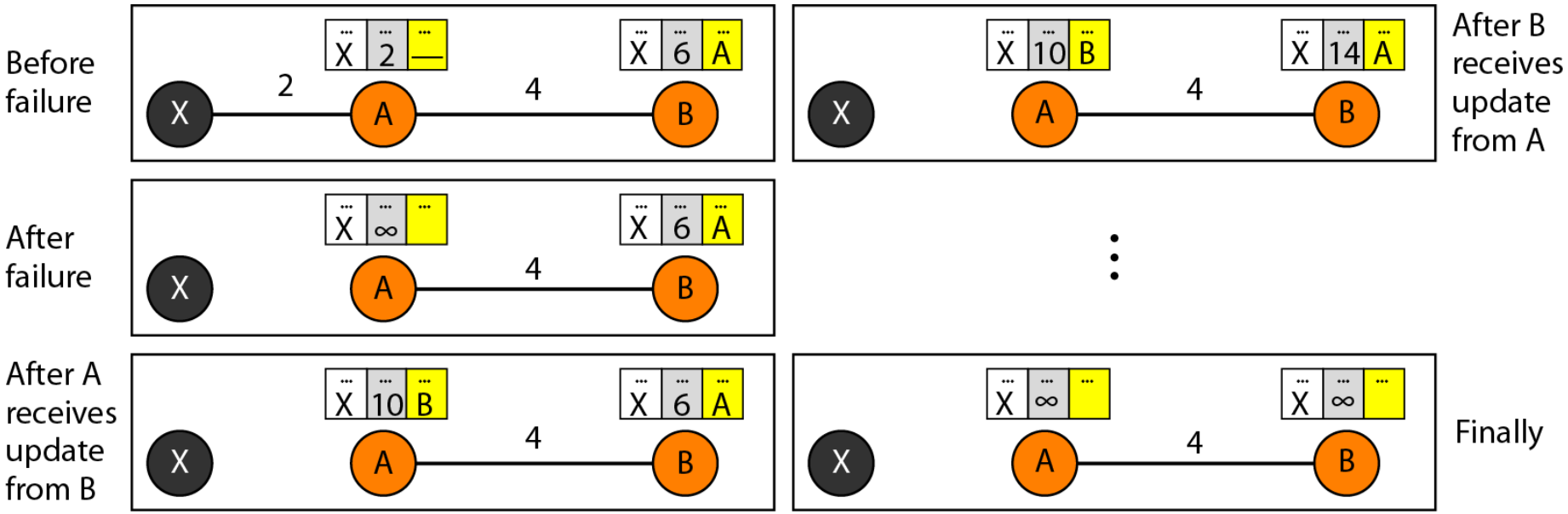
Distance vector routing tables



Count to Infinity Problem

- Any decrease in cost (good news) propagates quickly.
- Any increase in cost (bad news) will propagate slowly.
- if a link is broken (cost becomes infinity), every other router be aware about it after some time, not immediately.
- The problem is referred to as count to infinity.

Two-node instability



Count to Infinity Solution

Split Horizon

- Never send routing information back in the direction from which it was received.

Poison Reverse

- Replace the distance with infinity for a route to source of information router.

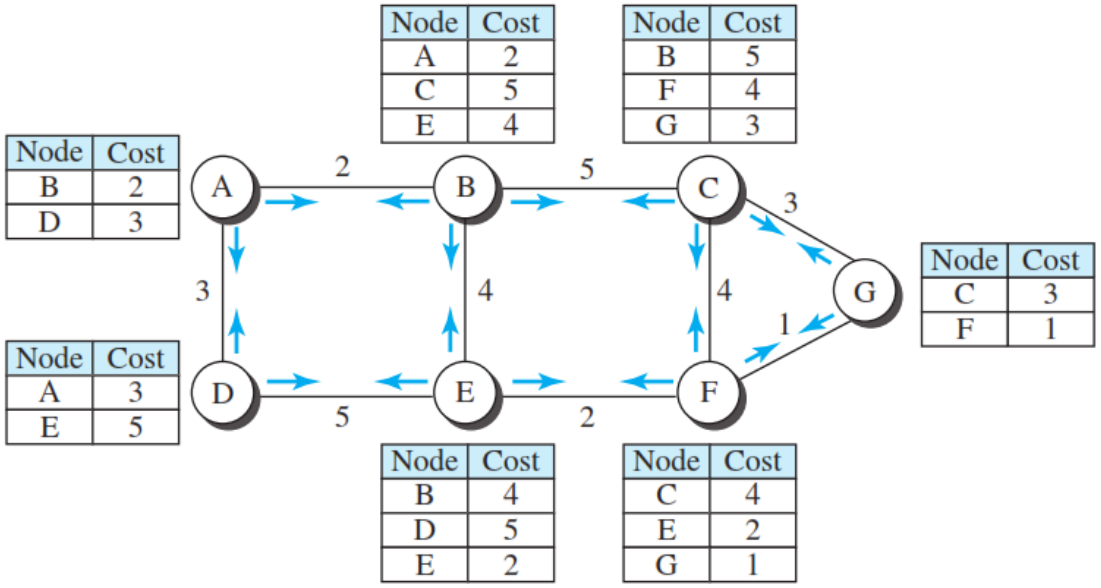
Link State Routing

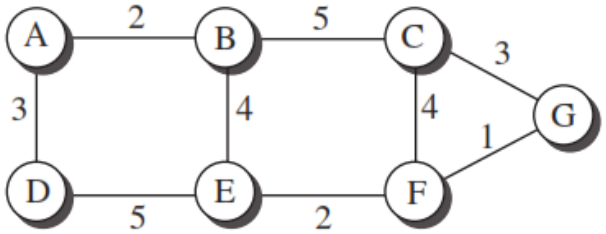
- The cost associated with an edge defines the state of the link.
- Links with lower costs are preferred.
- Each node needs to have a complete map of the network.
- The collection of states for all links is called the link-state database (LSDB).
- There is only one LSDB for the whole internet.
- The LSDB can be represented as a two-dimensional array(matrix).
- LSDB is created by a process called flooding.
- Formation of Least-Cost Trees using Dijkstra Algorithm.
- Using the least cost tree, node updates its routing table.

Link State Routing : Flooding

- Each node send some greeting messages to all its immediate neighbors to collect identity and cost information.
- Create a LSP (Link State Packet) from these two information.
- The LSP is sent out of each interface.
- When a node receives an LSP from one of its interfaces, it compares the LSP with the copy it may already have.
- If it is newer, the node discards the old LSP and keeps the received one.
- It then sends a copy of it out of each interface except the one from which the packet arrived.
- After receiving all new LSPs, each node creates the comprehensive LSDB.

Link State Routing : Flooding



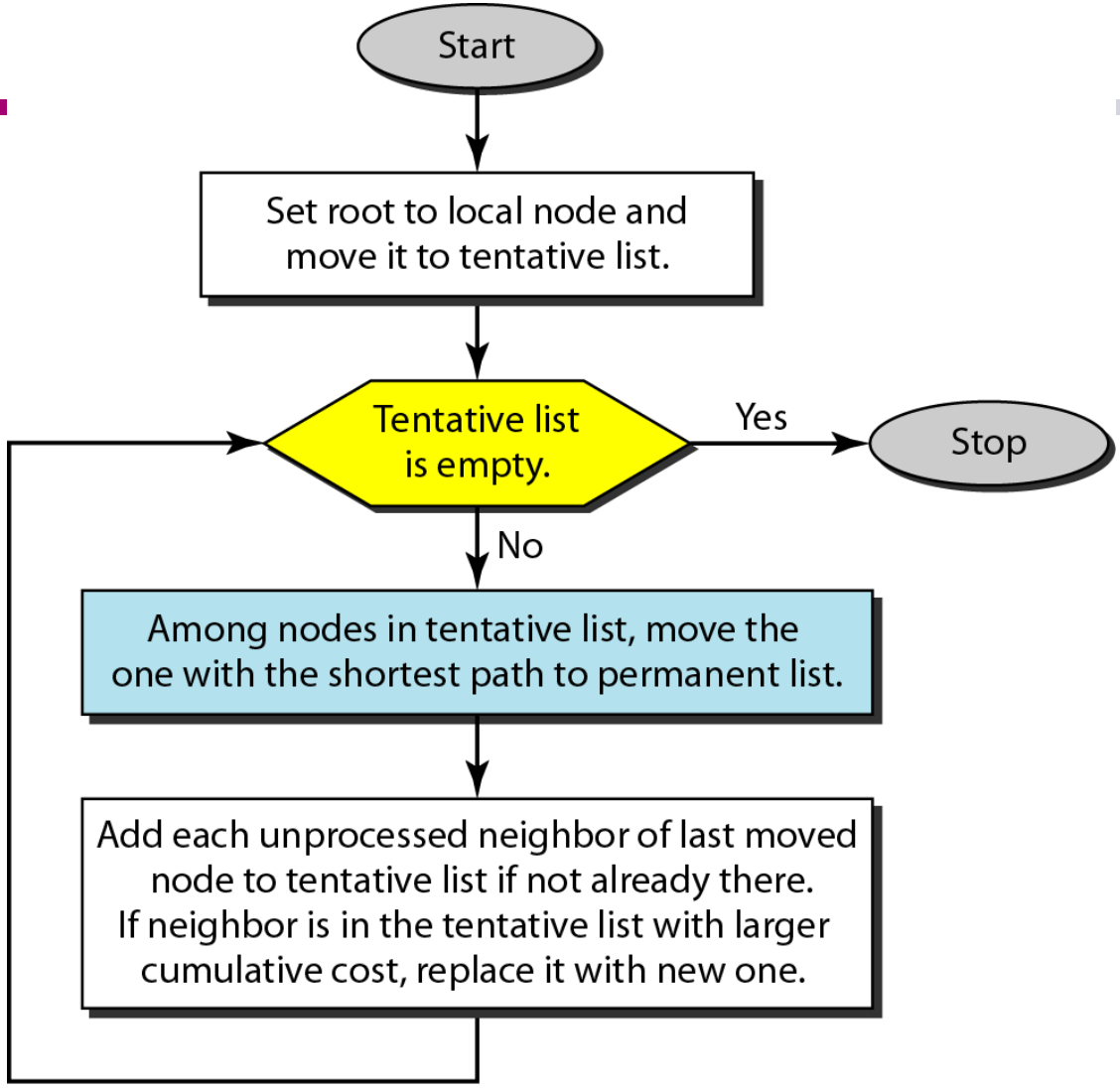


a. The weighted graph

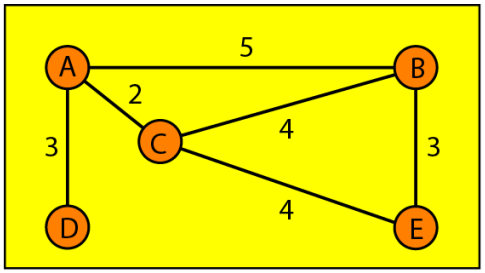
	A	B	C	D	E	F	G
A	0	2	∞	3	∞	∞	∞
B	2	0	5	∞	4	∞	∞
C	∞	5	0	∞	∞	4	3
D	3	∞	∞	0	5	∞	∞
E	∞	4	∞	5	0	2	∞
F	∞	∞	4	∞	2	0	1
G	∞	∞	3	∞	∞	1	0

b. Link state database

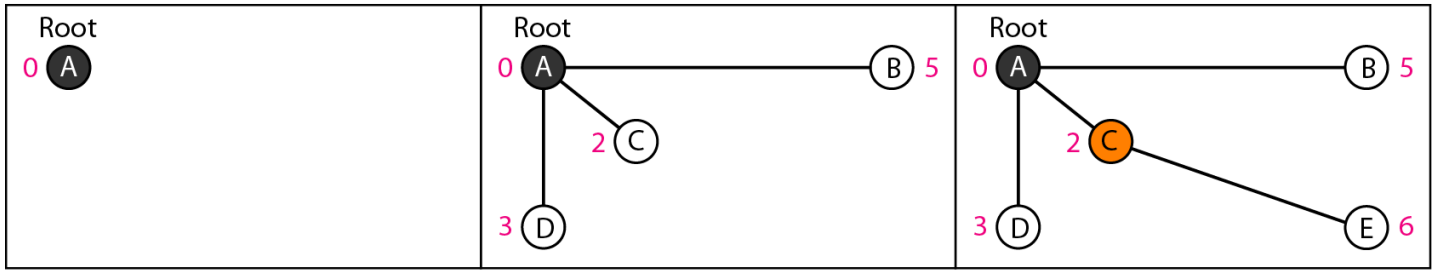
Dijkstra algorithm



Example of formation of shortest path tree



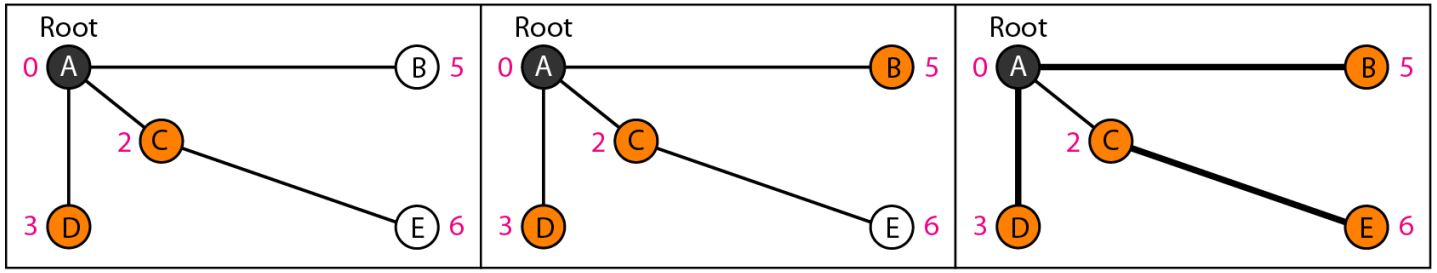
Topology



1. Set root to A and move A to tentative list.

2. Move A to permanent list and add B, C, and D to tentative list.

3. Move C to permanent and add E to tentative list.



4. Move D to permanent list.

5. Move B to permanent list.

6. Move E to permanent list (tentative list is empty).

Routing table for node A

Node	*Cost*	*Next Router*
A	0	—
B	5	—
C	2	—
D	3	—
E	6	C

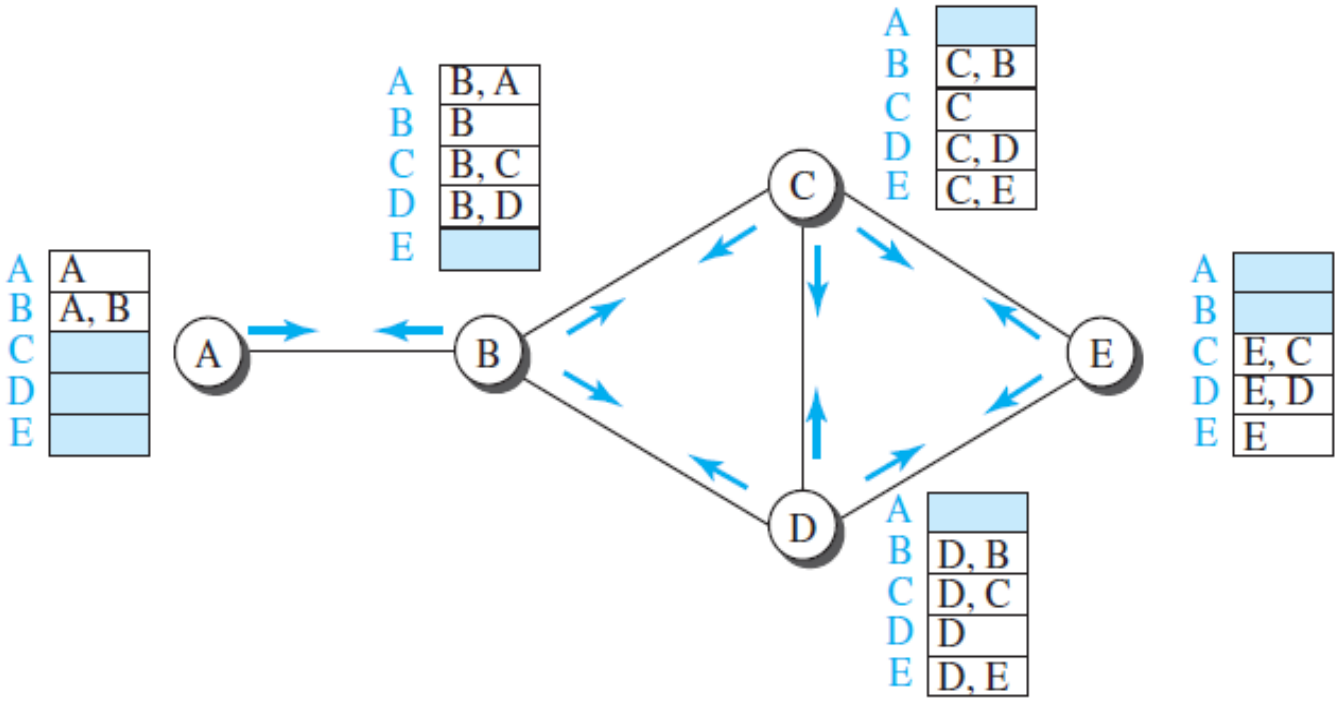
Path Vector Routing

- Path vector routing is only Interdomain routing.
- In path vector routing, the routing table contains <destination address>, <next router> and <the path to reach the destination>
- Path is defined as an ordered list of ASs that the packet need to travel through.
- Policy based routing protocols can be implemented using path vector routing.
- It is a distributive routing approach.
- Each node create a initial path vector at booting time.
- Shares the path vector with immediate neighbors.
- Nodes updates path vectors based on new information received from other nodes.

Path Vector Routing

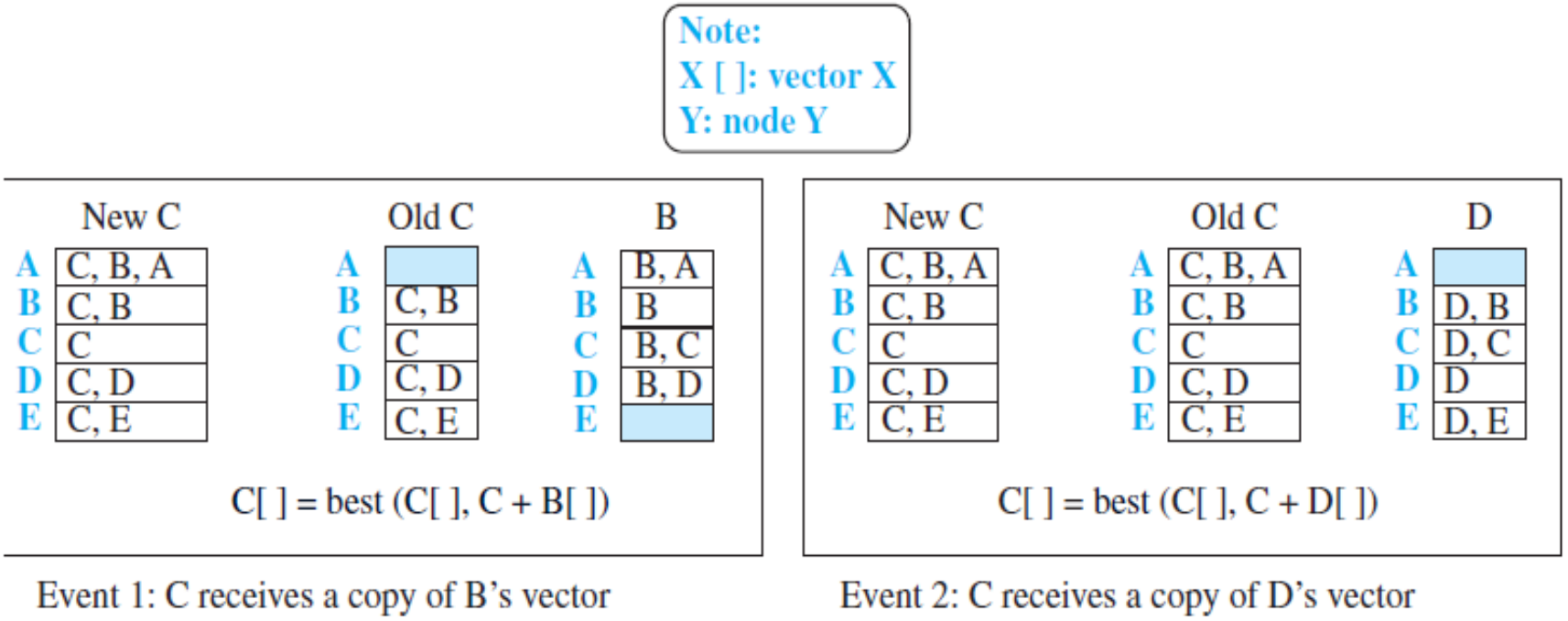
- In path-vector routing, the path from a source to all destinations is determined by the *best spanning tree*
- The best spanning tree, however, is not the least-cost tree; it is the tree determined by the source when it imposes its own policy

Path vectors made at booting time



Path Vector Routing

Updating path vectors



Event 1: C receives a copy of B's vector

Event 2: C receives a copy of D's vector

The update process is repeated forever in path vector routing

Internetworking and Internet Protocol (IP)

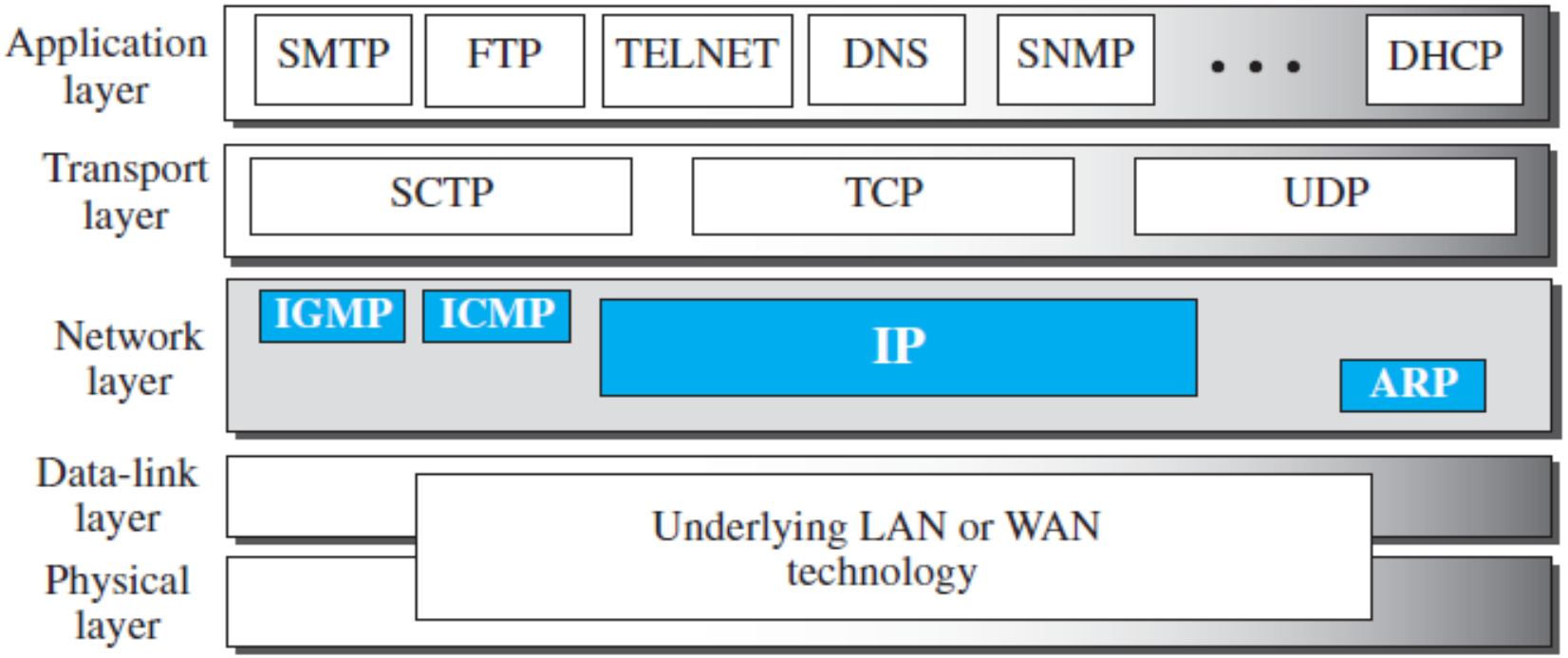
Internetworking

- **Internetworking:** Connecting networks together to make an internetwork or an internet
- Switching at the network layer in the Internet uses the datagram approach to packet switching
- Communication at the network layer in the Internet is connectionless

Internet Protocol (IP)

- The network layer in version 4 can be thought of as one main protocol and three auxiliary ones
- The main protocol, Internet Protocol version 4 (IPv4)
  - It is the delivery mechanism used by the TCP/IP protocols
  - It is responsible for packetizing, forwarding, and delivery of a packet at the network layer
- The Internet Control Message Protocol version 4 (ICMPv4) helps IPv4 to handle some errors that may occur in the network-layer delivery
- The Internet Group Management Protocol (IGMP) is used to help IPv4 in multicasting
- The Address Resolution Protocol (ARP) is used to glue the network and data-link layers in mapping network-layer addresses to link-layer addresses

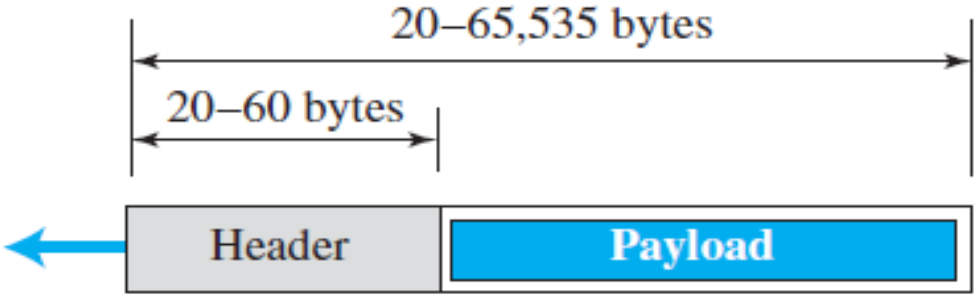
Position of IP and other network-layer protocols in TCP/IP protocol suite



Internet Protocol Version 4 (IPv4)

- IPv4 is an unreliable datagram protocol
- IPv4 is also a connectionless protocol
- It uses the datagram approach
  - Each datagram is handled independently, and
  - Each datagram can follow a different route to the destination
- Packets used by the IP are called *datagrams*
- A datagram is a variable-length packet consisting of two parts: header and payload (data).
- The header is 20 to 60 bytes in length
  - It contains information essential to routing and delivery

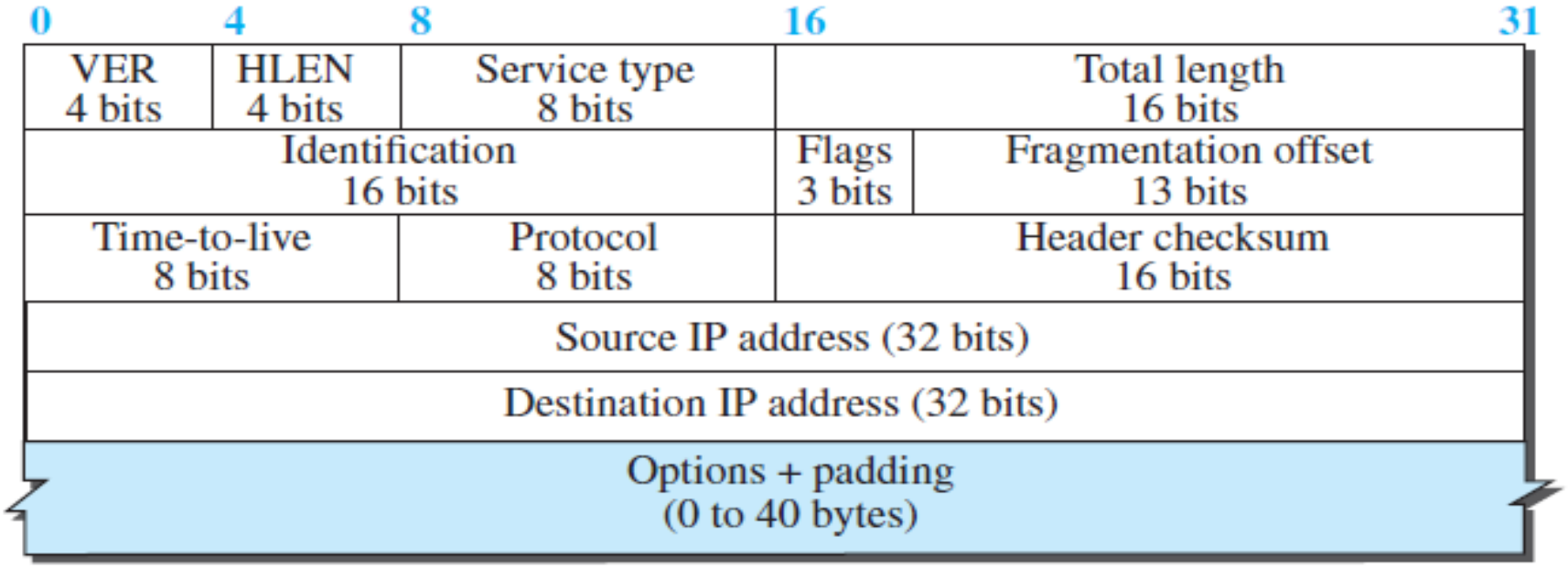
IPv4 Datagram Format



a. IP datagram

Legend

VER: version number
HLEN: header length
byte: 8 bits

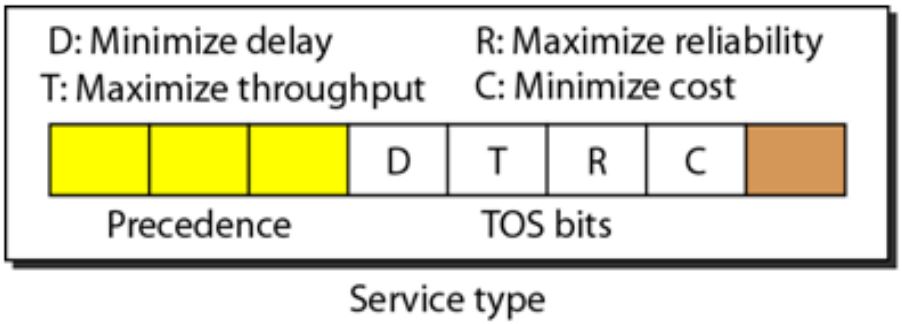


b. Header

IPv4 Datagram Format

- **VER:** The 4-bit version number (VER) field defines the version of the IPv4 protocol, here its value is 4 (0100).
- **HLEN:** The 4-bit header length (HLEN) field defines the total length of the datagram header in 4-byte words.
- **Service Type:** In the original design of the IP header, this field was referred to as type of service (TOS), which defined how the datagram should be handled

Service type or differentiated services



The precedence subfield was part of version 4, but never used

IPv4 Datagram Format

Types of service

TOS Bits	*Description*
0000	Normal (default)
0001	Minimize cost
0010	Maximize reliability
0100	Maximize throughput
1000	Minimize delay

IPv4 Datagram Format

- **Total Length:** This 16-bit field defines the total length (header plus data) of the IP datagram in bytes
  - A 16-bit number can define a total length of up to 65,535 (when all bits are 1s)
- **Identification, Flags, and Fragmentation Offset:** These three fields are related to the fragmentation
- **Time-to-live:** Due to some malfunctioning of routing protocols a datagram may be circulating in the Internet, visiting some networks over and over without reaching the destination
  - The time-to-live (TTL) field is used to control the maximum number of hops (routers) visited by the datagram
  - Each router that processes the datagram decrements this number by one
  - If this value, after being decremented, is zero, the router discards the datagram

IPv4 Datagram Format

- **Protocol:** The Internet authority has given any protocol (TCP, UDP or any other transport layer protocol) that uses the service of IP a unique 8-bit number which is inserted in the protocol field
  - When the payload is encapsulated in a datagram at the source IP, the corresponding protocol number is inserted in this field

Some protocol values

ICMP	01
IGMP	02
TCP	06
UDP	17
OSPF	89

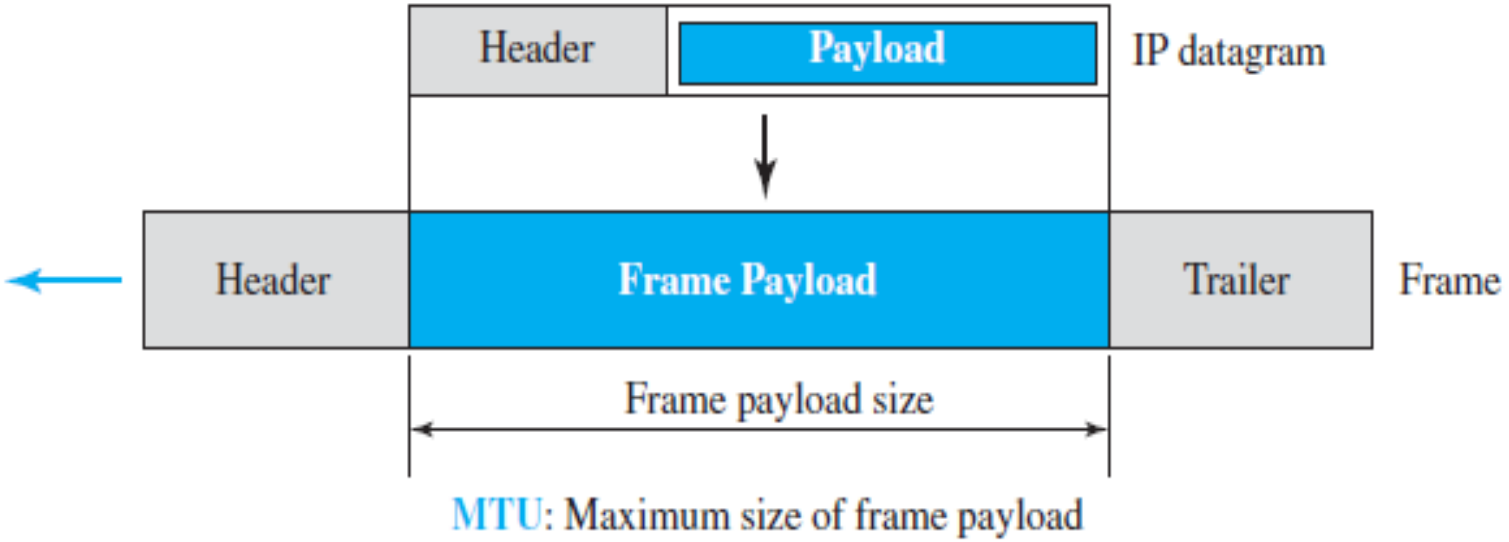
IPv4 Datagram Format

- **Header checksum:** IP adds a header checksum field to check the header because if the header field is corrupted, the payload may be delivered to the wrong destination.
- **Source and Destination Addresses:** These 32-bit source and destination address fields define the IP address of the source and destination respectively.
- **Options:** A datagram header can have up to 40 bytes of options. Options can be used for network testing and debugging.
- **Payload:** Payload, or data, is the main reason for creating a datagram. Payload is the packet coming from other protocols (from upper layer) that use the service of IP.

IPv4 Fragmentation

Maximum Transfer Unit (MTU)

- Each link-layer protocol has its own frame format. One of the features of each format is the maximum size of the payload that can be encapsulated



IPv4 Fragmentation

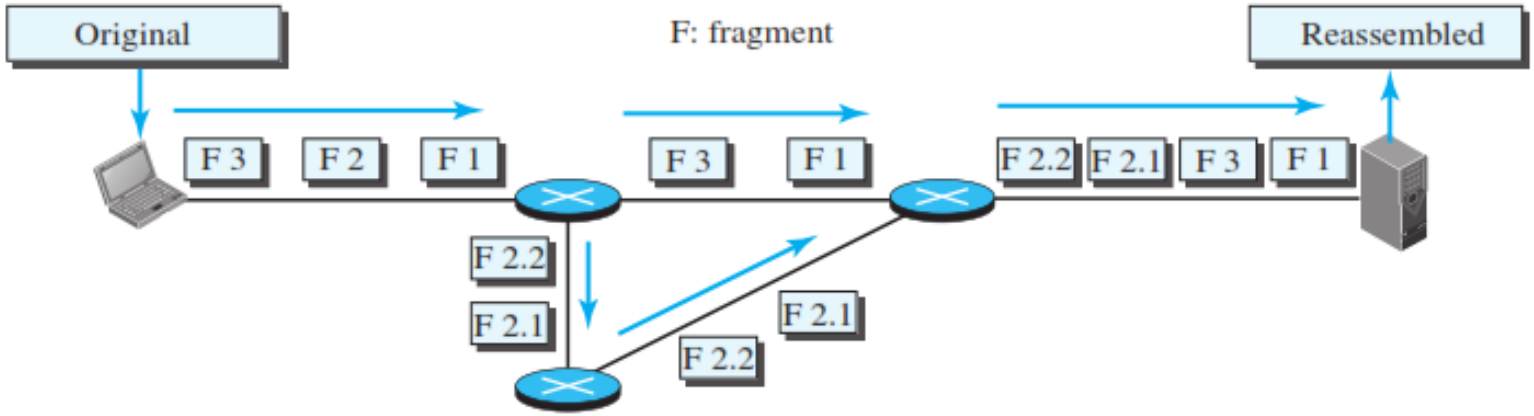
- **Fragmentation** is done by the network layer when the size of a datagram is greater than maximum size of data that can be held a frame i.e., its Maximum Transfer Unit (MTU)
- The network layer divides the datagram received from transport layer into fragments so that data flow is not disrupted.
- **Fragmentation:** Dividing the datagram into smaller parts
- When a datagram is fragmented, each fragment has its own header with most of the fields repeated, but some have been changed.
- A datagram can be fragmented by the source host or any router in the path.
- The reassembly of the datagram, however, is done only by the destination host
- The host or router that fragments a datagram must change the values of the fields: identification, flags, fragmentation offset, and total length

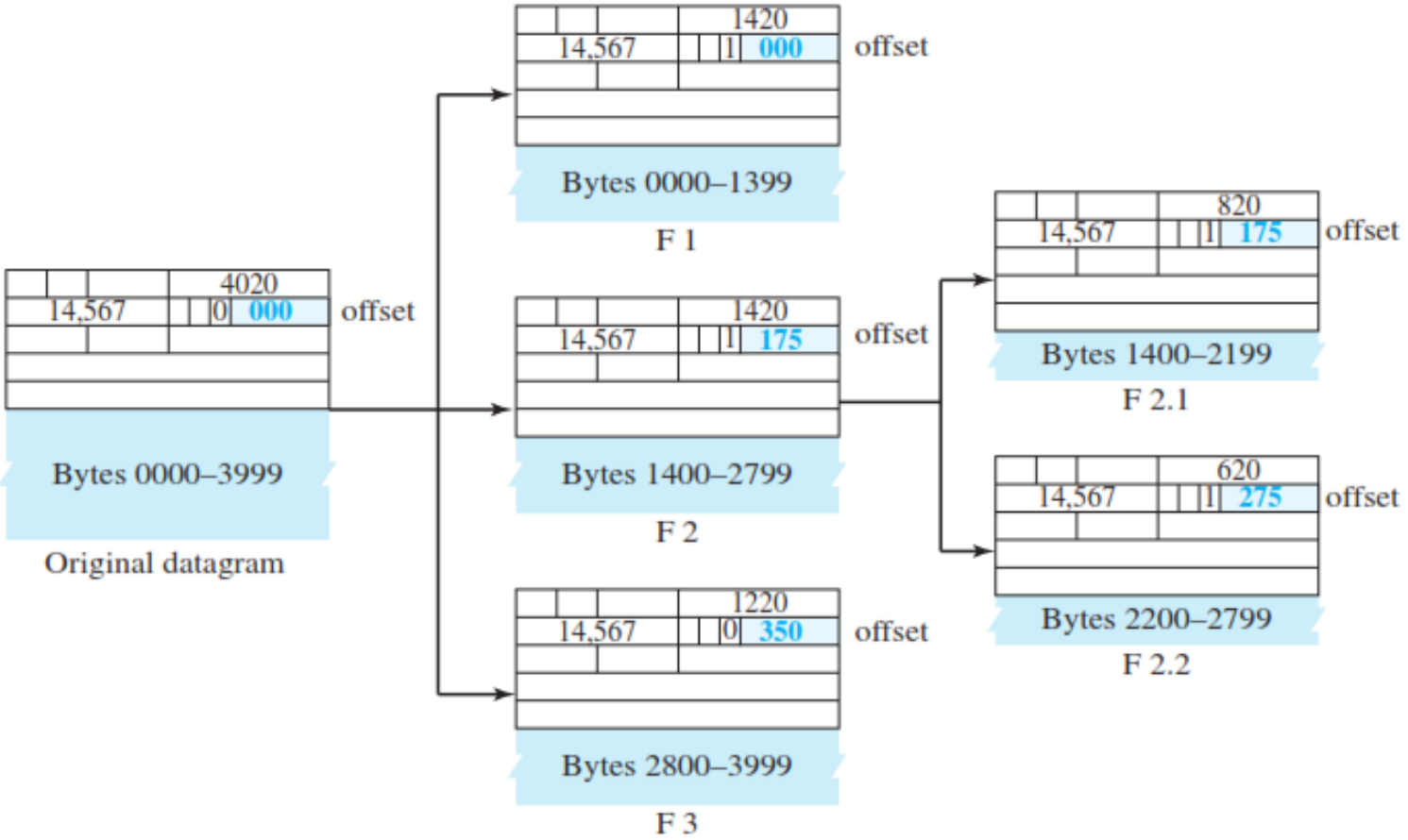
IPv4 Fragmentation

Fields Related to Fragmentation

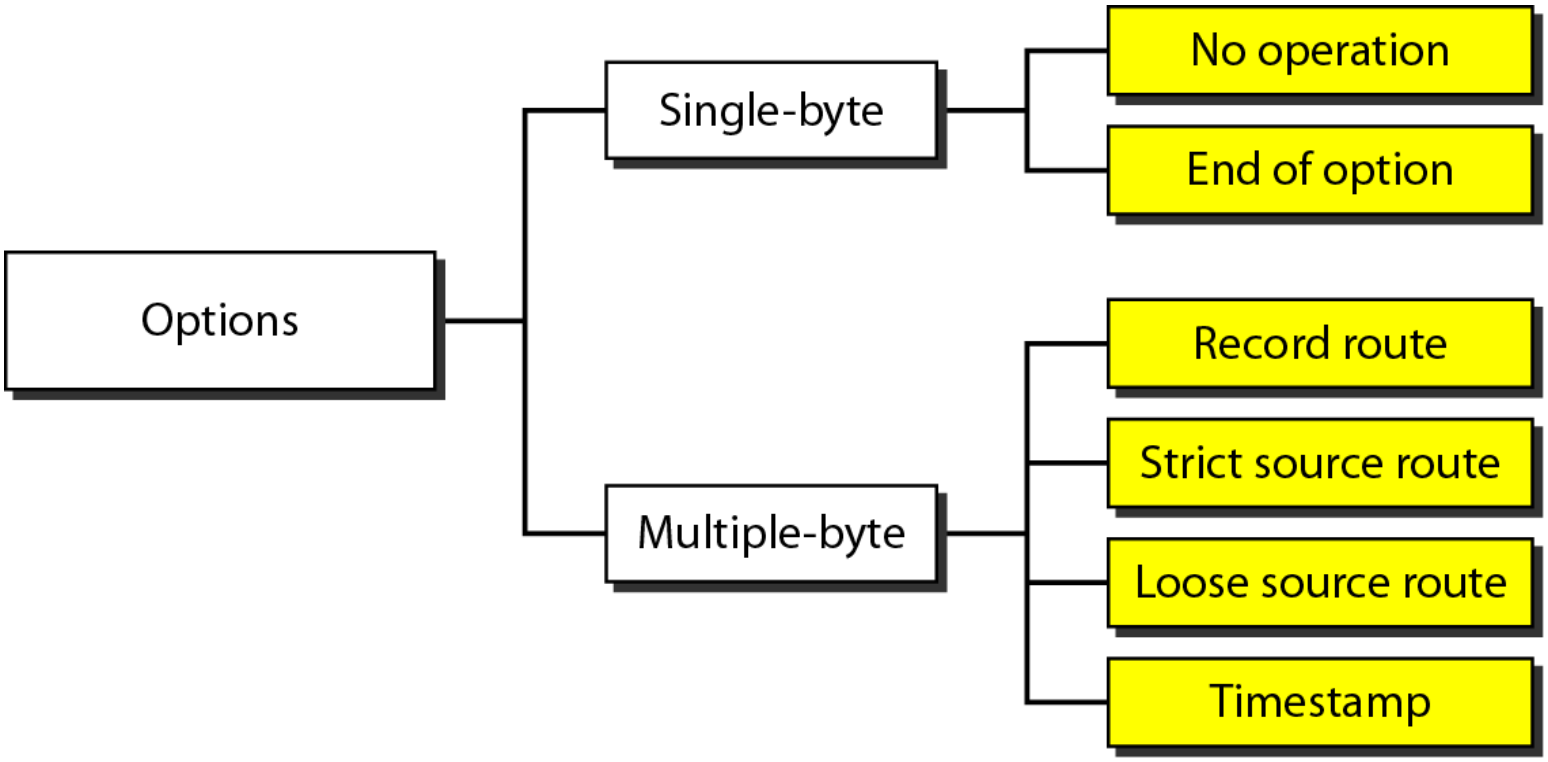
- Identification
  - 16 bit Unique ID for a datagram
  - All fragments have the same identification number
- Flags
  - The 3-bit flags field defines three flags
  - The leftmost bit is reserved (not used).
  - (D bit) is called the do not fragment bit.
  - (M bit) is called the more fragment bit.

- Fragmentation Offset
  - shows the relative position of this fragment with respect to the whole datagram.
  - measured in units of 8 bytes.


Original
F: fragment
Reassembled
F 3
F 2
F 1
F 3
F 1
F 2.2
F 2.1
F 3
F 1
F 2.2
F 2.1
F 2.1
F 2.2


4020
14,567
0
000
offset
Bytes 0000–3999
Original datagram
1420
14,567
1
000
offset
Bytes 0000–1399
F 1
1420
14,567
1
175
offset
Bytes 1400–2799
F 2
1220
14,567
0
350
offset
Bytes 2800–3999
F 3
820
14,567
1
175
offset
Bytes 1400–2199
F 2.1
620
14,567
1
275
offset
Bytes 2200–2799
F 2.2

Taxonomy of options in IPv4



An IPv4 packet has arrived with the first 8 bits as $(01000010)_2$ The receiver discards the packet. Why?

In an IPv4 packet, the value of HLEN is $(1000)_2$. How many bytes of options are being carried by this packet?

In an IPv4 packet, the value of HLEN is 5, and the value of the total length field is $(0028)_{16}$. How many bytes of data are being carried by this packet?

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?