

GATE

2019

**COMPUTER
NETWORK**

**COMPUTER SCIENCE &
INFORMATION TECHNOLOGY**



ECG
Publications



A Unit of **ENGINEERS CAREER GROUP**

Head Office: S.C.O-121-122-123, 2nd Floor, Sector-34/A, Chandigarh-160022

Website: www.engineerscareergroup.in **Toll Free:** 1800-270-4242

E-Mail: ecgpublications@gmail.com | info@engineerscareergroup.in

GATE-2019: Computer Network | Detailed theory with GATE previous year papers and detailed solutions.

©Copyright @2016 by ECG Publications

(A unit of **ENGINEERS CAREER GROUP**)

All rights are reserved to reproduce the copy of this book in the form storage, introduced into a retrieval system, electronic, mechanical, photocopying, recording, screenshot or any other form without any prior written permission from ECG Publications (A Unit of **ENGINEERS CAREER GROUP**).

First Edition: 2016

Price of Book: INR 310/-

ECG PUBLICATIONS (A Unit of **ENGINEERS CAREER GROUP**) collected and providing data like: theory for different topics or previous year solutions very carefully while publishing this book. If in any case inaccuracy or printing error may be found or occurred then **ECG PUBLICATIONS** (A Unit of **ENGINEERS CAREER GROUP**) owes no responsibility. The suggestions for inaccuracies or printing error will always be welcome by us.

CONTENTS

CHAPTER	PAGE
1. BASIC CONCEPTS.....	1-23
2. PHYSICAL LAYER.....	24- 48
3. DATA LINK LAYER.....	49-100
4. NETWORK LAYER.....	101-157
5. TRANSPORT LAYER.....	158-175
6. APPLICATION LAYER	176-187
7. NETWORK SECURITY.....	188-210

CHAPTER - 1

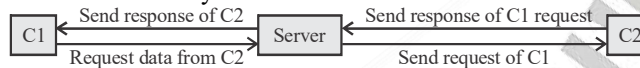
BASIC CONCEPTS

1.1 COMPUTER NETWORK

It is connection of computer used for exchanging data such as text, audio, video etc.

1.1.1 Network

It is set of devices/nodes connected by communication link. Its basic structure is as follows:

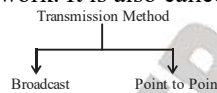


C1, C2 are Machine-1, Machine-2

Server is not visible to anyone but everyone request or respond through server. For example, we can download through youtube only when youtube server has uploaded data on website.

1.2 TRANSMISSION METHODS

It is the way of exchanging data on network. It is also called transmission technology.



1.2.1 Broadcast

1. Whenever any message is sent, it is broadcast to all machines on the network. But finally received by the intended recipient, whose address gets matched with the address field of the message and all other machines whose address does not match, they just ignore the message.
2. It has single communication channel among all the machines on the network.
3. Smaller network usually uses broadcast transmission method.

Example

Dish TV provide us access of all channels at a time but we can access one channel at a time.



Ethernet (IEEE 802.3), Token ring (IEEE 802.05) LAN uses broadcast technology.

1.2.2 Point-to-Point

1. The message is passed through one or more intermediate machines when it is sent by sender to reach the destination.
2. It consists of many connections between individual pair of machines.
3. Entire capacity of link is shared between two communicating devices.
4. Point to point transmission is called as unicasting.
5. Larger network usually uses point to point transmission method.

Example

- (i) Remote & TV uses point to point transmission technology.
- (ii) Telephony communication

ASSIGNMENT

1. Flow control is the responsibility of
(a) Data link layer
(b) Transport layer
(c) Both (a) and (b)
(d) Application layer
2. Which of the following address cannot be changed?
(a) Hardware address
(b) Logical address
(c) Both (a) and (b)
(d) None of these
3. What is the Protocol Data (PDU) employed at the Data Link Layer?
(a) Bits
(b) Frames
(c) Packets
(d) Segments
4. Host to Host or Hop to Hop delivery is related to
(a) Data link layer
(b) Network layer
(c) Transport layer
(d) All of these
5. Which one of the following OSI layers performs error checking of data?
(a) Network (b) Transport
(c) Data link (d) Physical
6. The following are names of data units in each layer, which choice is not a correct match?
(a) Frame: Data Link
(b) Packet: Network
(c) Segment: Transport
(d) Frame: Network
7. Which layer of the OSI model is responsible for routing packets from one network to another?
connectionless and connection-oriented communication in the transport layer.
8. In the TCP/IP protocol suite, the physical layer is responsible for
(a) Exchange of programs over the physical medium
(b) The movement of bits over the physical medium.
(c) Operations of protocol over the physical medium
(d) Addressing over physical medium
9. Which of the following layer is responsible for moving frames from one hop to the next?
(a) Transport layer (b) Network layer
(c) Data - link layer (d) Physical layer
10. In the TCP/IP protocol suite, which of the following is a network layer protocol?
(a) The Secure shell (SSH)
(b) The internet Protocol (IP)
(c) The Stream Control Transmission Protocol (SCTP)
(d) User Datagram Protocol (UDP)
11. In the TCP/IP protocol suite, at which of the following layer a logical address is an identifier?
(a) Data link layer (b) Application layer
(c) Network layer (d) Transport layer
12. The Internet Protocol (IP) is
(a) A connection-oriented protocol
(b) A reliable protocol
(c) An unreliable protocol
(d) A reliable and connection-oriented protocol
13. Consider the following two tasks:
T₁: Dividing the transmitted bit stream into frames
T₂: Determining which route through the subnet to use

GATE QUESTIONS

1. In the following pairs of OSI protocol layer/sub-layer and its functionality, the incorrect pair is

[GATE - 2014]

- (a) Network layer and Routing
- (b) Data Link layer and Bit synchronization
- (c) Transport layer and End - to - end process communication
- (d) Medium Access Control sub-layer and Channel sharing

2. Assume that source S and destination D are connected through two intermediate routers labels R. Determine how many times each packet has to visit the network layer and the data link layer during a transmission from S to D.



[GATE - 2013]

- (a) Network layer-4 times and Data link layer-4 times
- (b) Network layer-4 times and Data link layer-3 times
- (c) Network layer-4 times and Data link layer-6 times
- (d) Network layer-2 times and Data link layer-6 times

3. The Protocol Data Unit (PDU) for the application layer in the Internet stack is

[GATE - 2012]

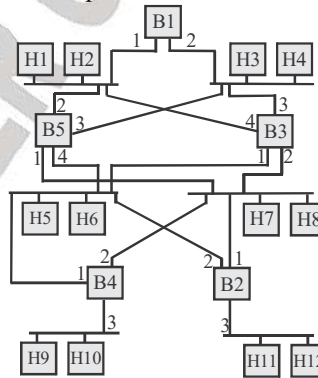
- (a) Segment
- (b) Datagram
- (c) Message
- (d) Frame

Linked Statement for Q. 4 & Q.5

Consider the diagram shown below where a number of LANs are connected by (transparent) bridges. In order to avoid packets looping through circuits in the graph, the bridges organize themselves in a spanning tree. First, the root bridge is identified as the bridge with the least serial number. Next, the root sends out

(one or more) data units to enable the setting up of the spanning tree of shortest paths from the root bridge to each bridge.

Each bridge identifies a port (the root port) through which it will forward frames to the root bridge. Port conflicts are always resolved in favour of the port with the lower index value. When there is a possibility of multiple bridges forwarding to the same LAN (but not through the root port), ties are broken as follows: bridges closest to the root get preference and between such bridges, the one with the lowest serial number is preferred.



4. For the given connection of LAN by bridges, which one of the following choices represents the depth first traversal of the spanning tree of bridges

[GATE - 2006]

- (a) B1, B5, B3, B4, B2
- (b) B1, B3, B5, B2, B4
- (c) B1, B5, B2, B3, B4
- (d) B1, B3, B4, B5, B2

5. Consider the correct spanning tree for the previous question. Let host H1 send out a broadcast ping packet. Which of the following options represents the correct forwarding table on B3?

SOLUTIONS

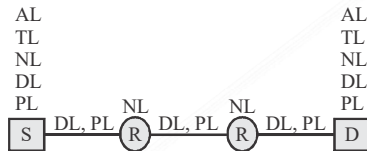
Sol 1. (b)

Bit synchronization is associated with physical layer because physical layer deals with bits.

Sol 2. (c)

As the connectivity shown include router between source and destination. And router does processing up to the network layer.

So layers accessed by source, destination and router and links are shown below:

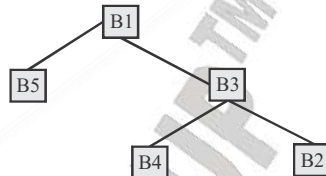


So data link layer is visited 6 times and network layer is visited 4 times.

Sol 3. (c)

Sol 4. (a)

Minimum spanning tree is formed by eliminating edges that will create loops/cycle. So, minimum spanning tree is



So by using depth first traversal of spanning tree of bridges gives traversal in order B1B5B3B4B2.

Sol 5. (a)

Based on above traversal Forwarding table is correctly given in option (a).

Sol 6. (a)

Sol 7. (d)

The transport protocol provides an end to end connectivity that shields network layer protocol from the details of the intervening network or networks. A transport protocol can be either connection oriented such as TCP, or connectionless such as UDP.

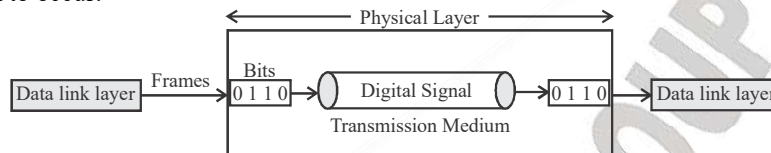
ENGINEERS CAREER GROUP™

CHAPTER - 2

PHYSICAL LAYER

2.1 INTRODUCTION

Physical layer coordinates the functions required to carry a bit stream over a physical medium. It defines the procedures and functions that physical devices and interfaces have to perform for transmission to occur.



Physical layer is concerned with

1. Line coding
2. Transmission Impairment
3. Data Rate limits
4. Performance
5. Transmission Mode
6. Transmission Medium

2.2 LINE CODING

1. It is the process of converting digital data to digital signals.
2. Data, in the form of text, numbers, graphical images, audio or video are stored in computer memory as sequence of bits.
3. At the sender, digital data are encoded into digital signal and at the receiver, digital data are recreated by decoding the digital signal.

2.2.1 Characteristics

1. Signal Element Versus Data Element

(i) Data Element

It is the smallest entity that can represent a piece of information i.e. a bit. It means what we need to send is data element.

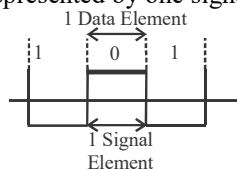
(ii) Signal Element

What we can send over communication line is signal element which carries our data element to be sent.

We define, $\text{ratio}(r) = \text{No. of data elements carried by each signal element}$.

Several situations with different value of r as follows

- (a) $r = 1$ when one data element is represented by one signal element

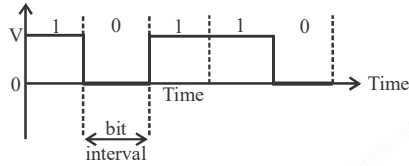


WORKBOOK

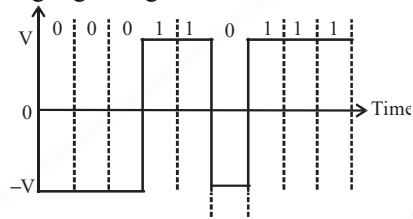
1. Draw the digital signal for the following data bit stream

Solution.

(a) If the data bit stream is 10110 then what will be its NRZ coding/digital signal.



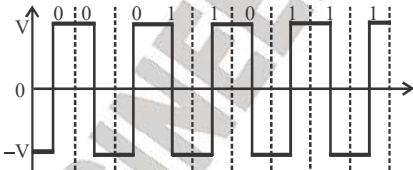
(b) If the data bit stream is 000110111 its NRZ-L coding/digital signal will be



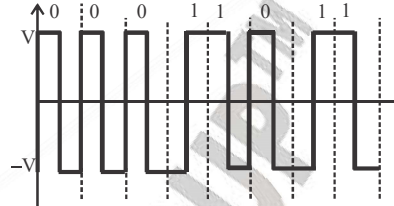
(c) If the data bit stream is 000110111 its NRZ-I coding/digital signal will be



(d) Manchester encoding of bits stream 000110111



(e) Differential Manchester encoding of bits stream 000110111



2. If a signal travels through an amplifier, and its power is increased 10 times. Calculate amplification of signal.

Solution.

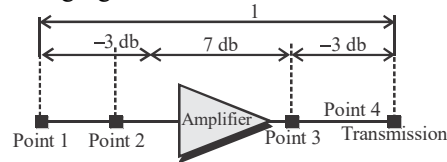
Let P_1 and P_2 are power of signals at the point 1 and point 2

$P_2 = 10P_1$ (given). The amplification (gain of power) can be calculated as

$$10 \log_{10} \frac{P_2}{P_1} = 10 \log_{10} \frac{10P_1}{P_1} = 10$$

$$\log_{10} 10 = 10(1) = 10 \text{ dB}$$

3. Calculate the resultant decibel value for the following figure.



Solution.

Here signal is suffering from attenuation and amplification.

During attenuation there are negative decibel values and during amplification there is positive decibel value.

The resultant dB = $-3 + 7 - 3 = +1$ it means the signal has gained in power

4. Calculate the power of a signal if its $\text{dB}_m = -30$.

Solution.

We can calculate the power in milliwatt of the signal as

ASSIGNMENT

1. As the data packet moves from a lower layer to higher layer, the headers are
(a) Added (b) Removed
(c) Rearranged (d) Modified
2. Session layer is included in
(a) MAC layer
(b) Data link layer
(c) Transport layer
(d) Application layer
3. The upper layers of the OSI model are, in correct order:
(a) Session, application, presentation
(b) Session, Presentation, application
(c) Application, presentation, session
(d) None of these
4. The part of OSI where one most commonly finds data encryption, compression, and other encoding for network communication is:
(a) Application (layer seven)
(b) Session (layer five)
(c) Presentation (layer six)
(d) None of these
5. Which of these network devices belong at the OSI data link layer (layer two)?
(a) Router (b) Bridge
(c) TVPN (d) None of these
6. Which of these network devices belongs at the OSI physical layer (layer one)?
(a) Repeater (b) Router
(c) Switch (d) Bridge
7. What is the Protocol Data Unit (PDU) employed at the physical layer?
(a) Bits (b) Frames
(c) Packets (d) Segments
8. The bottom layer of the OSI model is about electrical and mechanical aspects networking. What is this layer known as?
(a) Transport (b) Data link
(c) Physical (d) Session
9. What network topology implements at least two paths to and from each node?
(a) Bus (b) Ring
(c) Star (d) Mesh
10. What type of network topology is depicted by a single cable where device connect using 'T' connectors?
(a) Star (b) Bus
(c) Ring (d) 10 BaseT
11. The physical layer is responsible for the transmission of _____ over the physical medium.
(a) Packets (b) Bits
(c) Message (d) All of these
12. Which of the following statements best describes a hub?
(a) All connected systems are in the same broadcast domain, but different collision domains.
(b) All connected systems are in the same collision domain, but different broadcast domains.
(c) All connected systems are in the same broadcast and collision domains
(d) All connected systems are in their own broadcast and collision domains
13. A hub is a
(a) Router (b) Bridge
(c) Repeater (d) All of these

GATE QUESTION

1. How many 8-bit characters can be transmitted per second over a 9600 baud serial communication link using asynchronous mode of transmission with one start bit, eight data bits, and one parity bit?

- (a) 600
(c) 876

[GATE - 2004]

- (b) 800
(d) 1200

SOLUTION

Sol.1 (a)

Total number of bits = 12
Modulation Rate = 9600 baud

Number of bit (8 bit character) transmitted =

$$\frac{9600}{12 \text{ bits}} = 800$$

CHAPTER - 3**DATA LINK LAYER****3.1 INTRODUCTION**

1. It attaches its own header to the packet delivered by network layer and makes it frame.
2. It moves frames over each of individual links in the end to end path.
3. It uses MAC address to deliver frames to recipient machine. MAC address is of 6 bytes (48 bits) and is represented in hexadecimal format such as 0E-E9-06-14-BB-4B .
4. Its protocol defines the format of the frames to be exchanged between the nodes at the ends of links and defines the action taken by these nodes on receiving and sending of frames.
5. Link layer protocols are Ethernet, Token Ring, 8.02.11 wireless LAN and PPP.
6. Data link layer is divided into two sub-layers named as Logical Link Control (LLC) and Medium Access Control (MAC)
7. LLC is responsible for Framing, Flow Control, Error Control and where as MAC is responsible for link access control i.e. coordinating the multiple users using shared channel.

3.1.1 FRAMING

1. Data-link layer performs function of framing the packets that are received from network layer
2. It breaks the bit stream up into discrete frames and tells the size of each frame to the receiver. It can be done using following methods.

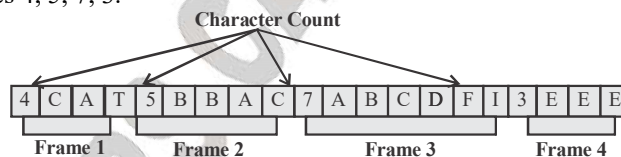
- (i) Character count
- (ii) Character stuffing
- (iii) Bit stuffing

3.1.1.1 Character Count

It specifies the number of characters to be in one frame in its one of the header field of frame.

Example.

Let 3 frames of sizes 4, 5, 7, 3.

**1. Limitation**

Any error in count field can change the boundaries of the frame and receiver will receive indifferent frames that are not of the sender.

3.1.1.2 Character Stuffing

1. It specifies the start and end of the frame using character sequence DLE STX and DLE ETX (DLE is Data link Escape, STX is start of Text, ETX is end of text)
2. If the characters for DLE STX and DLE ETX are itself present as part of the data, it inserts an ASCII character DLE before each incident DLE character in the data.

Example.

Data sent by Network Layer

(a) A DLE B

Data after being character stuffed

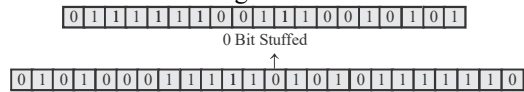
(a) STX DLE A DLE DLE B DLE ETX

WORKBOOK

1. If flag byte is 01111110 and Original data is 01110010101010100011111101. Compute the new data after bit stuffing to be sent.

Solution.

Data after bits stuffing is



2. Find the hamming distance of codewords 000110 and 110010

Solution.

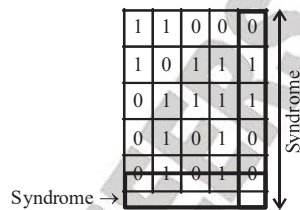
XOR of the two given codewords is calculated as

$$\begin{array}{r} 000110 \\ 110010 \\ \hline 110100 \leftarrow \text{result} \end{array}$$

The number of 1's in result 110100 are 3. So, their Hamming distance is 3 that implies the two codewords differ by 3 bits.

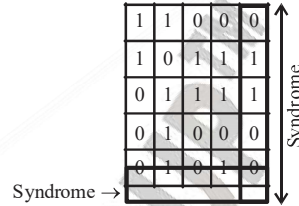
3. Find the syndrome for dataword 1100101101110101 using two-dimensional even parity-check code with column of 4 bits.

Solution.

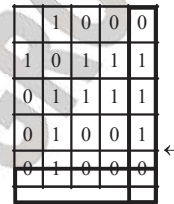


If an error occurred in of the 3rd digit from the left in 4th word.

Errored data received with syndrome calculated at Sender's side



New syndrome for received data is



Paring the recalculated syndrome and received syndrome, we get error in 3rd digit of the 4th word. As It is binary data, changing 0 to 1.

4. If a CRC generator uses a divisor as polynomial $x^3 + x + 1$ and dataword is 1001010. Find the codeword to be send to the receiver.

Solution.

Divisor is $x^3 + x + 1$ (given) in Binary form, it can be written as 1011 and dataword is 1001010 At Sender's side

ASSIGNMENT - I

1. A 10 Base 2 Ethernet network uses what type of cable?
 (a) Thin coaxial (b) Twisted-Pair cable
 (c) Thick coaxial (d) None of these
2. Which of the following is a product of LLC sub layer?
 (a) 802.3 frame
 (b) 802.5 frame
 (c) PDU (Protocol data unit)
 (d) Preamble
3. Consider building a CSMA/CD network running at 10 Mbps over a cable with no repeaters. If the signal speed in the cable is 10^6 Km/sec and minimum frame size is 1500 bytes then what is the cable length?
 (a) 600 km (b) 1200 Km
 (c) 12 Km (d) 120 Km
4. How many bits a 550 meter ring cable containing 60 equally spaced stations (assuming each has one bit delay) can occupy? Assume 4 Mbps data rate and propagation speed of 250 meters per micro sec
 (a) 60 bits (b) 69 bits
 (c) 70 bits (d) 75 bits
5. One Gbps CSMA/CD LAN is to be designed over 1Km cable without repeater. The cable supports signal speed of 200, 000 Km/sec. What is the Minimum frame size that Data Link layer should consider?
 (a) 10^9 bits (b) 5000 bits
 (c) 10000 bits (d) 10 bits
6. Consider a token ring LAN in which time on the ring will alternate between data frame transmission and token. In a single instance of data frame followed by a token as a cycle and defined as follows
 t is average time for one cycle
 t_1 is average time to transmit a data frame
 t_2 is average time to pass a token
 If u is the maximum utilization of the channel then the value of u is
 (a) $\frac{t}{t_1 + t_2}$ (b) $\frac{t_1}{t_1 + t_2}$
 (c) $\frac{t-1}{t_1 + t_2}$ (d) $\frac{1}{t - (t_1 + t_2)}$
7. MTU stands for
 (a) Minimum Transfer Unit
 (b) Minimum Transmission Unit
 (c) Maximum Transmission Unit
 (d) Maximum Transfer Unit
8. Synchronization of bits is done by
 (a) Data link layer (b) Network layer
 (c) Transport layer (d) All of these
9. Bridges operate at the _____ of the OSI model.
 (a) The Data-Link Layer
 (b) Top Layer
 (c) The Transport Layer
 (d) The Network Layer
10. Which of the following options is not an useful property of Manchester line code for an Ethernet?
 (a) Continuous Energy
 (b) Continuous Clock transition
 (c) No DC component
 (d) No signal change at a 1 to 0 transition
11. At which layer of the OSI model does a switch exist?
 (a) Physical (b) Data link
 (c) Network (d) Session
12. If a frame enters a bridge and the MAC address is not found in the MAC address table, what will the bridge do with the frame?
 (a) Drop it

GATE QUESTIONS

1. A computer network uses polynomials over GF(2) for error checking with 8 bits as information bits and uses $x^3 + x + 1$ as the generator polynomial to generate the check bits. In this network, the message 01011011 is transmitted as

[GATE - 2017]

- (a) 01011011010 (b) 01011011011
(c) 01011011101 (d) 01011011100

1. The values of parameters for the Stop-and-Wait ARQ protocol are as given below:
Bit rate of the transmission channel = 1 Mbps.
Propagation delay from sender to receiver = 0.75 ms.
Time to process a frame = 0.25 ms.
Number of bytes in the information frame = 1980
Number of bytes in the acknowledge frame = 20
Number of overhead bytes in the information frame = 20
Assume that there are no transmission errors. Then, the transmission efficiency (expressed in percentage) of the Stop-and-Wait ARQ protocol for the above parameters is _____ (correct to 2 decimal places).

[GATE - 2017]

2. Consider two hosts X and Y, connected by a single direct link of rate 10^6 bits/sec. The distance between the two hosts is 10,000 km and the propagation speed along the link is 2×10^8 m/sec. Host X send a file of 50,000 bytes as one large message to host Y continuously. Let the transmission and propagation delays be p milliseconds and q milliseconds, respectively. Then the values of p and q are

[GATE - 2017]

- (a) p=50 and q=100 (b) p=50 and q=400
(c) p=100 and q=50 (d) p=400 and q=50

3. Consider a binary code that consists of only four valid code words as given below:
00000, 01011, 10101, 11110

Let the minimum Hamming distance of the code be p and the maximum number of erroneous bits that can be corrected by the code by q. Then the values of p and q are

[GATE - 2017]

- (a) p=3 and q=1 (b) p=3 and q=2
(c) p=4 and q=1 (d) p=4 and q=2

4. A sender uses the Stop-and-Wait ARQ protocol for reliable transmission of frames. Frames are of size 1000 bytes and the transmission rate at the sender is 80 Kbps (1Kbps = 1000 bits/second). Size of an acknowledgement is 100 bytes and the transmission rate at the receiver is 8 Kbps. The one-way propagation delay is 100 milliseconds. Assuming no frame is lost, the sender throughput is _____ bytes/second.

[GATE - 2016]

5. In an Ethernet local area network, which one of the following statements is TRUE?

[GATE - 2016]

- (a) A station stops to sense the channel once it starts transmitting a frame.
(b) The purpose of the jamming signal is to pad the frames that are smaller than the minimum frame size.
(c) A station continues to transmit the packet even after the collision is detected.
(d) The exponential backoff mechanism reduces the probability of collision on retransmissions.

6. A network has a data transmission bandwidth of 20×10^6 bits per second. It uses CSMA/CD in the MAC layer. The maximum signal propagation time from one node to another node is 40 microseconds. The minimum size of a frame in the network is _____ bytes.

SOLUTIONS

Sol 1. (c)

1011)01011011000(01000011

$$\begin{array}{r} 1011 \\ \underline{1100} \\ 1110 \\ \underline{1011} \\ 101 \\ \underline{0101} \\ \text{CRC} \end{array}$$

Sol 2. (89.33)

B = 1 Mbps
 $T_p = 0.75$ ms
 $T_{proc} = 0.25$ ms
 Payload = 1980 B
 Ack = 20 B
 OH = 20 B
 $L = \text{Payload} + \text{OH} = 1980 + 20 = 2000$ Bytes
 $T_x = \frac{L}{B} = \frac{2000 \times 8}{1 \times 10^6} = 16$ ms
 $T_{ax} = \frac{20 \times 8}{1 \times 10^6}$
 $= 160$ μ sec
 $= 0.16$ msec
 Total time = $T_x + T_p + T_{proc} + T_{ax} + T_p + T_{aproc}$
 $= 16$ ms + 0.75 ms + 0.25 ms + 0.16 ms + 0.75 ms
 $= 17.91$ ms
 $\eta = \frac{T_x}{\text{Total Time}} = \frac{16}{17.91} = 89.33\%$

Sol 3. (d)

Given
 $B = 10^6$ bps
 Distance = 10000 km
 $T_p = 2 \times 10^8$ m/s
 $L = 50000$ B
 $p = \frac{T_x}{B} = \frac{L}{100 \times 10^4} = \frac{4}{10} \times \frac{10^3}{10^3} = \frac{4000}{10} = 400$ nsec

$$q = \frac{d}{v} = \frac{10000 \times 10^3}{2 \times 10^8} = \frac{1}{20} = \frac{1}{20} \times \frac{10^3}{10^3} = \frac{1000}{20} \text{ ms} = 50 \text{ ms}$$

Sol 4. (a)

The number of bit positions in which two code words differ is called Hamming Distance.

00000	00000	00000	00000
01011	10101	11110	10101
⊕---	⊕---	⊕---	⊕---
= 3	3	4	4

01011	10101
11110	11110
⊕---	⊕---
= 3	3

So minimum distance = 3 \Leftarrow p
 To connect d errors, you need a distance 2d + 1 code
 So 3 = 2d + 1
 Hence, d = 1 \Leftarrow q
 p = 3, q = 1

Sol 5. (2500)

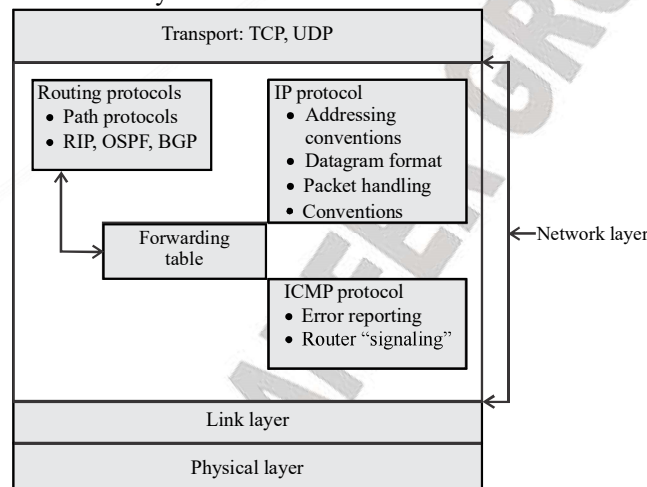
Bandwidth(R) = 80 kbps
 Frame size = 1000 bytes
 Propagation Time(PT) = 100 ms
 Acknowledgement size = 100byte
 Transmission Rate at receiver = 8kbps
 Transmission Time (TT) = $\frac{\text{Framesize}}{\text{Bandwidth}}$
 $= 100$ ms
 Transmission Time of acknowledgement
 $(TT_{Ack}) = \frac{\text{Acknowledgement size}}{\text{Bandwidth}} = 100$ ms
 Efficiency = $\frac{TT}{TT + 2PT + TT_{Ack}}$

CHAPTER - 4

NETWORK LAYER

4.1 INTRODUCTION

1. Network layer provides host to host communication. It receives data from transport layer and forms the packets of received data.
2. It works on IP address of host to route the packets to receipt network
3. It has three major components: IP protocol, Routing, ICMP.
4. IP protocol includes Addressing convention, Datagram format and Packet handling conventions
5. Routing determines the path a datagram should follow from source to destination. It includes path selection and RIP, OSPF, BGP protocols.
6. ICMP (Internet control message protocol) is a facility to report errors in a datagram and respond to requests for certain network layer information



4.2 IP ADDRESSING

1. IP address is technically associated with the interface rather than with the host or router containing that interface.
2. Each IP address is 32 bits long which contains the network and host identifier on that particular network.
3. It is written in dotted decimal notation. For example 193.32.216.164
4. They are logical addresses.
5. They are assigned by Internet Corporation for Assigned Names and Numbers (ICANN). It assigns IP addresses to ISPs (Internet Service Provider) and ISP handles allocation or management of addresses within their regions. This address 255.255.255.255 is Broadcast IP address.
6. Each IP address has two parts network id and host id. Network id has all host bits 0's.
7. There are two types of Addressing:-
 - (i) Classful Addressing
 - (ii) Classless Addressing

WORKBOOK

1. Determine the netid and direct broadcast address for IP address 201.16.139.149.

Solution.

IP address = 201.16.139.149

(Netid is calculated by taking ANP of IP address and m/w mask)

So, Network address is as follows.

201.16.139.149

AND 255.255.255.0

201.16.139.0

Netid: 201.16.139.0

Direct Broadcast address of 201.16.139.0 is 201.16.139.255.

2. Determine the netid and direct broadcast for IP address 144.16.19.159.

Solution.

IP address = 144.16.19.159

It is class B

So its network mask is 255.255.0.0

Network address will be as

144.16.19.159

AND 255.255.255.0

144.16.0.0

Net ID: 144.16.0.0

Direct Broadcast → 144.16.255.255

3. Determine netid and Broadcast address for IP address 37.159.87.120.

Solution.

IP address = 37.159.87.120

It Belongs to class A

So its network mask is 255.0.0.0

37.159.87.120

AND 255. 0. 0. 0

Netid → 37.0.0.0

Direct Broadcast Address → 37.255.255.255

4. In class C, if subnet mask is 255.255.255.224 Calculate number of subnets and no. of host in each subnet

Solution.

Subnet mask is 255.255.255.224

In class C, Network mask is 255.255.255.0

In creating subnets, we use fewer bits of hosted and in this subnet mark last byte is 224 (11100000).

So it uses 3 bits to create subnet.

Total no. of subnets = $2^3 - 2 = 6$

Here 2 subnets are created, one of which gives network mask address and other gives subnets mask.

Here, 3 bits of host are used to create subnet and so 5 bits are used to define host.

∴ No. of hosts per subnet = $2^5 - 2 = 30$ hosts

5. In the class C if subnet mask is 255.255.255.240. Calculate the no. of subnets and no. of hosts in each subnet.

Solution.

Subnet mask = 255.255.255.240

240 = 11110000

Here 4 bits of host id bytes are used to define subnets.

So no. of subnets = $2^4 - 2 = 14$

No. of hosts in each subnet = $2^4 - 2 = 14$

Because 4 bits of host id bytes are left to define host.

6. IP address 203.16.17.119 and subnet mask is 255.255.255.224 is given. Calculate subnet IP subnet number first, last host and direct broadcast address of that subnet.

Solution.

IP address = 203.16.17.119

It belongs to class C

IP: 203.16.17.119

AND 255.255.255.224

First 3 bytes of subnetid is same as IP address because they are AND with 255. Last byte is calculated as follows

119 01110111

224 11100000

ASSIGNMENT - I

1. Network layer activities are:

- (a) Logical addressing
- (b) Port addressing
- (c) Access control
- (d) All of these

2. Consider the following statements

S1: Protocols define how corresponding layers on separate machines communicate,

S2: Services are used by adjacent layers on the same machine to communicate. Which one of the following is true?

- (a) S1 is true, S2 is false
- (b) S2 is true, S1 is false
- (c) Both are true
- (d) Both are false

3. Routing is done in

- (a) Network layer
- (b) Physical layer
- (c) Data link layer
- (d) Transport layer

4. What types of routing connection would typically fragment data packets?

- (a) Connection Oriented
- (b) Cut-Through
- (c) Reliable
- (d) Connectionless

5. The Internet Protocol (IP) generally corresponds to which OSI layer?

- (a) Network (layer three)
- (b) Transport (layer four)
- (c) Data Link (layer two)
- (d) Session (layer five)

6. Which of these network devices primarily functions at the OSI network layer (layer 3)?

- (a) Switch
- (b) Gateway
- (c) Router
- (d) All of these

7. Which Protocol Data Unit (PDU) is employed at Network Layer?

- (a) Bits
- (b) Frames
- (c) Packets
- (d) Segments

8. Consider the following statements

S1: Network layer dividing the transmitted bit stream into frames.

S2: Network layer determining which route through the subnet to use.

Which one of the following is true?

- (a) S1 is true, S2 is false
- (b) S2 is true, S1 is false
- (c) Both are true
- (d) Both are false

9. State true (T) and false (F) of the following statements

1. A straight-through cable should be used to connect a router to a switch.

2. A switch builds its MAC table based on Destination MAC addresses

3. Routers separate broadcast domains.

4. A bridge will always forward all broadcast traffic to all ports.

- (a) TFFT
- (b) FTTF
- (c) TFFT
- (d) FTFT

10. In a networking system, a session must be established before data may be transmitted. What OSI layer is responsible for this function?

- (a) Presentation
- (b) Session
- (c) Transport
- (d) Data Link

11. To establish a session between network communicators, Session layer protocols are required. Choose two protocols that work in this layer.

- (i) RPC
- (ii) PICT
- (iii) ASP
- (iv) JPEG

(a) (i) and (iii)

(b) (ii) and (iii)

(c) (ii) and (iii)

(d) (iii) and (iv)

12. What kind of device might you need if you wanted to connect your corporate LAN to the internet?

GATE QUESTIONS

1. For a host machine that uses the token bucket algorithm for congestion control, the token bucket has a capacity of 1 megabyte and the maximum output rate is 20 megabytes per second. Tokens arrive at a rate to sustain output at a rate of 10 megabytes per second. The token bucket is currently full and the machine needs to send 12 megabytes of data. The minimum time required to transmit the data is _____ seconds.

[GATE - 2016]

2. An IP datagram of size 1000 bytes arrives at a router. The router has to forward this packet on a link whose MTU (maximum transmission unit) is 100 bytes. Assume that the size of the IP header is 20 bytes. The number of fragments that the IP datagram will be divided into for transmission is _____.

[GATE - 2016]

3. Which one of the following protocols is **NOT** used to resolve one form of address to another one?

[GATE - 2016]

- (a) DNS (b) ARP
(c) DHCP (d) RARP

4. Which one of the following fields of an IP header is **NOT** modified by a typical IP router?

[GATE- 2015]

- (a) Checksum (b) Source address
(c) Time to Live (TTL) (d) Length

5. Consider the following routing table at an IP router:

[GATE- 2015]

Network No.	Net Mask	Next Hop
128.96.170.0	255.255.254.0	Interface 0
128.96.1680	255.255.254.0	Interface 1
128.96.166.0	255.255.254.0	R2
128.96.164.0	255.255.252.0	R3
0.0.0.0	Default	R4

For each IP address in Group-I identify the correct choice of the next hop from Group-II using the entries from the routing table above.

List-I

- A. 128.96.171.92
B. 128.96.167.151
C. 128.96.163.151
D. 128.96.165.121

Group-II

- (i) Interface 0
(ii) Interface 1
(iii) R2
(iv) R3
(v) R4

Codes:

- (a) A-i, B-iii, C-v, D-iv
(b) A-i, B-iv, C-ii, D-v
(c) A-ii, B-iii, C-iv, D-v
(d) A-ii, B-iii, C-v, D-iv

6. Host A sends a UDP datagram containing 8880 bytes of user data to host B over an Ethernet LAN. Ethernet frames may carry data up to 1500 bytes (i.e. MTU = 1500 bytes). Size of UDP header is 8 bytes and size of IP header is 20 bytes. There is no option field in IP header. How many total number of IP fragments will be transmitted and what will be the contents of offset field in the last fragment?

[GATE - 2015]

- (a) 6 and 925 (b) 6 and 7400
(c) 7 and 1110 (d) 7 and 8880

7. Two hosts are connected via a packet switch with 107 bits per second links. Each link has a propagation delay of 20 microseconds. The switch begins forwarding a packet 35 microseconds after it receives the same. If 10000 bits of data are to be transmitted between the two hosts using a packet size of 5000 bits, the time elapsed between the transmission of the first bit of data and the

SOLUTIONS

Sol 1. (1.2)

Given

Maximum burst rate, $M = 20$ MBToken arrival rate, $P = 10$ MBConstant rate (bucket o/p), $P = 10$ MBBucket capacity, $C = 1$ MB

$$\text{Time for 1 MB, } S = \frac{C}{M-P} = \frac{1}{20-10} = 0.1 \text{ sec}$$

For the total message of 12 MB is 1.2 sec

Sol 2. (13) $L = 1000$ bytes

MTU = 100 bytes

IP header = 20 bytes

So MTU payload is $100 - 20 = 80$ bytesData = $1000 - 20 = 980$ Number of fragments = $980 / 80 = 13$ **Sol 3. (c)**

DHCP is dynamic host configuration protocol: allocates one of the unused IP address.

Sol 4. (a)

Router cannot change the source address because it is the IP address of sender.

Sol 5. (a)

To find the interface for IP address, we have to AND IP address and Netmask, if the answer matches with the network address then its corresponding interface is chosen for next hop.

A:

IP address 128.96.171.92

AND Netmask 255.255.254.0

= 128.96.10101011.01011100

AND 255.255.11111110.00000000

128.96.170.0

Therefore, A select interface 0

B:

IP address 128.96.167.151

AND Netmask 255.255.254.0

= 128.96.10100111.10010111

AND 255.255.11111110.00000000
128.96.166.0Therefore, B select interface R_2 **C:**

IP address 128.96.163.121

AND Netmask 255.255.254.0

= 128.96.10100001.01111001

AND 255.255.11111110.00000000

128.96.162.0

Therefore, C select interface R_4 **D:**

IP address 128.96.165.121

AND Netmask 255.255.254.0

= 128.96.10100011.01111001

AND 255.255.11111110.00000000

128.96.164.0

Therefore, D select interface R_3

So, option (a) is correct.

Sol 6. (c)

UDP datagram = 8880 byte

UDP header = 8 bytes

IP header = 20 bytes

MTU = 1500 bytes

Actual data in MTU = $1500 - 20 = 1480$ bytesActual data in IP packet = $8880 + 8$

= 8888 bytes

So, size of first fragment = $(1480 + 20)$ bytessize of 2nd fragment = $(1480 + 20)$ bytessize of 3rd fragment = $(1480 + 20)$ bytessize of 4th fragment = $(1480 + 20)$ bytessize of 5th fragment = $(1480 + 20)$ bytessize of 6th fragment = $(1480 + 20)$ bytessize of 7th fragment = $(56 + 20)$ bytes

offset of 7 fragment is 1110

Sol 7. (1575 μ sec)Bandwidth = 10^7 bpsPropagation delay = 20 μ secForward time by switch = 35 μ sec

Packet size = 5000 bits

CHAPTER - 5

TRANSPORT LAYER

5.1 INTRODUCTION

1. A transport layer provides logical communication between application processes running on different hosts. Logical communication means hosts running the process are directly connected.
2. Transport layer protocols are implemented in the end system but not in network routers.
3. It provides data to the network layer.
4. Services that transport protocol can provide are often constrained by the services model of network-layer protocol. But there are also certain services that are offered by transport protocol but not by network layer protocol.
5. Transport layer has 16 bit port number to identify a process to which data is to be delivered (destination port number) and by which data is sent (source Port number).
6. The port number ranging from 0 to 1023 are well known such as HTTP's port no. 80, FTP's port no. 21.

Example.

Consider two houses one in the East Delhi and other in West Delhi having four kids in each. The kids in the two houses write letters to each other. Let kids A1 and A2 are responsible for letter collection from their respective house kids and their distribution to postal service mail carrier. Now, it can be understood about the transport layer of computer network by the following analogy

Letter in envelopes	Application layer Message
Kids	Process
Houses	Hosts (end systems)
Kids A1, A2	Transport Layer Protocol
Postal Service	Network layer Protocol

5.1.1 Multiplexing and De-Multiplexing

1. A process can have one or more sockets for communication between the network and processes.
2. The transport layer in the receiving host delivers data to an intermediate socket. Because at any given time, there can be more than one socket in the receiving host, each socket has a unique identifier.
3. The job of gathering data chunks from different sockets and encapsulating header information with them to create segment and passing segments to the network layer is called multiplexing.
4. The job of delivering the data in the transport layer segment to the correct socket is called De-multiplexing.

5.2 USER DATAGRAM PROTOCOL (UDP)

1. UDP takes message from the application process, attaches source and destination port number field for the multiplexing/ de-multiplexing service and passes the resulting segment to the network layer.
2. When segment arrives at the receiving host, UDP uses the destination port number to deliver the segment's data to the correct application process.
3. UDP is connectionless as there is no handshake between sender and receiver.
4. DNS (Domain Name Server), RIP (Routing Information Protocol), and SNMP (Simple Network management Protocol) uses UDP.

ASSIGNMENT - I

1. The transport layer provides
- Node-to-node communication
 - Host- to-host communication
 - Process-to-process communication
 - All of the above
2. Why some transport-layer packets may be lost in the Internet?
- Because a valid destination address is not known to these packets
 - A socket address is not defined for these packets.
 - The router through which the datagram need To pass to reach their destination may be congested
 - Because the time has been cleared out.
3. Which of the following features is not common in both data link layer and transport layer?
- Recovery from transmission errors
 - Flow control
 - Multiplexing
 - Framing
4. Choose the correct statement
- A client program normally uses a well-known port number. A server program normally uses an ephemeral port number.
 - A client program normally uses an ephemeral port number. A server program normally uses a well-known port number.
 - A client program normally uses an private port number. A server program normally uses a well-known port number.
 - None of these
5. Match List-I (Range of Port numbers) with List-II (corresponding port) and choose the correct answer using the codes given below.
- List-I**
- A. 0 to 1023
- B. 1024 to 49, 151
- C. 49, 152 to 65, 353
- List-II**
- Registered ports
 - Well-known ports
 - Dynamic ports
- Codes**
- A-ii, B-i, C-iii
 - A-i, B-ii, C-iii
 - A-iii, B-ii, C-i
 - A-ii, B-iii, C-i
6. A port number is
- 8 bits long
 - 16 bits long
 - 32 bits long
 - 64 bits long
7. In DHCP, the ports used by the client and server are respectively
- An ephemeral; a well-known
 - A well-known; a well-known
 - A well-known; an ephemeral
 - An ephemeral; an ephemeral
8. Suppose that a new organization needs to create a new server process and allows its customers to access the organization site using that process. How should the port number for the server process be selected?
- The organization can choose any real number as a port number.
 - The organization can choose a port number from 0 to 1023.
 - This will depend on the IP address of the organization.
 - The organization needs to select a port number from the registered range. 1024 to 49,151 and register that port number with ICANN.
9. A client residing on a host with IP address 122.45.12.7 sends a message to the corresponding server residing on a host with IP

GATE QUESTIONS

1. Consider socket API on a Linux machine that supports connected UDP sockets. A connected UDP socket is a UDP socket on which connect function has already been called. Which of the following statements is/are CORRECT?

I. A connected UDP socket can be used to communicate with multiple peers simultaneously.

II. A process can successfully call connect function again for an already connected UDP socket.

[GATE - 2017]

- (a) I only (b) II only
(c) Both I and II (d) Neither I nor II

2. Consider the following statement about the routing protocols. Routing Information Protocol (RIP) and Open Shortest Path First (OSPF) in a IPv4 network.

I. RIP uses distance vector routing

II. RIP packets are sent using UDP

III. OSPF packets are sent using TCP

IV. OSPF operation is based on link-state routing

Which of the statements above are CORRECT?

[GATE - 2017]

- (a) I and IV only
(b) I, II and III only
(c) I, II and IV only
(d) II, III and IV only

3. Suppose two hosts use a TCP connection to transfer a large file. Which of the following statements is/are False with respect to the TCP connection?

1. If the sequence number of a segment is m , then the sequence number of the subsequent segment is always $m + 1$

2. If the estimated round trip time at any point of time is t sec, the value of the retransmission timeout is always set to greater than or equal to t sec.

3. The size of the advertised window never changes during the course of the TCP connection.

4. The number of unacknowledged bytes at the sender is always less than or equal to the advertised window.

[GATE - 2015]

- (a) 3 only (b) 1 and 3 only
(c) 1 and 4 only (d) 2 and 4 only

4. Assume that the bandwidth for a TCP connection is 1048560 bits/sec. Let α be the value of RTT in milliseconds (rounded off to the nearest integer) after which the TCP window scale option is needed. Let β be the maximum possible window size with window scale option. Then the values of α and β are

[GATE - 2015]

- (a) 63 milliseconds 65535×2^{14}
(b) 63 milliseconds 65535×2^{16}
(c) 500 milliseconds 65535×2^{14}
(d) 500 milliseconds 65535×2^{16}

5. Consider the following statements.

(i) TCP connections are full duplex.

(ii) TCP has no option for selective acknowledgment

(iii) TCP connections are message streams.

[GATE - 2015]

- (a) Only i is correct
(b) Only i and ii are correct
(c) Only ii and iii are correct
(d) All of i, ii and iii are correct.

6. Which one of the following socket API functions converts an unconnected active TCP socket into a passive socket

[GATE - 2014]

- (a) Connect (b) Bind
(c) Listen (d) Accept

7. Let the size of congestion window of a TCP connection be 32 KB when a timeout occurs.

CHAPTER - 6

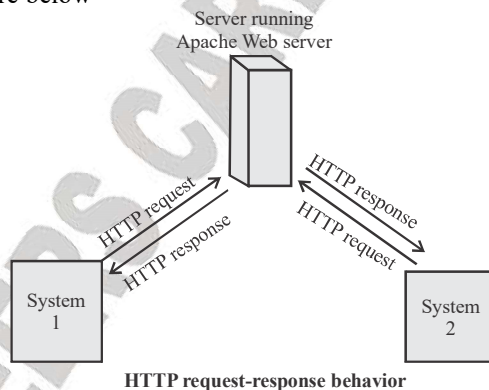
APPLICATION LAYER

6.1 INTRODUCTION

1. It is related to client and server architecture, there is an always one host called the server, which services requests from many other hosts, called clients. A client host can be either sometimes on or always on.
2. No two clients can directly communicate with each other. It is only possible having in between server/ servers for their communication.
3. Web Application, file transfer, Remote Login, and e-mail all are application layers application.
4. An application-layer protocol defines how an each application's processes, running on different end systems, pass messages to each other.
5. Its different protocols are HTTP, FTP, SMTP, POP, DNS etc.

6.1.1 HTTP (Hyper Text Transfer Protocol) (Web page Display)

1. It is the web's application layer protocol and is the heart of the web.
2. It is implemented in both client and server program, executing on different end systems, by exchanging HTTP messages.
3. HTTP uses TCP as its underlying transport protocols.
4. It runs on port number 80.
5. A browser is a user agent for the web; it displays the requested web page to the user.
6. Thus, when a user requests a web page the browser sends HTTP request message for the objects in the page to server. The server receives the requested message and responds with HTTP response message as shown in figure below



7. HTTP is said to be stateless protocol, as HTTP server maintains no information about the client.
 8. HTTP is pull protocol i.e., someone loads information on a web server and users use HTTP to pull the information from the server at their convenience.
- It uses both non persistent and persistent connection (in default mode).

(i) Non-persistent connection

- (a) Here, TCP connection transports exactly one request message and one response message.
- (b) TCP connection is closed after the server sends the requested data.

ASSIGNMENT

1. Match List-I(internet application) with List-II (application-layer protocol that the application uses) and choose the correct answer using the codes given below

List-I (Application)

- A. Web
- B. File Transfer
- C. Remote Login
- D. Email

List-II (Protocol used in the application)

- (i) FTP
- (ii) HTTP
- (iii) SMTP
- (iv) Telnet

Codes:

- (a) A-ii, B-iv, C-i, D-iii
- (b) A-ii, B-ii, C-i, D-iv
- (c) A-ii, B-i, C-iv, D-iii
- (d) A-iv, B-i, C-ii, D-iii

2. **Assertion (A):** HTTP, FTP, SMTP, POP3, and IMAP run on top of TCP rather than UDP.

Reason (R): TCP provides a reliable data transfer service.

- (a) Both A and R are true and R is the correct explanation of A.
- (b) Both A and R true but R is not the correct explanation of A.
- (c) A is true but R is false.
- (d) A is false but R is true.

3. In a centralized P2P network,

- (a) The directory system uses the client-server paradigm; the storing and downloading of the files are also done using client-server paradigm.
- (b) The directory system uses the peer-to-peer paradigm: the storing and downloading of the files are done using client-server paradigm.
- (c) The directory system uses the client-server paradigm; the storing and downloading of the files are done using peer-to-peer paradigm.

(d) The directory system uses the peer-to-peer paradigm; the storing and downloading of the files are also done using peer-to-peer paradigm.

4. In a client-server network architecture

- (a) The server provides service and the client receives service
- (b) The client provides service and the server receives service
- (c) Both the client and server can provide and receive service
- (d) None of these

5. Which of the following application is not based on client-server network architecture?

- (a) Web
- (b) File transfer
- (c) Remote-login
- (d) Instant messaging

6. For a communication session between a pair of processes, which process is the client and which is the server?

- (a) The process which requests to establish the connection is client; the process which requests to terminate the connection is the server.
- (b) The process which sends the first packet into the network is the client and the process which sends the last packet into the network is server.
- (c) The process which sends first three packets consecutively is the client and the process which receives last two packets consecutively is the server.
- (d) The process which initiates the communication is the client; the process that waits to be contacted is the server.

7. What information is used by a process running on one host to identify a process running on another host?

- (a) The IP address of the destination host and the port number of the destination

GATE QUESTIONS

1. Consider a TCP client and a TCP server running on two different machines. After completing data transfer, the TCP client calls close to terminate the connection and a FIN segment is sent to the TCP server. Server-side TCP responds by sending an ACK, which is received by the client-side TCP. As per the TCP connection state diagram (RFC 793), in which state does the client-side TCP connection wait for the FIN from the server-side TCP?
[GATE - 2017]
- (a) LAST-ACK (b) TIME-WAIT
(c) FIN-WAIT-1 (d) FIN-WAIT-2
2. Which of the following is/are example(s) of stateful application layer protocols?
(i) HTTP (ii) FTP
(iii) TCP (iv) POP3
[GATE - 2016]
- (a) (i) and (ii) only
(b) (ii) and (iii) only
(c) (ii) and (iv) only
(d) (iv) only
3. Identify the correct sequence in which the following packets are transmitted on the network by a host when a browser requests a webpage from a remote server, assuming that the host has just been restarted.
[GATE - 2016]
- (a) HTTP GET request, DNS query, TCP SYN
(b) DNS query, HTTP GET request, TCP SYN
(c) DNS query, TCP SYN, HTTP GET request
(d) TCP SYN, DNS query, HTTP GET request
4. In one of the pairs of protocols given below, both the protocols can use multiple TCP connections between the same client and the server. Which one is that?
[GATE - 2015]
- (a) HTTP, FTP
(b) HTTP, TELNET
(c) FTP, SMTP
(d) HTTP, SMTP
5. Which one of the following statements is NOT correct about HTTP cookies?
[GATE - 2015]
- (a) A cookie is a piece of code that has the potential to compromise the security of an Internet user
(b) A cookie gains entry to the user's work area through an HTTP header
(c) A cookie has an expiry date and time
(d) Cookies can be used to track the browsing pattern of a user at a particular site
6. Which of the following transport layer protocols is used to support electronic mail?
[GATE - 2012]
- (a) SMTP (b) IP
(c) TCP (d) UDP
7. Consider different activities related to email
m1 : Sends an email from a mail client to a mail server
m2: Download an e-mail from mailbox server to a mail client
m3: Checking e- mail in a web browser
Which is the application layer protocol used in each activity?
[GATE - 2011]
- (a) m1 : HTTP m2: SMTP m3 POP
(b) m1:SMTP m2:FTP m3:HTTP
(c) m1 : SMTP m2:POP m3:HTTP
(d) m1:POP m2:SMTP m3:IMAP
8. A layer-4 firewall (a device that can look at all protocol headers up to the transport layer) CANNOT
[GATE - 2011]
- (a) Block entire HTTP traffic during 9 : 00 pm and 5:00am
(b) Block all ICMP traffic

GATE

2019

DATABASE MANAGEMENT SYSTEM

**COMPUTER SCIENCE &
INFORMATION TECHNOLOGY**



ECG
Publications



A Unit of **ENGINEERS CAREER GROUP**

Head Office: S.C.O-121-122-123, 2nd Floor, Sector-34/A, Chandigarh-160022

Website: www.engineerscareergroup.in **Toll Free:** 1800-270-4242

E-Mail: ecgpublishers@gmail.com | info@engineerscareergroup.in

GATE-2019: Database Management System | Detailed theory with GATE previous year papers and detailed solutions.

©Copyright @2016 by ECG Publishers
(A unit of **ENGINEERS CAREER GROUP**)

All rights are reserved to reproduce the copy of this book in the form storage, introduced into a retrieval system, electronic, mechanical, photocopying, recording, screenshot or any other form without any prior written permission from ECG Publishers (A Unit of **ENGINEERS CAREER GROUP**).

First Edition: 2016

Price of Book: INR 270/-

ECG PUBLICATIONS (A Unit of **ENGINEERS CAREER GROUP**) collected and providing data like: theory for different topics or previous year solutions very carefully while publishing this book. If in any case inaccuracy or printing error may be found or occurred then **ECG PUBLICATIONS** (A Unit of **ENGINEERS CAREER GROUP**) owes no responsibility. The suggestions for inaccuracies or printing error will always be welcome by us.

CONTENTS

CHAPTER	PAGE
1. DATA BASE AND ITS FUNDAMENTALS.....	1-52
2. TRANSACTION PROCESSING CONCEPTS.....	53- 100
3. INDEXING.....	101-125
4. QUERY LANGUAGE.....	126-177

CHAPTER - 1***DATA BASE AND ITS FUNDAMENTALS*****1.1 INTRODUCTION**

1. Database is a collection of files containing related information.
2. It is managed by popular software systems called database management system (DBMS)
Example: University database contain entities such as students, faculty, courses and classrooms which related to each other by relationship such as student enrollment in courses, faculty, teaching courses.

1.1.1 Features of Database

1. It is logically related
2. It is accessible indifferent orders
3. It is stored only once

1.2 TERMINOLOGY OF DATABASE SYSTEM**1. Data**

It can be defined as representation of facts, concepts or instruction in a formalized manner suitable for interpretation or processing by human or electronic machine. It can be represented with the help of characters like alphabet (A – Z, a – z) digits (0 – 9) or special characters (+, -, %, <, > etc)

(i) Data Item (File)

It is a set of character which is used together to represent specific data element e.g. name of student can be represented by NAME.

(ii) Record

A record is a collection of related data items e.g. a payroll record of employee, profile of a student in college.

(iii) File (Data File)

File is a collection of related record e.g. a payroll file might consist of the employees pay records for a company.

For example:

Roll No.	Name	Marks
101	Rachit	85
105	Rohit	75

Here, given table is a file and Roll No., Name, Marks are fields.

(iv) Information

Information is organized or classified data and it has some meaningful value.

Information has following characteristics:

- (a) Timely
- (b) Accurate
- (c) Complete
- (d) Given to right person

ASSIGNMENT - I

1. Find the number of candidate keys of relation R (A, B, C, D, E, F, G, H, I, J) with F.D set $ABD \rightarrow E, C \rightarrow J, AB \rightarrow G, CI \rightarrow J, B \rightarrow F, G \rightarrow H$.
2. How many super keys are possible with candidate by having m attributes if original relation has T member of attributes?
3. Find the Prime attributes in relation R(A, B, C, D, E) with F.D set = $\{AB \rightarrow C, C \rightarrow D, D \rightarrow E, C \rightarrow A\}$.
4. How many total candidate keys are possible for relation R(A, B, C, D) with F.D set = $\{A \rightarrow B, B \rightarrow C, C \rightarrow D, D \rightarrow A\}$?
5. Find the candidate keys of relation R(A, B, C, G, H, I) with F.D set $\{A \rightarrow B, A \rightarrow C, CG \rightarrow A, CG \rightarrow I, B \rightarrow H\}$.
6. Find number of candidate keys of relation R(A, B, C, D, E, F) with F.D set = $\{A \rightarrow C, B \rightarrow D, C \rightarrow E, D \rightarrow E, E \rightarrow A, F \rightarrow B\}$.
7. Find the set of decomposed relation R(A, C, D) using F.D set = $\{AB \rightarrow C, C \rightarrow A, C \rightarrow D\}$ of relation R(A, B, C, D).
8. Find the candidate keys of relation R(A, B, C, D, E, F) having F.D set $\{AB \rightarrow CD, C \rightarrow D, D \rightarrow E, E \rightarrow F\}$.
9. Check whether decomposition $D = \{\{A, B, C\}, \{A, C, D, E\}\}$ of relation R (ABCDE) is Lossless or Lossy if F.D set of R is $\{A \rightarrow BC, CD \rightarrow E, B \rightarrow D, E \rightarrow A\}$.
10. Check whether decomposition set $D = \{ABC, CD, DE\}$ is dependency preserving or not for original relation R(ABCDE) with F.D set = $\{AB \rightarrow CD, C \rightarrow D, D \rightarrow E\}$.
11. Consider a relation R(A, B, C, D, E, F) with F.D set $F = \{AB \rightarrow CD, C \rightarrow D, D \rightarrow E, E \rightarrow F\}$ and decomposition set $D = \{AB, CDE, EF\}$. Check whether it is dependency preserving or not.
12. What is the Highest Normal Form of following relations
 (a) R (ABCD) with F.D set = $\{AB \rightarrow C, BC \rightarrow D\}$
 (b) R(A, B, C, D, F, E) with F.D set $\{AB \rightarrow C, C \rightarrow DE, E \rightarrow F, D \rightarrow A\}$
 (c) R(ABCDEFGH) with F.D set $\{AB \rightarrow CD, D \rightarrow EG, F \rightarrow H, C \rightarrow EF, H \rightarrow A, G \rightarrow B, A \rightarrow B\}$
13. Check the Highest normal form of relation R(A, B, C, D) with F.D set $\{A \rightarrow B, B \rightarrow C, C \rightarrow D, D \rightarrow A\}$ and decompose it, if it is not in BCNF.
14. What is the HNF of the Relation R (ABCDEFGHI) with F.D set = $\{AB \rightarrow CD, C \rightarrow DF, E \rightarrow GH, C \rightarrow E, H \rightarrow I, F \rightarrow B\}$.
15. Consider the relation schema EMP_DEPT with following set G of functional dependencies $G = \{ENO \rightarrow \{ENAME, DOB, ADDRESS, DNUMBER\}, DNUMBER \rightarrow \{DNAME, DMGRENO\}\}$. Calculate the closures of $\{ENO\}^+$ and $\{DNUMBER\}^+$ with respect to G.
16. Consider the relation R, which has attributes that hold schedules of courses and sections t an university; $R = \{CourseNo, SecNo, OfferingDept, Credit-Hours, CourseLevel, InstructorENO, Semester, Year, Days_Hours, RoomNo, NoOfStudents\}$. Suppose that the following functional dependencies hold on R:
 $\{CourseNo\} \rightarrow \{OfferingDept, CreditHours, CourseLevel\}$
 $\{CourseNo, SecNo, Semester, Year\} \rightarrow \{InstructroENO, CourseNo, SecNo\}$
 Try to determine which sets of attributes form keys of R. How would you normalize this relation?

SOLUTIONS

Sol. 1.

R(A, B, C, D, E, F, G, H, I, J)
 $ABD \rightarrow E, C \rightarrow J, AB \rightarrow G, CI \rightarrow J, B \rightarrow F,$
 $G \rightarrow H.$
 $ABD^+ = ABDEGFH$
 $C^+ = CJ$
 $AB^+ = ABGH$
 $CI^+ = CIJ$
 $B^+ = BF$
 $G^+ = GH$

Actually there is no candidate key, But we have to make any of the above closure as candidate key in which we have to do less inclusion of attribute. So, $ABCDI^+$ will be candidate key.

Sol. 2.

2^{T-m}

Sol. 3.

R(ABCDE)
 $FD = \{ AB \rightarrow C, C \rightarrow D, D \rightarrow E, C \rightarrow A \}$
 $AB^+ = ABCDE$ $CB^+ = ABCDE$
 $C^+ = CDEA$

AB and CB are candidate keys. So, A, B and C are prime attributes

Sol. 4.

R(A, B, C, D)
 $A^+ = ABCD$
 $C^+ = ABCD$
 $C^+ = ABCD$
 $D^+ = ABCD$

So total candidate keys = 4

Sol. 5.

Its candidate key is CG only as its closure
 $(CG)^+ = \{CGAIBH\}$

Sol. 6.

R(A, B, C, D, E, F)
 $F.D = \{ A \rightarrow C, B \rightarrow D, C \rightarrow E, D \rightarrow E, E \rightarrow A,$
 $F \rightarrow B \}$
 $A^+ = ACE$

$B^+ = BDEAC$

$C^+ = ACE$

$D^+ = DEAC$

$E^+ = EAC$

$C^+ = FBDEAC = ABCDEF$

Number of Candidate keys = 1.

Sol. 7.

R(A, C, D) is decomposed relation of R(A, B, C, D)

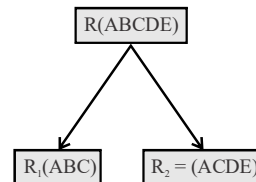
So Functional dependency will become as
 $F = \{ C \rightarrow A, C \rightarrow D \}$

Sol. 8.

R = (A, B, C, D, E, F)
 $\{ AB \rightarrow CD, C \rightarrow D, D \rightarrow E, E \rightarrow F \}$
 $AB^+ = ABCDEF$
 $C^+ = CDEF$
 $D^+ = DEF$
 $E^+ = EF$

∴ Candidate key is AB.

Sol. 9.



1. $R_1 \cup R_2 = R$
2. $R_1 \cap R_2 = AC$
3. $(AC)^+ = ABCDE$

This will act key for both relations. So, It is a lossless decomposition.

GATE QUESTIONS

1. Consider the following tables T₁ and T₂.

T ₁	
P	Q
2	2
3	8
7	3
5	8
6	9
8	5
9	8

T ₂	
R	S
2	2
8	3
3	2
9	7
5	7
7	2

In table T₁, P is the primary key and Q is the foreign key reference R in table T₂ with on-delete cascade and on-update cascade. In table T₂, R is the primary key and S is the foreign key referencing P in table T₁ with on-delete set NULL and on-update cascade. In order to delete record (3, 8) from table T₁, the number of additional records that need to be deleted from table T₁ is _____.

[GATE - 2017]

2. An ER model of a database consists of entity types A and B. These are connected by a relationship R which does not have its own attribute. Under which of the following conditions, can the relational table for R be merged with that of A?

[GATE - 2017]

- (a) Relationship R is one-to-many and the participation of A in R is total
- (b) Relationship R is one-to-many and the participation of A in R is partial
- (c) Relationship R is many-to-one and the participation of A in R is total
- (d) Relationship R is many-to-one and the participation of A in R is partial

3. Which of the following is NOT a superkey in a relational schema with attributes V, W, X, Y, Z and primary key V Y?

- [GATE - 2016]
- (a) VXYZ
 - (b) VWXZ
 - (c) VWXY
 - (d) VWXYZ

4. A database of research articles in a journal uses the following schema.

(VOLUME, NUMBER, STARTPAGE, ENDPAGE, TITLE, YEAR, PRICE)

The primary key is (VOLUME, NUMBER, STARTPAGE, ENDPAGE) and the following functional dependencies exist in the schema.

(VOLUME, NUMBER, STARTPAGE, ENDPAGE) → TITLE

(VOLUME, NUMBER) → YEAR

(VOLUME, NUMBER, STARTPAGE, ENDPAGE) → PRICE

The database is redesigned to use the following schemas.

(VOLUME, NUMBER, STARTPAGE, ENDPAGE, TITLE, PRICE)

(VOLUME, NUMBER, YEAR)

Which is the weakest normal form that the new database satisfies, but the old one does not?

[GATE - 2016]

- (a) 1NF
- (b) 2NF
- (c) 3NF
- (d) BCNF

5. Consider an Entity-Relationship (ER) model in which entity sets E₁ and E₂ are connected by an m : n relationship R₁₂, E₁ and E₃ are connected by a 1 : n (1 on the side of E₁ and n on the side of E₃) relationship R₁₃. E₁ has two single-valued attributes a₁₁ and a₁₂ of which a₁₁ is the key attribute. E₂ has two single-valued attributes a₂₁ and a₂₂ of which a₂₁ is the key attribute. E₃ has two single-valued attributes a₃₁ and a₃₂ of which a₃₁ is the key attribute. The relationships do not have Key attributes. The relationship do not have any attribute. If a relational model is derived from the above ER model, then the minimum number of relations that would be generated if all the relations are in 3NF is _____.

[GATE - 2015]

CHAPTER - 2**TRANSACTION PROCESSING CONCEPTS****2.1 INTRODUCTION**

A database system is classified on the basis of number of users who can use the system at the same time.

DBMS is classified into

1. Single user DBMS
2. Multiple user DMBS

2.1.1 Single user DBMS

A DBMS is single-user if at most one user at a time can use the system. Single user DBMS are mostly restricted to personal computer system.

2.1.2 Multi user DBMS

1. A DBMS is Multi-user if at most one user at a time (concurrency) can use the system.
2. Concurrency can be achieved either using single processor or multiprocessors.
3. In Single processor, concurrency is achieved using interleaving technique while multiprocessors use parallel processing technique for concurrency.

Example.

- (i) An Airline reservation system that is used by number of travel agents and reservation clerks concurrently.
- (ii) Online Banking System where number of transactions are processed of customers concurrently.

2.2 TRANSACTION**2.2.1 Two important assumptions about transaction:**

1. Transaction interacts with each other only via database read and write operations.
2. A database is fixed collection of independent objects. When objects are added to or deleted from a database or there are relationships between database objects that we want to exploit for performance, some additional issue arise.
3. A transaction is an executing program that forms a logical unit of database processing. It is an atomic unit of work that is either completed entirely or not done at all.
4. A transaction includes one or more database access operations (insertion, deletion, modification, or retrieval).
5. The transactions boundaries are specified by explicit begin and end statements.
6. If the database operations in a transaction do not update the database but only retrieve data, the transaction is called a read-only transaction.
7. Basic database access operations of transactions are
 - (i) Read_item(x): It reads a database item named X. It can be written as $r(x)$.
 - (ii) Write_item(x): It writes a database item named X. It can be written as $w(x)$.
8. A transaction includes read and write operations to access and update the database.
9. Execution of read operation command
10. Find the address of the disk block that contains items X.
11. Copy that disk block into a buffer in main memory (if that disk block is not already in some main memory buffer).
12. Copy item X from the buffer to the program variable named X.

ASSIGNMENT

1. The concept of locking can be used to solve the problem of

- (a) Lost update
- (b) Uncommitted dependency
- (c) Inconsistent data
- (d) Deadlock

2. What are the potential problems when a DBMS executes multiple transitions concurrently?

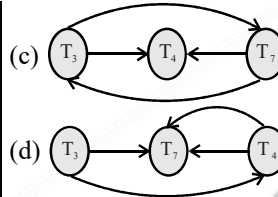
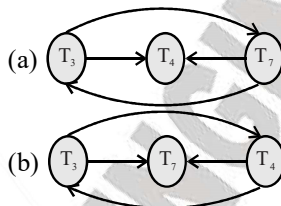
- (i) The lost update problem
 - (ii) The dirty read problem
 - (iii) The unrepeatable problem
 - (iv) The phantom problem
- (a) 3 and 4 only
 - (b) 1, 2 and 4 only
 - (c) 2 and 3 only
 - (d) All of 1, 2, 3 and 4

3. Assume transaction A holds a shared lock R. If transaction B also request for a shared lock on R, it will

- (a) Result in a deadlock situation
- (b) Immediately be granted
- (c) Immediately be rejected
- (d) Be granted as soon as it is released by A

4. Consider the following transaction schedule

T ₃	T ₄	T ₇
R(Q)		
	W(Q)	
W(Q)		R(Q)
		W(Q)



5. Consider the following three schedules of transactions T1, T2 and T3. [Notation: In the following NYO represent the action Y (Y or read, write) performed by transaction N on object O.]

(S1)	(S2)	(S3)
2RA	3RC	2RA
2WA	2RA	3RC
3RC	2WA	3WA
2WB	2WB	2WA
3WA	3WA	2WB
3WC	1RB	3WC
1RA	1RB	1RA
1RB	1WA	1RB
1WA	1WB	1WA
1WB	3WC	1WB

Which of the following statements is TRUE?

- (a) S1, S2 and S3 are all conflict equivalent to each other
- (b) No two of S1, S2 and S3 are conflict equivalent to each other
- (c) S2 is conflict equivalent to S3, but not to S1
- (d) S1 is conflict equivalent to S2 but not to S3

6. In case of timestamp ordering R-timestamp (Q) denotes

- (a) The largest timestamp of any transaction that execute read (Q) successfully.
- (b) The average timestamp of any transaction that execute read (Q) successfully
- (c) The average timestamp of any transaction that execute read (Q) unsuccessfully

(d) The smallest timestamp of any transaction that execute read (Q) successfully

7. Locking was introduced into databases so that

- (a) Keys can be provided to maintain security
- (b) All simultaneous transactions are prevented
- (c) Passwords can be provided to maintain security
- (d) Consistency can be enforced.

8. Which level of locking provides the highest degree of concurrency flight in a relational database?

- (a) Page
- (b) Table
- (c) Row
- (d) Page, table and row level locking allow the same degree of concurrency

9. Which of the following is true for two-phase locking?

- (a) Lock acquisition is the second phase
- (b) Locks can be acquired at any time
- (c) Locks are acquired in the first phase
- (d) None of the above

10. Consider the following statements

- (a) S1 : Entire database cannot be locked
- (b) S2 : Entire relation can be locked

Which of the above statements is/are true?

- (a) S1 only
- (b) S2 only
- (c) both S1 and S2 are true
- (d) both S1 and S2 are false

11. Choose the false statement.

- (a) Timestamp protocol is deadlock free
- (b) Two phase locking guarantees serializability
- (c) Strict two phase locking is deadlock free
- (d) Timestamp protocol may not result recoverable schedule

12. In DBMS without concurrency control, what consistency problem does the following transaction schedule depict?

Time	Transaction A	Transaction B
T1	read R	

T2		
T3		Read R
T4		write R

- (a) Dirty Read
- (b) Uncommitted Dependency
- (c) Inconsistent Analysis
- (d) Lost Update

13. For the schedule given below, which of the following is correct?

- (i) Read A
- (ii) Read B
- (iii) Write A
- (iv) Read A
- (v) Write A
- (vi) Write B
- (vii) Read B
- (viii) Write B

(a) This schedule is serializable and can occur in a scheme using 2PL protocol

- (b) This schedule is serializable but cannot occur in a scheme using 2PL protocol
- (c) This schedule is not serializable but can occur in a scheme using 2PL protocol
- (d) This schedule is not serializable and cannot occur in a scheme using 2PL protocol

14. When n transactions are run concurrently and in an interleaved manner, the number of possible schedule are _____.

- (a) Much larger than n!
- (b) Much lower than n!
- (c) Much larger than (n - 1)!
- (d) Much lower than (n - 1)!

15. Consider the following schedules involving two transactions.

Which one of the following statements is TRUE?

S₁ : r₁ (X); r₁ (Y); r₂ (X); r₂ (Y); w₂ (Y); w₁ (X)
 S₂ : r₁ (X); r₂ (X); w₂ (Y); w₂ (Y); r₁ (Y); w₁ (X)

- (a) Both S1 and S2 are conflict serializable
- (b) S1 is conflict serializable and S2 is not conflict serializable
- (c) S1 is not conflict serialization and S2 is conflict serializable

GATE QUESTIONS

1. Two transactions T_1 and T_2 are given as

$T_1 : r_1(X)w_1(X)r_1(Y)w_1(Y)$

$T_2 : r_2(Y)w_2(Y)r_2(Z)w_2(Z)$

where $r_i(V)$ denotes a read operation by transaction T_i on a variable V and $w_i(V)$ denotes a write operation by transaction T_i on a variable V . The total number of conflict serializable schedules that can be formed by T_1 and T_2 is _____.

[GATE - 2017]

2. Consider the following database schedule with two transactions, T_1 and T_2 .

$S = r_2(X); r_1(X); r_2(Y); w_1(X); r_1(Y); w_2(X); a_1; a_2$

where $r_i(Z)$ denotes a read operation by transaction T_i on a variable Z , $w_i(Z)$ denotes a write operation by T_i on a variable Z and a_i denotes an abort by transaction T_i .

Which one of the following statements about the above schedule is TRUE?

[GATE - 2016]

- (a) S is non-recoverable
- (b) S is recoverable, but has a cascading abort
- (c) S does not have a cascading abort
- (d) S is strict

3. Suppose a database schedule S involves transactions T_1, \dots, T_n . Construct the precedence graph of S with vertices representing the transactions and edges representing the conflicts. If S is serializable, which one of the following orderings of the vertices of the precedence graph is guaranteed to yield a serial schedule?

[GATE - 2016]

- (a) Topological order
- (b) Depth-first order
- (c) Breadth-first order
- (d) Ascending order of transaction indices

4. Consider the following two phase locking protocol. Suppose a transaction T accesses (for read or write operations), a certain set of objects

$\{O_1, \dots, O_k\}$. This is done in the following manner.

Step 1: T acquires exclusive locks to O_1, \dots, O_k in increasing order of their addresses.

Step 2: The required operations are performed

Step 3: All locks are released.

This protocol will

[GATE - 2016]

- (a) Guarantee serializability and deadlock-freedom
- (b) Guarantee neither serializability nor deadlock-freedom
- (c) Guarantee serializability but not deadlock-freedom
- (d) Guarantee deadlock-freedom but not serializability

5. Which one of the following is not a part of the ACID properties of database transactions?

[GATE - 2016]

- (a) Atomicity
- (b) Consistency
- (c) Isolation
- (d) Deadlock-freedom

6. Consider the following transaction involving two bank accounts x and y .

read (x) ; $x := x - 50$; write (x) ; read (y) ;
 $y := y + 50$; write (y)

The constraint that the sum of the accounts x and y should remain constant is that of

[GATE - 2015]

- (a) Atomicity
- (b) Consistency
- (c) Isolation
- (d) Durability

7. Consider a simple check pointing protocol and the following set of operations in the log.

(start, T_4); (write, T_4 , y , 2, 3); (start, T_1);
(commit, T_4); (write, T_1 , z , 5, 7);

(checkpoint);

(start, T_2); (write, T_2 , x , 1, 9); (commit, T_2);
(start, T_3); (write, T_3 , z , 7, 2);

SOLUTIONS

Sol 1. (54)

There is only one conflict serializable schedule as

$T_1 \rightarrow T_2$, because last operation of T_1 and first operation of T_2 conflicts each other.

Number of schedules that are conflict serializable to $T_2 \rightarrow T_1$ is 53.

Proof: The operations of T_1 is

$R_1(x) W_1(x) R_1(y) W_1(y)$

The first operation of T_2 that conflicts with operation of T_1 is $W_2(y)$ but not $R_2(z)$, $W_2(z)$.

The number of places where $W_2(y)$ can appear is

Case-1. $W_2(y) R_1(x) W_1(x) R_1(y) W_1(y)$

Case-2. $R_1(x) W_2(y) W_1(x) R_1(y) W_1(y)$

Case-3. $R_1(x) W_1(x) W_2(y) R_1(y) W_1(y)$

Case-1. The number of positions that $R_2(z)$ $W_2(z)$ can come before $W_2(y)$ is ${}^5C_1 + {}^5C_2 = 15$ (either both can take same space or two different spaces).

Case-2. The number of positions that $R_2(z)$ $W_2(z)$ can come before $W_2(y)$ is ${}^4C_1 + {}^4C_2 = 10$

For each of these 10 positions $R_2(y)$ can take 2 positions before $W_2(y)$ therefore total possible schedules are $10 \times 2 = 20$

Case-3. The number of positions that $R_2(z)$ $W_2(z)$ can come before $W_2(y)$ is ${}^3C_1 + {}^3C_2 = 6$

For each of these 6 positions $R_2(y)$ can take 3 positions before $W_2(y)$ therefore total possible schedules are $6 \times 3 = 18$.

The total conflict serializable schedules as $T_2 \rightarrow T_1 = 15 + 20 + 18 = 53$

\therefore Total conflict serializable schedules = $1 + 53 = 54$

Sol 2. (c)

As there is no dirty-read in the given schedule, the schedule is both recoverable and cascadeless.

Sol 3. (a)

If a schedule is serializable, the topological order of a graph (precedence graph) yields a serial schedule.

Sol 4. (a)

Two phase locking protocol ensures serializability, but does not ensures freedom from deadlock.

Sol 5. (d)

ACID properties of database transaction is defined as:

A: Atomicity

C: Consistency

I: Isolation

D: Durability

Sol 6. (b)

x and y are bank accounts

Read(x) ; $x = x - 50$

Write(x) ; read(y) ; $y = y + 50$; write(y)

It is the property of consistency directly as it says that sum should remain constant before and after the transaction.

For example, $A = 1000$ $B = 2000$

and we want to transfer 500 from A to B .

then $A = 500$, $B = 2500$

But sum before and after will remain same.

Sol 7. (a)

Need to undo T_3 and T_1 as they are not committed and redo only T_2

Sol 8. (b)

Schedule is non-recoverable because transaction T_2 commits before T_1 gets failed. So it is non-recoverable even if T_1 go to initial state T_2 can not go because it has committed and committed transaction cannot go back to original position.

Sol 9. (d)

CHAPTER - 3
INDEXING

3.1 INTRODUCTION

1. It is a map to locate records of the database file on disk storage space.
2. It is used to speed up the retrieval of records in response to certain search conditions.
3. It provides secondary access paths, which provide alternative ways of accessing to records without affecting the physical placement of records on disk.
4. To find a record or records in database file, initially index or index file is accessed to get the block address of the search records.
5. Any field of the database file stored on disk can be used to create an index. This field is called indexing field.
6. Any file can have multiple indexes on its different fields.
7. Every index file contains two fields (Searching value, Block pointer) in each entry.

3.2 REQUIREMENT OF INDEXING

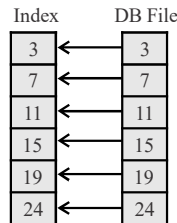
The data is shifted from the secondary memory (disk) to the main memory block by block. In the worst case, whole database file will have to be transferred to the main memory. And if binary search is applied for searching any record, $\log_2 N$ blocks will need to be transferred, if file is in N-blocks. So, to minimize the number of blocks to transfer from secondary memory to main memory, indexing is used because index file has very small size.

3.2.1 Important points of Indexing

1. It reduces I/O cost
2. It provides alternative path to access records without affecting the physical records on the disk
3. It is used to step access the desired data.
4. Any field can used create index.
5. Multiple index can exist for same file.
6. Self of attributes which is used to look up records in a file.
7. Index is classified into two categories
 - (i) Dense Index
 - (ii) Sparse Index

3.2.2 Dense Index

1. It contains block address of each record of database file.
2. It contains number of entries equal to the number of records in database file.



Example.

If database file contains records of students according to their names in alphabetical order But its Dense index file contains field student_id to locate records as follows

ASSIGNMENT

1. Suppose that the search field is a non-ordering key field, and we construct a B-tree on this field. Where search field is of 9 bytes, the disk block size is 512 bytes, a record pointer is of 7 bytes and a block pointer is of 6 bytes. Assume each node of the B-tree is 69 percent full. Then calculate the numbers of nodes in this B-tree of level 4.
2. Consider a disk with block size $B = 512$ bytes. A block pointer is $P = 6$ bytes long, and a record pointer is $P_R = 7$ bytes long. A file has $r = 30,000$ EMPLOYEE records of fixed length. Each record has PHONE (9 bytes), DOB (8 bytes), ENO (9 bytes), DEPARTMENTCODE (9 bytes), ADDRESS (40 bytes), PHONE (9 bytes), DOB (8 bytes), SEX (1 bytes), JOBCODE (4 bytes), SALARY (4 bytes, real number). An additional byte is used as a deletion marker:
- Calculate the record size R in bytes.
 - Calculate the block factor and the number of file blocks b , assuming an un-spanned organization.
 - Suppose that the file is ordered by the key field ENO and we want to construct a primary index on ENO. Calculate:
 - The index blocking factor bfr ; (which is also the index fan-out fo);
 - The number of first-level index entries and the number of first-level index blocks;
 - The number of levels needed if we make it into a multilevel index;
 - The total number of blocks required by the multilevel index; and
 - The number of block accesses needed to search for and retrieve a record from the file-given its ENO value-using the primary index.
3. Consider a B^+ tree with fan out (the number of block pointers per node) equal to 3 for the following set of key values 80, 50, 10, 70, 30, 100, 90. Assume that the tree is initially empty and the values are added in the order given.
- Show the tree after insertion of 10, 30, and 90.
 - The key values 30 and 10 are now deleted from the tree in the order. Show the tree after each deletion.
4. A parts file with Part# as key field includes records with the following Part# values: 8, 5, 1, 7, 3, 12, 9, 6. Suppose that the search field values are inserted in the given order in a B^+ tree with leaf order $p_{leaf} = 3$; show the final tree will-look-like.

SOLUTIONS

Sol 1.

Apply B-tree structure formula
 $P(B.P) + (P - 1) \text{ key} + (P - 1) R.P \leq 512$
 Let order of B-Tree is P
 Given, B.P= 6 bytes, Key = 9 bytes and R.P=7 bytes.
 $6P + (P-1)(9+7) \leq 512$
 $\Rightarrow P = 24$
 As B-Tree node is 69 % full $\therefore P = 0.69 * 24 = 17$
 So, each B-Tree node has 17 pointers and 16 keys

Level	Nodes
1	1
2	17
3	17*17
4	17*17*17

At level 4 , there are 4913 nodes

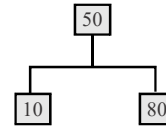
Sol 2.

(i) Record size R
 $= 9+8+9+9+40+9+8+1+4+4+1$
 $= 102 \text{ Byte}$
 (ii) Block factor = Number of records in one block.
 Given,
 Block size = 512 bytes
 Each record is of size = 102 Bytes
 $\therefore \text{Number of records in a Block} = \frac{512}{102} = 5$ records.
 $\Rightarrow \text{Block factor} = 5$
 No. of records in file = 30,000
 $\therefore \text{Number of blocks for a file} = \frac{30000}{5} = 6000$ blocks
 (a) Entry of primary index file contains key and Block pointer only
 $\therefore \text{Record size of Primary index file} = 9 + 6 = 15 \text{ bytes}$

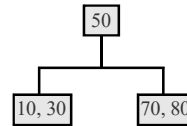
$\therefore \text{Index blocking factor } bfr_i = \frac{\text{Block size}}{\text{Size of index file's Record}} = \frac{512}{15} = 34$
 (b) No. of entries in Primary index = 30,000
 Each block can store 34 entries of primary index.
 $\therefore \text{Number of blocks at the first level of primary index} = \frac{30,000}{34} = 883s$
 (c) Number of blocks at Ist level = $\frac{30000}{34} = 883$
 Number of blocks at 2nd level = $\frac{883}{34} = 26$
 Number of blocks at 3rd level = $\frac{26}{34} = 1 \text{ block.}$
 \therefore There are 3 levels of multilevel Primary index.
 (d) Total number of blocks required for primary index = $883 + 26 + 1 = 910$
 (e) Number of blocks needed to search or retrieve a record from file includes 3 index blocks and 1 data block
 \therefore Number of blocks required to access a record = 4 blocks

Sol 3.

(i) B⁺ tree after insertion of 10.



(ii) B⁺ tree after insertion of 30.



(iii) B⁺ tree after insertion of 90.

GATE QUESTIONS

1. In a B⁺ tree, if the search-key value is 8 bytes long, the block size is 512 bytes and the block pointer size is 2 bytes, then the maximum order of the B⁺ tree is _____.

[GATE - 2017]

2. B⁺ Trees are considered BALANCED because

[GATE - 2016]

- (a) The lengths of the paths from the root to all leaf nodes are all equal.
- (b) The lengths of the paths from the root to all leaf nodes differ from each other by at most 1.
- (c) The number of children of any two non-leaf sibling nodes differ by at most 1.
- (d) The number of records in any two leaf nodes differ by at most 1.

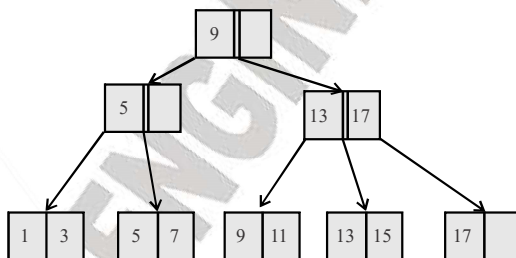
3. A file is organized so that the ordering of data records is the same or close to the ordering of data entries in some index. Then that index is called

[GATE - 2015]

- (a) Dense
- (b) Sparse
- (c) Clustered
- (d) Unclustered

4. With reference to the B⁺ tree index of order 1 shown below, the minimum number of nodes (including the Root node) that must be fetched in order to satisfy the following query "Get all records with a search key greater than or equal to 7 and less than 15" is _____.

[GATE - 2015]



5. In a B⁺ tree in which the search key is 12B long, block size is 1024B, record pointer is 10B long and block pointer is 8B long. The maximum number of keys that can be accommodate in each non-leaf node of the tree is _____.

[GATE - 2015]

6. An index is clustered, if

[GATE - 2013]

- (a) It is on a set of fields that form a candidate key.
- (b) It is on a set of fields that include the primary key.
- (c) The data records of the file are organized in the same order as the data entries of the index.
- (d) The data records of the file are organized not in the same order as the data entries of the index.

7. Consider a B⁺ - tree in which the maximum number of keys in a node is 5. What is the minimum number of keys in any non-root node?

[GATE - 2010]

- (a) 1
- (b) 2
- (c) 3
- (d) 4

8. The following key values are inserted into a B⁺ - tree in which order of the internal nodes is 3, and that of the leaf nodes is 2, in the sequence given below. The order of internal nodes is the maximum number of tree pointers in each node, and the order of leaf nodes is the maximum number of data items that can be stored in it. The B⁺ - tree is initially empty.

10, 3, 6, 8, 4, 2, 1

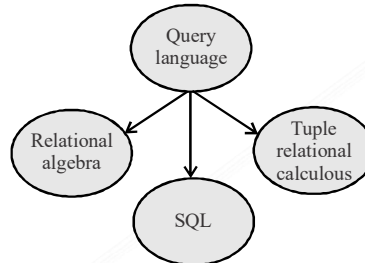
The maximum number of times leaf nodes would get split up as a result of these insertions is

[GATE - 2009]

- (a) 2
- (b) 3
- (c) 4
- (d) 5

CHAPTER - 4
QUERY LANGUAGE

4.1 INTRODUCTION



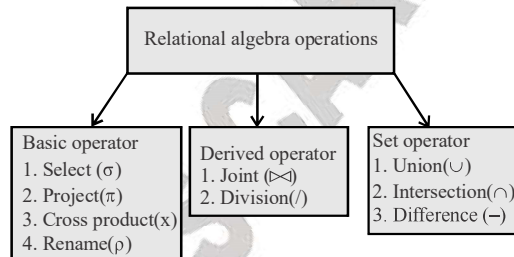
Query is always executed tuple by tuple and one tuple at a time.

4.2 RELATIONAL ALGEBRA (RA)

The relational algebra is a procedural query language. It consists of set of operators that take one or two relations as input and product a new relation as their result. RA forms the core component of a relational query engine. RA provides a framework for query optimization. SQL queries are internally translated into RA expressions.

Relation algebra by default eliminates the duplicate tuples from the result.

4.2.1 RA Operations



4.2.1.1 Basic Operator

(i) The Select Operation (σ)

It is a unary operator and is used to select those tuples of a relation that satisfy a given condition.

(a) Notation

$$\sigma_{\theta}(r)$$

Where σ is select operator (sigma)

θ is selection condition, r is relation

(b) Result

A relation with the same schema as r consisting of the tuples in r that satisfy condition θ .

(c) Properties

It is commutative as $\sigma_{C1}(\sigma_{C2}(r)) = \sigma_{C2}(\sigma_{C1}(r))$.

(d) Select Condition

Atomic or composite condition. Composite condition is atomic conditions combined with logical operators AND, OR and NOT.

ASSIGNMENT

1. Which of the following operations is not part of the five basic set operations in relational algebra?

- (a) Union (b) Division
(c) Cartesian Product (d) Set Difference

2. Which of the following relational algebraic operation is not a commutative operation?

- (a) Union (b) Intersection
(c) Selection (d) Projection

3. Which of the following is wrong?

- (a) $\pi_{L1} \cup L2 (E1 \bowtie \theta E2) = (\pi_{L1} (E1)) \bowtie \theta (\pi_{L1} (E2))$
(b) $\sigma_P (E1 - E2) = \sigma_P (E1) - \sigma_P (E2)$
(c) $\sigma_{\theta_1 \wedge \theta_2} (E) = \sigma_{\theta_1} (\sigma_{\theta_2} (E))$
(d) $E1 \bowtie \theta E2 = E2 \bowtie \theta E1$

4. Which of the following is correct?

- (a) An SQL query automatically eliminates duplicates
(b) An SQL query will not work if there are no indexes on the relations
(c) SQL permits attribute names to be repeated in the same relation
(d) None of the above

5. In SQL, relations can contain null values, and comparisons with null values are treated as unknown. Suppose all value are treated as false. Which of the following pairs is not equivalent?

- (a) $x = 5$ not (not($x = 5$))
(b) $x = 5$ $x > 4$ and $x < 6$, where x is an integer
(c) $x \neq 5$ not ($x = 5$)
(d) None of the above

6. Consider the following statements:

- S1: we can use IN in place of = ANY
S2: we can use IN in place of = ALL
Which one of the following is true?
(a) S1 is true
(b) S2 is true

- (c) Both S1 and S2 are true
(d) None of the above

7. Constraints are specified as a part of

- (a) Data definition
(b) Data manipulation
(c) Data control
(d) None of the above

8. Which of the following tuple relational calculus finds all customers who have a loan amount of more than 1200?

- (a) $\{t \mid (\text{Customer_name}) \mid t \in \text{borrow} \wedge t[\text{amount}] > 1200\}$
(b) $\{t \mid (\text{Customer_name}) \in \text{borrow} \wedge t[\text{amount}] > 1200\}$
(c) $\{t \mid \exists s \in \text{borrow} (t[\text{Customer_name}] = s[\text{Customer_name}] \wedge s[\text{amount}] > 1200)\}$
(d) None of the above

9. The following tables gives details of employees in a company department

Emp_ID	Job	Salary	Des-ID
110	Designing	25000	SW
115	Calibrating	19000	QA
120	Programming	26000	SW
135	Quality Assuring	18000	QA
150	Consulting	45000	CO
168	Consulting	35000	CO
188	Analysis and Design	22000	SW

Which of the following SQL statement gives the average Salary for each designation ID (Des_ID)?

- (a) `SELECT AVG (Salary) FROM Employee`
(b) `SELECT Des-ID, AVG (Salary) FROM Employee ORDER BY Des_ID`
(c) `SELECT Des_ID, Salary FROM Employee GROUP BY Des-ID`
(d) `SELECT Des_ID, AVG (Salary) FROM Employee GROUP BY Des_ID`

GATE QUESTIONS

1. Consider a database that has the relation schema EMP (EmpID, EmpName, and DeptName). An instance of the schema EMP and a SQL query on it are given below.

EMP		
EmpID	EmpName	DeptName
1	XYA	AA
2	XYB	AA
3	XYC	AA
4	XYD	AA
5	XYE	AB
6	XYF	AB
7	XYG	AB
8	XYH	AC
9	XYI	AC
10	XYJ	AC
11	XYK	AD
12	XYL	AD
13	XYM	AE

```
SELECT AVG(EC.Num)
FROM EC
WHERE (DeptName, Num) IN
  (SELECT DeptName, COUNT (EmpId) AS
    EC(DeptName,
    Num)
  FROM EMP
  GROUP BY DeptName
```

The output of executing the SQL query is _____.

[GATE - 2017]

2. Consider the following database table named top-scorer.

top scorer		
player	country	goals
Klose	Germany	16
Ronaldo	Brazil	15
G Muller	Germany	14
Fontaine	France	13
Pele	Brazil	12

Klinsmann	Germany	11
Kocsis	Hungary	11
Batistuta	Argentina	10
Cubillas	Peru	10
Lato	Poland	10
Lineker	England	10
T Muller	Germany	10
Rahn	Germany	10

Consider the following SQL query:
 SELECT ta.player FROM top_scorer AS ta
 WHERE ta.goals > ALL (SELECT tb.goals
 FROM top_scorer AS tb
 WHERE tb.country = 'Spain')
 AND ta.goals > ANY (SELECT tc.goals FROM
 top_scorer AS tc
 WHERE tc.country = 'Germany')
 The number of tuples returned by the above
 SQL query is _____.

[GATE - 2017]

3. Consider the following database table named water_schemes:

Water schemes		
Scheme_no	District_name	Capacity
1	Ajmer	20
1	Bikaner	10
2	Bikaner	10
3	Bikaner	20
1	Churu	10
2	Churu	20
1	Dungargarg	10

The number of tuples returned by the following
 SQL query is _____.
 with total(name, capacity) as
 select district_name, sum(capacity)
 from water_schemes
 group by district_name
 with total_avg(capacity) as
 select avg(capacity)
 from total
 select name

SOLUTIONS

Sol 1. (26)

Result of inner query

Dept Name	Number
AA	4
AB	3
AC	3
AD	2
AE	1

Sol 2. (7)

The output of the query is ta.player
 Klose
 Ronaldo
 G Muller
 Fontaine
 Pele
 Klismann
 Koesis

Sol 3. (2)

Total	
Name	Capacity
Ajmer	20
Bikaner	40
Churu	30
Dungargargh	10

Total avg Capacity
25

The result of the query is: name
 Bikaner, Churu

Sol 4. (b)

Because in SQL SELECT command retains duplicates by default. In order to eliminate those duplicates we have to write DISTINCT Keyword i.e SELECT DISTINCT

Sol 5. (a)

Sol 6. (a)

Sol 7. (c)

Sol 8. (a)

Optimized version is $\pi_{A_1}((\sigma_{F_1 \wedge F_2})(r))$

Sol 9. (d)

Sol 10. (d)

So, an employee whose ALL customers gives him GOOD rating is chose. All such employees are chose.

Sol 11. (b)

Sol 12. (a)

All the four queries will select the Sname of students with no duplication having Roll number same in both R and S Table and course number as 107 and percentage greater than 90.

Sol 13. (c)

Q and R are True about SQL Query.

Sol 14. (a)

$(A \cup B) \bowtie_{A.id > 40 \vee C.id < 15} C$

$(A \cup B) \bowtie C \Rightarrow$

A.id	Name	Age	C.id	Name	Age
12			10		
12			99		
15			10		
15			99		
25			10		
25			99		
98			10		
98			99		
99			10		
99			99		

We have to deal with id only. So, tuples with A.id > 40 or C.id < 15 will be selected.