INDIAN STATISTICAL INSTITUTE

Mid Semestral Examination

M. Tech (CS) - I Year, 2018-2019 (Semester - I)

Probability and Stochastic Processes

Date: 03.09.2018 Maximum Marks: 90 Duration: 3 Hours

Note: This is a 2-page question paper. The question paper is of 100 marks.

Answer as much as you can, but the maximum you can score is 90.

(Q1) Let there be n sticks each of which is broken into one long and one short part. The $2n$ parts are arranged into n pairs from which new sticks are formed. Find the probability that

(a) the parts will be joined in the original order.

(b) that all long parts are paired with short parts.

[6+6=12]

(Q2) Airlines find that each passenger who reserves a seat fails to turn up with probability $\frac{1}{10}$ independently of the other passengers. So, Indigo Airlines always sells 10 tickets for their 9 seater aeroplane while Air India always sells 20 tickets for their 18 seater aeroplane. Which is more often over-booked? [5]

(Q3) For the following, let X_1 and X_2 be two independent random variables. $\text{Bin}(\cdot,\cdot)$ and $\text{Poi}(\cdot)$ denote the Binomial and Poisson distributions, respectively. "pdf" means the probability distribution function.

(i) Let $X_i \sim \text{Bin}(n_i, p), i = 1, 2$. Find out the pdf of $X_1 + X_2$.

(ii) Let $X_i \sim \text{Poi}(\lambda_i), i = 1, 2$. Find out the pdf of $X_1 + X_2$.

[6+6=12]

(Q4) **(i)** Let X and Y be independent discrete random variables, and let $g, h : \mathbb{R} \to \mathbb{R}$. Show that $g(X)$ and $h(Y)$ are independent.

(ii) Let X and Y be independent Bernoulli random variables with parameter $\frac{1}{2}$. Show that $X + Y$ and $|X - Y|$ are dependent though uncorrelated.

[4+6=10]

(Q5) Let X be a random variable defined over a sample space Ω such that $\mathrm{E}[X] = \mu$. Show that $\Pr(X \geq \mu) > 0$ and $\Pr(X \leq \mu) > 0$. [4+4=8]

[Hints: Can you try to prove using contradiction?]

(Q6) Independent trials, each resulting in a success with probability p or a failure with probability $q = 1 - p$, are performed. Compute the probability that a run of n consecutive successes occurs before a run of m consecutive failures. [15]

(Q7) A person has a matchbox in his left pocket and another one in his right pocket. Both the matchboxes initially contained N match sticks. Whenever the person needs a match stick, he is equally likely to take the match box from either pocket. Consider the moment when the person first finds that one of his matchboxes is empty. What is the probability that there are exactly i matchsticks in the other box, $i = 0, 1, \ldots, N$? [10]

(Q8) Let X and Y be independent random variables taking positive integer values and having the same mass function $f(x) = 2^{-x}$ for $x = 1, 2, \ldots$. Find (i) $\Pr(\min\{X, Y\} \leq x)$; and (ii) $\Pr(X \text{ divides } Y)$. [5+5=10]

(Q9) Suppose we roll a standard fair die 200 times. Let X be the sum of the numbers that appear over the 200 rolls. Use Chebyshev's inequality to bound $\Pr[X \geq 750]$. [8]

(Q10) If X is a random variable with mean 0 and finite variance σ^2, then for any $a > 0$, show that $\Pr(X \geq a) \leq \frac{\sigma^2}{\sigma^2 + a^2}$. [10]

[Hints: The above is a different form of Chebyshev's inequality. Use Markov's inequality to prove it by observing $\Pr(X \geq a) = \Pr(X + b \geq a + b)$, for $b > 0$. You will obtain an expression involving σ, a and b. Now try to find a suitable b.]

INDIAN STATISTICAL INSTITUTE

MIDTERM EXAMINATION
M.TECH(CS) 1 YEAR

DATA AND FILE STRUCTURES

Date: 04.09.2018 Maximum marks: 60 Duration: 2.5 hours.

1. (a) Given the roots of two binary trees write a recursive routine to determine if the two trees are same. Two trees are same if they have the same structure and the corresponding nodes contains the same keys.

   (b) Given a binary min heap, describe a procedure which when given x as input outputs all nodes whose values are less than x. Your algorithm should take $O(N)$ time where N is the number of nodes which your algorithm outputs.

   $[3 + 7 = 10]$

2. (a) Suppose we have a pointer to a node in a singly linked list that is guaranteed not to be the last node in the list. We do not have pointers to any other nodes (except by following links). Describe an $O(1)$ algorithm that logically removes the value stored in such a node from the linked list, maintaining the integrity of the linked list.

   (b) You are given a priority queue with the following procedures: Insert() and ExtractMin(). Show how to implement an ordinary stack and an ordinary queue using the given priority queue data structure.

   $[2 + (4 + 4) = 10]$

3. (a) Show how to implement a queue with two ordinary stacks so that the amortized cost of each ENQUEUE and DEQUEUE operations is $O(1)$.

   (b) Consider an array $A[0, 1, \ldots, k-1]$ containing only zeros and ones. Thus A represents the number $x = \sum_{i=0}^{k-1} A[i]2^i$. To add 1 modulo 2^k to x we use the following procedure:

```
INCREMENT(A, k)
1. i ← 0;
2. while i < k AND A[i] = 1
3.     A[i] ← 0;
4.     i ← i + 1;
5. end while
6. if i < k,
7.     A[i] ← 1
8. end if
```

Given a number x, define the potential $\Phi(x)$ of x as the number of ones in the binary representation of x. Use a potential function argument to show that the amortized cost of an increment as described above is $O(1)$. Assume that the initial value of the counter x is zero.

[5 + 5 = 10]

4. A $n \times n$ matrix A is called GOOD if the following are true:

   - Each entry of A is either a zero or a 1.
   - Each row and each column of A contains exactly one 1.

   Propose an efficient data structure to store a GOOD matrix such that any two GOOD matrices can be multiplied in $O(n)$ time. Give the multiplication algorithm.

   [10]

5. A d-ary heap is like a binary heap but with one possible exception, non leaf nodes have d children instead of two children.

   (a) Discuss an array implementation of a d-ary heap. In particular, write procedures for finding the children and parent of a given node.

   (b) What is the height of a d-ary heap of n elements in terms of n and d.

   (c) Give an efficient procedure to implement Extract-Max in a d-ary max-heap. Analyze the running time in terms of d and n.

   [2 + 2 + 6 = 10]

6. Let $\mathcal{H}$ be an universal family of hash function where each $h \in \mathcal{H}$ maps U to $\{0, 1, \ldots, m_1\}$, where U is the set of all possible keys and m a fixed positive integer. Let T be a hash table (with chaining) which is constructed using a hash function uniformly selected from $\mathcal{H}$, thus T has m slots. Suppose T contains n elements ad let $\alpha = n/m$. Prove that for any key $k \in U$, the expected length of the chain in T to which k is hashed is at most $(1 + \alpha)$.

   [10]

INDIAN STATISTICAL INSTITUTE
Mid-Semester Examination
M. Tech. (CS) I year (1st Sem): 2018–2019
Algebraic Structures

Date: 05. 09. 2018 Total Marks : 70 Time : 2.5 Hours

Answer as much as you can. Maximum you can score is 60.

1. (a) What are the utilities of Euclid's algorithm and extended Euclid's algorithm?
   (b) Clearly specify both the algorithms with the input parameters, the outputs and the internal steps of computation.

   [(2+2)+(4+6)=14]

2. (a) How many elements in a group of order n has their order as 1?
   (b) Suppose a is a generator of the cyclic group G. Prove that a^{-1} is also a generator.

   [4+6=10]

3. List all subgroups of $\mathbb{Z}_{30}$. Find their orders and generators. [8+3+3=14]

4. (a) Find all orbits of $\left(\begin{smallmatrix}1&2&3&4&5&6&7&8\\2&3&5&1&4&6&8&7\end{smallmatrix}\right)$.
   (b) Find the even and odd permutations of S_3.
   (c) Show that S_1 and S_2 are cyclic, but S_3 is not.

   [6+6+8=20]

5. (a) Find all cosets of the subgroup $4\mathbb{Z}$ of $\mathbb{Z}$.
   (b) Suppose a is an arbitrary element of order m in a group G of order n. Show that
      i. $m|n$.
      ii. n prime $\Rightarrow G$ is cyclic.
      iii. $a^n = e$.

   [4+4+4=12]

INDIAN STATISTICAL INSTITUTE

Semestral Examination: 2018-19

Course Name: M. Tech. I Year

Subject Name: Database Introduction to Programming

Date: 06/09/2018 **Maximum Marks: 60** **Duration: 3hours**

Note: Answer all questions

1. Explain/compare Data Abstraction and Encapsulation in OOP.

10

2. What does static mean in C. Explain its benefits (if any).

5

3. Consider the following sample structure declaration in C:

```
struct stNODE {
        int iRollNo;
        char * cpName;
        struct NODE * stNODEpNext;
} *stNODEpStart;
```

Assume stNODEpStart is used to store the starting address of a linked list (with each node as struct NODE type) or NULL (when the list does not exist).

Write the following C functions with proper parameter passing, when ever required. Use minimum number of variable.

i) Create a new node with name and roll number collected from the key board inside the function.

ii) Find a given roll no entered from keyboard in the main program and if it exist in the given list (whose first node address is kept in stNODEpStart), delete it.

12+18=30

4. You have a softcopy of a story book stored in a file named MyStory.txt. Write a C program to find the word(s) that appeared least (non-zero) in this book.

15

INDIAN STATISTICAL INSTITUTE
M. Tech. (CS) - I year: 2018–2019
Computer Organization
Mid-Semestral Examination

Date: 06. 09. 2018 Marks: 50 Time: 3 Hours

Answer any part of any question. The maximum marks you can get is 50. The question is of 55 marks. Please write the part answers of a question at the same place. You may also evaluate your script with a pencil once you finish answering the paper. You are welcome to evaluate your own answers properly. However the final marks will be decided by the examiner.

1. (a) Explain 2's complements in detail with examples of 8-bit binary numbers.
   (b) What is 'int' in C programming language on a 32-bit environment?
   (c) Write a C program to justify your claim above and provide necessary examples with proper input and output to your program.
   (d) Describe a circuit that exploits full-adder to implement addition as well as subtraction of two n-bit integers.

   $2+2+3+3=10$

2. Consider two 2-bit integers y_3y_2 and y_1y_0.
   (a) How many bits are required to store the product z of these two integers?
   (b) Construct the truth table(s) where y_3, y_2, y_1, y_0 are the input bits and the output bits are z mod 4.
   (c) Describe relevant Karnaugh maps to obtain simplified Boolean circuits to provide the result.
   (d) Implement the circuit with two-input one-output logic gates of your choice. More credit will be given for implementation with less number of gates.

   $1+3+3+3=10$

3. (a) Briefly explain the IEEE 754 floating point format with examples.
   (b) Given two such floating point numbers, describe a method to multiply them.
   (c) Provide an outline of the combinational circuit for such a floating point multiplier.

   $2+4+4=10$

P. T. O.

4. Consider a 4-bit counter that can count cyclically from 1111 to 0000 (in the decreasing order) corresponding to every clock pulse.

   (a) Describe each step of your design to implement the counter.

   (b) Draw the complete circuit with Flip-Flops and logic gates of your choice.

   $5 + 5 = 10$

5. Consider that we need to study the evolution of the recurrence relation

$$x_{n+4} = x_{n+1} \oplus x_n, \text{ where } x_i \in \{0, 1\}.$$

   (a) How will you implement a combinational/sequential circuit to study this?

   (b) Consider that you start with $x_0 = 1, x_1 = x_2 = x_3 = 0$. Clearly explain how your circuit evolves.

   $5 + 5 = 10$

6. Prove that all n-input m-output Boolean functions can be implemented with two input NOR gates. 5

Midterm
MTech CS Discrete Mathematics, 2018

7th September, 2018

Time: 3 hours, Maximum Marks: 100

Answer as many questions as possible but the maximum possible marks one can obtain is 100. Your answers should be well-written and you should explain your arguments properly.

1. (5 marks) Let $x \in \mathbb{R}$ and $x > -1$. Prove that $(1+x)^n \geq 1 + nx$ for all natural numbers n.

2. Prove or disprove the following set of asymptotic relations: (3 marks each)

   (a) $(2.9)^{\log_2 n} = \Theta(n^{\log_2 3})$

   (b) $\log\log n = \Omega((\log\log\log n)^{\log\log\log n})$

   (c) $n^4 \sim (1 - 1/n)^n n^3$

   (d) $2^{(\log n) - (\log\log n)} \sim 2^{(1-1/n)\log n}$

   (e) $n^{10(\log\log n)^{100}} = \Theta((\log n)!)$

3. (10 marks) The *Lucas Sequence* $1, 3, 4, 7, 11, 18, 29, \ldots$ is defined by $a_1 = 1, a_2 = 3, a_n = a_{n-1} + a_{n-2}$. Prove that $a_n = O(1.75^n)$.

4. (7+7 marks) For natural number p and q, the Ramsey number $R(p, q)$ is defined as the smallest integer n so that among any n people, there exist p of them who know each other, or there exist q of them who don't know each other. Prove that Note that $R(p, 1) = R(1, q) = 1$. Prove that:

   (a) $R(p+1, q+1) \leq R(p, q+1) + R(p+1, q)$

   (b) $R(p, q) \leq C_{p-1}^{p+q-2}$

5. (6 Marks) Write the negation of the following statement:

$$\forall x \geq 0 \; \exists y \in \mathbb{N} \;\; (y \geq x) \wedge (y \text{ is a prime})$$

6. (5 + 5 Marks)

   (a) We know that $\sqrt{3}$ is not rational. Using this prove that $\sqrt{3} + \sqrt{24}$ is not rational.

   (b) If m is a positive integers such that $\sqrt{m}$ is not rational then prove that for any positive integer n the number $\sqrt{m} + \sqrt{n}$ is not rational.

7. (10 Marks) In the Towers of Hanoi problem, there are three posts and n disks of different sizes. Each disk has a hole through the center so that it fits on a post. At the start, all n disks are on post #1. The disks are arranged by size so that the smallest is on top and the largest is on the bottom. The goal is to end up with all n disks in the same order, but on a different post. In a single move one can move one disk from one post and place in another post. At no point can one place a bigger disk over a smaller disk in any post. How many steps will be taken to move the n disk from the Post #1 to any other post.

8. (10 Marks) Prove that at a party with at least two people, that there are two people who know the same number of people there (not necessarily the same people - just the same number) given that every person at the party knows at least one person. (Note that nobody can be his or her own friend.)

9. (3+ 7 Marks) A tournament is a directed graph (digraph) obtained by assigning a direction for each edge in an undirected complete graph. That is, it is an orientation of a complete graph, or equivalently a directed graph in which every pair of distinct vertices is connected by a single directed edge.

   (a) For any given n, give an example of a tournament which has no directed cycle.

   (b) Prove that a tournament has a directed 3-cycle if and only if it has a directed cycle.

10. (10 marks) How many non-increasing functions be there from $\{1, 2, \ldots, n\}$ to $\{1, 2, \ldots, k\}$.

11. (10 marks) If G is a labeled complete graph, K_n, on n vertices and u, v, w be three distinct vertices in the vertex set of G. How many different paths are there from u to v passing through w.

12. (10 marks) If $G = (V, E)$ is a graph on n vertices such that all the vertices have even degree. Show that the edge set E can be partitioned into pairwise disjoint sets $C_1, C_2, \ldots, C_k$ such that for all $1 \leq i \leq k$ the subgraphs (V, C_i) is a cycle and a collection of isolated vertices.

13. (10 marks) Let c_n^k be the number of ways to distribute n distinguishable balls can be k distinct buckets where the order of balls in a bucket does not count. Set up an ordinary generating function for c_n^k.

Endterm

MTech CS Discrete Mathematics, 2018

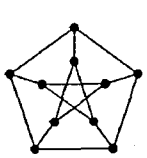
First year

12 November, 2018

Time: 3 hours, Maximum Marks: 100

Answer as many questions as possible but the maximum possible marks one can obtain is 100.
Your answers should be well-written and you should explain your arguments properly.
Note that this is a closed book exam.

1. (7 marks) Let $a_n, b_n \to \infty$. Show if $a_n = \Theta(b_n)$ then $\log a_n \sim \log b_n$.
2. (8 marks) Prove that a tournament always has a Hamiltonian path.
3. (10 marks) In how many ways can five distinct books be tied up in at most three bundles. (Here the order of books in the bundle does not matter and the bundles are not to be distinguished from one another.)
4. (10 marks) Prove or disprove: The following graph is planar



5. (10 marks) If G is a 2-vertex-connected graph and u, v are two vertices in G then there is a cycle in G passing through both u and v.
6. (10 marks) Prove that a planar graph (with $|V| \geq 4$) has at least 4 vertices of degree ≤ 5.
7. (10 marks) Let G be a graph with minimum degree 2. Show that there exists a connected graph with same degree sequence.
8. (10 marks) Let G be an arbitrary planar graph with v vertices, e edges, f regions, and m connected components. Prove that $v - e + f = m - 1$.

9. (10 Marks) Prove: every tournament on 2^{k-1} vertices contains a subtournament on k vertices which is a Directed Acyclic Graph (DAG). (A tournament is a directed complete graph: every pair of vertices is directed in exactly one direction.)

10. (10 marks) In a Discrete Math class there are 35 Mtech students and 35 JRF students. The instructor has decided that every Mtech student should team up with another JRF student and each team should do a different project. Ofcourse every student should be in exactly one team. To form the teams each student has been asked to present a set of exactly 4 students with whom he/she would want to team up. Thus every Mtech student must give a list of 4 JRF student of his/her choice and similarly every JRF student should give a list Mtech students of his/her choice. But the students have been asked to ensure that if student A is in the list of student B then student B must also be in the list of student A, so that it is mutually agreeable. Prove that whatever the submitted lists of the students may be the instuctor can surely form 35 teams such that everybody is happy. That is, for every team the Mtech student is in the list submitted by the JRF student and the JRF student is in the list submitted by the Mtech student.

11. (10 marks) Given an undirected graph the distance between any two vertices is the smallest length of the path that connects the two vertices. *Diameter* of an undirected graph is the maximum distance among all the pairs of vertices. So if the diameter of a graph is k that means any vertex can be reached from any other vertex by a path of length at most k.

    Among all the graphs of diameter 2 with n vertices, which one has the largest independent set? How large is it? Give proofs.

12. (10 marks) The *Lucas Sequence* $1, 3, 4, 7, 11, 18, 29, \ldots$ is defined by $a_1 = 1, a_2 = 3, a_n = a_{n-1} + a_{n-2}$. Give a closed form expression for the nth term of the Lucas Sequence.

13. (10 Marks) For a graph G, the chromatic number of G, denoted as $\chi(G)$, is the minimum number of colors required for coloring the vertices of G such that no two adjacent vertices have the same color.

    The independent number of G, denoted as $\alpha(G)$, is the maximum number of vertices of G that form an independent set, that is, no two vertices in the independent set are adjacent.

    Prove that: $\alpha(G)\chi(G) \geq n$.

14. (10 marks) Let $p^*(k)$ be the number of ways to partition the integer k into distinct integers, and $p_o(k)$ be the number of ways to partition integer k into odd integers. Prove that $p^*(k) = p_0(k)$.

INDIAN STATISTICAL INSTITUTE

ENDTERM EXAMINATION
M.TECH(CS) I YEAR

DATA AND FILE STRUCTURES

Date: 16.11.2018 Maximum marks: 100 Duration: 3 hours.

The paper contains 110 marks. Answer as much as you can, the maximum you can score is 100.

1. Give the tightest possible upper bound for the worst case running time for each of the following operations in terms of n. Give a very short explanation for your answer.
   (a) Enqueue a value onto a queue containing n values implemented as a circular array.
   (b) Finding the maximum value in a binary min heap of size n.
   (c) Finding the minimum value in a binary search tree of size n.
   (d) Given a binary search tree containing n integers, create an AVL tree containing the same values. You should not destroy the original binary search tree in the process.
   (e) Given an AVL tree containing n positive integers, print out all the even values contained in the tree in descending order.
   (f) Performing ExtractMin in a binomial min heap.

   $[6 \times 3 = 18]$

2. Give short answers to the following questions.
   (a) Describe a procedure to implement two stacks in a single array.
   (b) Write the pseudocode of a recursive procedure to find the number of nodes along the longest path from the root node down to the farthest leaf node. For an empty tree this number is zero, and for a single node tree it is 1.
   (c) Describe a procedure to find the third minimum in a binary min heap in constant time.
   (d) Suppose you are given a binary search tree in which each node contains an additional field, size, which contains the number of keys in the subtree rooted at this node. (Thus the size field of a leaf is 1, and the size field of the root is the total number of nodes in the tree). Using this size field, give pseudocode for a procedure findKth(t, k), which returns the k-th smallest element in the binary search tree t.
   (e) Consider a splay tree with nodes x, p and g where p is a left child of g and x a left child of p. For splaying x, what elementary operation would be required at this step? Assuming that each node has the fields left, right and parent, write a pseudo code to implement the elementary operation. Note that x may not be the leaf node and p, g can possibly have right children.

P.T.O.

(f) Consider a node x of degree k in a Fibonacci heap. Let $y_1, y_2, \ldots, y_k$ be children of x added in that order. Prove that $\deg(y_j) \geq j - 2$, for $1 < j \leq k$.

$[6 \times 7 = 42]$

3. We want to store intervals in a data structure, i.e., each element to be stored is a tuple (x, y) where x, y are real numbers and $x < y$. The following operations are to be supported:

   - Insert(x, y): Inserts the interval (x, y) in the data structure.
   - Count(z): Find the number of intervals (α, β) currently in the data structure such that $\alpha \leq z \leq \beta$. For example, after inserting the five intervals (3,10), (4,5), (6,12), (8,15), and (19,30) into the data structure, Count(9.1) is 3 and Count(17.2) is 0

   Propose a data structure which support both these operations in $O(\log(n))$ time. Describe the data structure and procedures precisely in English. You can use any data structure and/or procedures discussed in class without detailed description.

   [10]

4. We have to implement a phone book using a hash table which would have the facility of looking up a phone number given a name.

   (a) Design a hash function for the given purpose. Assume that names are strings over the alphabet which contains only uppercase English letters and blank. Also assume that there is only a single phone number associated with any name.

   (b) Compute the collision probability of the hash function you designed.

   (c) Assuming you use chaining, what would be the expected time required to lookup a phone number given a name.

   $[4 + 3 + 3 = 10]$

5. (a) Find the insertion order of the keys 19, 5, 1, 18, 3, 8, 24 that leads to a 2-3 tree of height 1.

   (b) Draw the corresponding left leaning red black tree (LLRBT) of the final tree obtained. Mark the red links appropriately. What is the height of the resulting LLRBT?

   (c) Draw the LLRBT tree obtained after inserting 2, 17 (in this order) to the previous tree.

   $[4 + 3 + 3 = 10]$

6. Suppose we have a skip list with n entries in which, rather than promoting each node to the next higher level with probability 1/2, we promote each node with probability p, for $0 < p < 1$.

   (a) Given a skip list with n entries, show that the expected number of list entries of height $\log_{\frac{1}{p}} n$ is $O(1)$. Briefly explain.

P. T. O.

(b) Show that (excluding the sentinel nodes) the total number of nodes in such a skip list is expected to be at most $\frac{n}{1-p}$.

[5 + 5 = 10]

7. Let $S = \{x_1, x_2, \ldots, x_n\}$ be a set of keys in a splay tree where $x_1 < x_2 < \cdots < x_n$. Prove that if you perform a sequence of lookups of $x_1, x_2, \ldots, x_i$ in that order, then the resulting splay tree will have the following structure:

   - x_i will be the root; and
   - the left subtree will be a linear chain of nodes containing the nodes $x_1, x_2, \ldots, x_{i-1}$.

   Hint: Use induction.

[10]

INDIAN STATISTICAL INSTITUTE
M. Tech. (CS) - I year : 2018-2019
Computer Organization
Semestral Examination

Date: 19. 11. 2018 Marks: 100 Time: 3 Hours

Answer any five questions. Each question is of 20 marks. Please write all the part answers of a question at the same place.

1. (a) How can you implement a 4-input XOR gate with 2-input NAND gates only? Try to minimize the number of gates.
   (b) Simplify the Boolean function $F(x, y, z) = \sum(0, 2, 3, 5, 6, 7)$ using Karnaugh map.
   (c) Prove that all n-input 1-output Boolean functions can be implemented with two input NAND gates only.

   $5 + 5 + 10 = 20$

2. (a) How can you implement a Full Adder?
   (b) How this circuit can be used to construct a 4-bit adder?
   (c) How can you modify the above circuit for both addition and subtraction of 4-bit numbers?
   (d) An 8×1 multiplexer has inputs A, B, and C connected to the selection inputs s_2, s_1, s_0 respectively. The data inputs are as follows: $I_1 = I_2 = I_7 = 0$, $I_3 = I_5 = 1$, $I_0 = I_4 = D$ and $I_6 = D'$. Write down the truth table of the Boolean function implemented by this multiplexer.

   $4 + 4 + 4 + 8 = 20$

3. Consider a 4-bit counter that can count the odd numbers cyclically from 1111 to 0001 (in the decreasing order) corresponding to every clock pulse.
   (a) Describe each step of your design to implement the counter.
   (b) Draw the complete circuit diagram with Flip-Flops and logic gates of your choice.

   $10 + 10 = 20$

4. (a) Explain the circuit of a multi-bit Tri-state Buffer.
   (b) Consider four identical RAM chips, each of 8-bit data width. How memory interleaving can be exploited to use those chips in a 32-bit environment?
   (c) Explain how I/O devices can be served using interrupts.
   (d) In what situation Direct Memory Access (DMA) is useful for data transfer?

   $5 + 5 + 5 + 5 = 20$

PTO

5. (a) "In the context of cache memory, write-back is convenient for Single CPU scenario." Do you agree? Briefly comment on this.

   (b) Consider a Fully Associative Cache with 2^8 slots. Each slot contains 2^4 bytes. Considering 32-bit address, identify the bits that will be used for tag. What is the total size of this cache memory?

   (c) Consider an implementation of the Set Associative Cache. There are 256 slots, 64 bytes per slot and there are 16 slots per set. Given the address A_{31-0}, how can one locate a specific slot in cache memory (if it is indeed in the cache)?

$6 + (6 + 2) + 6 = 20$

6. (a) What are the three broad types of instructions in MIPS? Provide an example for each one.

   (b) Identify the difference between MIPS 'add' and 'addi' instructions with proper examples.

   (c) How the sign extension is achieved in the MIPS 'addi' instruction? Briefly explain with an example.

   (d) Deduce what will be the content of register $r1 after execution of the following two instructions:
   lui $r1, 0x4567
   ori $r1, 0x89ef

$6 + 4 + 5 + 5 = 20$

7. (a) "Division by the decimal number 16 is always possible to implement using MIPS bit shift operators" – Justify.

   (b) You like to subtract two 32-bit integers stored in certain memory locations and further store the result in some other memory location. What should be the sequence of instructions in MIPS ISA?

   (c) Consider the MIPS instruction "beq $rs, $rt, offset". This allows to jump anywhere between the addresses (PC$-x$) to (PC$+y$). Explain what are the values of x, y.

   (d) Consider that each branch instruction requires 2 cycles, each store instruction requires 4 cycles and other instructions require 5 cycles each. The average clocks per instruction is 4.6. If the branch frequency is 12%, deduce the store frequency.

$5 + 5 + 5 + 5 = 20$

8. (a) Explain a five stage pipelined architecture in detail.

   (b) Write three MIPS instructions such that data hazard occurs in last two. Calculate total number of cycles to complete these three instructions.

$10 + 10 = 20$

INDIAN STATISTICAL INSTITUTE

First Semester Examination: 2018-19

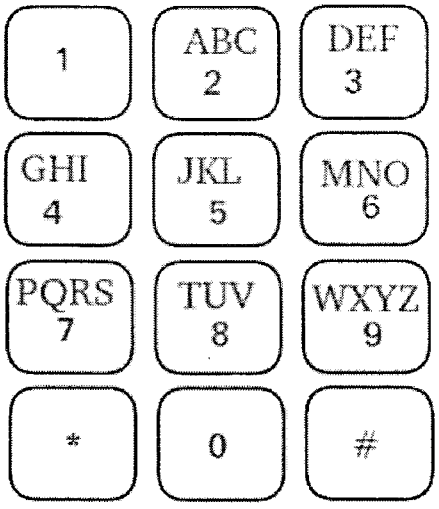
Course Name: M. Tech. I Year

Subject Name: Introduction to Programming

Date: 20/11/2018 **Maximum Marks: 100** **Duration: 3 hours**

Note: Answer all questions

1. Consider the following keypad, where texts and numbers are placed on the same key. For example, 2 has "ABC". If we want to write anything starting with 'A' we need to type key 2 once. If we want to type 'B', press key 2 twice and thrice for typing 'C



Given a keypad as shown above, and single digit number (less than 7), write a C program to list all words that are possible by pressing these numbers on the keypad.

25

2. What is the difference between Functional and Logic Programming? Explain Symbolic and Propositional Logic. Explain tail recursion.

25

3. There are two very long integer numbers store in two files - number1.txt and number2.txt. Write a C program which will read these two numbers from the respective files, and then add them and write the result to another file – result.txt.

25

P.T.O

4. Consider the following C code segment:

```
typedef struct _Person
{
    char* pFirstName;
    char* pLastName;
}Person;

new_Person(const char* const pFirstName,
           const char* const pLastName);
delete_Person(Person* const pPersonObj);

void Person_DisplayInfo(Person* const pPersonObj);
void Person_WriteToFile(Person* const pPersonObj,
                        const char* const pFileName);

person* new_Person(const char* const pFirstName,
                   const char* const pLastName)
{
    Person* pObj = NULL;
    pObj = (Person*)malloc(sizeof(Person));
    if (pObj == NULL)
    {
        return NULL;
    }
    pObj->pFirstName = malloc(sizeof(char)*(strlen(pFirstName)+1));
    if (pObj->pFirstName == NULL)
    {
        return NULL;
    }
    strcpy(pObj->pFirstName, pFirstName);
    pObj->pLastName = malloc(sizeof(char)*(strlen(pLastName)+1));
    if (pObj->pLastName == NULL)
    {
        return NULL;
    }
    strcpy(pObj->pLastName, pLastName);
    pObj->Delete = delete_Person;
    pObj->Display = Person_DisplayInfo;
    pObj->WriteToFile = Person_WriteToFile;
    return pObj;
}
```

Check for syntax error. Explain all the functions' intension with proper use (parameter, initialization etc).

INDIAN STATISTICAL INSTITUTE
Semester Examination
M. Tech. (CS) I year (1st Sem): 2018–2019
Algebraic Structures

Date: 22. 11. 2018 Total Marks : 110 Time : 3 Hours

Answer as much as you can. Maximum you can score is 100.

1. (a) Prove that an infinite cyclic group has exactly two generators.
   (b) We know that every group of prime order is cyclic. Is the converse true? That is, is the order of every cyclic group necessarily prime? Justify.

   [8+6=14]

2. (a) The *characteristic* of a ring R is the smallest positive integer n (if it exists) such that $na = 0$ for all $a \in R$. If no such integer exists, the characteristic of R is said to be zero. Prove that if R is an integral domain, then the characteristic of R is 0 or prime.
   (b) Which elements in $\mathbb{Z}_p$ are their own multiplicative inverses?
   (c) Show that a subring of a principal ideal domain need not be a principal ideal domain.

   [8+6+4=18]

3. (a) Show that for any ring R, $R[x]$ can never be a field.
   (b) Show that the ideal $< x^2 + 1 >$ is maximal in $\mathbb{R}[x]$.
   (c) Let $I =< 2 >$. Prove that $I[x]$ is not a maximal ideal of $\mathbb{Z}[x]$, though I is a maximal ideal of $\mathbb{Z}$.

   [6+10+8=24]

4. (a) If A and P are both $n \times n$ matrices and P is non-singular, then prove that A and $P^{-1}AP$ have the same eigen values.
   (b) Is the mapping $T(x_1, x_2, \ldots, x_n) = x_1$ linear? Justify.
   (c) Is the mapping $T(x_1, x_2, x_3) = (x_1 + 1, x_2 + 1, x_3 + 1)$ linear? Justify.

   [8+(2+4)+(2+4)=20]

5. What are the subfields of $GF(p^n)$? Do you find any analogy of this result with a similar result in group theory? Justify. [10+(2+4)=16]

6. What is the relationship between the zero of a polynomial in $\mathbb{F}[x]$ and a generator of the multiplicative subgroup of the extension field of $\mathbb{F}$? [8]

7. Can the same linear mapping can be represented by two different matrices? If no, justify. If yes, what is their relationship to each other? [10]

INDIAN STATISTICAL INSTITUTE

End Semestral Examination

M. Tech (CS) - I Year, 2018-2019 (Semester - I)

Probability and Stochastic Processes

Date: 26.11.2018 Maximum Marks: 100 Duration: 3.5 Hours

Note: This is a 3-page question paper. Apart from these 3 pages, there is a page for standard normal distribution table.

The question paper is of 100 marks. Answer as much as you can, but the maximum you can score in Group-A is 40, and in Group-B is 60.

$E[X]$ and var$[X]$ denote the expectation and variance of the random variable X, respectively.

Group A

(QA1) (i) Let X be a random variable such that $E[X] = \mu$. Let f be a convex function and there exists a value c such that

$$f(x) = f(\mu) + f'(\mu)(x - \mu) + \frac{f''(c)(x-\mu)^2}{2}.$$

Show that $E[f(X)] \geq f(E[X])$. [5]

(ii) Let X be a non-negative continuous random variable. Show that

$$E[X] = \int_0^\infty \Pr(X > x)\ dx$$

[5]

[5+5=10]

(QA2) (i) Define moment generating function for a continuous random variable. [2]

(ii) Let X be a normal random variable with mean μ and variance σ^2, i.e. $X \sim N(\mu, \sigma^2)$. Find the moment generating function corresponding to X. [4]

(iii) Let $X \sim N(\mu_x, \sigma_x^2)$ and $Y \sim N(\mu_y, \sigma_y^2)$. Calculate, using moment generating function, the pdf of $X + Y$. [4]

[2+4+4=10]

(QA3) Consider a program that includes one call to a function $\mathcal{F}$. Each call to $\mathcal{F}$ recursively generates new calls to $\mathcal{F}$, where the number of new functions generated is a Poisson random variable with parameter $\lambda > 0$. Assume that these random variables are independent for each call to $\mathcal{F}$.

(i) What is the expected number of copies of the function $\mathcal{F}$ generated?

(ii) Comment on the behaviour of the expected value calculated vis-a-vis the value of λ.

[8+2=10]

(QA4) (i) Let X be a random variable that assumes only nonnegative values. Show that for all $a > 0$, $\Pr(X \geq a) \leq \frac{E[X^m]}{a^m}$, where m is a positive integer ≥ 2. [5]

(ii) Let $X_1, \ldots, X_n$ be independent Poisson trials and $X = \sum_{i=1}^{n} X_i$ and $\mu = E[X]$. Then show that for $R \geq 6\mu$, $\Pr(X \geq R) \leq 2^{-R}$. [5]

[5+5=10]

(QA5) (i) State the weak law and strong law of large numbers. [1+1=2]

(ii) Point out the difference in the statements of weak law of large numbers and strong law of large numbers. Be as analytical as possible in your answer. [2]

(iii) Let $X_1, X_2, \ldots$ be a sequence of independent identically distributed random variables and assume that $E[X_i^4]$ is finite. Under the above assumption, prove the strong law of large numbers. [6]

[2+2+6=10]

Group B

(QB1) Let $X_1, \ldots, X_n$ be independent random variables such that $E[X_i] \neq 0$. Show that

$$\frac{\text{var}\left(\prod_{i=1}^{n} X_i\right)}{\prod_{i=1}^{n} E[X_i]^2} = \prod_{i=1}^{n} \left(\frac{\text{var}(X_i)}{E[X_i]^2} + 1\right) - 1$$

[12]

(QB2) X and Y are independent random variables, each distributed normally, as $N(0,1)$. Show that, for any fixed θ, the random variables $U = X\cos\theta + Y\sin\theta$ and $V = -X\sin\theta + Y\cos\theta$ are independent and find their distributions. [12]

(QB3) A *tournament* on a set V of n players is an orientation $T = (V, E)$ of the edges of the complete graph on the set of vertices V. Thus for every two distinct elements x and y of V, either (x, y) or (y, x) belongs to E, but not both. A simple interpretation of *tournament* is in terms of games where each distinct pair x, y of players, $x, y \in V$, play a single match; the outcome of the games are either win or loss. (x, y) is in the *tournament* if and only if x beats y.

T has the property S_k if for every set of k players there is one who beats them all.

Show that if $\binom{n}{k}(1-2^{-k})^{n-k} < 1$, then there is a tournament on n vertices that has the property S_k. [12]

[**Hints:** Can you use probabilistic methods? You can consider a random tournament on V by choosing either edge (i,j) or (j,i), where each of these two choices is equally likely. Consider a fixed subset K of V, $|K| = k$ and let $\mathcal{E}_K$ denote the event that there is no vertex that beats all the members of K. What is $\Pr(\mathcal{E}_K)$?]

(QB4) (i) A sequence X_n of random variables is said to converge to a number c in the mean square, if $\lim_{n\to\infty} E[(X_n - c)^2] = 0$. Show that convergence in the mean square implies convergence in probability. [6]

(ii) Let $N, X_1, X_2, \ldots$ be independent random variables, where X_i's are Bernoulli with parameter p, and N takes nonnegative integral values. Let $Y = X_1 + \ldots + X_N$ and $Y = 0$ when $N = 0$. Show that if N is binomial with parameters m and q, then Y is binomial with parameters m and pq. [6]

[6+6=12]

(QB5) Find the expected time taken by a random walk in visiting all the vertices of K_n, a complete graph on n vertices. [12]

(QB6) Jobs are processed, one at a time. The processing times of jobs are independent random variables, uniformly distributed in $[1, 10]$. Find or approximate the probability that the number of jobs processed within 200 time units is at least 75. You can use standard normal distribution table. [12]

(QB7) A *queue* is a line where customers wait for service. Time is divided into steps of equal length. At each time step, exactly one of the following occurs:

- If the queue has fewer than n customers, then with probability p, a new customer joins the queue.
- If the queue is not empty, then with probability q, the head of the line is served and leaves the queue.
- With the remaining probability, the queue is unchanged.

Let X_t be the number of customers in the queue at time t.

(i) Show that X_t qualifies to be a finite-state Markov chain. [2]

(ii) Find the transition matrix corresponding to the above Markov chain. [2]

(iii) Show that this Markov chain is irreducible and aperiodic. [2]

(iv) Find the unique stationary distribution for the above Markov chain. [6]

[2+2+2+6=12]

STANDARD NORMAL DISTRIBUTION: Table Values Represent AREA to the LEFT of the Z score.

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.50000	.50399	.50798	.51197	.51595	.51994	.52392	.52790	.53188	.53586
0.1	.53983	.54380	.54776	.55172	.55567	.55962	.56356	.56749	.57142	.57535
0.2	.57926	.58317	.58706	.59095	.59483	.59871	.60257	.60642	.61026	.61409
0.3	.61791	.62172	.62552	.62930	.63307	.63683	.64058	.64431	.64803	.65173
0.4	.65542	.65910	.66276	.66640	.67003	.67364	.67724	.68082	.68439	.68793
0.5	.69146	.69497	.69847	.70194	.70540	.70884	.71226	.71566	.71904	.72240
0.6	.72575	.72907	.73237	.73565	.73891	.74215	.74537	.74857	.75175	.75490
0.7	.75804	.76115	.76424	.76730	.77035	.77337	.77637	.77935	.78230	.78524
0.8	.78814	.79103	.79389	.79673	.79955	.80234	.80511	.80785	.81057	.81327
0.9	.81594	.81859	.82121	.82381	.82639	.82894	.83147	.83398	.83646	.83891
1.0	.84134	.84375	.84614	.84849	.85083	.85314	.85543	.85769	.85993	.86214
1.1	.86433	.86650	.86864	.87076	.87286	.87493	.87698	.87900	.88100	.88298
1.2	.88493	.88686	.88877	.89065	.89251	.89435	.89617	.89796	.89973	.90147
1.3	.90320	.90490	.90658	.90824	.90988	.91149	.91309	.91466	.91621	.91774
1.4	.91924	.92073	.92220	.92364	.92507	.92647	.92785	.92922	.93056	.93189
1.5	.93319	.93448	.93574	.93699	.93822	.93943	.94062	.94179	.94295	.94408
1.6	.94520	.94630	.94738	.94845	.94950	.95053	.95154	.95254	.95352	.95449
1.7	.95543	.95637	.95728	.95818	.95907	.95994	.96080	.96164	.96246	.96327
1.8	.96407	.96485	.96562	.96638	.96712	.96784	.96856	.96926	.96995	.97062
1.9	.97128	.97193	.97257	.97320	.97381	.97441	.97500	.97558	.97615	.97670
2.0	.97725	.97778	.97831	.97882	.97932	.97982	.98030	.98077	.98124	.98169
2.1	.98214	.98257	.98300	.98341	.98382	.98422	.98461	.98500	.98537	.98574
2.2	.98610	.98645	.98679	.98713	.98745	.98778	.98809	.98840	.98870	.98899
2.3	.98928	.98956	.98983	.99010	.99036	.99061	.99086	.99111	.99134	.99158
2.4	.99180	.99202	.99224	.99245	.99266	.99286	.99305	.99324	.99343	.99361
2.5	.99379	.99396	.99413	.99430	.99446	.99461	.99477	.99492	.99506	.99520
2.6	.99534	.99547	.99560	.99573	.99585	.99598	.99609	.99621	.99632	.99643
2.7	.99653	.99664	.99674	.99683	.99693	.99702	.99711	.99720	.99728	.99736
2.8	.99744	.99752	.99760	.99767	.99774	.99781	.99788	.99795	.99801	.99807
2.9	.99813	.99819	.99825	.99831	.99836	.99841	.99846	.99851	.99856	.99861
3.0	.99865	.99869	.99874	.99878	.99882	.99886	.99889	.99893	.99896	.99900
3.1	.99903	.99906	.99910	.99913	.99916	.99918	.99921	.99924	.99926	.99929
3.2	.99931	.99934	.99936	.99938	.99940	.99942	.99944	.99946	.99948	.99950
3.3	.99952	.99953	.99955	.99957	.99958	.99960	.99961	.99962	.99964	.99965
3.4	.99966	.99968	.99969	.99970	.99971	.99972	.99973	.99974	.99975	.99976
3.5	.99977	.99978	.99978	.99979	.99980	.99981	.99981	.99982	.99983	.99983
3.6	.99984	.99985	.99985	.99986	.99986	.99987	.99987	.99988	.99988	.99989
3.7	.99989	.99990	.99990	.99990	.99991	.99991	.99992	.99992	.99992	.99992
3.8	.99993	.99993	.99993	.99994	.99994	.99994	.99994	.99995	.99995	.99995
3.9	.99995	.99995	.99996	.99996	.99996	.99996	.99996	.99996	.99997	.99997

INDIAN STATISTICAL INSTITUTE

BACKPAPER EXAMINATION
M.TECH(CS) I YEAR

DATA AND FILE STRUCTURES

Date: 08.02.2019 Maximum marks: 100 Duration: 3 hours.

The paper contains 110 marks. Answer as much as you can.

1. Give the tightest possible upper bound for the worst case running time for each of the following operations in terms of n. Give a very short explanation for your answer.

   (a) Pushing a value onto a stack containing n values, implemented as a linked list.

   (b) Printing all values in the leaves of a balanced binary search tree (containing n nodes) in ascending order.

   (c) Finding the maximum value in a binary min heap of size n.

   (d) Finding and deleting the minimum value in a binary search tree of size n.

   (e) Finding and deleting the minimum value in a balanced binary search tree of size n.

   (f) Performing ExtractMin in a binomial min heap.

   $[6 \times 3 = 18]$

2. Consider that you have been given a stack interface with the following operations:

newStack():	Returns a new empty stack
isEmpty(s):	Returns 1 if the stack s is empty, 0 otherwise
push(s, a):	Pushes the element a in the stack s
pop(s):	Returns the top element of the stack s and deletes it from the stack
top(s):	Returns the top element of the stack s

   (a) Using the above interface, design a procedure moveMintoTop(s), which given a stack s finds the minimum element in s, removes it from its current position and puts it in the top of s keeping the remainder of the stack in its original order. In your design you are not allowed to use any other non-trivial operations/data structures except what is provided by the stack interface described above, you can use multiple instances of stacks as returned by newStack() if required.

   (b) Assume that in the specific implementation of the stack interface that you are provided, each of the operations above takes $O(1)$ time. If s contains n elements what would be the asymptotic complexity of the procedure moveMintoTop(s) that you designed in part (a) of this question.

   $[8 + 4 = 12]$

3. In an initially empty binary search tree the following elements are inserted in the given order: 3, 1, 10, 12, 4, 20, 14, 19, 22, 18, 2, 8

   (a) Draw the resulting tree.

   (b) Is the tree balanced? If the tree is not balanced, which nodes are not balanced, mark them in your answer to part (a).

   (c) In the tree of part (a) delete the nodes containing 12, 3, 10 in the same order and draw the resulting tree.

   [4 + 4 + 4 = 12]

4. Suppose there is a binary min-heap with exactly 4 nodes, containing items with keys 3, 9, 11, and 15.

   (a) Show every possible binary min-heap that could match this description. For each, draw the appropriate tree and the array representation.

   (b) For one of your answers to part (a), show what happens with three deleteMin operations. Clearly indicate which heap you are starting with and show the heap after each deleteMin. You can just draw the tree (not the array) after each step.

   [6 + 6 = 12]

5. Consider a binary search tree implementation where each node has a key, a parent pointer and left and right child pointers. Given a node X in the tree, write a routine which does single right rotation at X. Adjust all pointers (including the parent pointers) appropriately.

   [10]

6. (a) Find the insertion order of the keys 19, 5, 1, 18, 3, 8, 24 that leads to a 2-3 tree of height 1.

   (b) Draw the corresponding left leaning red black tree (LLRBT) of the final tree obtained. Mark the red links appropriately. What is the height of the resulting LLRBT?

   (c) Draw the LLRBT tree obtained after inserting 2, 17 (in this order) to the previous tree.

   [4 + 4 + 4 = 12]

7. Consider an array $A[0, 1, \ldots, k-1]$ containing only zeros and ones. Thus A represents the number $x = \sum_{i=0}^{k-1} A[i]2^i$. To add 1 modulo 2^k to x we use the following procedure:

```
INCREMENT(A, k)
1. i ← 0;
2. while i < k AND A[i] = 1
3.     A[i] ← 0;
4.     i ← i + 1;
5. end while
6. if i < k,
7.     A[i] ← 1
8. end if
```

Given a number x, define the potential $\Phi(x)$ of x as the number of ones in the binary representation of x. Use a potential function argument to show that the amortized cost of an increment as described above is $O(1)$.

[10]

8. Suppose we have a skip list with n entries in which, rather than promoting each node to the next higher level with probability $1/2$, we promote each node with probability p, for $0 < p < 1$.

   (a) Given a skip list with n entries, show that the expected number of list entries of height $\log_{\frac{1}{p}} n$ is $O(1)$. Briefly explain.

   (b) Show that (excluding the sentinel nodes) the total number of nodes in such a skip list is expected to be at most $\frac{n}{1-p}$.

[$6 + 6 = 12$]

9. (a) Define a binomial tree.

   (b) Draw a binomial tree with 16 nodes.

   (c) Prove that for a binomial tree with n nodes, the number of nodes at level i is $\binom{n}{i}$.

[$4 + 4 + 4 = 12$]

INDIAN STATISTICAL INSTITUTE

Back Paper Examination

M. Tech (CS) - I Year, 2018-2019 (Semester - I)

Probability and Stochastic Processes

Date: 08.02.2019 Maximum Marks: 100 Duration: 3.0 Hours

Note: This is a 3-page question paper. Apart from these 3 pages, there is a page for standard normal distribution table.

The question paper is of 100 marks. Answer as much as you can, but the maximum you can score in Group-A is 40, and in Group-B is 60.

$E[X]$ and var$[X]$ denote the expectation and variance of the random variable X, respectively.

Group A

(QA1) (i) For any random variables X and Y, show that

$$E[X] = \sum_y \Pr(Y = y)\, E[X \mid Y = y]$$

where the sum is over all values in the range of Y and all of the expectations exist.

(ii) Show that for a geometric random variable X with parameter p and for $n > 0$,

$$\Pr(X = n + k \mid X > k) = \Pr(X = n).$$

[5+5=10]

(QA2) Let $C_1, \ldots, C_n$ be disjoint events that form a partition of the sample space Ω. Let B be an event such that $\Pr(B \cap C_i) > 0, \forall i$, and also let A be an event. Show that

$$\Pr(A \mid B) = \sum_{i=1}^{n} \Pr(C_i \mid B)\, \Pr(A \mid B \cap C_i).$$

[10]

(QA3) (i) Define the CDF for a random variable X.

(ii) If X is discrete, then comment on the nature of the function related to the CDF of X.

(iii) If X is continuous, state how can the PDF and CDF be obtained from each other.

(iv) Let $X_1, \ldots, X_n$ be n continuous random variables. If the events $\{X_1 \leq x\}, \{X_2 \leq x\}, \ldots, \{X_n \leq x\}$ are independent for every x, then state how you can compute the CDF and PDF of $X = \max\{X_1, \ldots, X_n\}$.

[1+2+2+5=10]

(QA4) (i) State and prove Chebyshev's inequality.

(ii) State and prove the weak law of large numbers.

[(1+4)+(1+4)=10]

(QA5) (i) Show that it is possible to color the edges of a complete graph K_n on n vertices, with two colors so that it has no monochromatic K_k subgraph if $\binom{n}{k}2^{-\binom{k}{2}+1} < 1$.

(ii) Show that given an undirected graph G with m edges, there is a partition of V into two disjoint sets A and B such that at least $m/2$ edges connect a vertex in A to a vertex in B.

[6+4=10]

(QA6) (i) When is a Markov chain said to be *irreducible*?

(ii) Prove that a Markov chain is irreducible if all states belong to one communicating class.

(iii) Define *recurrent* and *transient* states in a Markov chain.

(iv) Define *positive recurrent* and *null recurrent* states of a Markov chain.

(v) Consider a Markov chain whose states are the positive integers. The transition probabilities are as follows:

- the probability of going from state i to state $i+1$ is $\frac{i}{i+1}$.
- the probability of going from state i to state 1 is $\frac{1}{i+1}$.

Show that the state 1 is null recurrent.

[1+3+1+1+4=10]

Group B

(QB1) (i) Suppose a candidate is participating in a game show. The candidate is given the choice of three doors (A, B and C) – behind one door is a car; behind the other two, goats. The candidate picks a door at random, say A, but the chosen door is not opened immediately. The host of the game show, who knows what is behind the doors, opens another door, say C, which shows a goat. The host then says to the candidate, "Do you want to pick door B?" Is it to the advantage of the candidate to switch his/her choice? Give proper arguments in favour of your answer.

(ii) Each member of a group of n players rolls a die. For any pair of players who throw the same number, the group scores 1 point. Find the mean and variance of the total score of the group.

[6+4=10]

(QB2) Bulbs are defective with probability $p = 0.001$, independently of one another.

(i) Show that the probability of a batch of n bulbs having at most one defective bulb is approximately, $e^{-np}(1 + np)$. Mention clearly the type of random variable you are using.

(ii) The manufacturer of the bulbs want to have at least 98% of the batch to contain at most one defective bulb. What is the largest allowable batch size?

[3+7=10]

(QB3) Let A and B be events with $\Pr(A) > 0$ and $\Pr(B) > 0$. We say that an event B *suggests* an event A if $\Pr(A \mid B) > \Pr(A)$, and *does not suggest* event A if $\Pr(A \mid B) < \Pr(A)$.

(i) Show that B suggests A if and only if A suggests B.

(ii) Assume that $\Pr(\overline{B}) > 0$. Show that B suggests A if and only if $\overline{B}$ does not suggest A.

[5+5=10]

(QB4) Show that the expected distance between two independently chosen points in the interval $0 < x < 1$, following uniform probability density, is $\frac{1}{3}$. [10]

(QB5) The success probability of a random experiment is p. Let X_n be the number of times the experiment needs to be performed to obtain a run of n consecutive successes. Show that $E[X_n] = \sum_{k=1}^{n} p^{-k}$. [10]

(QB6) Let X and Y have joint density function $f(x, y) = 2e^{-x-y}$, $0 < x < y < \infty$.

(i) Are they independent? [4]

(ii) Find their marginal density functions. [4]

(iii) Find the covariance of X and Y. [2]

[4+4+2=10]

(QB7) Forty eight numbers are chosen independently, according to the uniform distribution, on the interval $(-0.5, 0.5)$. Let S be the sum of the numbers.

(i) Find $E[S]$ and $\text{var}[S]$.

(ii) Find the value of $\Pr(|S| > 3)$. You may use the standard normal table.

[(2+2)+6=10]

STANDARD NORMAL DISTRIBUTION: Table Values Represent AREA to the LEFT of the Z score.

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.50000	.50399	.50798	.51197	.51595	.51994	.52392	.52790	.53188	.53586
0.1	.53983	.54380	.54776	.55172	.55567	.55962	.56356	.56749	.57142	.57535
0.2	.57926	.58317	.58706	.59095	.59483	.59871	.60257	.60642	.61026	.61409
0.3	.61791	.62172	.62552	.62930	.63307	.63683	.64058	.64431	.64803	.65173
0.4	.65542	.65910	.66276	.66640	.67003	.67364	.67724	.68082	.68439	.68793
0.5	.69146	.69497	.69847	.70194	.70540	.70884	.71226	.71566	.71904	.72240
0.6	.72575	.72907	.73237	.73565	.73891	.74215	.74537	.74857	.75175	.75490
0.7	.75804	.76115	.76424	.76730	.77035	.77337	.77637	.77935	.78230	.78524
0.8	.78814	.79103	.79389	.79673	.79955	.80234	.80511	.80785	.81057	.81327
0.9	.81594	.81859	.82121	.82381	.82639	.82894	.83147	.83398	.83646	.83891
1.0	.84134	.84375	.84614	.84849	.85083	.85314	.85543	.85769	.85993	.86214
1.1	.86433	.86650	.86864	.87076	.87286	.87493	.87698	.87900	.88100	.88298
1.2	.88493	.88686	.88877	.89065	.89251	.89435	.89617	.89796	.89973	.90147
1.3	.90320	.90490	.90658	.90824	.90988	.91149	.91309	.91466	.91621	.91774
1.4	.91924	.92073	.92220	.92364	.92507	.92647	.92785	.92922	.93056	.93189
1.5	.93319	.93448	.93574	.93699	.93822	.93943	.94062	.94179	.94295	.94408
1.6	.94520	.94630	.94738	.94845	.94950	.95053	.95154	.95254	.95352	.95449
1.7	.95543	.95637	.95728	.95818	.95907	.95994	.96080	.96164	.96246	.96327
1.8	.96407	.96485	.96562	.96638	.96712	.96784	.96856	.96926	.96995	.97062
1.9	.97128	.97193	.97257	.97320	.97381	.97441	.97500	.97558	.97615	.97670
2.0	.97725	.97778	.97831	.97882	.97932	.97982	.98030	.98077	.98124	.98169
2.1	.98214	.98257	.98300	.98341	.98382	.98422	.98461	.98500	.98537	.98574
2.2	.98610	.98645	.98679	.98713	.98745	.98778	.98809	.98840	.98870	.98899
2.3	.98928	.98956	.98983	.99010	.99036	.99061	.99086	.99111	.99134	.99158
2.4	.99180	.99202	.99224	.99245	.99266	.99286	.99305	.99324	.99343	.99361
2.5	.99379	.99396	.99413	.99430	.99446	.99461	.99477	.99492	.99506	.99520
2.6	.99534	.99547	.99560	.99573	.99585	.99598	.99609	.99621	.99632	.99643
2.7	.99653	.99664	.99674	.99683	.99693	.99702	.99711	.99720	.99728	.99736
2.8	.99744	.99752	.99760	.99767	.99774	.99781	.99788	.99795	.99801	.99807
2.9	.99813	.99819	.99825	.99831	.99836	.99841	.99846	.99851	.99856	.99861
3.0	.99865	.99869	.99874	.99878	.99882	.99886	.99889	.99893	.99896	.99900
3.1	.99903	.99906	.99910	.99913	.99916	.99918	.99921	.99924	.99926	.99929
3.2	.99931	.99934	.99936	.99938	.99940	.99942	.99944	.99946	.99948	.99950
3.3	.99952	.99953	.99955	.99957	.99958	.99960	.99961	.99962	.99964	.99965
3.4	.99966	.99968	.99969	.99970	.99971	.99972	.99973	.99974	.99975	.99976
3.5	.99977	.99978	.99978	.99979	.99980	.99981	.99981	.99982	.99983	.99983
3.6	.99984	.99985	.99985	.99986	.99986	.99987	.99987	.99988	.99988	.99989
3.7	.99989	.99990	.99990	.99990	.99991	.99991	.99992	.99992	.99992	.99992
3.8	.99993	.99993	.99993	.99994	.99994	.99994	.99994	.99995	.99995	.99995
3.9	.99995	.99995	.99996	.99996	.99996	.99996	.99996	.99996	.99997	.99997

INDIAN STATISTICAL INSTITUTE

Mid-Semester Examination, Second Semester 2018-2019

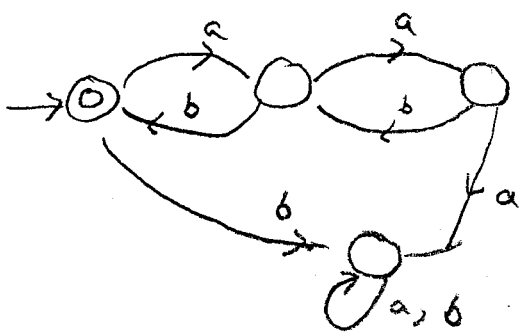
Automata, Languages and Computation, M.Tech –I year

Date:18 February, 2019 Maximum Marks: 30

Duration: 2 hours

1. Construct finite automaton that accept all binary strings containing odd numbers of 0 and even number of 1. 4

2. Describe informally the language accepted by the following finite automaton. 4



3. Draw state diagrams for finite automata of regular expressions given below following the steps of algorithm that convert regular expression to finite automaton. 4+4

   (ab)(ba)*/(aa*)*

   *((ab/aab)*a*)**

4. Draw state diagram of a non-deterministic finite automata accepting strings of *(ab/aba/aab)** and generate deterministic finite automata using the conversion algorithm. *6*

5. Let L be a subset of Σ^*. Suppose L is accepted by some finite automaton. Write which of the following languages are regular. Justify you answer.

   L^R={ω: reversed string of ω is in L}

   Max(L) ={ω: string of ω is in L, and for any string x of positive length ωx is not in L}

   P(L)={ω: $\omega = \omega^R$} 6+6+6

Indian Statistical Institute

M.Tech. (CS), First Year, Mid-Sem of Second Semester Examination, 2018-19

Database Management Systems

Full Marks: 50 Date: 19-02-2019 Time: 2 Hours

Answer any *five* of the following questions

$5 \times 10 = 50$

1. Considering that R_1, R_2 and R_3 are relations in relational algebra and the other notations have their usual meanings, prove the following things.
   (a) $R_1 \bowtie R_2 \equiv R_1 \bowtie R_2 \cap R_1 \bowtie R_2$.
   (b) $(R_1 \bowtie R_2) \bowtie R_3 \equiv R_1 \bowtie (R_2 \bowtie R_3)$.
   (c) The set $\{\sigma, \pi, \cup, -, \times\}$ is complete.

   3+3+4

2. Let there be two different relations R_1 and R_2 having the same number of attributes but $t1$ and $t2$ tuples, respectively. Find the minimum and maximum possible number of tuples that may occur in the resulting relations provided by the following operations.
   (a) A projection operation on R_1.
   (b) A selection operation on R_2.
   (c) $R_1 - R_2$.
   (d) $R_1 \times R_2$.
   (e) $R_1 \bowtie_\theta R_2$.

   2+2+2+2+2

3. (a) Express the natural join operation in relational algebra in terms of Cartesian product and other operations. Justify your answer.
   (b) Consider a schema with the following two relations.

   REL1 = < *attribute*1: integer, *attribute*2: integer>
   REL2 = <*attribute*2: integer, *attribute*3: integer>

   Without making any assumptions on the keys, justify which of the following two relational algebra expressions are equivalent to each other.
   (i) $\pi_{attribute1,\ attribute3}$ ($\pi_{attribute1}$ REL1 $\times$ $\sigma_{attribute2 > 0}$ REL2)
   (ii) $\pi_{attribute1,\ attribute3}$ (REL1 $\bowtie$ $\sigma_{attribute2 > 0}$ REL2)
   (iii) $\pi_{attribute1}$ ($\sigma_{attribute2 > 0}$ REL1) $\times$ $\pi_{attribute3}$ ($\sigma_{attribute2 > 0}$ REL2)
   [Hint: You can justify through examples too.]

   (2+2)+6

4. Consider the following schema representing details about the online customers who place orders in different outlets of Domino's Pizza.

   CUSTOMER = <*name*: string, *age*: integer, *gender*: string, road-*address*: string>

ORDERS = <*name*: string, *pizza*: string>
DELIVERS = <*outlet*: string, *pizza*: string, *cost*: real>

The keys corresponding to each relation are underlined. Write the following queries in SQL.

(a) Find the names of all teenager customers whose name consist of at least 5 characters and start with "A". Assume that a teenager is less than 18 years of age.

(b) Find the names of all males who live at "203 B. T. Road" and orders both "Margherita" and "Chicken Tikka" pizza.

(c) Find the names of all customers who have never placed an order. Note that, this might be a *null* value.

(d) Find all the outlets that delivers at least one pizza that can be ordered by Malay in less than ₹300.

(e) Find the outlets delivering the cheapest "Capsicum" pizza. In case of ties, return all the cheapest "Capsicum" pizza delivering outlets.

[Hint: The standard form of an SQL query is as follows: select <ATTRIBUTE NAMES SEPARATED BY COMMA> from <TABLE NAME> where <CONDITIONS ON ATTRIBUTES>]

2+2+2+2+2

5. (a) State the Armstrong's axioms related to functional dependency. What do you mean by non-trivial functional dependency?
(b) Given the decomposition of a relational schema R = <VWXYZ>, which is not in BCNF, into <VWX> and <VYZ>, verify whether the decomposition satisfies BCNF if the set of functional dependencies {XY → Z, W → Y, Z → V, V → WX} hold. What is the possible foreign key of the decomposed relations?

(3+1)+(4+2)

6. Cite appropriate examples for the following things with proper justifications.
(a) A relation that comprises a derived attribute.
(b) An E-R diagram that contains a generalization of entities.
(c) An E-R diagram that includes total participation of an entity set in a relationship set and the entity participates in at most one relationship.
(d) A Venn diagram encompassing the sets of relations in 1NF, 2NF, 3NF and BCNF.
(e) A relation that is in 2NF but not in 3NF.

2+2+2+2+2

Indian Statistical Institute

Semester-II 2018-2019

M.Tech.(CS) - First Year

Mid-semester Examination (20 February, 2019)

Subject: Operating Systems

Total: 36 marks Maximum marks: 30 Duration: 2.5 hrs.

INSTRUCTIONS

0. **You may carry one A4 sized cheatsheet (written / printed on both sides) with you. Please write your name clearly on the sheet, and get it signed along with your answer booklet by the invigilator.**
1. **For each question, please write your answer in the space provided after that question.**
2. **You may use answer sheets for rough work (only). Please submit the answer sheets along with this question paper.**
3. **Please avoid changing your answer. If you *have* to, please put a line through the old answer and write the new answer clearly. Do NOT overwrite.**
4. **Please keep your answers brief and to the point.**

Name / Roll no.: _____

1. (5 marks)

(a) What is the purpose of the `idtr` register on x86 processors? [2]

(b) Can this register be written in user mode? YES/NO [1]

(c) Justify your answer. [2]

2. (8 marks) Suppose that the following is an excerpt from the definition of the `proc` structure in recent Linux kernels.

```
1  typedef struct task_struct {
2          long        state;
3          pid_t       pid;
4          int         prio;
5          int         static_prio;
6          ...
7          struct task_struct   *parent, *eldest_child, *prev_sibling, *next_sibling;
8          ...
9  } proc;
```

Further suppose that a process P has called `exit()`. Write C code (or pseudo-code for partial credit) corresponding to what the kernel must do to the (i) parent, (ii) siblings and (iii) children of P. You may assume that

- `p` is a pointer to P's `proc` structure, `init` is a pointer to the *init* process' `proc` structure (i.e., `p` and `init` are both of type `struct task_struct *`);
- for a zombie process, the `state` field has the value `EXIT_ZOMBIE`;
- the function `kill(X, SIGCHLD)` can be used to send the `SIGCHLD` signal to a process with pid `X`;
- the function `waitpid(X, NULL, 0)` can be used to release the `proc` structure with pid `X`.

If you need more functions, you may make up your own names, but please clearly state what each such function does (see `kill` and `waitpid` above for examples). [2+2+4 = 8]

3. (9 marks) Suppose $P_1, P_2, \ldots, P_n$ are n processes each having a single CPU burst of duration T_o units, and no I/O bursts. The creation times of $P_1, P_2, \ldots, P_n$ are $0, \varepsilon, 2\varepsilon, \ldots, (n-1)\varepsilon$, where $0 < \varepsilon \ll T_0$ (i.e., in your calculations, you may take $\varepsilon = 0$). Suppose these processes are scheduled using a round-robin scheme, with a time quantum of x units. Let the average turnaround time for the n processes be y units.

(a) Draw the Gantt chart for $x \geq T_0$. [1]

(b) Compute the average turnaround time for $x \geq T_0$. [2]

(c) Draw the Gantt chart for $\varepsilon \ll x < T_0$. [3]

(d) Compute the average turnaround time for $\varepsilon \ll x < T_0$. [3]

4. (8 marks) Consider the following set of conventional processes being managed by version 2.4 of the Linux kernel at some time $t = t_0$.

Process	Becomes ready at	`nice`	`counter`	CPU burst 1 duration	I/O burst 1 duration
P_1	t_0	0	6	4	3
P_2	$t_0 + 2$	-3	12	1	7
P_3	$t_0 + 3$	+2	3	2	5

Draw the Gantt chart for the **current** epoch. [8]

5. (6 marks) Consider an application in which N co-operating processes access a shared buffer. To synchronize their operations on the shared buffer, the i-th process P_i executes the following code:

```
1  shared int token[N] = {0, ..., 0};
2
3  P_i()
4  {
5    while (1) {
6      token[i] = MAX(token[0], token[1], ..., token[N-1]) + 1;
7      for (j = 0; j < N; j++) {
8         while ((token[j] != 0) &&
9                (token[j] < token[i] || (token[j] == token[i] && j < i))) ;
10     }
11     access_shared_buffer();
12     token[i] = 0;
13     other_processing();
14   }
15 }
```

Fill in the table below to construct a scenario that shows that processes running the above code may execute `access_share_buffer()` concurrently (i.e., not in a mutually exclusive manner). [6]

Process	Completes executing (provide line number(s))	Remarks

INDIAN STATISTICAL INSTITUTE

Mid Semestral Examination

M. Tech (CS) - I Year, 2018-2019 (Semester - II)

Design and Analysis of Algorithms

Date: 22.02.2019 Maximum Marks: 100 Duration: 3.0 Hours

Note: This is a 2-page question paper.

The question paper is of 100 marks. Answer as much as you can, but the maximum you can score in Group-A is 40, and in Group-B is 60.

It is imperative that if you design an algorithm, you should compute its space and time complexity, and give its proof of correctness.

Group A

(QA1) Show that, on the assumption that all permutations of a sequence of n elements are equally likely to appear as input, any decision tree that sorts n elements has an expected depth of at least $\log n!$. [10]

(QA2) Let P be a set of n points in $\mathbb{R}^2$. Design and analyze an algorithm to compute the closest pair of points in P in $O(n \log n)$ time. [5+5=10]

(QA3) Let $G = (V, E)$ be a directed graph, $v_0 \in V$ is a source vertex and $w : e \mapsto \mathbb{R}^+$, be a function that maps each edge e of G to a positive real weight.

(i) Describe Dijkstra's shortest path algorithm to find for each vertex $v \in V$, the least weighted path from v_0 to $v \in V$ in G. Deduce the time complexity of the algorithm.

(ii) Give the proof of correctness of the above algorithm.

[(3+2)+5=10]

(QA4) Given a sequence of matrices $A_1, \ldots, A_n$ and dimensions $p_0, \ldots, p_n$ where A_i is of dimension $p_{i-1} \times p_i$, design an algorithm that determines the multiplication order that minimizes the number of operations (i.e., multiplications), along with the minimum number of operations required. Explain your steps properly, and find the space and time complexities. [10]

(QA5) You are given an alphabet set and a set of frequencies for the letters.

(i) Define an optimal prefix code for the above input.

(ii) Design and analyze an algorithm to compute such an optimal prefix code that minimizes the average number of bits per letter.

(iii) Prove the correctness of your algorithm.

[2+4+4=10]

[P.T.O.]

Group B

(QB1) Let $\mathcal{A} = \{a_1, \ldots, a_n\}$ be a sequence of n distinct numbers. We say that two indices $i < j$ form an *inversion* if $a_i > a_j$. Design and analyze an efficient algorithm to find out the number of inversions in $\mathcal{A}$. [7+3=10]

(QB2) Let $\mathcal{A}$ be an array of n integers. Each $a_i \in \mathcal{A}$ lies in the range $[0, n^4 - 1]$. Design an efficient algorithm to sort $\mathcal{A}$. [10]

[Hints: Sorting in $O(n \log n)$ time using any known algorithm will not fetch any credit. Try using any linear time sorting!]

(QB3) Your problem is to sort n distinct elements using a comparison based sort. So, your input can be any one of the $n!$ permutations of the distinct elements. Now, prove or disprove the following statement: *There can exist a comparison based sort whose running time is $O(n)$ for at least a constant fraction of the $n!$ inputs of length n.* [10]

(QB4) In the selection algorithm studied in the class, we worked with 'groups of 5' and deduced that the algorithm runs in $O(n)$ time. Find out what happens, if we work with 'groups of 3' and 'groups of 7'. [5+5=10]

(QB5) (i) $X = \{x_1, x_2, \ldots, x_n\}$ is an array of n distinct integers and x is an integer. Design an efficient algorithm to determine whether there are two elements in X whose sum is exactly x. Find the time complexity of the algorithm you designed.

(ii) $X = \{x_1, x_2, \ldots, x_n\}$ is an array of n distinct integers. Design an efficient algorithm to determine whether there are three elements in X whose sum is exactly 0. Find the time complexity of the algorithm you designed.

[4+6=10]

(QB6) (i) A *binary heap* is an *almost-complete* binary tree with each node satisfying the *heap property:* If v and $\text{par}(v)$ are a node and its parent, respectively, then the key of the item stored in $\text{par}(v)$ is not greater than the key of the item stored in v. A binary heap supports the following operations: (i) deleting the minimum element in $O(\log_2 n)$ time, and (ii) inserting/shifting an element up the heap in $O(\log_2 n)$ time. A d-ary heap is a generalization of a binary heap in which each internal node in the almost-complete d-ary rooted tree has at most d children instead of 2, where $d > 2$ can be arbitrary.

Design and analyze efficient algorithms for the following operations in a d-ary heap. Explicitly mention the time complexity.

- deleting the minimum element
- inserting/shifting an element up the heap

(ii) Let $G = (V, E)$, with $|V| = n$ and $|E| = m$, be a directed graph as in the setting of (QA3). Let $m \geq n^{1+\epsilon}$, for some $\epsilon > 0$ that is not too small, i.e. the graph is dense. Show that one can obtain a time complexity of $O(\frac{m}{\epsilon})$ for Dijkstra's algorithm if a d-ary heap, with a suitable choice of d as a function of m and n, is used.

[(5+5=10)]

(QB7) Let $G = (V, E)$ be a directed acyclic graph with weight $w(u, v)$ on edge $(u, v) \in E$. Design and analyze an efficient algorithm to find the average path length from a source vertex $s \in V$ to a destination vertex $t \in V$. The average path length is defined as the total weight of all paths from s to t divided by the total number of distinct paths. [10]

[Hints: It would be easy to give a dynamic programming solution. The graph G being a directed acyclic graph admits a topological ordering on its set of vertices.]

INDIAN STATISTICAL INSTITUTE

End Semestral Examination

M. Tech (CS) - I Year, 2018-2019 (Semester - II)

Design and Analysis of Algorithms

Date: 18.04.2019 Maximum Marks: 100 Duration: 4.0 Hours

Note: Read the instructions very carefully.

This is a 4-page question paper. The question paper has **6** questions in Group-A and **9** questions in Group-B, each of 12 marks. You can answer at most **4** questions from Group-A and the maximum you can score in Group-A is 40. You can answer at most **6** questions from Group-B and the maximum you can score in Group-B is 60. Read the questions carefully to choose the questions you want to answer.

Mark explicitly in the first page of your answer script the questions that you have attempted. Only those answers will be checked.

If you design an algorithm, compute its space and time complexity, and give its proof of correctness.

Whenever we say that, $\mathcal{P}$ is a linear program (LP), we mean $\mathcal{P}$ is of the form

$$\begin{array}{lllll} \text{Maximize} & \mathbf{c}^T\mathbf{x} & & \\ \text{subject to} & A\mathbf{x} & \leq & \mathbf{b} \\ & \mathbf{x} & \geq & \mathbf{0} \end{array}$$

Group A

(QA1) Let A be a set of n elements. The *Union-Find* data structure allows us to maintain disjoint sets by building a forest-like data structure τ of trees for supporting two set operations – FIND(a) that returns the name of the set containing a, $\forall a \in A$; and UNION(A, B) is an operation that takes two sets A and B and merges them into a single set.

(i) Describe the union-by-rank and path-compression heuristic algorithm for building τ for Union-Find.

(ii) Show that for any node x in τ, rank$(x) <$ rank$(\Gamma(x))$, where $\Gamma(x)$ denotes the parent of x in τ.

(iii) Show that any root node of rank k has at least 2^k nodes in its tree.

(iv) Show that there can be at most $\frac{n}{2^k}$ nodes of rank k.

(v) Show that if there are m FIND operations using the path-compression heuristic, the total number of operations is at most $O(m \log^* n)$, where $\log^* n$ is defined to be the number of successive log operations that need to be applied to n to bring it down to less than or equal to 1.

[4+1+2+1+4=12]

(QA2) Given two vectors $\mathbf{a} = (a_0, a_1, \ldots, a_{n-1})$ and $\mathbf{b} = (b_0, b_1, \ldots, b_{n-1})$, define the convolution of $\mathbf{a}$ and $\mathbf{b}$ as the vector $\mathbf{c} = \sum_{(i,j):i+j=k; i,j<n} a_i b_j$. Design an $O(n \log n)$ time algorithm to compute $\mathbf{c}$, by showing the following steps explicitly: (i) the algorithm; (ii) time complexity and (iii) proof of correctness. [5+2+5=12]

(QA3) Let $\sigma = \langle a_1, \ldots, a_i, \ldots, a_m \rangle$ be a stream of m elements in the streaming model, where each $a_i \in \{1, 2, \ldots, n\}$. Let f_{a_i} denote the number of occurrences of item a_i in the stream σ. Consider the Frequency Estimation problem in the *streaming model* stated as follows. Given an error parameter $0 \leq \epsilon \leq 1$, the task in Frequency Estimation problem is to maintain for each item a_i in σ, an estimate $\widehat{f}_{a_i}$, such that $f_{a_i} - \epsilon \cdot m \leq \widehat{f}_{a_i} \leq f_{a_i}$.

(i) Design an algorithm to solve the Frequency Estimation problem.

(ii) Estimate the (extra) space taken by your streaming algorithm.

(iii) Prove that your algorithm returns an estimate within the stated bounds, for each $a_i \in \sigma$.

[5+2+5=12]

(QA4) Let $G = (V, E)$ be a *flow network* with a *source vertex* $s \in V$ and a *sink vertex* $t \in V$. Each edge $e \in E$ has a positive integral capacity value c_e.

(i) Describe the *maximum flow* problem in G by defining the *capacity* and *conservation* conditions.

(ii) State and prove the *max-flow min-cut* theorem.

(iii) Show how the Ford-Fulkerson algorithm can be used to solve the problem of *bipartite matching* in polynomial time.

[2+5+5=12]

(QA5) (i) Define the Hamiltonian Cycle problem.

(ii) Show that the Hamiltonian Cycle problem is NP-complete.
[Hints: You can try a reduction from 3SAT.] [2+10=12]

(QA6) Answer the following: (i) Define *polynomial reducibilty*; (ii) Define the class NP; (iii) Define the class NP-complete; (iv) Define the class co-NP; (v) Show that the problem of Linear Program belongs to the class NP $\cap$ co-NP; (vi) Suppose Π is an NP-complete problem. Show that Π is polynomially solvable if and only if P=NP. [2 × 6 = 12]

Group B

(QB1) (i) Use SIMPLEX method to solve the following LP:

$$\begin{aligned} \text{Maximize} \quad & x_1 + 2x_2 - x_3 \\ \text{subject to} \quad & 2x_1 + x_2 + x_3 \leq 14 \\ & 4x_1 + 2x_2 + 3x_3 \leq 28 \\ & 2x_1 + 5x_2 + 5x_3 \leq 30 \\ & x_1, x_2, x_3 \geq 0 \end{aligned}$$

(ii) Deduce the conditions on A, b and c so that the primal LP $\mathcal{P}$ and its dual $\mathcal{D}$ are the same LP. [8+4=12]

(QB2) In the *set cover* problem, we have a universe $\mathcal{U} = \{u_1, \ldots, u_n\}$ of n elements. Let $\mathcal{S} = \{S_1, \ldots, S_m\}$ be a set of m sets, where each set $S_i \subseteq \mathcal{U}$. Each set S_i has a weight $w_i \geq 0$. The problem in *set cover* is to find a minimum weight collection of subsets of $\mathcal{S}$ that covers all elements of $\mathcal{U}$.

(a) Write an integer linear program (ILP) for the *set cover problem* using decision variables x_i to indicate whether the set S_i is included in the solution or not.

(b) Relax the above ILP and round the optimal solution of the linear program as follows: given the optimal solution $\mathbf{x}^*$ of the linear program, we include the subset S_i in our solution if and only if $x_i^* \geq \frac{1}{f}$, where f is the maximum number of sets in which any element appears and x_i^* is the i-th component of $\mathbf{x}$.

For this rounding scheme, show that the set generated is a set cover and is an f-factor approximation algorithm. [4+8=12]

(QB3) Consider the *vertex cover* problem, where the input is a graph $G = (V, E)$ and a positive integer k. The problem is to find if there exists a set $V'(\subseteq V)$ of at most k vertices such that all edges in G are incident to vertices in V'. Now, suppose it is known that $k = O(\log_2 n)$. Describe an efficient algorithm to find if G has a vertex cover of size at most k. Provide a proof of correctness and analyze the time complexity. [5+4+3=12]

(QB4) In the k-center problem, the input is a set of points $P = \{p_1, \ldots, p_n\}$ in $\mathbb{R}^2$ and a positive integer $k \leq n$; the distance between two points $p_i, p_j \in P$ is measured using the usual ℓ_2 distance. The output of the problem is a *partition* of P into k clusters $C_1, C_2, \ldots, C_k$ such that the diameter of the clusters is minimized. The diameter of the clusters is defined as

$$\max_j \max_{p_a, p_b \in C_j} d(p_a, p_b)$$

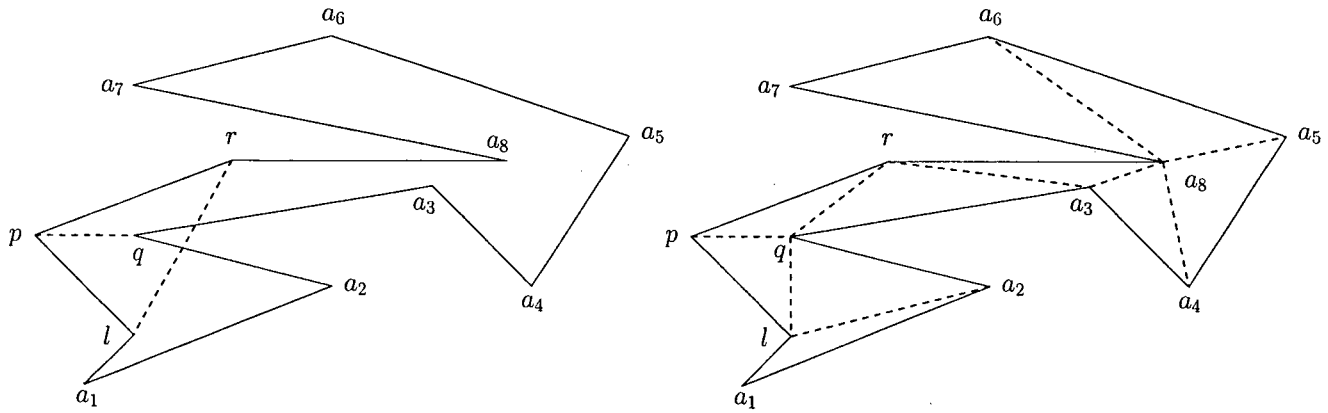
where $d(p_a, p_b)$ denotes the ℓ_2 distance between p_a and p_b.

Show that there is no ρ-approximation algorithm for the k-centre problem for $\rho < 2$ unless NP=P.

[Hints: You can prove this by using a reduction to the *Dominating Set* problem. In the *Dominating Set* problem for an undirected graph $G = (V, E)$, we want to find if there exists a set $D \subseteq V$ of size at most k so that every vertex $v \in V \setminus D$ has a neighbor in D.] [12]

(QB5) Let P be a simple polygon (edges of the polygon intersect only at vertices, no holes inside the polygon) of n vertices; the vertices are given in a clockwise order. Your problem is to design and analyze the time complexity of an efficient algorithm to compute the area of P. You can either design an algorithm of your own, or read the text below to build towards the solution.

The solution consists of *triangulating* P, i.e., breaking P into triangles by drawing non-crossing diagonals. A diagonal of P is a line segment $\overline{pq}$ that stays completely inside P, where p and q are vertices of P. See Figure 1 and read the caption for hints of the solution. The diagonal $\overline{pq}$ divides P into two simple polygons – $pra_8a_7a_6a_5a_4a_3qp$ and pqa_2a_1lp. All the dotted lines in Figure 1(b) are non-crossing diagonals of P.



(a) $\overline{pq}$ is a diagonal, but $\overline{rl}$ is not.

(b) A triangulation of P using diagonals.

Figure 1: A simple polygon P, its diagonals and a triangulation. If there was no segment of the polygon cutting the line segment $\overline{rl}$, then $\overline{rl}$ is a diagonal. How do you identify points like p, r and l? If $\overline{rl}$ is not a diagonal, then you have to identify a point q inside $\triangle plr$ such that $\overline{pq}$ is a diagonal. There can be many points of the polygon P inside $\triangle plr$. From them, how do you identify the point q?

Now, answer the following sub-questions to build towards the solution.

- Design an algorithm to compute a diagonal of P that divides P into two simple polygons.
- If P is divided into two smaller simple polygons, can you recurse to find the triangulation of P, and hence, its area?

- To estimate the time complexity, one needs to find the number of non-crossing diagonals and the number of triangles in P that the diagonals generate. Show, using induction or otherwise, that the number of non-crossing diagonals and the number of triangles in P is $n-3$ and $n-2$, respectively. [12]

(QB6) Consider a 2SAT expression involving n variables and m clauses where each clause is a disjunction of exactly two literals. Show that 2SAT is polynomially solvable.

[**Hints:** Can you reduce it to a path finding problem in a suitably constructed graph?] [12]

(QB7) Your input is a set of n points $P = \{p_1, \ldots, p_n\}$, $p_i \in \mathbb{R}^2$, and a positive integral parameter $k \leq n$. We define a k-clustering of P to be a *partition* of P into k non-empty sets $C_1, \ldots, C_k$, such that the *spacing*, defined to be the minimum distance between any pair of points lying in different clusters/partitions, is maximized. The *spacing* between two clusters C_i and C_j, $i \neq j$, is

$$\min_{p_a \in C_i, p_b \in C_j, i \neq j} d(p_a, p_b)$$

where $d(p_a, p_b)$ denotes the usual ℓ_2-distance between p_a and p_b. Thus the *spacing* of a k-clustering is

$$\min_{1 \leq i < j \leq n} \min_{p_a \in C_i, p_b \in C_j} d(p_a, p_b)$$

Design an algorithm for computing a k-clustering of P. Analyze its time complexity. Provide a proof of correctness.

[**Hints:** Do not allow yourself to be fooled into thinking that it is only a geometric problem. You can also think graphs and consider the fact that minimum distances are to be maximized! You can think in terms of a polynomial solution by constructing a particular graph structure.] [5+2+5=12]

(QB8) Let $T = (V, E)$ be a tree with $|V| = n$; each vertex $v \in V$ has a weight $w_v > 0$. The *maximum weight independent set* problem is to find an *independent set* S in T such that the total weight $\sum_{v \in S} w_v$ is maximized. Design and analyze an efficient algorithm for the above problem. Prove its correctness.

[**Hints:** Can you locate overlapping subproblems and use dynamic programming? If you use dynamic programming, you have to think of two issues — (i) for a vertex $u \in V$, think whether u will/will not be in the optimum solution. If u is in the optimum solution, none of its neighbor can be in the optimum solution; (ii) in which order would you handle the vertices; which vertex do you start from and which vertex you end at, i.e., what is your base condition?] [8+4=12]

(QB9) The problem is about scheduling students to work during the vacation period!

The specific problem is about finding a schedule for n students during m vacation periods where each vacation period consists of a certain number of consecutive days. Let D_j be the set of days included in the vacation period j, $1 \leq j \leq m$. So, $\bigcup_j D_j$ is the union of all these vacation days. Design an algorithm to prepare the work schedule keeping the following constraints under consideration:

- A student i, $1 \leq i \leq n$, has a set of vacation days S_i when he is available for work. S_i's will obviously be spread across D_j's.
- Each student should be assigned to work at most c vacation days in total among the days in which (s)he is available.
- For each vacation period j, each student should be assigned to work at most one of the days in the set D_j.

[**Hints:** This is a feasibility question. Can you formulate a network flow/circulation problem?] [12]

Indian Statistical Institute
I year 2nd semester examination: 2019
Course Name: M. Tech in Computer Science
Subject Name: Computer Networks

Date: 22-04-2019 Maximum Marks: 110 Duration: 3 hours

Instructions:

You **may** attempt **all** questions which carry a total of **110** marks. However, the maximum marks you can score is only **100**.

1. (a) A channel has a data rate of 4 kbps and one way propagation delay of 20 ms. For what range of frame sizes does stop-and-wait give an efficiency of at least 50%? [5]

   (b) Suppose you want to transfer a file from Berkeley to Los Angeles. For this problem, assume the following:

   - The size of file is 116800 bytes. It will be transferred in 1500-byte data packets, of which 40 bytes are taken up with headers. The size of acknowledgement packets is 40 bytes including header. Every packet is acknowledged.
   - The communication is bidirectional and the bandwidth is 12 Mbps (megabits/sec) in each direction.
   - The propagation delay between Berkeley and Los Angeles is 12 ms in each direction.
   - Assume that time to process a packet is very small, so the receiver can send an acknowledgement as soon as it receives a data packet entirely.
   - Likewise, for sliding window the sender can send packets back-to-back (to the degree that the window size permits).
   - Assume no packet loss.

   i. Using sliding window with Send Window Size = Receive Window Size = 8 packets, how long will it take to transfer the file? [5]

   ii. If Send Window Size and Receive Window Size are set to the bandwidth-delay product, then how long does it take to transfer the file? [5]

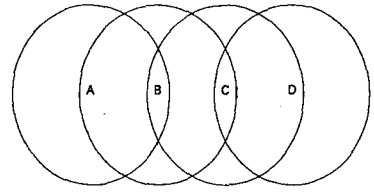
2. (a) A large population of Aloha users manages to generate 50 requests/sec, including both new requests and retransmissions. Time is slotted in units of 40 msec. What is the chance of success on the first attempt? [5]

   (b) In a CDMA/CD network with a data rate of 10 Mbps, the minimum frame size is found to be 512 bits for the correct operation of the collision detection process. What should be the minimum frame size if we increase the data rate to 100 Mbps? [5]

   (c) How long does the station having largest ID has to wait in the worst case before it can start transmitting its data frame over a LAN that uses bitmap protocol. Assume that there are N stations in this multiaccess network and the network is heavily loaded (every station always has data to send). Also assume that each data frame is d bits long. [5]

   (d) Consider the following topology of wireless laptops A, B, C and D. The circles indicate the range of wireless transmissions for each node. For example, B is within range of A, A & C are within range of B, B & D are within range of C and only C is within range of D.

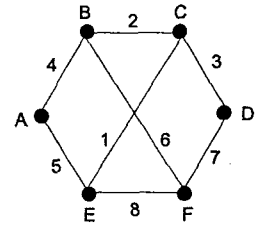
P.T.O.



i. Using the above figure, explain hidden terminal problems. [4]

ii. Assume that each node uses an RTS/CTS based MAC protocol. Explain how RTS/CTS can help solve hidden terminal problem. [4]

3. (a) Consider the subnet shown below where communication delays are shown as labels on the bidirectional links. Distance vector routing is used, and the following vectors have just come into router C: from B: $(5, 0, 8, 12, 6, 2)$; from D: $(16, 12, 6, 0, 9, 10)$; and from E: $(7, 6, 3, 9, 0, 4)$. The measured delays to B, D, and E, are 6, 3, and 5, respectively. What is C's new routing table? Mention both the outgoing link to be used and expected delay. [10]



(b) A leaky bucket is at the host network interface. The data rate on the network is 2MByte/s and the data rate on the link from the host to the bucket is 2.5Mbyte/s. Suppose the host has 250 Mbytes to send onto the network and it sends the data in a burst. What should be the minimum capacity of the bucket (in bytes) in order that no data is lost? [5]

(c) A computer on a 6-Mbps network is regulated by a token bucket. The token bucket is filled with a rate 1 Mbps. The bucket is initially filled to capacity with 1 Mb. How long can the computer transmit at the full 6 Mbps? [5]

(d) What is the purpose of DF, MF and Fragment Offset field in the IPv4 header. [3]

4. (a) Do the IP addresses 128.114.48.62 and 169.232.56.135 belong to the same class of IP address? [4]

(b) A network on the Internet has a subnet mask of 255.255.240.0. What is the maximum number of hosts it can handle? [4]

(c) Indicate whether each of the following subnet masks is valid or invalid? For each invalid one, briefly indicate why it is invalid. [4]

i. 255.255.255.252

ii. 255.255.255.112

(d) Indicate whether each of the following IP addresses is a valid host address? For each invalid one, briefly indicate why it is invalid. [4]

i. 197.15.136.64/26

ii. 110.212.80.127/27

P.T.O.

(e) A large number of consecutive IP address are available starting at 198.16.0.0. Suppose that four organizations, A, B, C and D request 4000, 2000, 4000, and 8000 address, respectively, and in that order. Suppose also that each organization is assigned the lowest address values possible at the time of its request. For each of these, give the first IP address assigned, the last IP address assigned, and the mask in the w.x.y.z/s notation. [10]

5. (a) What is NAT box and how does it work? [7]

(b) Suppose we have a 10-gigabit/second network with a roundtrip time of 100 milliseconds. The Maximum Segment Size (MSS) is 1500 bytes.

i. How long will it take for a TCP connection to reach maximum speed after starting in slow-start mode? [4]

ii. Assume that current window size is 1 MSS. Now consider a scenario where first segment was successfully sent and acknowledged but the second segment was lost. What will be the window size after time out on the acknowledgement of the second segment? After this scenario, how long will it take the TCP connection to achieve maximum speed, assuming no further packet loss? [4]

(c) Describe the method by which the appropriate value of retransmission timeout is determined in TCP. [8]

Indian Statistical Institute

Semester-II 2018-2019

M.Tech.(CS) - First Year

End-semester Examination (26 April, 2019)

Subject: Operating Systems

Duration: 3.5 hrs.

INSTRUCTIONS

1. **You are permitted to use one A4-sized 'cheatsheet' that is written / printed on both sides.**
2. **There are 5 questions in this paper, carrying a total of 58 marks. You may answer as many questions as you like. The maximum that you can score is 50.**
3. **For each question, please write your answer in the space provided after that question. Please keep your answers brief and to the point.**
4. **You may use answer sheets for rough work (only). Please submit the answer sheets along with this question paper.**

1. (a) Recall that the integer value of a binary semaphore can only be 0 or 1. Write pseudo-code to show the functioning of the **blocking** version of the *wait* and *signal* operations on a *binary* semaphore S.

[You do not need to worry about the actual implementation of these operations that ensures their atomicity. It is enough to show how these operations behave.] [2.5 + 2.5 = 5]

```
wait(S) {

}
```

```
signal(S) {

}
```

(b) Given a kernel that is completely unaware of threads, show how you would implement the above functions within a user-level thread library. You may use either C or pseudo-code for your implementation. You may also assume that the following types and functions with self-explanatory names have already been defined within the thread library.

```
typedef struct {
    int thread_id;
    int state; // BLOCKED, READY or RUNNING.
    ...
} THREAD;

THREAD *curr_thread;        // pointer to THREAD structure for currently running thread
THREAD *thread_scheduler(); // selects the next thread to run
```

```
int thread_save(THREAD *t); // saves "context" of THREAD t
int thread_restore(THREAD *t); // restores context of t, and resumes its execution
```

You may use any additional types/functions as needed, but please briefly state the intended use / behaviour of these types/functions in 1-2 lines each. [5 + 5 = 10]

`wait(S)`

`signal(S)`

(c) Very briefly describe what high-level change you would have to make in order to implement these functions within a kernel that is thread-aware (i.e., the kernel is capable of switching between threads). [3]

2. On a certain computer, the memory is byte-addressable, and is organised into words of 4 bytes each. Addresses are 48-bits long. The memory management unit uses paging; the page/frame size is 2^{16} bytes.

(a) Give a reasonable estimate of the size of each page table entry. Justify your answer. [1]

(b) What is the size of the complete first-level (primary) page table for a process? [3]

(c) If page tables are also stored using paging, how many levels of paging will be required? [3]

(d) Suppose the memory management hardware on this computer used segmentation instead of paging. Give a reasonable estimate of the size of each segment table entry. Justify your answer. [2]

3. Recall that a sequence of the page numbers referenced by a process when it runs is called a *reference string*. In a reference string, any sequence of successive references to the same page is generally replaced by only one reference. Suppose that a process is allocated 4 frames of physical memory. Construct reference strings of the form mentioned above for which the page fault rate

   (a) remains the same when the page replacement algorithm is either LRU or FIFO;

   (b) is more for LRU than for the FIFO page replacement algorithm;

   (c) is more for FIFO than for the LRU page replacement algorithm.

$[2 + 3 + 3 = 8]$

4. You are given an SVR2 filesystem with the following parameters:

   - disk block size = 1024 bytes;
   - space required to store an inode on disk = 128 bytes;
   - space required to store a disk block number = 4 bytes.

   A directory on this filesystem (`/x`, say) contains only 700 sub-directories `/x/y1`, `/x/y2`, ..., `/x/y700`, but no regular files.

   (a) Compute the total number of data blocks that the directory `/x` itself will occupy. Do **not** include the data blocks occupied by `/x/y1`, `/x/y2`, etc., in your count. [5]

   (b) What will be the link count in the inode for `/x`? Explain your answer. [2]

   (c) Assume that the directories `/x/y1`, `/x/y2`, ..., `/x/y700`, have inode numbers 101, 102, 103, ..., 800, respectively. Calculate the total number of disk blocks that will have to be read if a user runs the command `ls -l` in the directory `/x`. [6]

5. (a) Define a *safe sequence* for a resource allocation state (RAS). [2]

(b) Consider the following RAS involving 5 processes and 4 resources. $Max[i,j]$ specifies the maximum number of instances that process i may request of resource j. $Alloc[i,j]$ gives the number of instances of resource j currently allocated to process i. $Total[i]$ specifies the total number of instances of resource i that are available in the system.

$$Max = \begin{bmatrix} 2 & 1 & 8 & 6 \\ 2 & 2 & 4 & 4 \\ 5 & 3 & 5 & 3 \\ 0 & 2 & 6 & 6 \\ 4 & 1 & 2 & 1 \end{bmatrix} \quad Alloc = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix} \quad Total = [\ 9\ 3\ 8\ 6\ ]$$

Calculate whether the system is in a safe state. Hence, conclude whether the system uses the Banker's algorithm for resource management.

[6 + 2 = 8]

INDIAN STATISTICAL INSTITUTE

Semestral Examination

M. Tech (CS) - I Year (Semester - II)

Automata, Languages and Computation

Date : 30/4/19 Maximum Marks : 100 Duration : 3:30 Hours

Note : You may answer any part of any question, but maximum you can score is 100.

1. Write a deterministic finite automaton accepting the following language over the alphabet $\{0,1\}$.

   Set of all strings such that second symbol from right end is 1. [6]

2. Let an equivalence relation R^L be defined by xR^Ly if and only if for all $z \in \Sigma^*$, xz is in language L exactly when yz is in L (as defined in Myhill-Nerode theorem). Here, Σ denotes the alphabet set. What are the equivalence classes of R^L for $L = \{0^n1^n | n \geq 1\}$. Use your answer to show that L is not regular. [7+5=12]

3. Construct a push down automaton equivalent to the following grammar G.

   G : Start symbol $S \rightarrow aAA$
   $A \rightarrow aS|bS|a$ [10]

4. Let G be the grammar

   $S \rightarrow aB|bA$
   $A \rightarrow a|aS|bAA$
   $B \rightarrow b|bS|aBB$

   Find (i) leftmost, (ii) rightmost derivation and (iii) parse tree for the string *aaabbabbba*.

   Is the grammar unambiguous? Justify. [(3+3+3)+8=17]

5. Is the language $\{a^ib^jc^k | i < j < k\}$ a context free language? Justify your answer. [10]

6. Consider $\Sigma = \{0,1\}$. Design a Turing machine M that starts with 0^m10^n on its tape for integers $m, n \geq 1$ and ends with $0^{m.n}$ surrounded by blanks on its tape. [10]

7. Show that the collection of decidable languages is closed under the operations (i) union, (ii) intersection and (iii) concatenation (A formal construction is necessary). [5+5+5=15]

8. Prove that some languages are not Turing recognizable. [15]

9. Prove that any context free language is a member of P. [15]

10. Show that NP is closed under closure operation (Kleene closure). [10]

Indian Statistical Institute

M.Tech. (CS), First Year, Final of Second Semester Examination, 2018-19

Database Management Systems

Full Marks: 100 Date: 03-05-2019 Time: 3 Hours

Answer any *ten* of the following questions

$10 \times 10 = 100$

1. For any arbitrary relations R_1, R_2, R_3, R_4, where $R_2 \supseteq R_1$ and $R_4 \supseteq R_3$, prove or disprove the following:
   (a) $R_2 \times R_4 \supseteq R_1 \times R_3$.
   (b) $R_2 - R_4 \supseteq R_1 - R_3$.

5+5

2. (a) Express the division operation in relational algebra in terms of Cartesian product and other operations. Justify your answer.
   (b) Consider a schema with the following three relations, with the primary keys underlined therein, as follows.

   REL1 = <*attribute1*: integer, *attribute2*: string>
   REL2 = <*attribute3*: integer, *attribute4*: string>
   REL3 = <*attribute1*: integer, *attribute3*: integer>

   Decompose the following query into single-variable queries.

   select REL1.attribute2 from REL1, REL2, REL3 where REL1.attribute1 = REL3.attribute1 and REL2.attribute3 = REL3.attribute3 and REL2.attribute4 = "DBMS";

(2+3)+5

3. Consider the following schema of an online code repository system like GitHub.

   CONTRIBUTOR = <*contributor_id*: integer, *contributor_name*: string>
   CODE-GROUP = <*contributor_id*: integer, *code_group*: string, *num_submit*: integer>

   The keys corresponding to each relation are underlined. Write the following queries in SQL.
   (a) Find all the contributors who have made at most one submission within the R code group.
   (b) Find all the contributors who have made at least two submissions within the Python code group.

5+5

4. Consider a relational schema R = <STUVWXYZ>, where all the fields in R contain atomic values. Let F = {UZ → Y, S → TU, T → UXZ, W → S, X → WY} be a set of functional dependencies such that F^+ is exactly the set of functional dependencies that hold for R.

(a) Find out the candidate keys of R.

(b) What is the highest possible normal form that R satisfies? Justify your answer.

5+(1+4)

5. (a) Justify the statement "A *conflict serializable* schedule is always *view serializable* but not the vice versa."

(b) Let there be a relation with data items A and B. Derive the *precedence graph* from the following schedule of transactions working on the said relation. Conclude whether the given schedule is *conflict serializable* or not.

Transaction T_1	**Transaction T_2**	**Transaction T_3**
read(A)		
	read(B)	
	write(A)	
	Commit	
write(A)		
Commit		
		write(A)
		Commit

3+(6+1)

6. (a) Explain the concept of *uncommitted dependency* (or *dirty read*), which might happen while applying a two-phase locking protocol on databases aiming concurrency control.

(b) What is phantom phenomenon? How can it be solved during concurrency control?

4+(2+4)

7. Let there be a relation with data items A and B having initial values 0 and 100, respectively. A *log-based recovery* scheme is used with *immediate database modification* to keep control over the updates in this relation. A pair of transactions T_1 and T_2 works on this relation as follows.

Transaction	**Instruction**
T_1	read(A)
T_1	A ← A + 20
T_1	read(B)
T_1	B ← A
T_1	write(A)
T_1	write(B)
T_2	read(A)
T_2	A ← A * 2
T_2	write(A)

Write down the entries to the log record file and recovery steps to be carried out against each instruction shown above.

10

8. Consider a distributed dynamic environment with 3 nodes and 25 data objects. The nodes and objects are labeled as Ni (i = 100, 200, 300) and Oi (i = 0, 1, ..., 24), respectively. A circular array C (with index 0, 1, ... 49) is used to implement *consistent hashing* for load balancing among the said nodes while processing the data objects. The node Ni is hashed at

$C[i\%80]$ and the object Oi is hashed at $C[2i+1]$. Let there be the following data objects processed on those nodes at a given instance of time.

$$O11, O2, O15, O7, O3, O19, O8, O13$$

Suppose an existing node departs and a new node $N400$ comes in with no effect on the current loads. Then identify the node that departed with appropriate justifications.

10

9. Suppose a company is willing to develop a mobile based application, say GREET. The purpose of this application is to study public reactions during musical reality shows (like Indian Idol). The account holders on GREET can only post audios and can only reply through Emojis (e.g., ☺). Perform a case study on the database systems required to support this. You are expected to suggest the data model, schema, and storage implementation details, with appropriate justifications.

10

10. Give appropriate justifications of the following things:
    (a) Settling transaction logs and temporary spaces properly are crucial for database tuning.
    (b) Thomas' write rule for timestamp-based protocols reduces the amount of transaction rollbacks.

5+5

11. Give appropriate examples of the following things:
    (a) Total and partial participation of entity sets in an entity-relationship model.
    (b) System R style optimization of queries.

5+5

Indian Statistical Institute

Semester-II 2018-2019

M.Tech.(CS) - First Year

Backpaper Examination

Subject: Operating Systems

Maximum marks: 60 Duration: 3 hrs. Date: 24-7-2019

Please keep your answers brief and to the point.

1. What is the difference between the type of administrative information stored by the system in the *u area* and the *proc* structure? Use one example of each kind of information to explain your answer. [4]

2. How is the static priority of a newly created process determined by the Linux 2.6 kernel? For conventional processes: (a) what is the default value? (b) how can the value be changed? (c) what possible values can the static priority take? [1 + 1 + 2 = 4]

3. Consider the following procedure to remove a node p from a doubly-linked list:

```
void remove(NODE *p)
{ NODE *prev, *next;
  prev = p->prev; next = p->next;
  if (next == NULL) tail = prev;
  else next->prev = prev;
  if (prev == NULL) head = next;
  else prev->next = next;
}
```

Suppose two processes P_1 and P_2 share a doubly-linked list.

(a) Using an example, show how the list may get corrupted if the two processes execute `remove` concurrently.

(You should show the initial state of the list, the calls to `remove` by the processes, the intended final state of the list, the interleaving of operations by the two processes, and the actual final state of the list.)

(b) Show how you would re-design `remove` by using semaphores to prevent such a possibility.

[5 + 4 = 9]

4. (a) Define internal and external fragmentation.

(b) For each of the following memory management schemes, briefly explain whether internal and external fragmentation are possible: (i) paging; (ii) segmentation.

(c) What is the disadvantage of having a large page / frame size in a page based memory management system? What is the disadvantage of having a small page / frame size?

(d) Consider an operating system that uses a paging-based memory management scheme. Logical addresses are 24 bits long. Each page table entry occupies 4 bytes. Calculate the minimum and maximum page sizes for which exactly 2 levels of paging will be required.

[NOTE: In a 2-level paging scheme, the second-level page table must fit in a single page.]

(e) Calculate the page fault rate for the following reference string generated by a process, if the proce is allocated five frames of memory and FIFO page replacement strategy is used.

1, 2, 3, 4, 1, 2, 5, 1, 2, 3, 4, 5

[2+8+2+8+7=2

Consider an SVR2 file system with the following parameters:

- disk block size = 2048 bytes
- space required to store an inode on disk = 128 bytes
- space required to store a disk block number = 4 bytes

(a) Assuming that the inode list on disk starts from block number 3, calculate the exact disk locati (block number, byte offset) of inode number 549.

(b) How is *file size* defined for an SVR2 file system?

(c) Using an example, explain how it is possible for a file of size 10,000 to fit in one disk block.

(d) When can the reference count become greater than one for entries in the (i) inode cache and (global file table?

[4+1+3+3=1

(a) Consider the following Resource Allocation State involving 5 processes and 4 resources. $Max[i,$ specifies the maximum number of instances that process i may request of resource j. $Alloc[i,j]$ giv the number of instances of resource j currently allocated to process i. $Avail[i]$ specifies how ma instances of resource i are currently available.

$$Max = \begin{bmatrix} 4 & 2 & 3 & 1 \\ 6 & 2 & 3 & 4 \\ 3 & 2 & 1 & 3 \\ 1 & 2 & 1 & 3 \\ 3 & 0 & 3 & 2 \end{bmatrix} \quad Alloc = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix} \quad Avail = [\ 4\ 4\ 1\ 2\ ]$$

Calculate whether the system is in a safe state.