INDIAN STATISTICAL INSTITUTE

Mid-Semester Examination

M. Tech (CS) - I Year, 2017-2018 (Semester - I)

Course: *Probability and Stochastic Processes*

Instructor: Arijit Ghosh (ACMU)

Teaching assistant: Gopinath Mishra (ACMU)

Date: 04.09.2017

Duration: 3 hours

Note: Maximum one can score from Group A and Group B is 24 and 36 marks respectively. Answer as much as you can.

Group A

Each question in this group carry 6 marks each.

(Q1) Consider an infinite sequence of trials. The probability of success at the i-th trial is p_i. Let I be the event that there is an infinite number of success. Prove that if $\sum_{i=1}^{\infty} p_i = \alpha < \infty$ then $P(I) = 0$.

(Q2) Let X and Y be two discrete random variables. Prove that

$$\mathbb{E}\left[\mathbb{E}\left[X \mid Y\right]\right] = \mathbb{E}[X].$$

(Q3) Let X be a continuous random variable. Then show that

$$\mathbb{E}\left[X\right] = \int_0^{\infty} P(X > x)\, dx \; - \; \int_0^{\infty} P(X < -x)\, dx$$

(Q4) Let $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$ be two independent normal distributions, and let $Z = X + Y$. Show that $Z \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

(Q5) Let $X_1, \ldots, X_n$ be n identically distributed independent continuous random variables with probability density function (pdf) $f(x)$ and cumulative distribution function (cdf) $F(x)$. Let $Y_1 = \min\{X_1, \ldots, X_n\}$ and $Y_n = \max\{X_1, \ldots, X_n\}$. Compute pdf and cdf of Y_1 and Y_n, and the joint pdf and cdf of Y_1 and Y_n in terms of $f(x)$ and $F(x)$.

Group B

Questions Q6, Q7 and Q8 carry 8 marks each, and questions Q9, Q10 and Q11 carry 10 marks each.

(Q6) Let X be a random experiment with outcomes $\{1, \ldots, n\}$ such that for all $i \in [n]$ we have $P(X = i) = p_i$ and $\sum_{i=1}^{n} p_i = 1$. In expectation, how many independtdent trials of the random experiment needs to be done to get the following sequence $123 \ldots n$?

(Q7) If X and Y are independent and identically distributed uniform random variables on $(0, 1)$. Compute the joint density of the random variables $U = X + Y$ and $V = \frac{X}{X+Y}$.

(Q8) There are n rabbits in a forest and m hunters trying to hunt them. Each hunter chooses one rabbit uniformly from the n rabbits and the probability that the i-th hunter can kill its choosen is p. In expectation, how many rabbits will the hunters kill?

(Q9) (a) Imagine a particle performing a random walk on the integer points of the real line, where in each step the particle moves to the next interger on the right with probability p and to the next integer on the left with probability $1 - p$ and the particle starts at the origin. More formally, for all $t \in \mathbb{N} \cup \{0\}$, let X_t be the location of particle at the time instant t, and the random variables evolve according to the following rule:

$$P(X_{t+1} = X_t + 1) = p, \quad P(X_{t+1} = X_t - 1) = 1 - p \text{ and } X_0 = 0.$$

For $r \in \mathbb{N} \cup \{0\}$, compute the following

i. $P(X_n = r)$.

ii. $P(|X_n| \leq r)$.

(This part of Q9 carries 6 marks)

(b) For all $i \in [m]$, let X_i be a discrete random variable taking values uniformly from the set $\{1, \ldots, n\}$ and let $X = \sum_{i=1}^{n} X_i$. Show that if $k_1 + k_2 = (n + 1)m$, then

$$P(X = k_1) = P(X = k_2).$$

(This part of Q9 carries 4 marks)

(Q10) Let $v_1, \ldots, v_n$ denotes n unit vectors in $\mathbb{R}^D$. Show there exists real numbers $a_i \in \{-1, +1\}$, for all $i \in [n]$, such that the following is true

$$\|v\| \leq \sqrt{n}, \quad \text{where } v = \sum_{i=1}^{n} a_i v_i.$$

Also, show that there exists $b_i \in [-1, +1]$ such that

$$\|u\| \geq \sqrt{\frac{n}{3}}, \quad \text{where } u = \sum_{i=1}^{n} b_i v_i.$$

(Q11) A stick of unit length is broken in two places, independently and uniformly along its length. What is the probability that the three pieces will make a triangle?

Indian Statistical Institute

Mid-Semester Examination (2017-2018)

M.Tech. (CS) First Year

Discrete Mathematics

Date: September 5, 2017 Maximum Marks: 60 Time: 2 hours

Answer as much as you can. This question has two pages and carries 85 marks. The maximum you can score is 60 marks. Marks alloted to each question are indicated within square brackets near the right margin.

1. (a) Prove that for all integers $n \geq 8$, Rs. n in postage can be made using only Rs. 3 and Rs. 5 stamps. Mention the proof techique that you have used.

   (b) What is wrong with this famous supposed proof that $1 = 2$?
   "Proof:" We use these steps, where a and b are two equal positive integers.

Step	**Reason**
1. $a = b$	Given
2. $a^2 = ab$	Multiply both sides of (1) by a
3. $a^2b^2 = abb^2$	Subtract b^2 from both sides of(2)
4. $(ab)(a + b) = b(ab)$	Factor both sides of (3)
5. $a + b = b$	Divide both sides of (4) by ab
6. $2b = b$	Replace a by b in (5) because $a = b$ and simplify
7. $2 = 1$	Divide both sides of (6) by b.

[6+4=10]

2. In how many ways can n distinguishable marbles be distributed in k distinguishable boxes such that first box has n_1 marbles, second box has n_2 marbles, and ... k^{th} box has n_k marbles where $\sum_{i=1}^{k} n_i = n$. [5]

3. Prove that if each of the 21 points on an integer grid $\mathcal{N}_7 \times \mathcal{N}_3$ are coloured either red or blue, then there exist 4 points, all having the same colour, lying at the vertices of a rectangle with sides parallel to the coordinate axes. [10]

4. (a) Prove that for $a \geq 2$, $R(a, a) < 4^{a-1}$.

   (b) Find the Ramsey number $R(4, 5)$.

[6+6=12]

5. Consider a double-storeyed building having a staircase with n stairs. In how many ways can a person climb the staircase, if she can climb by 1 or by 2 stairs in each step? Find out a closed form expression in terms of n. [8]

6. Using generating functions, solve the following:

   (a) show that $\sum_{i=0}^{n} F_{2(n-i)} = F_{2n+1} - 1$ and $\sum_{i=0}^{n-1} F_{2(n-i)-1} = F_{2n}$.

   (b) A box contains 9 blue balls, 21 red balls, 7 green balls, 16 yellow balls and 30 white balls. How many balls must we choose to ensure that we have 12 balls of the same colour?

[6+5=11]

7. (a) Count the number of primes less than or equal to 100. Mention the combinatorial technique applied.

   (b) Count the number of distinct ways in which subsets of k elements can be formed from the set $\{1, 2, \ldots, n\}$ such that in each subset there are no consecutive integers.

[4+6=10]

8. Recall the problem of placing k rooks in $n \times n$ board with a certain number of forbidden squares. Show that if a board B' is obtained from board B by deleting rows or columns with no acceptable squares, then $r_k(B) = r_k(B')$. [4]

9. Solve or give tight bounds for the following recurrences:

   (a) $a_n = -3a_{n-1} - 3a_{n-2} - a_{n-3}$, and $a_0 = 1$, $a_1 = -2, a_2 = -1$..

   (b) $T(n) = 7T(n/2) + 15n^2/4$ for n even and $T(1) = 1$.

[4+5=9]

10. (a) Recall that a pair of partitions of an integer n are called conjugates if the transpose of the Ferrers diagram for one is the Ferrers diagram of the other. A partition is said to be self-conjugate if its conjugate is same as itself. Show that $p_{sc}(n)$, the number of self-conjugate partitions of an integer n is equal to $p_{OD}(n)$, the number of partitions of integer n into distinct odd parts.

[6]

INDIAN STATISTICAL INSTITUTE

Mid-Semestral Examination: 2017-18

Course Name: M. Tech. (CS) I Year

Subject Name: Introduction To Programming

Date: 06/09/2017 **Maximum Marks: 60** **Duration: 3hours**

Note: Answer all questions

1. Write a C program to list the word(s) that has maximum number of vowels in it. Ignore those words in the list which contains only vowels.

15

2. The C library function *strtok* searches *s* for tokens delimited by characters from *ct*.

   char **strtok*(char **s*, const char **ct*);

   A sequence of calls of *strtok*(*s*,*ct*) splits a string *s* into tokens, each delimited by a character from *ct*. The first call in a sequence has a non-NULL string *s*, it finds the first token in *s* consisting of characters not in *ct*; it terminates that by overwriting the next character of *s* with '\0' and returns a pointer to the token. Each subsequent call, indicated by a NULL value of *s*, returns the next such token, searching from just past the end of the previous one. strtok returns NULL when no further token is found. The string *ct* may be different on each call.

   What is meant by constant pointer? Explain why *ct* is declared as constant character pointer. Design your own *strtok* (say *myStrtok*) function without using *strtok*.

25

3. Write a command line C program which will act like basic calculator for two real and / or integer numbers (will be able to perform basic +, - * and / on two numbers and print result).

20

INDIAN STATISTICAL INSTITUTE

MIDTERM EXAMINATION
M.TECH(CS) I YEAR

ELEMENTS OF ALGEBRAIC STRUCTURES

Date: 07.09.2016 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. Prove or disprove the following:
   (a) Let R and S be equivalence relations on a set X, then $R \setminus S$ is also an equivalence relation.
   (b) Let S be an arbitrary set closed under an associative operation $\circ$ where both cancellation laws with respect to $\circ$ hold. $(S, \circ)$ is a group.
   (c) Let H, K be subgroups of a group G, then HK is a subgroup of G.
   (d) Let H, K be subgroups of a group G, then $H \cup K$ is a subgroup of G.
   (e) Every subgroup of a cyclic group is cyclic.
   (f) Let S_n A_n be the symmetric group and the alternating group of degree n respectively. A_n is a normal subgroup of S_n
   (g) If Ha and Hb are distinct right cosets of H in G then aH and bH are also distinct left cosets of H in G.

[$5 \times 7 = 35$]

2. (a) Define a normal subgroup.
   (b) Let H, K be normal subgroups of a group G, such that $H \cap K = \{e\}$. Prove that for every $h \in H$ and every $k \in K$, $hk = kh$.
   (c) Let H, K be as above, show that $H \times K$ is isomorphic to HK.

[4+8+8=20]

3. (a) Define the following: (a) center of a group (b) Group homomorphism (c) Kernel of a homomorphism.
   (b) Find a homomorphism $\psi : G \to A(G)$ such that $\text{Ker}_\psi = Z(G)$, the center of G.
   (c) Prove that if G be a group such that $o(G) = p^n$ for a prime p, then G has a non trivial center.

[6+7+7=20]

4. (a) State and prove the first isomorphism theorem for groups.

(b) Let G be a group, for $g \in G$ define $\phi_g : G \to G$ as $\phi_g(x) = gxg^{-1}$ for every $x \in G$. Show that ϕ is an isomorphism of G onto itself.

(c) Let ψ be an isomorphism of a group G onto G. If H is a subgroup of G then $\psi(H) = \{\psi(h) : h \in H\}$ is also a subgroup of G.

(d) Let G be a finite group and K be the only non-trivial subgroup of G. Show that K is normal in G.

[10+4+4+7=25]

5. Let G be a finite group and H a subgroup of G. For A, B subgroups of G, define $A \equiv_H B$ (read as A conjugate to B relative to H), if $B = x^{-1}Ax$ for some $x \in H$.

(a) Prove that $\equiv_H$ is an equivalence relation on all subgroups of G.

(b) Let $[A]$ denote the equivalence class containing A. Prove that $|[A]|$ is equal to the index of $N(A) \cap H$ in H, where $N(A) = \{x \in G : xAx^{-1} = A\}$.

[3+7=10]

INDIAN STATISTICAL INSTITUTE

Mid-Semestral Examination

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

Data and File Structure

Date : Sept. 08, 2017 Maximum Marks : 50 Duration : 2:30 Hours

1 Give brief answer to the following questions. [6*5=30]

(a) We know that we can find i^{th} smallest element in a set in just O(1) time if the set is stored in a sorted array. Suppose the set is stored in two sorted arrays A and B with no element common between them. Size of array A is m and size of array B is n. Give a tight bound on the worst case time complexity of finding i^{th} smallest element of the set?

(b) Find the number of different binary search trees whose inorder traversal outputs $1, 2, 3, \ldots, n$.

(c) We say the point (x_1, y_1) dominates the point (x_2, y_2) if $x_1 > x_2$ and $y_1 > y_2$. A point is called maximal if no other point dominates it. Given a set of n points sorted according to their x-coordinates, give a tight bound on the time complexity of the algorithm to find all maximal points.

(d) A binary tree is defined to be a *2-height balance* tree if the balance factor of each node lies in $\{-2, -1, 0, 1, 2\}$. Let T be a 2-height balance tree and the height of the tree is 9. What is the minimum number of nodes T must have?

(e) In a binary search tree, some organizational defect has happened due to some mishandling. We call a node v defective when addresses of left and right child pointers in node v are swapped. We know that the binary search tree T has some defective nodes and if we swap the addresses of defective nodes, T will turn out to be a valid search tree. Can we identify the number of defective nodes in the defective tree T from the output of the inorder traversal of T? If yes, give a tight bound on time complexity or justify why it is not possible.

(f) Give an example of a tree which is not height balanced but the tree is in WB[$\frac{1}{3}$].

2 Consider a binary heap storing n real numbers (the root stores the greatest number). You are given a positive integer $k < n$ and a real number x. You have to determine whether the k^{th} largest element of the heap is greater than x or not. Write an algorithm that takes $O(k)$ time. You may use $O(k)$ extra storage. [15]

3 There is a doubly linked list storing 0s and 1s, and what you are provided with is just a reference p to a node in this list storing 0. You need to find out the nearest node containing 1. Your algorithm must run in time $O(x)$, where x is the distance to the nearest node containing 1 in the linked list. Note that you know neither x nor the direction (from p) of the nearest node storing 1. You can use only one variable for reference and initially it is assigned p. You may use a constant number of integer variables. [15]

INDIAN STATISTICAL INSTITUTE

First Semester Examination: 2017-18

Course Name: M. Tech. (CS) I Year

Subject Name: Introduction to Programming

Date: 17/11/2017 **Maximum Marks: 100** **Duration: 3 hours**

Answer any four questions (4x25=100).

1. Variable argument lists (stdarg.h) provides facilities for stepping through a list of function arguments of unknown number and type. Why it has provided macros instead of functions to handle variable arguments?

25

2. The C library function strtok searches s for tokens delimited by characters from ct.

```
char *strtok(char *s, const char *ct);
```

A sequence of calls of strtok(s,ct) splits a string s into tokens, each delimited by a character from ct. The first call in a sequence has a non-NULL string s, it finds the first token in s consisting of characters not in ct; it terminates that by overwriting the next character of s with '\0' and returns a pointer to the token. Each subsequent call, indicated by a NULL value of s, returns the next such token, searching from just past the end of the previous one. strtok returns NULL when no further token is found. The string ct may be different on each call.
Design your own *strtok* (say *myStrtok*) function without using *strtok*.

25

3. Explain the implementation of a storage allocator (ANSI C) like malloc, in which allocated blocks may be freed in any order.

25

4. In mathematics, a Hofstadter sequence is a member of a family of related integer sequences defined by non-linear recurrence relations as below:
$F(0) = 1$
$M(0) = 0$
$F(n) = n - M(F(n-1)), n > 0$
$M(n) = n - F(M(n-1)), n > 0$
Write an ANSI C program to implement Hofstadter sequence using recursion.

25

5. Answer the following questions with proper examples:

a. Explain Polymorphism and Encapsulation in OOP.

b. List the advantages of Functional programming. Explain the call by value and call by name evaluation strategy of Functional Programming.

c. Explain with example – symbolic logic and propositional logic

25

INDIAN STATISTICAL INSTITUTE

First Semester Examination: 2017-18

M. Tech (CS) - I Year, 2017-2018 (Semester - I)

Probability and Stochastic Processes

Instructor: Arijit Ghosh (ACMU)

Teaching assistant: Gopinath Mishra (ACMU)

Date: 20.11.2017

Marks: 100

Duration: 4 hours

Note: Maximum one can score from Group A and Group B is 40 and 60 marks respectively. Answer as much as you can.

Notations

- $P(E)$ denotes the probability of the event E.
- $\mathbb{E}(X)$ denotes the expectation of the random variable X
- $\mathbb{V}[X]$ denotes the variance of the random variable X
- $\mathcal{N}(\mu, \sigma^2)$ denotes normal distribution with expectation μ and variance σ^2
- Let $a = (a_1, \ldots, a_m)$ and $b = (b_1, \ldots, b_m)$ be two vectors in $\mathbb{R}^m$. Then $\langle a, b \rangle = \sum_{i=1}^{m} a_i b_i$.

Group A

Each question in this group carries 10 marks.

(Q1) Let X, Y and Z be discrete random variables, and let $g : \mathbb{R} \to \mathbb{R}$. Prove

(a) $\mathbb{E}[Xg(Y)] = \mathbb{E}[\mathbb{E}[X \mid Y] g(Y)]$

(b) $\mathbb{E}[\mathbb{E}[X \mid Y, Z] \mid Z] = \mathbb{E}[X \mid Z]$

(Marks: 5+5)

(Q2) Let the sequences $X_1, X_2, \ldots$ and $Y_1, Y_2, \ldots$ converge in probability to some constants. Show that the following sequences also converge in probability to some fixed constants:

(a) $X_1 + Y_1, X_2 + Y_2, \ldots, X_n + Y_n, \ldots$

(b) $X_1 Y_1, X_2 Y_2, \ldots, X_n Y_n, \ldots$

(Marks: 5+5)

(Q3) Let $X_1, \ldots, X_n, \ldots$ be a sequence of random variables satisfying the following properties:

- For all i, $|X_i| < 1$.
- For all subsequences $i_1 < i_2 < \ldots < i_k$, we have $\mathbb{E}[X_{i_1} \ldots X_{i_k}] = 0$.

Prove that for all $\lambda > 0$ and all n, we have

$$P\left(\left| \sum_{i=1}^{n} X_i \right| \geq \lambda \right) \leq 2 \exp\left(\frac{-\lambda^2}{2n} \right).$$

(Q4) (a) State the Martingale Stopping Theorem.

(b) Let $X_1, X_2, \ldots$ be nonnegative, independent, and identically distributed random variables with distribution X. Let T be the stopping time for this sequence. If T and X have bounded expectation, then show that

$$\mathbb{E}\left[\sum_{i=1}^{T} X_i\right] = \mathbb{E}[T]\,\mathbb{E}[X].$$

(Hint: Use Martingale Stopping Theorem.)

(Marks: 2+8)

(Q5) Let $(\Omega, \mathcal{F}, P)$ be a probability space, and let $A_1, \ldots, A_n$ be subsets of Ω with $A_i \in \mathcal{F}$ for all $i \in \{1, \ldots, n\}$. Using, indicator random variables, prove the following

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{k=1}^{n}\left((-1)^{k-1} \sum_{i_1<\ldots<i_k} P\left(A_{i_1} \cap \ldots \cap A_{i_k}\right)\right).$$

(Q6) (a) State the strong law of large numbers.

(b) Prove the strong law of large numbers for random variables with bounded 4th moments.

(Marks: 2+8)

Group B

Each question in this group carries 10 marks.

(Q7) (a) Prove that communicating relation in a Markov chain defines an equivalence class.

(b) Prove that if one state of a communicating class of a Markov chain is recurrent then all the states in that class is recurrent.

(Marks: 5+5)

(Q8) Let $X_0, X_1, \ldots$ be independent and identically distributed random variables with expectation 0 and variance $\sigma^2 < \infty$. Let

$$Z_n = \left(\sum_{i=1}^{n} X_i\right)^2 - n\sigma^2.$$

Show that $Z_0, Z_1, \ldots$ is a martingale.

(Q9) Suppose we throw m balls independently and uniformly at random into n bins. Let the random variable X denotes the number of empty bins. For all $0 < \delta < 1$, show that

$$P\left(\left|X - n\left(1 - \frac{1}{n}\right)^m\right| \leq \sqrt{2m \log \frac{2}{\delta}}\right) \geq 1 - \delta.$$

(Q10) Let $\vec{X} = (X_1, \ldots, X_n)$ where the random variables $X_1, \ldots, X_n$ are independent, and $X_i \sim \mathcal{N}(0, 1)$ for all $i \in \{1, \ldots, n\}$. Compute the probability density function, expectation and variance of the random variable $Y = \langle \vec{X}, \vec{a} \rangle$ where $\vec{a} = (a_1, \ldots, a_n)$ is fixed vector in $\mathbb{R}^n$.

(Q11) Let X be a nonnegative integer-valued random variable, taking finitely many values, with positive expectations. Show that

$$\frac{\mathbb{E}[Y]^2}{\mathbb{E}[Y^2]} \leq P(Y \neq 0) \leq \mathbb{E}[Y].$$

(Hint: You may use Cauchy-Schwartz inequality for expectations together with an indicator random variable.)

(Q12) Let $S \subseteq 2^{[n]}$ where $[n] = \{1, \ldots, n\}$ satisfying the following properties: (i) $|S| = m$, and (ii) $\forall \sigma \in S$ we have $|\sigma| = k$. Show that for all $0 < p < 1$, there exist a set $A_p \subseteq [n]$ such that, for all $\sigma \in S$, $A_p \cap \sigma \neq \emptyset$ and

$$|A_p| \leq np + m(1-p)^k.$$

If $n \leq mk$, then using the above result show that there exist a set $A \subseteq [n]$ such that, for all $\sigma \in S$, $A \cap \sigma \neq \emptyset$ and $|A| = O\left(\frac{n}{k} \log\left(\frac{mk}{n}\right)\right)$.

(Q13) Let $X_0, X_1, \ldots$ be a sequence for random variables evolving according to the following rule: $X_0 = 0$, and for all $i \geq 0$

$$P(X_{i+1} = X_i + 1) = \frac{1}{2^{X_i}}, \quad \text{and} \quad P(X_{i+1} = X_i) = 1 - \frac{1}{2^{X_i}}.$$

(a) Prove that $\mathbb{E}\left[2^{X_n}\right] = n + 1$.

(b) Using the above result show that $\mathbb{E}[X_n] \leq \log_2(n+1)$.

(Hint: You may use induction and conditional expectation.) (Marks: 7+3)

(Q14) Let $\overrightarrow{X} = (X_1, \ldots, X_n)$, where for all $i \in \{1, \ldots, n\}$ we have

$$P(X_i = 1) = P(X_i = 0) = \frac{1}{2}.$$

(a) Let $a = (a_1, \ldots, a_n)$ be a fixed vector where $a_i \in \{0, 1\}$ for all $i \in \{1, \ldots, n\}$, and $a_i \neq 0$ for all $i \in \{1, \ldots, n\}$. Then show that

$$P\left(\langle \overrightarrow{X}, a\rangle \equiv 0 \bmod 2\right) = \frac{1}{2}.$$

(b) Let $a = (a_1, \ldots, a_n)$ and $b = (b_1, \ldots, b_n)$ be two fixed vectors with $a_i, b_i \in \{0, 1\}$ for all $i \in \{1, \ldots, n\}$. Using the above result, or otherwise, show that if $a \neq b$ then

$$P\left(\langle \overrightarrow{X}, a\rangle \not\equiv \langle \overrightarrow{X}, b\rangle \bmod 2\right) = \frac{1}{2}.$$

(Marks: 6+4)

(Q15) We plan to conduct an opinion poll to find out the percentage of people in a course who want its course instructor removed. Assume that every student answers either "yes" or "no". Let the actual fraction of the students who want the course instructor removed is p and we want to find an estimate X of p as follows. We query N students chosen uniformly at random from the class and output the fraction of them who want the instructor removed. For a given $\epsilon, \delta \in (0, 1)$, how large should N be such that $P(|X - p| \leq \epsilon p) > 1 - \delta$?

(Hint: You may use Chernoff bounds for the sum of indicator random variables.)

INDIAN STATISTICAL INSTITUTE

Semestral Examination

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

Data and File Structure

Date : 23.11.2017 Maximum Marks : 100 Duration :3:00 Hours

1 Give brief answer to the following questions. [6*5=30]

(a) Given an array A of n numbers (not necessarily integers), what is the maximum value of k for which it is possible to find all k smallest numbers of the array in $O(n)$ time using the best algorithm/data-structure you can think of?

(b) What is the worst case number of rotations required in deleting an element from a red-black tree of n nodes? Justify your claim.

(c) We wish to maintain the most efficient data structure for maintaining a set of numbers under the following operations

i. **insert**(x, S) : insert number x into the set S.

ii.**Find-min**(S) : report the value of the smallest number in the set S.

iii. **Find-max**(S) : report the value of the maximum number in the set S.

What is the time complexity of each of these operations using the most efficient data structure you can design?

(d) In a B-tree, search key field is 9 bytes long and block size is 512 bytes and data pointer is 7 bytes. Here block pointer is 7 bytes long. How many maximum tree pointers can you accomodate in the B-tree? If the number of nodes in the B-tree is 200, find the maximum height of the tree.

(e) Suppose the edges of an n-vertex graph comes sequentially. Let the operation $Cycle(e)$ detect if e is part of a cycle in the graph so formed. Design an efficient $O(n)$ space data structure such that it is capable of executing k $Cycle(e)$ operations in $O(k \log k)$ times.

2 Binary search in a sorted array takes logarithmic search time, but the time to insert a new element is linear in the size of the array. We can improve the time for insertion by keeping several sorted arrays. Specifically, suppose that we wish to support Search and Insert on a set of n elements. Let $k = \lfloor \log(n+1) \rfloor$, and let the binary representation of n be $< n_{k-1}, n_{k-2}, ..., n_0 >$. We have k sorted arrays $A_0, A_1, ..., A_{k-1}$, where for $i = 0, 1, ..., k-1$, the length of array A_i is 2^i. Each array is either full or empty, depending on whether $n_i = 1$ or $n_i = 0$, respectively. The total number of elements held in all k arrays is therefore $n = \sum_{i=0}^{k-1} n_i 2^i$. Although each individual array is sorted, there is no particular relationship between elements in different arrays.

(a) Describe how to perform the Search operation for this data structure. Analyze its worst-case running time.

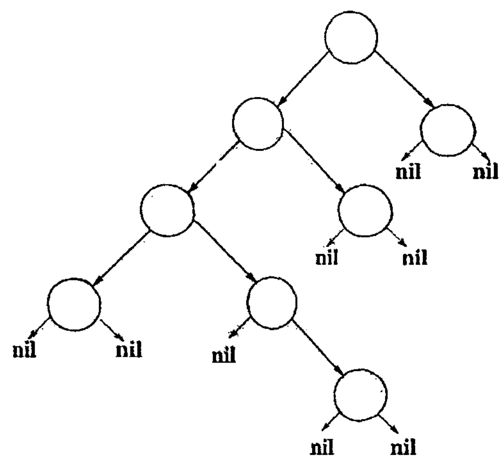
(b) Describe how to insert a new element into this data structure. Analyze its worst-case and amortized running time. (Amortized analysis determines the upper bound $T(n)$ on the total cost of a sequence of n operations, then calculates the amortized cost to be $T(n)/n$.)

(c) Extend the algorithm to incorporate deletion of elements as well. Assume that the number of elements present in the set at any time is at most n. You may add additional fields to the arrays if you wish. You may use more arrays, but the space requirement should never exceed $O(n)$.

[9+9+9=27]

3 Two height balanced binary search trees are given where largest element of one tree is smaller than smallest element of the second tree. Give an efficient algorithm to merge these two trees to generate a single height balanced binary search tree. Write the time complexity of your algorithm. [10]

4 (a) Assign the keys 2, 3, 5, 7, 11, 13, 17, 19 to the nodes of the binary search tree below so that they satisfy the binary-search-tree property.



(b) Explain briefly why this binary search tree cannot be colored to form a legal red-black tree.

(c) The binary search tree in (a) can be transformed into a red-black tree by performing a single rotation. Draw the red-black tree that results, labeling each node with red or black. Include the keys from part (a). [4+4+7=15]

5 Suppose that you are given a collection S of n intervals $[a_i, b_i]$, for $i = 1, 2, \ldots, n$. Present an efficient data structure to store these intrvals for solving the following query. Given a query interval $Q = [s, t]$, report the number of intervals of S that have a nonempty intersection with Q. Your query algorithm should run in $O(\log n)$ time. Explain how your algorithm works, and derive its running time. [15]

6 Let U be any input set. A family of functions $\mathcal{H} = \{h : U \to [m]\}$ is called a universal family if, $\forall x, y \in U$, $x \neq y$: $Pr_{h \in \mathcal{H}}[h(x) = h(y)] \leq 1/m$, where m is nonnegative integer and $[m] = \{0, 1, \ldots, m-1\}$. Let $Z_p = \{0, 1, \ldots, p-1\}$ and $Z_p^* = \{1, 2, \ldots, p-1\}$, where p is a prime number.

We now define the hash function $h_{a,b}(k) = ((ak + b) \mod p) \mod m$, for any $a \in Z_p^*$ and any $b \in Z_p$. The family of all such hash functions is $\mathcal{H}_{p,m} = \{h_{a,b} : a \in Z_p^* \text{ and } b \in Z_p\}$.

Prove that the class $\mathcal{H}_{p,m}$ of hash functions is universal. [15]

INDIAN STATISTICAL INSTITUTE

First Semester Examination: 2017-18

M.TECH(CS) I YEAR

ELEMENTS OF ALGEBRAIC STRUCTURES

Date: 27.11.2017 Maximum marks: 100 Duration: 3.5 hours.

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

1. Prove or disprove the following:
   (a) Every finite cyclic group is of prime order.
   (b) A group of order 21 always have a normal subgroup of order 7.
   (c) Let R be a commutative ring with unit element and $f(x), g(x) \in R[x]$, then $\deg(f(x)g(x)) = \deg(f(x)) + \deg(g(x))$.
   (d) Let R be a Euclidean ring and $a \in R$, such that $d(a) = d(1)$, then a is a unit in R.
   (e) An integral domain with finite characteristic is a field.
   (f) The polynomial $x^4 - 2x^3 + x + 1$ is irreducible in $\mathbb{Q}[x]$.
   (g) If $f(x) \in F[x]$ does not have multiple roots, then $f(x)$ and $f'(x)$ have no common factors of positive degree.

   $[7 \times 5 = 35]$.

2. (a) State the first homomorphism theorem for groups.
   (b) Let G be a group. Prove that the map $\psi : G \to G$ defined as $\psi(g) = g^2$ is a homomorphism if and only if G is abelian.
   (c) Let G be a finite group and $\phi : G \to G$ be an isomorphism such that for every $x \in G$, $x \neq e$, $\phi(x) \neq x$. Prove any $g \in G$ can be represented as $g = x^{-1}\phi(x)$ for some $x \in G$.
   (d) Let G be a finite cyclic group of order n with generator a. Prove that a^k generates G if and only if $\gcd(k, n) = 1$.

   $[4 + 4 + 6 + 6 = 20]$

3. (a) Define the following: (i) Integral Domain (ii) Maximal ideal (iii) Unique factorization domain
   (b) Let R be an integral domain and $a, b \in R$. Prove that
      i. $a|b$ if and only if $(b) \subset (a)$.
      ii. a, b are associates if and only if $(a) = (b)$.

(c) Let I be the subset of $\mathbb{Q}$ such that the numerator of each element in I is divisible by a prime p. Show that I is an ideal of $\mathbb{Q}$ and $\mathbb{Q}/I$ is isomorphic to $\mathbb{Z}_p$.

[6+8+6=20]

4. (a) Define the following: (i) Algebraic Extension (ii) Splitting field

(b) Let F be a field and $f(x) \in F[x]$ be of degree $n \geq 1$. Prove that $f(x)$ can have at most n roots in any extension of F.

(c) Prove that every finite extension of a field is an algebraic extension.

(d) Let K be an extension of F and $a, b \in K$ be algebraic over F with degrees m and n respectively. If $\gcd(m, n) = 1$ then prove that the degree of extension of $F(a, b)$ is mn.

(e) Let p be a prime and n a positive integer, let $q = p^n$. Prove that the splitting field of $f(x) = x^q - x \in \mathbb{F}_p[x]$ is $\mathbb{F}_q$.

(f) Explicitly construct the finite field $\mathbb{F}_4$. Provide the addition and multiplication tables of this field. How many generators does $\mathbb{F}_4^*$ have and what are they?

[4+8+4+6+3+10=35]

Indian Statistical Institute

First Semester Examination (2017-2018)

M.Tech. (CS) First Year

Discrete Mathematics

Date: November 30, 2017 Maximum Marks: 100 Time: 3.5 hours

This question paper has two pages and carries 120 marks.Answer as many questions as you can. The maximum credit you may get is 100 marks.

Marks allotted to each question are indicated within square brackets near the right margin.

1. If a, b, c are odd integers, show that $ax^2 + bx + c = 0$ does not have a rational root. [6]

2. (a) Write a recurrence for each of the following: (i) T_n, the number of distinct simple ordered rooted binary trees, (ii) D_n, the number of derangements of n items.

   (b) Apply the Principle of Inclusion-Exclusion to count the number of integral solutions to the equation $a + b + c + d = 100$, given that $1 \leq a \leq 10$, $b \geq 0$, $c \geq 2$, $20 \leq d \leq 30$.

   [2*5++10=20]

3. How many distinct ~~strings~~ necklaces of beads can be made from 3 black beads and 5 white beads? [10]

4. A wolf, a goat, and a cabbage are on one bank of a river. A ferry man wants to take them across, but his boat is too small to accommodate more than one of them. Evidently, he can neither leave the wolf and the goat, or the cabbage and the goat behind. Using graph theory, prove whether the ferryman can still get all of them across the river.

   (Hint: Each vertex representing a configuration is a 3-tuple for the three entities with binary value as there are two banks of the river.) [10]

5. (a) Show that every cycle in a connected graph G has an even number of edges in common with any cut-set in G.

   (b) Show that every cut-set in a connected graph G must contain at least one edge of every spanning tree of G. [5+5=10]

(P.T.O.)

6. (a) Consider G, the cartesian product of two graphs G_1 and G_2. Prove or disprove that $\overline{G} = \overline{G}_1 \times \overline{G}_2$.

   (b) Draw a tree having one vertex in its center and two vertices in its centroid. [6+4=10]

7. (a) Prove that if a graph $G = (V, E)$ has a Hamiltonian cycle, then for each non-empty subset $S \subseteq V$, the graph $G - S$ has at most $|S|$ components.

   (b) Define a tournament. Prove that every tournament has a spanning path.

   (c) Derive the chromatic polynomial of the graph $K_{1,n}$. [5+(2+10)+3= 20]

8. Write a logic expression for the following statements:

   (i) The sum of two number is even if and only if either both are even or both are odd.

   (ii) There is a number between 3 and 5.

   (iii) Erdos-Szekeres Theorem for subsequence [4+4+6=14]

9. Convert the formula $q \vee r \Rightarrow (\neg r \Rightarrow q)$ to Disjunctive Normal Form (DNF). [5]

10. (a) Use the tree method to show that $\neg p \vee q \Rightarrow r$ is logically equivalent to $(p \wedge \neg q) \vee r$.

    (b) With an example explain that first-order logic is undecidable. [7+8=15]

INDIAN STATISTICAL INSTITUTE

Backpaper for First Semester Examination: 2017-18

M. Tech (CS) - I Year, 2017-2018 (Semester - I)

Probability and Stochastic Processes

Instructor: Arijit Ghosh (ACMU)

Teaching assistant: Gopinath Mishra (ACMU)

Date: - 02.1.2018

Maximum marks: 100

Duration: 4 hours

Note: Maximum one can score is 100. Answer as much as you can.

Notations

- $P(E)$ denotes the probability of the event E.
- $\mathbb{E}(X)$ denotes the expectation of the random variable X
- $\mathbb{V}[X]$ denotes the variance of the random variable X
- $\mathcal{N}(\mu, \sigma^2)$ denotes normal distribution with expectation μ and variance σ^2
- Let $a = (a_1, \ldots, a_m)$ and $b = (b_1, \ldots, b_m)$ be two vectors in $\mathbb{R}^m$. Then $\langle a, b \rangle = \sum_{i=1}^m a_i b_i$.

Questions

Each question carries 10 marks each.

(Q1) What is a probability space?

(Q2) The events A and B are independent.

(a) Show that $\bar{A}$ and B are independent.

(b) Using the above result show that $\bar{A}$ and $\bar{B}$ are independent.

(Marks: 6+4)

(Q3) Let X and Y be two discrete random variables. Then show that $\mathbb{E}[\mathbb{E}[X \mid Y]] = \mathbb{E}[X]$.

(Q4) Let $X \sim \mathcal{N}(0, \sigma_1^2)$ and $Y \sim \mathcal{N}(0, \sigma_2^2)$ be two independent normal distributions, and let $Z = X + Y$. Show that $Z \sim \mathcal{N}(0, \sigma_1^2 + \sigma_2^2)$.

(Q5) For a random variable X with expectation μ and variance $\sigma^2 < \infty$, define the function

$$f(x) := \mathbb{E}\left[(X - x)^2\right].$$

Show that $\mathbb{E}[f(X)] = 2\sigma^2$.

(Q6) (a) What is a Martingale?

(b) Let $X_0,\ X_1,\ \ldots$ be a sequnce of independent random copies of the random variable X, where

$$P(X = -1) = P(X = 1) = \frac{1}{2}.$$

Show that the sequence $Z_0,\ Z_1,\ \ldots$, where $Z_i = \sum_{j=0}^{i} X_j$, is a martingale with respect to the sequence $X_0,\ X_1,\ \ldots$.

(Marks: 5+5)

(Q7) (a) Let X be a nonnegative discrete random variable. For all $\lambda > 0$, show that $P(X \geq \lambda) \leq \frac{\mathbb{E}(X)}{\lambda}$.

(b) Let X be a discrete random variable. Then for all $\lambda > 0$, show that $P\left(\mid X - \mathbb{E}[X] \mid \geq \lambda\right) \leq \frac{\mathbb{V}[X]}{\lambda^2}$.

(Marks: 5+5)

(Q8) Let X be a discrete random variable.

(a) Show that for all $t \in (0,\ \infty)$ and $\lambda \in (0,\ \infty)$, $P\left(\mid X \mid \geq \lambda\right) \leq \mathbb{E}\left[\mid X \mid^t\right] \lambda^{-t}$.

(b) Show that, for all $t > 0$ and $\lambda \in \mathbb{R}$, $P\left(X \geq \lambda\right) \leq \mathbb{E}\left[\exp\left(tX\right)\right] \exp\left(-\lambda t\right)$.

(Marks: 5+5)

(Q9) State and prove weak law of large numbers.

(Q10) (a) What is a Doob martingale construction?

(b) Show that it is a martingale.

(Marks: 5+5)

(Q11) Let $X_1,\ X_2,\ \ldots,\ X_n$ be a sequence of independent copies of a random variable with mean μ and variance $\sigma^2 < 0$. Also, let $Y_i = \sum_{j=1}^{n} X_j$, and N be a random variable taking values in the set $\{1,\ \ldots,\ n\}$ and N is independent of the random variables $X_1,\ X_2,\ \ldots,\ X_n$.

(a) Prove that $\mathbb{E}(Y_N \mid N = n) = n\mu$, and therefore $\mathbb{E}(Y_N) = \mu\mathbb{E}(N)$.

(b) Show that $\mathbb{E}(Y_N^2 | N = n) = n\sigma^2 + n^2\mu^2$, and therefore $\mathbb{V}\left[Y_N\right] = \sigma^2\mathbb{E}[N] + \mu^2\mathbb{V}[N]$.

(Marks: 4+6)

(Q12) Suppose that each box of cereal contains one of n different coupons, and coupon in each cereal box is chosen independently and uniformly at random from the n possibilities. In expectation, how many boxes of cereals must you buy to get at least one coupon of each type?

Indian Statistical Institute

First Semester Examination (2017-2018)

M.Tech. (CS) First Year

Discrete Mathematics

BACK PAPER

Date: 03.01.18 Maximum Marks: 100 Time: 3.5 hours

Answer as many as you can. Marks alloted to each question are indicated within square brackets near the right margin. The maximum you can score is 100.

1. Prove that the expression $3^{3n+3} - 26n - 27$ is a multiple of 169 for all natural numbers n. [5]

2. If a given convex polyhedron has six vertices and twelve edges, prove that every face is a triangle. [5]

3. In how many ways may 8 identical markers be placed on an 8 x 8 square grid (up to rotation of the grid)? [10]

4. A connected planar graph has 24 edges and it divides the plane into 13 regions. How many vertices does it have? [4]

5. (a) Suppose a tree has a vertex of degree k. Show that it has at least k end-vertices (leaves).

   (b) Let G be a graph with 9 vertices, each of degree 5 or 6. Prove that it has at least 5 vertices of degree 6, or at least 6 vertices of degree 5.

   (c) Show that in an Eulerian graph on 6 vertices, a subset of 5 vertices cannot form a complete subgraph. [6+8+6=20]

6. (a) Prove that any outerplanar graph is 3-colorable.

   (b) Derive the chromatic polynomial for a tree having n vertices. [5+5=10]

7. (a) State and prove Konig's theorem for bipartite graphs.

   (b) A graph G is *perfect* if its chromatic number $\chi(G) = \omega(G)$, its clique number. Using the result in (a), prove that the complement of a bipartite graph is perfect. [10+10=20]

(P.T.O.)

8. (a) Define the cycle rank of a graph.

   (b) For a given spanning tree of a graph G, let $D = (e_1, e_2, \ldots, e_k)$ be a fundamental cut-set in which e_1 is a branch (edge of the spanning tree) and $e_2, e_3, \ldots e_k$ are chords (edge joining two non-adjacent vertices of the spanning tree).

   Then show that e_1 is contained in each of the fundamental cycles corresponding to each of the chords $e_i, i = 2, 3, \ldots, k$.

   Moreover, prove that e_1 is not contained in any other fundamental cycles. [4+(6+5)=15]

9. (a) State Vizing's theorem and derive bounds for the vertex chromatic and edge chromatic numbers of a graph.

   (b) Give an example to illustrate the limitations of propositional logic. [(2+2*4)+5=15]

10. (a) Translate the following sentences in propositional logic:

    (i) Unless Ram comes to the party, Anjaneya will not be happy.

    (ii) If Usha jumps and Anju does not make a leap, Sidhu will have to take a gigantic step.

    (b) Write the negation of $(\forall n \in \mathcal{N})(\exists x \in (0, +\infty))(nx < 1)$.

    (c) Derive whether the following formulae are tautologies or not:

    (i) $(p \to q) \to (\bar{q} \to \bar{p})$

    (ii) $(((p \to q) \to p) \to p)$

    (iii) $(p \to (q \to p))$

    (d) Put the following formula in Disjunctive Normal Form: $(p \Leftrightarrow q) \wedge (\bar{r} \to p)$

    [2*3 + 5 + 3*3 + 5 = 25]

11. Prove by tableau method that "If everyone has a smartphone, then someone has a smartphone."

[10]

INDIAN STATISTICAL INSTITUTE

BACK PAPER EXAMINATION
M.TECH(CS) I YEAR

ELEMENTS OF ALGEBRAIC STRUCTURES

Date: 04.01.2018 Maximum marks: 100 Duration: 3 hours.

Each question carries 5 marks.

1. State whether the following statements are true or false do not justify.
   (a) If $ab \equiv ac \bmod n$, then $b \equiv c \bmod n$.
   (b) G is a group and H a subgroup of G, where $o(G) = 25$ and $o(H) = 17$.
   (c) G is a group such that $o(G) = 17$, G is abelian.
   (d) S is a set closed under a associative product, and both cancellation laws hold in S. S is a group.
   (e) G is a cyclic group with 17 elements. The number of generators of G is 15.
2. State Lagrange's Theorem.
3. Give an example of an element of order 3 in a group of order 6.
4. Give an example of two subgroups H and K in S_3 such that HK is not a subgroup.
5. Define normal subgroup.
6. Let H, K be normal subgroups of a group G, such that $H \cap K = \{e\}$. Prove that for every $h \in H$ and every $k \in K$, $hk = kh$.
7. Let $\phi : A \to B$ be a group homomorphism. Show that ϕ is injective if $\mathrm{Ker}_\phi = \{e_A\}$, where e_A is the identity element of A.
8. Let G be a group, for $g \in G$ define $\phi_g : G \to G$ as $\phi_g(x) = gxg^{-1}$ for every $x \in G$. Show that ϕ is an isomorphism of G onto itself.
9. Let G be a finite group and H the only non-trivial subgroup of G. Prove that H is normal in G.
10. Define center of a group.
11. Prove that if G be a group such that $o(G) = p^n$ for $n \geq 2$ and a prime p, then G has a non trivial center.

12. Let D be an integral domain and $F(D) = \{(a,b) : a, b \in D, b \neq 0\}$. Let $\sim$ be a relation on $F(D)$ defined as $(a,b) \sim (c,d)$ if $ad = bc$. Prove that $\sim$ is an equivalence relation.

13. Define an ideal of a ring.

14. Let R, S be rings and $\phi : R \to S$ be a ring homomorphism, prove that the kernel of ϕ is an ideal of R.

15. Define a basis of a vector space.

16. Let V be a vector space and W a subspace of V. Prove that $\dim(W) \leq \dim(V)$.

17. Let V and W be vector spaces and $\phi : V \to W$ an isomorphism. Prove that ϕ maps a basis of V into a basis of W.

18. Let $F \subset K$ be fields and $a \in K$ is algebraic over F. Let $p(x) \in F[x]$ be a non zero polynomial of lowest degree such that $p(a) = 0$. Prove that $p(x)$ is irreducible in $F[x]$.

19. Let F be a field of characteristic $p \neq 0$. Prove that $x^{p^n} - x \in F[x]$ have distinct roots (i.e., have no roots with multiplicity $m > 1$).

20. Explicitly construct a field of 8 elements, write the addition and multiplication tables of this field.

Indian Statistical Institute
Semester-II 2017-2018
M.Tech.(CS) - First Year
Mid-term Examination (26 February, 2018)
Subject: Operating Systems

Total: 35 marks Maximum marks: 30 Duration: 2.5 hrs.

INSTRUCTIONS

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.**

1. Are the following statements true or false? Write 1-2 lines to justify / explain your answer in each case. You will not get any credit if your justification / explanation is incorrect. [1 + 2 + 2 = 5]

   (a) Multiple processes can correspond to the same program / executable file *at the same time*. TRUE / FALSE

   (b) The same process can correspond to multiple programs / executable files (not necessarily at the same time). TRUE / FALSE

   (c) In user mode, the processor status word (PSW) may be modified, but arbitrary write operations on the PSW are not permitted. TRUE / FALSE

2. Explain how many distinct output(s) the following code fragment can generate, if **buffer** is a character array containing the string **"abcdef"**. Assume for this question that **rand()** generates a truly random, uniformly distributed positive integer. [6]

```
...
fork();
for (i = 0; i < 3; i++) {
    fprintf(stderr, "%c", buffer[i]);
    sleep(rand() % 5);
}
```

3. (a) Define *starvation* in the context of process scheduling. [1]

(b) Let n denote the number of ready processes at any point in an operating system. For each of the following scheduling schemes, write

(i) the (asymptotic) time complexity of adding a process to the ready queue;

(ii) the (asymptotic) time complexity of selecting a process during scheduling;

(iii) whether starvation is possible.

Justify each answer in 1-2 lines. [$3 \times 3 = 9$]

FCFS .

(i)

(ii)

(iii)

Priority-based scheduling .

(i)

(ii)

(iii)

Round-robin .

(i)

(ii)

(iii)

4. (a) Consider 2 processes P_1 (`counter` = 4, `nice` = 0) and P_2 (`counter` = 6, `nice` = 1) running under Linux kernel version 2.4.

   (i) If the current epoch has just ended, what time quantum will each process be assigned in the next epoch? Justify your answer. [2]

   (ii) Assuming that the processes are identical in all other respects, which process will be scheduled first in the next epoch, and why? [2]

(b) Explain what happens to a process under the Linux 2.6 kernel when it exhausts its quantum. [3]

5. Consider the bounded-buffer producer-consumer problem. Write the algorithm/pseudo-code for the two processes to show how you would use an atomic TEST-AND-SET instruction to synchronise their operation so that no race conditions occur. Please make sure that all variables are properly initialised. For full credit, your critical section should be as small as possible. [7]

INDIAN STATISTICAL INSTITUTE

Semestral Examination: 2017-18

Course Name: M. Tech. I Year

Subject Name: Database Management Systems

Date: 27/02/2018 **Maximum Marks: 60** **Duration: 2hours**

Note: Answer all questions

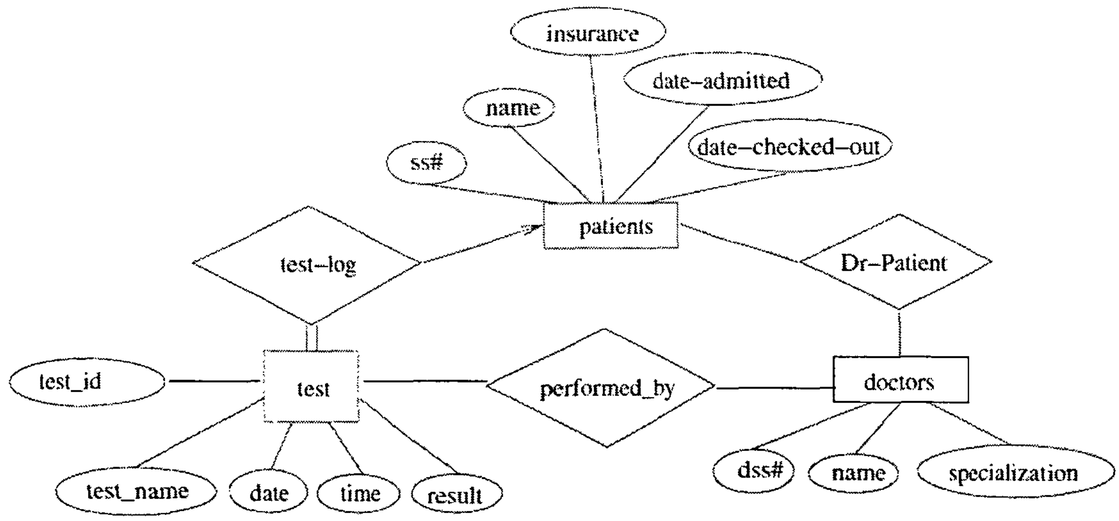
1. The data the structure of a database is described through metadata which is also stored in the database. Explain through example.

20

2. Design a generalization – specialization hierarchy for a motor vehicle sales company. The company sells motorcycles, autos, passenger cars, buses and trucks. Justify your placement of attributes at each level of the hierarchy. Explain why they should not be placed higher or lower levels.

15

3. Consider the following sample ER Diagram for a hospital:



a) Write the problem statement that will lead the above diagram.
b) Map the diagram to appropriate tables.
c) Write appropriate SQL stamen for creating these tables.

10+5+5

4. Consider the following relations:

Suppliers(<u>supplier_no</u>: int, supplier_name: char(50), status: int, city: char(40))
Parts(<u>part_no</u>: int, part_name: char(100), color: char(10), weight: real)
Projetcs(<u>project_no</u>: int, project_name: char(20), city: char(40))
Shipment(<u>supplier_no</u>: int, <u>part_no</u>: int, <u>project_no</u>: int, quantity: int)
Give SQL formulation for the following updates:

a) Find the average price of each part.
b) How many different "red" coloured "nut" are their in the part table.
c) Insert a new supplier "New Sup" from "Kolkata". His supplier_no will be 1 more than the existing maximum supplier_no, but status is yet known.

5

Indian Statistical Institute
Mid-Semester Examination: 2018
Course Name: M. Tech in Computer Science
Subject Name: Computer Networks

Date: 28-02-2018 Maximum Marks: 60 Duration: $2\frac{1}{2}$ hours

Instructions:
You **may** attempt **all** questions which cary a total of **65** marks. However, the maximum marks you can score is only **60**.

1. Consider a bit stuffing framing method where both start and end of a frame is indicated by the flag 01^k0 where 1^k denotes k consecutive ones.
   (a) What should be the bit stuffing rules at the transmitter? [2]
   (b) What should be the bit destuffing rules at the receiver? [2]
   (c) Is it necessary to stuff a 0 in $01^{k-1}0$? [2]
   (d) Assuming all bit patterns are equally likely, compute the expected overhead for a data packet of length L. Your expression for overhead calculation should include the flag length as well as the stuffed bits. [4]

2. (a) For the bit stream 1 0 0 1 1 1 1 1 0 0 0 1 0 0 0 1, sketch the waveforms for each of the following encoding schemes.
      i. NRZI
      ii. Manchester. [3+3=6]
   (b) Consider a communication link with bandwidth $B = 4000$ Hz and $S/N = 30$ dB. Calculate its maximum data rate in bits per second according to Shannon's theorem. Nyquist's theorem cannot be applied here because a factor is unknown. Point out what factor is unknown. [3+1=4]
   (c) Consider an audio signal with frequency components below 4000 Hz. Consider generating a PCM signal out of it using 5-bit PAM samples. If samples are taken according to the sampling theorem, what data rate is achieved? [3]
   (d) Briefly describe the effect of delta on quantizing noise and slope overload noise in delta modulation. [4]

3. (a) State the balance property and run property of a maximum length pseudo noise sequence? [1+2=3]
   (b) Consider an MFSK scheme with carrier frequency f_c equal to 250 kHz, difference frequency f_d equal to 25 KHz, number of different signal elements M equal to 8, and number of bits per signal element L equal to 3.
      i. Make a frequency assignment for each of the eight possible 3-bit data combinations. [4]

P.T.O

ii. Suppose we wish to apply FHSS to this MFSK scheme with $k = 2$; that is, the system will hop among four different carrier frequencies. Let T_c be the period at which the MFSK carrier frequency changes and T_s is the duration of a signal element. Consider a **slow FHSS** with T_c being $2T_s$. Show the sequence of frequencies used, and the times the frequency changes occur, for transmitting the bit string 011110001. Assume that the PN sequence for generating the frequency hops is 0011. [5]

4. (a) Consider a CRC code with the generator polynomial $g(X) = 1+X+X^4$. Determine if the codeword described by the polynomial $c(X) = 1 + X + X^3 + X^7$ is a valid codeword for this generator polynomial. [5]

   (b) The $(4,3)$ odd parity code is a code where 1 odd parity bit is appended to 3 message bits to produce 4-bit codewords. The error detecting capability of a code is defined as the maximum value of t such that all error patterns with t or less erroneous bits can be detected by the code. A code is linear if, and only if, the sum of any two codewords is another codeword.

      i. Is $(4,3)$ odd parity code a linear block code? [3]

      ii. Find the error detecting capability of the $(4,3)$ odd parity code. [3]

   (c) Suppose that a parity check code has a minimum Hamming distance d. Show that if the Hamming distance between a codeword and a given string is less than $d/2$, the Hamming distance between any other codeword and the given string must exceed $d/2$. [5]

5. (a) Suppose two nodes communicate with each other using a stop-and-wait protocol. The data packet size is 10000 bits. The total round-trip time (RTT) between the nodes is equal to 0.2 milliseconds (this includes the time to process the packet, transmit an ACK, process the ACK and transmit the ACK) **plus** the transmission time of the 10000 bit packet over the link. Let l be the bi-directional packet loss probability and R be the data rate of the link. Suppose you have two options to configure your connection with the following properties: 1) if you choose $R = 10$ Megabits/s, then l will be $1/11$, 2) if you choose $R = 20$ Megabits/s, then l will be $1/4$. For both bit rates, the retransmission timeout (RTO) is 2.4 milliseconds.

      i. For each bit rate, calculate the expected time, in milliseconds, to successfully deliver a packet and get an ACK for it. [6]

      ii. Suppose your goal is to select the bit rate that provides the higher throughput for a stream of packets that need to be delivered reliably between the nodes. Which bit rate would you choose to achieve your goal? [2]

   (b) A Go-back-N ARQ uses a window of size 15. How many bits are needed to define the sequence number. [2]

INDIAN STATISTICAL INSTITUTE

Periodical Examination

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

Design and Analysis of Algorithms

Date : March 02, 2018 Maximum Marks : 60 Duration : 3 Hours

Note : You may answer any part of any question, but maximum you can score is 60.

1. Solve the recurrences:

   (i)

$$T(n) = \begin{cases} \sqrt{n} \cdot T(\sqrt{n}) + n & \text{if } n \geq 1; \\ 1 & \text{if } n = 0. \end{cases}$$

   (ii)

$$T(n) = \begin{cases} 1 & \text{if } n = 1 \\ T(n/5) + T(7n/10) + n, & \text{otherwise.} \end{cases}$$

[6+4=10]

2. (a) Tell which one of the following two relations (i) and (ii) (or both) is true, and justify:
   (i) $2^{n+1} = O(2^n)$ and (ii) $2^{2n} = O(2^n)$

   (b) If $\lim_{n\to\infty} \frac{f(n)}{g(n)} = \infty$ then tell which one of the following is true, and justify:
   (i) $f(n) = o(n)$, (ii) $f(n) = \omega(n)$, (iii) $f(n) = \theta(n)$.

[5+5=10]

3. Consider two sums, $X = x_1 + x_2 + \ldots + x_n$ and $Y = y_1 + y_2 + \ldots + y_m$, where both the sequences $\{x_1, x_2, \ldots, x_n\}$ and $\{y_1, y_2, \ldots, y_m\}$ are given in increasing order. Give a linear time algorithm to find indices i and j such that swapping x_i with y_j makes the two sums equal, that is, $X - x_i + y_j = Y - y_j + x_i$, if such a pair (x_i, y_j) exist. Justify the correctness and time complexity of your algorithm. [10+3+2=15]

4. We are given a read-only array S of n points in $\mathbb{R}^2$. Each point $p_i \in S$ is attached with its x and y coordinates. You need to compute the convex hull of the points in S. As S is read only, you are not allowed to write on this array, but you can read the elements of the array S as many times as required.

   Using at most 10 number of extra memory locations, write an efficient algorithm to compute the convex hull of the points in S. Note that, you need not have to store the convex hull vertices in the memory. As soon as you identify a vertex of the convex hull, you print it.

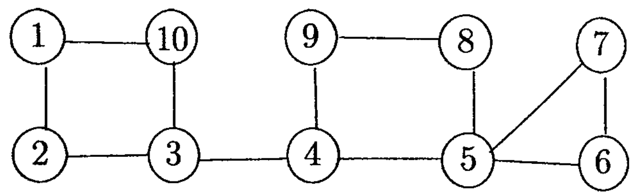
   Analyze the time complexity of your proposed algorithm.

   State whether your proposed algorithm can report the vertices of the convex hull in (clockwise or counter-clockwise) order? [8+4+3=15]

P.T.O

5. (a) You are given a connected undirected graph $G = (V, E)$, where the weight attached to each edge is one of the following values $\{1.1, 2.2, 3.3, 4.4, 5.5\}$. Present an $O(|E|+|V|)$ time algorithm for computing the minimum spanning tree of the graph G. Explain the correctness of your algorithm and justify the time complexity result.

(b) Write down the pseudo-code of the algorithm for computing the biconnected components of a connected graph. Show the execution trace (using DFS-number and low-number defined in the class) with the connected undirected graph, given below.



[8+(7+3)=10]

6. Define Fibonacci heap with an example. Show that, the following operations can be done in a Fibonacci heap in the time stated along with it.

- **Inserting** an element – – – $O(1)$ amortized time,
- **Deleting** the element with minimum key-value – – – $O(\log n)$ amortized time,

[3+2+5=10]

INDIAN STATISTICAL INSTITUTE
Mid-Semester Examination: 2017-2018
Course Name: M.Tech. In Computer Science
Subject Name: ***Automata Languages and Computation***

Date: 05.03.2018 | *Maximum Marks: 60* | *Duration: 2.5 hours*

Answer question1 and any 3 from the rest

1. For each of the following claims, indicate with short justification whether it is true or false. For the false ones (if any) provide a counter-example. For the true ones (if any) give a proof outline. No credit will be given for just writing true or false. **Answer <u>any 5</u> questions**: **[5X6]**
   a) **Claim**: Union of two non-regular languages is always non-regular.
   b) **Claim**: Union of a regular language with a disjoint non-regular language is always non-regular.
   c) **Claim:** The set of finite-length strings over a countably infinite alphabet is countably infinite.
   d) **Claim**: $L((ab^*ba^*) \cap (ba^*ab^*)) = \{\varepsilon\}$.
   e) **Claim**: The symmetric difference SD of 2 languages (over some alphabet Σ) is always context-free. *SD(L,M) of languages L and M is the set of strings in exactly one of L and M.*
   f) **Claim**: The intersection of two context free languages is never context free.
   g) **Claim**: Every finite subset of Σ^* is a regular set.
   h) Let $L = \{1^n : n \leq 1000 \text{ and } n \text{ is prime}\}$. **Claim**: A DFA accepting L may have less than 900 states.

2. (i) Let $\Sigma := \{a,b\}$ and $L := \{ \alpha \in \Sigma^* \mid \alpha$ ends with *bb* but does not start with *bb*$\}$.
   (a) Write a regular expression (over Σ) to represent L.
   (b) Design a NFA (or ε-NFA) whose language is L. **[3+2]**

   (ii) Let α be a string over some alphabet Σ. By odd (α), we refer to the string obtained by deleting symbols at all even positions of α. For example, if $\alpha = \alpha_1\alpha_2\alpha_3\ldots\alpha_n$, then odd $(\alpha) = \alpha_1\alpha_3\alpha_5\ldots\alpha_{n'}$, where n' is n or n-1, depending on whether n is odd or even. For a language $L \subseteq \Sigma^*$, define odd $(L) = \{odd(\alpha) \mid \alpha \in L\}$. Prove that if L is regular, then odd(L) is regular too. **[5]**

3. (i) Let L be a finite language over the binary alphabet {0,1}. Assume that $|L| = m$. Let D be a DFA with n states such that $L(D) = L$. Show that $n \geq \log(m+1)$, where the log is in base 2. **[5]**

   (ii) Consider the grammar $G := \langle \{S\}, \{a, b\}, \{S \rightarrow b \mid Sa \mid aS \mid SS\}, S \rangle$.
   (a) Does aabbaa $\in$ L(G)?

P.T.O

(b) The CFG given above is ambiguous. Show an example string for which there are two leftmost derivations.

(c) We claim that the regular expression for the language L(G) is a*bb*a*. Provide a short proof that indeed it is, or a counterexample if it is not. **[1+2+2]**

4. Let $\Sigma := \{a, b, c\}$ and $L := \{ \alpha c \alpha^R c \alpha \mid \alpha \in \{a, b\}^* \}$.

   (a) Show that L is not context-free.

   (b) Represent L as the intersection of two context-free languages (over Σ). **[5+5]**

5. A *wiggle string* is defined to be a nonempty string of 0's and 1's such that each 0 is followed by a 1, and each 1 is followed by a 0. A *0-wiggle string* is a wiggle string that begins and ends in 0, and a *1-wiggle string* is a wiggle string that begins and ends with 1. For example, 01010 is a 0-wiggle string, 1 is a 1-wiggle string, 0101 is a wiggle string that is neither a 0-wiggle string nor a 1-wiggle string, and 010010 is not a wiggle string. The language of the grammar G consisting of productions

S -> SAS | 0
A -> ASA | 1

(S is the start symbol) is the set of 0-wiggle strings, as you will prove. In this proof, you can assume the true fact that the concatenation of two wiggle strings is a wiggle string, as long as the first ends in a symbol different from the symbol by which the second begins. You do not have to state this fact in the proof. Give your proof by answering the following very specific questions.

Prove first that every 0-wiggle string is in L(G). To do so, we need to prove a more general statement inductively: (1) if w is a 0-wiggle string then S =>* w, and (2) if w is a 1-wiggle string, then A =>* w.

a) On what is your induction?
b) What is the basis case?
c) Prove the basis.
d) For the inductive part, first show that if w is a 0-wiggle string, then S =>* w.

The second part of the induction is that if w is a 1-wiggle string, then A =>* w. You do not have to provide this part, since its proof is essentially like that of (d), with 0 and 1 interchanged.

Conversely, to show that every string in L(G) is a 0-wiggle string, we shall show that (1) if S =>* w, then w is a 0-wiggle string, and (2) if A =>* w, then w is a 1-wiggle string.

a) On what is your induction?
b) What is the basis case?
c) Prove the basis.
d) For the inductive part, first show that if S =>* w, then w is a 0-wiggle string of length n.

To complete the inductive part, you also need to show that if A =>* w, then w is a 1-wiggle string. You do not need to provide this part, since it is essentially (h), with 0 and 1 interchanged. **[(1+1+1+2) + (1+1+1+2)]**

INDIAN STATISTICAL INSTITUTE

Mid-Semester Examination

M. Tech. (CS) I year (2nd Sem) and JRF (CS): 2017–2018

Data and File Structures

Date: 09. 03. 2018 Total Marks : 72 Time : 2 Hours

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

1. (a) Define worst case, best case and average case time complexities of an algorithm.

   (b) What is the difference between probabilistic analysis of algorithms and randomized algorithms?

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

2. (a) What is a heap?

   (b) Explain the Build-Heap procedure.

   (c) Can it be done in linear time? Explain.

   (d) Write a $O(\log n)$ time algorithm (in pseudocode format) for Heap-Insert.

   [2 + 4 + 8 + 8 = 22]

3. (a) What are the maximum and minimum number of comparisons in a binary search tree? Explain.

   (b) How can we reduce the gap between the above two numbers?

   (c) Write a C function to implement the recursive version of binary search.

   [(2 + 2) + 6 + 8 = 18]

4. (a) What is Fibonacci search? What are its advantages over other search algorithms?

   (b) Can you devise an algorithm for a $O(1)$ worst-case time-complexity of searching?

   [(4 + 2) + 6 = 12]

5. Write a function in C to implement merging two sorted array into a single sorted array. [10]

INDIAN STATISTICAL INSTITUTE

Semestral Examination

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

Design and Analysis of Algorithms

Date : April 23, 2018 Maximum Marks : 75 Duration : 3 Hours

Note : You may answer any part of any question, but maximum you can score is 75.

1. You are to organize a tournament involving n competitors, who are numbered as $\{1, 2, \ldots, n\}$. Each competitor must play exactly once against each of the $n-1$ opponents. Each competitor is to play at most one match per day.

   Show that if n is a power of 2, then it is possible to design an optimal scheduling of the games in the tournament that takes exactly $n-1$ days. Do this by giving an algorithm which takes n as input, and outputs the list of player-pairings for each of the $n-1$ days. State the time complexity of your algorithm in terms of the length of the input (i.e., the number of bits required for giving the input). [8+2 = 10]

2. Let $B = \{b_1, b_2, \ldots, b_m\}$ and $W = \{w_1, w_2, \ldots, w_n\}$ be two sets of points in $I\!R^2$, colored with *black* and *white*, respectively. Consider a bipartite graph $G = (B, W, E)$, where an edge (b_i, w_j) between a black point $b_i = (x_i, y_i) \in B$ and a white point $w_j = (x_j, y_j) \in W$ exists if $x_i \geq x_j$ and $y_i \geq y_j$. In a matching of G, every point in $B \cup W$ has at most one matching edge incident on it. A matching $M = \{(b_{i_1}, w_{j_1}), (b_{i_2}, w_{j_2}), \ldots, (b_{i_k}, w_{j_k})\}$ is said to be maximum if and only if k is maximum among all possible matchings. Design an $O(N \log N)$ time algorithm to get a maximum matching among the points in B and W, where $N = m + n$. Justify the time complexity of your algorithm. [8+4 = 12]

3. For a weighted planar graph $G = (V, E)$, can you design an algorithm for computing the all pair shortest paths in $O(|V|^2 \log |V|)$ time. You need to justify your answer. [8]

4. Give the *fail* indices generated by the KMP-flowchart construction algorithm for the following patterns:
   (i) AAAB,
   (ii) AABAACAABABA,
   (iii) ABRACADABRA,
   (iv) ASTRACASTRA. [3× 4 = 12]

5.(a) A *simple network* has either indegree 1 or outdegree 1 for each node.

Consider a simple *unit capacity* network. Show that the distance ℓ between the source s and the sink t cannot exceed $\frac{|V|}{f}$, where V is the set of vertices in the network, f is the value of the maximum flow, and the distance between a pair of nodes u and v is the number of nodes on the shortest path from u to v.

(b) Decide whether the following statement is *true* or *false*. If it is true, give proper justification, and if it is false, give a counter-example.

Let $G = (V, E)$ be an arbitrary flow network, with a source s and a sink t. Each edge $e \in E$ has a positive integer capacity c_e. Let (A, B) be the minimum $s-t$ cut in graph G with respect to the edge capacities $\{c_e, e \in E\}$. Now, if we add 1 to every edge capacity, then (A, B) still remains the minimum $s-t$ cut in the revised flow network.

[8+8 = 16]

6. Suppose that you wish to route flow through a network of pipes. We model the network as a connected, undirected graph $G = (V, E)$, in which each edge has a numeric value $c(u, v)$, which represents the *capacity* of the edge $(u, v) \in E$, that is, the amount of flow it can take. Given any path $P = u_1, u_2, \ldots, u_k$, its *bottleneck capacity* is defined to be the minimum capacity of any edge on the path, that is $cap(P) = \min\{c(u_1, u_2), c(u_2, u_3), \ldots, c(u_{k1}, u_k)\}$. By convention, $c(u, u) = \infty$, for all $u \in V$. For every $u, v \in V$, define $cap(u, v)$ to be the maximum *bottleneck capacity* over all paths from u to v.

Given a source vertex $s \in V$, present an algorithm that computes $cap(s, u)$ for all $u \in V$. Your algorithm should run in $O(|E| \log |V|)$ time. Note that, it is sufficient just to compute the *bottleneck capacity* of the paths. It is not required to compute the actual paths. [10]

7.(a) Let $P1$ and $P2$ be two problems such that $P1$ is polynomial time reducible to $P2$, the time complexity of this reduction is $O(n^2)$, and the time complexity of solving $P2$ is $O(n^4)$. Here n is the input size of the problem P1. What can you say about the time complexity of the problem $P1$? Justify your answer.

(b) Design an algorithm for computing the minimum vertex cover in time $O(2^k \times |V|^c)$, where k is the size of the minimum vertex cover, and c is a known constant. If it is not possble, then discuss the reasons with proper justifications. [5+5=10]

8. Let $G = (V, E)$ be a weighted directed graph. Show that the problem of getting a traveling salesman tour in G of cost $\rho \times opt$ is NP-complete, where opt indicates the cost of the optimum traveling salesman tour in G, and ρ is a given constant.

Also show that if the edge costs of G satisfy triangle inequality, then a traveling salesman tour of cost $2 \times opt$ can be found in polynomial time.

[8+7=15]

INDIAN STATISTICAL INSTITUTE

Second Semester Examination: 2017-18

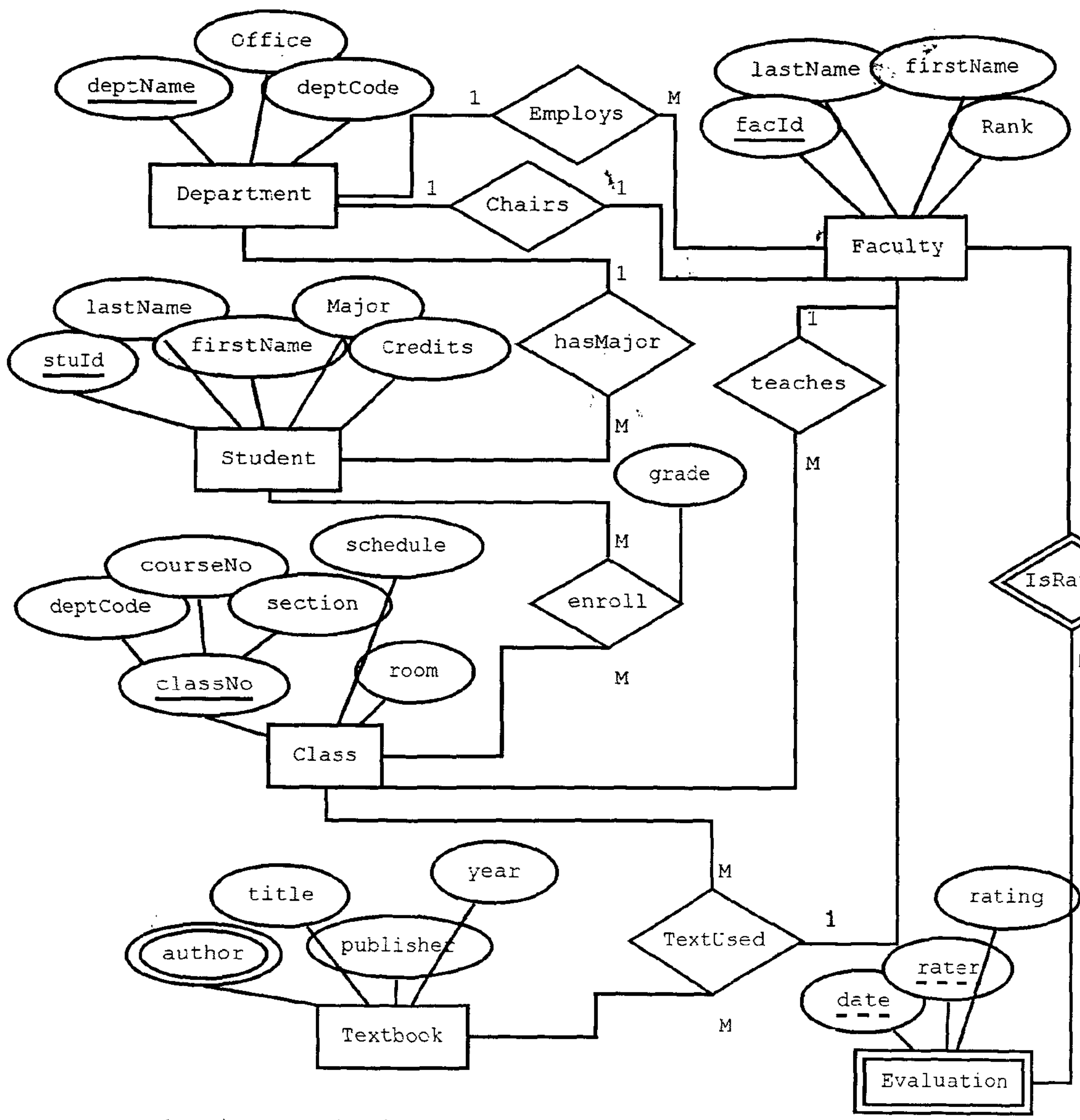
M. Tech. I Year (C.S)

Database Management Systems

Date: 25/04/2018 **Maximum Marks: 100** **Duration: 3hours**

Note: Answer any 5 questions.

1. Consider the following sample ER Diagram for an Institute:



a) Complete / correct the above ER Diagram
b) Write the problem statement that will lead **your** ER Diagram.

10+10=20

2. Compare:
   a) Sparse and Dense Indexing scheme for Insertion & Deletion of records
   b) Query Tree and Query Graph
   c) What is data fragmentation in distributed data storage system? Discuss the advantages of various types of data fragmentation.

7+6+7=20

P.T.O

3. a) Suppose a B+ tree node can hold up to 4 pointers and 3 keys. Show the tree formation **after every insertion** of the following:
   1, 3, 5, 7, 9, 2, 4, 6, 8, 10
   Show the tree after removing 9.
   b) What is Semi Join & Anti Join?
   c) Compare Immediate Update Deferred Update recovery protocol

10+5+5=20

4. Explain with example
   a) Conflict Serializable Schedule.
   b) View Serializable
   c) Table Lock & Row lock in Oracle.
   d) Join strategies that exploit parallelism.

5x4=20

5. a) Find the irreducible set of the following FDs of the relation schema R(A,B,C,D,E,F):

$AB \rightarrow C$
$C \rightarrow A$
$BC \rightarrow D$
$ACD \rightarrow B$
$BE \rightarrow C$
$CE \rightarrow FA$
$CF \rightarrow BD$
$D \rightarrow EF$

   Explain your answer.

   b) A schema is defined with the following attributes:

   D – Day of the week (1 to 5) P – Period within day (1 to 6)
   C – Class room number T – Teacher name
   L – Lesson name.

   Tuple (d,p,c,t,l) appears in this relation if and only if at time (d,p) lesson l is taught by teacher t in classroom c. List the functional dependencies and the candidate keys. Write your assumptions if required.

10+10=20

6. A) "Insulation between program and data can be achieved in database approach" Explain with example.
   b) Explain how system log helps recovery from crash?

15+5=20

7. Consider the following relations:
   Suppliers(<u>supplier_no</u>: int, supplier_name: char(50), city: char(40))
   Parts(<u>part_no</u>: int, part_name: char(100), color: char(10), weight: real)
   Projetcs(<u>project_no</u>: int, project_name: char(20), city: char(40))
   Shipment(<u>supplier_no</u>: int, <u>part_no</u>: int, <u>project_no</u>: int, quantity: int, unit_price: real, order_date: date)
   a) Write relational algebra/relational calculus formulation for the following updates:
      i. Suppliers who do not supply "Red" coloured parts.
      ii. The costliest "Yellow" coloured part.
      iii. Parts that were never supplied in various projects located at "Kolkata"
   b) Give SQL syntax to find the following
      i. Find the yearly average price of each part.
      ii. Find the top 2 costliest parts supplied by each supplier.
   c) Write an SQL query for the following and create a query tree
   Find the Supplier details who supply *High Quantity* (>100) "Blue" colour parts.

05+05+10=20

Indian Statistical Institute
Semester Examination: 2018
Course Name: M. Tech in Computer Science I-year
Subject Name: Computer Networks

Date: 27-04-2018 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) Consider a situation with eight parallel sessions using a 45 Mbps line. Each session generates Poisson traffic with $\lambda_i = 20$ packets/sec. Packet lengths are exponentially distributed with a mean of 8000 bits. There are two design choices:
   i. Each session is given a dedicated $\frac{45}{8}$ Mbps channel (via FDM or TDM).
   ii. Packets of all sessions compete for a single 45 Mbps shared channel.

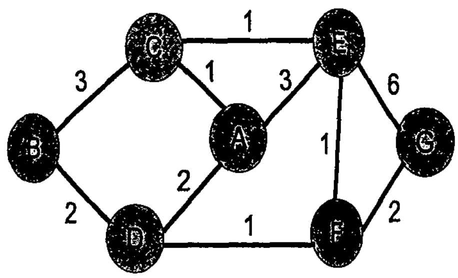
   Which choice is better with respect to average network delay? Explain you choice briefly. [8]

   (b) In CDMA/CD with a data rate of 10 Mbps, the maximum distance between any station pair is found to be 2500 m for correct operation of the collision detection process. What should be the maximum distance if we increase the data rate to 100 Mbps? [4]

   (c) Sixteen stations, numbered 1 through 16, are contending for the use of a shared channel by using the adaptive tree walk protocol. If stations 2, 3, 5, 9, 12, 14 suddenly become ready at once, compute how many bit slots are needed to resolve the contention. [6]

   (d) What problems would occur when CSMA is used in a wireless LAN? Describe the approaches that would solve the above problems. [2+8=10]

2. (a) Consider the network shown in the figure below. Using distance vector routing:
   i. Show the data that node A will receive in the first iteration of the algorithm.
   ii. Show the routing table for node A after the first iteration of the algorithm has been completed. [4+4=8]



   (b) What is Reverse Path Forwarding and how does it work? [8]

   (c) Suppose that all of the network sources are bursty-that they only occasionally have data to send. Would packet-switching or circuit switching be more desirable in this case? Justify your answer. [4]

P.T.O

(d) Consider a token bucket rate controller, used to control a reserved rate flow. Assume that the token bucket has a capacity of 10 tokens and a token fill rate of 100 tokens per second, and that every packet consumes one token. If no token is available for an arriving packet, it is *marked* for possible discarding. Suppose that at time 0, the token bucket is empty and the next token arrives at time 10 msec. If packets $P_1, P_2, P_3, P_4, P_5, P_6$ arrive at times 11, 13, 17, 19, 23 and 29 msec respectively,

i. Which packets (if any) are marked?

ii. How many tokens are in the token bucket at time 55 msec if no additional packets arrive?

iii. How many tokens are in the token bucket at time 195 msec?

iv. What is the largest number of packets that can be sent between time 201 msec and time 299 msec without any of the packets being marked? [3+3+3+3=12]

3. (a) Indicate whether each of the following subnet masks are valid or invalid. Justify your answer.

i. 255.255.32.0

ii. 255.255.224.0 [4+4=8]

(b) For the IP address 188.15.110.8/24, determine

i. the subnet address,

ii. directed broadcast address for the subnet,

iii. maximum number of hosts on that subnet, and

iv. maximum number of subnets, if the same subnet mask is used for all the subnets in the network. [2+2+2+2=8]

(c) Aggregate the following four /24 IP addresses to the highest degree possible.

i. 212.56.132.0/24

ii. 212.56.133.0/24

iii. 212.56.134.0/24

iv. 212.56.135.0/24 [4]

(d) A large number of consecutive IP addresses are available starting at 198.16.0.0. Suppose that four organizations, A, B, C and D request 4000, 2000, 4000, and 8000 addresses, respectively, and in that order. 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. [8]

4. (a) If the TCP round-trip time, RTT, is currently 30 msec and the following acknowledgements come in after 26, 32, and 24 msec, respectively, what is the new RTT estimate? Assume smoothing factor $\alpha = 0.9$. [6]

(b) At some point of time, a TCP connection is in slow-start phase with a congestion window of 4000 bytes. The maximum segment size used by the connection is 1000 bytes. What is the congestion window after it sends out 4 packets and receives ACKs for all of them before timeouts? [4]

(c) Why TCP and UDP use port numbers instead of using process IDs to identify the destination entity when delivering a message? [4]

(d) What is RPC and how does it work? [8]

INDIAN STATISTICAL INSTITUTE

Second-Semester Examination: 2017-18

Course Name: M. Tech. CS II Yr.

Subject Name: Computational Molecular Biology and Bioinformatics

Date: 27.04.18 Maximum Marks: 70 Duration: 2 Hrs 15 mins

Question no 1 is compulsory and answer any 5 questions from the rest.

1. Define bioinformatics and Computational Biology. [2.5+2.5=5]

2. Outline the basic steps of any method for predicting gene functions with necessary equations by integrating microarray gene expression, protein transitive homologues and KEGG pathway profiles as databases. [13]

3. Write the algorithm and construct the perfect binary phylogenetic tree for the character state matrix shown below. [13]

	Character					
Object	c_1	c_2	c_3	c_4	c_5	c_6
A	0	0	0	1	1	0
B	1	1	0	0	0	0
C	0	0	0	1	1	1
D	1	0	1	0	0	0
E	0	0	0	1	0	0

4. Find a possible semi-global alignment between sequences "***AGC***" and "***AAAC***" through explaining the necessary algorithms and steps in bidimensional score matrix. [13]

5. Describe the BLAST algorithm for aligning two protein sequences with example. [13]

6. State the differences between a biological pathway and a biological network. Why is diseasome a network? Name five categories of metabolic pathways and some databases for storing them. [5+3+5=13]

7. What is microRNA (miRNA)? Briefly describe the algorithm with necessary equations for microRNA ranking, containing diseased and normal expressions, using fuzzy rough entropy. [2+11=13]

8. Describe the various regions of 'DNA sequence' and 'RNA secondary structure' with diagrams. What is messenger RNA (mRNA)? State the various steps for measuring mRNA expressions using microarray technology. [(3+3)+2+5= 13]

Indian Statistical Institute
Semester-II 2017-2018
M.Tech.(CS) - First Year
End-semester Examination (02 May, 2018)
Subject: Operating Systems
Duration: 3.5 hrs.

INSTRUCTIONS

1. **There are 4 questions in this paper, carrying a total of 60 marks. You may answer as many questions as you like. The maximum that you can score is 50.**
2. **For each question, please write your answer in the space provided after that question. Please keep your answers brief and to the point.**
3. **You may use answer sheets for rough work (only). Please submit the answer sheets along with this question paper.**
4. **For each question, a very rough indication of the amount of time you should expect to spend on that question has also been provided.**

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 **busy waiting** version of the *wait* and *signal* operations on a binary semaphore S. Note that increment (++) and decrement (- -) operators **do not** apply to binary semaphores. [3 marks, 10 mins.]

`wait(S) {`
 	
 `}`
	`signal(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.]

(b) Consider the following code. The functions `wait_c(C)` and `signal_c(C)` are supposed to implement the *wait* and *signal* operations for a *counting* semaphore C.

```
binary_semaphore S1 = 1, S2 = 0;    int C = 1;
```

```
wait_c(C) {
    wait(S1);
    C--;
    if (C < 0) {
        signal(S1);
        wait(S2);
    }
    else signal(S1);
}
```

```
signal_c(C) {
    wait(S1);
    C++;
    if (C <= 0) signal(S2);
    signal(S1);
}
```

The `wait` / `signal` functions used within the body of `wait_c(C)` and `signal_c(C)` work on binary semaphores only. Two processes, P_1 and P_2, each run the following code to synchronise access to `critical_section()`.

`... wait_c(C); critical_section(); signal_c(C); ...`

Complete the example scenario below to show that the maximum number of times that P_2 can enter `critical_section()` before P_1 gets a chance to enter `critical_section()` depends on the scheduler instead of being a fixed constant. (In other words, the above implementation does not guarantee bounded waiting.) [5 marks, 30 mins.]

Process	**Completes executing (provide line number(s) or function name)**
P_2	`wait_c(C)`, `critical_section()`
P_1	`wait_c(C)` line 3

(c) (i) Show how you would use the `semget()` system call to create a single counting semaphore. [2 marks, 5 mins.]

(ii) Write C functions to implement the *wait* and *signal* operations using the `semop()` system call. Your functions should take the semaphore descriptor (`semid`) as the only argument. [4+4=8 marks, 30 mins.]

```
wait(int semid) {

}
```

```
signal(int semid) {

}
```

2. Consider the following Resource Allocation State involving 5 processes and 5 resources.

$$Total = [\ 10\ 20\ 30\ 20\ 10\ ]$$

$$Max = \begin{bmatrix} 5 & 10 & 15 & 10 & 5 \\ 5 & 4 & 3 & 2 & 1 \\ 10 & 10 & 10 & 10 & 10 \\ 6 & 7 & 8 & 9 & 10 \\ 10 & 15 & 20 & 10 & 5 \end{bmatrix} \quad Alloc = \begin{bmatrix} 1 & 2 & 1 & 4 & 0 \\ 1 & 2 & 2 & 0 & 1 \\ 1 & 2 & 3 & 4 & 2 \\ 1 & 2 & 4 & 0 & 1 \\ 1 & 2 & 5 & 4 & 2 \end{bmatrix} \quad Req = \begin{bmatrix} 4 & 8 & 12 & 6 & 4 \\ 1 & 2 & 1 & 2 & 0 \\ 8 & 8 & 7 & 6 & 5 \\ 5 & 5 & 4 & 9 & 4 \\ 1 & 13 & 15 & 5 & 3 \end{bmatrix}$$

$Total[i]$ specifies the total number of instances of resource i that exist in the system (including both allocated and free instances). $Max[i,j]$, $Alloc[i,j]$, and $Req[i,j]$ denote, respectively:

- the maximum number of instances that process i may request of resource j,
- the number of instances of resource j currently allocated to process i, and
- the number of instances of resource j that process i is currently requesting.

Show that the system is not currently in a deadlocked state. Also show that deadlock could possibly happen in the future (i.e., the system is not in a safe state). [6+7=13 marks, 25 mins.]

System not currently deadlocked

Deadlock may happen in future

3. (a) Consider an SVR2 filesystem with the following parameters.

- disk block size = 4096 bytes
- space required to store an inode on disk = 128 bytes
- amount of space used to store inode list = 2^{10} blocks
- amount of space used to store data blocks = 2^{18} blocks

Suppose the filesystem is full. Assume that none of the files contain any "holes".

(i) Compute how much space would be wasted if the average file size is

A. 16 Kbytes:

B. 16 Mbytes:

(ii) For what value of the average file size will no space be wasted when the file system is full?

[4+4+2=10 marks, 45 mins.]

(b) Briefly describe *two* fields that are contained in a cached (in-core) copy of an SVR2 inode, but are not contained in the corresponding disk copy. [3 marks, 15 mins.]

Field 1:

Field 2:

(c) List the 4 fields contained in each global file table entry. [2 marks, 10 mins.]

Field 1:

Field 2:

Field 3:

Field 4:

(d) Describe the possible situations that can result in the indegree of a global file table entry being greater than 1. [3 marks, 10 mins.]

4. (a) Consider the following extract from a program, written in a C-like language, that computes the transpose of a matrix.

```
for (i = 0; i < N; i++)
    for (j = 0; j < N; j++)
        B[i,j] = A[j,i];
```

A and B are N×N matrices with floating point entries that are stored in memory in row-major order as shown in the example below.

A[0,0]	A[0,1]	A[0,2]	...	A[0,N-1]	A[1,0]	A[1,1]	...	A[N-1,N-1]

This program runs under an operating system that uses paging based memory management with the following parameters.

- Page size: 2^{12} bytes
- Number of frames allocated to the program: 4
- Page replacement policy: FIFO
- $N = 2^{15}$
- Size of each matrix entry: 8 bytes
- Each of A and B is stored starting from the beginning of a page.
- None of the pages allocated to A or B are initially in memory.

(i) How many pages do the matrices A and B occupy?

(ii) Considering **only** memory references to the matrix entries, compute the page fault rate for the matrix transposition code given above.

[1+8=9 marks, 25 mins.]

(b) (i) In operating systems running on i386 processors, can the following happen? Justify your answer in each case. [2 marks, 5 mins.]

A. Internal fragmentation: YES NO

B. External fragmentation: YES NO

Indian Statistical Institute
Second Semester Examination: 2017-2018

First
M.Tech.(CS) - ~~Second~~ Year
Automata, Languages and Computation

Date: May 4, 2018 Maximum marks: 100 Duration 3 hrs.

Answer any 4 out of 6 questions. Each question carries 25 marks.

1. (a) Comment which of the languages below are regular / non-regular with a formal justification.
   (i) $L_1 = \{1^k y \mid y \in \{0,1\}^*$ and y contains at least k 1s, for k $\geq 1\}$.
   (ii) $L_2 = \{1^k y \mid y \in \{0,1\}^*$ and y contains at most k 1s, for k $\geq 1\}$.

   (b) Let N_1 $(Q_1, \Sigma, \delta_1, q_1, F_1)$ recognize a regular language A_1. Construct $N = (Q_1, \Sigma, \delta, q_1, F)$ for recognizing A_1^* using the following steps.
   - The states of N are the states of N_1.
   - The start state of N is the same as the start state of N_1.
   - F = $\{q_1\} \cup F_1$. The accept states F are the old accept states along with its start state.
   - Define δ so that for any $q \in Q$ and any $a \in \Sigma_\epsilon$,

   $$\delta(q,a) = \begin{cases} \delta_1(q,a) & q \notin F_1 \text{ or } a \neq \epsilon \\ \delta_1(q,a) \cup \{q_1\} & q \in F_1 \text{ and } a = \epsilon \end{cases}$$

   Give a counterexample to show that the construction above fails to prove the closure of the class of regular languages under the star operation. In other words, you must present a finite automaton, N_1 for which the automaton N constructed using the rules above, does not recognize the star of $N_1's$ language.

   [(8+8)+9]

2. (a) Consider the non-Context Free Language $L = \{a^i b^j c^k d^l \mid i,j,k,l \geq 0$ and either $i = 0$ or $j = k = l\}$, with $\Sigma = \{$a,b,c,d$\}$. Show that L acts like a CFL in the pumping lemma for CFLs. Explain why this does not contradict the pumping lemma.

   (b) Let A be a CFL that is generated by the Context Free Grammar (CFG) $G = (V, \Sigma, P, S)$. Add the new rule $S \rightarrow SS$ and call the resulting grammar G'. Is G' the grammar to generate A^*? Provide a short proof that indeed it is, or a counterexample to show that it is not.

   (c) Show that, if G is a CFG in Chomsky Normal Form, then for any string $w \in L(G)$ of length $n \geq 1$, exactly 2n-1 steps are required for any derivation of w.

   [10+8+7]

3. (a) Design a push-down automaton to recognize the language L below, defined over $\Sigma = \{$a,b,c$\}$

   $$L = \{a^i b^j c^k \mid i,j,k \geq 0 \text{ and } i = j \text{ or } i = k\}$$

   Present an informal description of the automaton, along with the corresponding diagram.

   (b) Consider 2 regular languages A and B. Define $A \diamond B = \{xy \mid x \in A$ and $y \in B$ and $|x| = |y|\}$. Note that $|x|$ for a string x denotes the length of x. Prove $A \diamond B$ is a CFL. *Hint: Try constructing a PDA.*

   [(8+7)+10]

4. (a) Show that the Post Correspondence Problem (PCP) is decidable over the unary alphabet $\Sigma = \{1\}$.

   (b) Let $AMBIG_{CFG} = \{\langle G \rangle \mid G$ is an ambiguous CFG, $\langle G \rangle$ being the encoding of the grammar$\}$. Show that $AMBIG_{CFG}$ is undecidable, by providing a reduction from PCP. In other words, show that finding whether an arbitrary context free grammar is ambiguous is undecidable. Explain briefly your reduction.

(c) Show that the collection of Turing-recognizable languages is closed under intersection.

[(8+10)+7]

5. (a) Present an implementation-level description for a Turing machine that decides the language L over the alphabet {0,1}:

$$L = \{w \mid w \text{ contains an equal number of 0s and 1s}\}$$

Analyse the time and space complexity of your solution.

(b) Consider all single-tape Turing machines that are not allowed to write on the portion of the tape containing the input string. Show that such TMs can recognize only regular languages. [Note: the TM is also not allowed to copy the input string elsewhere on the tape and modify].

[(9+6)+10]

6. (a) For each of the following languages, state whether the language *is* or *is not* recursively enumerable and whether the complement of the language *is* or *is not* recursively enumerable. Give brief justification for your answers.

- The language of all TMs that accept nothing.
- The language of all TMs that accept everything.
- The language of all TMs that accept Regular sets.
- The language of all PDAs that accept everything.
- The language of all CFGs that are ambiguous

(b) Show that any PSPACE-hard language is also NP-hard.

[(5 X 4)+5]

INDIAN STATISTICAL INSTITUTE
JRF (CS) - 2017-2018
Computer Organization
Semestral Examination

Date: 07. 05. 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. Consider two 2-bit integers y_3y_2 and y_1y_0.
   (a) How many bits are required to store the product 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 as decided in (1a) by you.
   (c) Describe relevant Karnaugh maps to obtain simplified Boolean circuits to provide the multiplication result.
   (d) Implement the circuit with logic gates of your choice.
   (e) How would you design a circuit to connect one or more 7-segment display(s) to present the output of the multiplication circuit so that the output result can be viewed as a decimal number?

   $2+4+4+4+6=20$

2. Consider a 4-bit counter that can count cyclically from 0000 to 1111 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.
   (c) Suppose the counter value at any clock instance is $y_3y_2y_1y_0$. How to connect your circuit in (1) with these counter bits so that the product of y_3y_2 and y_1y_0 at that very clock instance will be displayed in the 7-segment display.

   $10+6+4=20$

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.

   $8+6+6=20$

4. We need to write a C program for reading and manipulating unsigned integers. Write pieces of C codes or a complete program for the following.
   (a) How to read an unsigned integer through keyboard?
   (b) Print the number of bytes n in an unsigned integer?
   (c) How to print the contents of those n bytes separately?
   (d) How to obtain the product of those n individual bytes and to print the final result in hexadecimal?

   $3+3+8+6=20$

P.T.O

5. (a) Using "add \$r5, \$r6, \$r7" describe how the instruction can be executed in 5 clock cycles. Explain clearly what is happening in each clock cycle.

   (b) Write 10 consecutive R-type instructions (you may use add, sub, and, or, nor, xor) that can be parallelized using 5-stage pipelining without any data hazard. What is the average number of clock cycles required to execute an instruction in this scenario?

   (c) Modify the sequence of instructions above to demonstrate data hazard. What is the average number of clock cycles required to execute an instruction with such hazard?

$6 + 8 + 6 = 20$

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) "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 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 one can locate a specific slot in cache memory (if it is indeed in the cache)?

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

8. (a) Is the three input gate $G(x, y, z) = x \oplus yz$ universal? Give reason to justify your answer.

   (b) What is the count of distinct 3-input 3-output Boolean functions? How many of them are reversible?

   (c) Take two 8-bit numbers M, N such that $M < N$. Show how one can calculate $M - N$ using two's complement.

   (d) How many address lines should be connected to a 65536 KByte RAM Chip?

   (e) A movie of 760 MByte (residing in Hard Disk) is to be played. Briefly explain why DMA (Direct Memory Access) is useful in such a scenario.

$4 + (2 + 2) + 4 + 4 + 4 = 20$

INDIAN STATISTICAL INSTITUTE
Semester Examination
M. Tech. (CS) I year (2nd Sem) and JRF (CS): 2017–2018
Data and File Structures

Date: 09. 05. 2018 Total Marks : 60 Time : 2.5 Hours

Answer as much as you can. Maximum you can score is 50.

1. (a) Write a simple algorithm to detect if a graph is cyclic or not. What is its running time?
   (b) What does the k-th power of the adjacency matrix of a graph denote?

   [(4 + 2) + 4 = 10]

2. (a) What are the worst case search complexities in a BST tree and in an AVL tree?
   (b) Write the pseudocode for insertion into an AVL tree.
   (c) Write the pseudocode for deletion from an AVL tree.

   [4 + 6 + 6 = 16]

3. Excluding the worst-case linear time algorithm, what are the other unoptimized algorithms to solve the k-th order statistics problem and what are their time complexities? [16]

4. (a) Solve the recurrence $T(n) = 2T(\sqrt{n}) + \log n$. Assume T(1) is a small constant.
   (b) Many identities in generating functions work under the assumption $|x| < 1$. Why doesn't this create a restriction in solving recurrence relations using generating functions?

   [4 + 4 = 8]

5. (a) What is the difference between a B^+ tree and a B^* tree?
   (b) What is an inverted file? Give an example.

   [4 + (2 + 4) = 10]