INDIAN STATISTICAL INSTITUTE

Mid-Semestral Examination

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

Discrete Mathematics

Date : Sept. 7, 2015 Maximum Marks : 50 Duration : 2:30 Hours

Note : You may answer any part of any question, but maximum you can score is 50.

1 One hundred twenty eight cars were assembled in a factory. The options available were a radio, an air conditioner, and white-wall tires. It is known that 60 of the cars have radios, 32 of them have air conditioners, and 24 of them have white-wall tires. Moreover 12 of them have all three options. Can we predict anything about the number of cars that do not have any options at all. Justify your answer. [10]

2 Note that a partial ordered set is said to be a lattice if every two elements in the set have a unique least upper bound and a unique greatest lower bound. Let $(A, \leq)$ be a partially ordered set. Let $\leq_R$ be a binary relation on A such that for a and b in A, $a \leq_R b$ if and only if $b \leq a$. Prove that $\leq_R$ is a partial order relation. If $(A, \leq)$ is a lattice, can we claim that $(A, \leq_R)$ is also a lattice? [5+5=10]

3 An induced subgraph of G is a subgraph H of G such that $E(H)$ consists of all edges of G whose endpoints belongs to $V(H)$. Suppose G is a simple graph having no vertex of degree 0 and no induced subgraph with three edges. Find maximum number of vertices G can have. Justify your claim. [10]

4 Suppose a simple graph G has n vertices an m edges. Use Turán's Theorem to give lower bounds on the number of vertices in the largest clique and the largest independent set in G. [10]

5 Consider a complete graph K_n having n vertices. For any edge (u, v), we give a direction either from u to v or v to u but not in both direction. Such a directed graph is called *tournament*. Prove that every tournament has a vertex from which every vertex is reachable by a directed path of length at most 2. [10]

6 The Ramsey number $R(s;t)$ is the minimum number n such that any graph on n vertices contains either an independent set of size s or a clique of size t. Prove that if $A = R(k; m-1)$ and $B = (k-1; m)$ are both even then $R(k;m) \leq A + B - 1$. [10]

Indian Statistical Institute

Semester-I 2015-2016

M.Tech.(CS) - First Year

Mid-term Examination (09 September, 2015)

Subject: Introduction to Programming

Total: 20 marks Maximum marks: 20 Duration: 2 hrs.

Please keep your answers brief and to the point.

1. Compute the values of `a`, `b`, `f`, `g`, `c` and `d` after the following code fragment is executed.

```
int a, b, c, d, x, y, z;
float f, g;
x = 2;  y = 5;  z = 10;
a = y / x;
b = z % y;
f = y / x;
g = (y + 0.0) / x;
c = (x++) * (--y);
d = y - x;
```

a = _____

b = _____

f = _____

g = _____

c = _____

d = _____

[0.5×4 + 1×2 = 4]

2. Compute the values of `x`, `*p`, `*q`, and `a[0]` after the following code fragment is executed.

```
int x, *p, *q, a[10];
a[0] = 10;  a[1] = 20;  a[2] = 30;
x = a[0]; p = &x; q = &(a[0]);
x++; *p = *p + a[2]; *q = *q + 10;
```

x = _____

*p = _____

*q = _____

a[0] = _____

[1×4 = 4]

ROUGH WORK

3. (a) A point in the X-Y plane is defined by its x and y coordinates. A circle in the X-Y plane is defined by its centre (a point), and its radius. Define C structures to represent a point and a circle in this manner. Use appropriate `typedefs` to give the names `POINT` and `CIRCLE` to these structure types.

(b) Show how you would use `malloc()` to allocate space for an array of `n` `CIRCLEs`.

(c) Fill in the skeleton below to write a C function to determine whether a given point lies within a given circle. The return value should be 1 if the point lies within the circle (or on the circumference), and 0 otherwise.

```
int isInside(POINT *p, CIRCLE *c)
{   /* declarations */



    /* code */
```

[(1+1.5) + 1.5 + 4 = 8]

4. The function given below **reverse** takes a string **s** as an argument and is supposed to reverse **s in place**.

```
void reverse(char *s)
{
   char c;
   int i, l = strlen(s);

   for (i = 0; i < l; i++) {
       c = s[i];
       s[i] = s[l-i];
       s[l-i] = c;
    }
    return;
}
```

(a) Explain why it does not work.

(b) What changes would you make to get the function to work correctly?

$[3 + 1 = 4]$

INDIAN STATISTICAL INSTITUTE

MID-SEMESTRAL EXAMINATION
M.TECH.(CS) I YEAR

ELEMENTS OF ALGEBRAIC STRUCTURES

Date: 11.09.2015 Maximum marks: 60 Duration: 2 hours

The paper contains 65 marks. Answer as much as you can, the maximum you can score is 60.

1. (a) Prove that a group can never be the union of two proper subgroups. [8]
   (b) Suppose G is a group and H is a normal subgroup of G such that both H and G/H are cyclic. Prove that G is generated by two elements. [8]
   (c) Let H be a cyclic subgroup of a group G, and H is normal in G. Prove that, every subgroup of H is normal in G. [8]
   (d) Let G be a finite cyclic group of order n. Prove that G has $\phi(n)$ many generators. [8]
2. Let G be a group. Let $\mathrm{Aut}(G)$ be the set of isomorphisms $\phi : G \to G$. The elements of $\mathrm{Aut}(G)$ are called automorphisms of G.
   (a) Prove that $\mathrm{Aut}(G)$ is a group under composition. [4]
   (b) Let $G =< a >$ be a cyclic group. Let ϕ be an automorphism of G. Prove that $\phi(a)$ is a generator of G. Furthermore, show that, if b is another generator of G, then there is a unique automorphism ψ of G, such the $\psi(a) = b$. [3+5]
   (c) Find two automorphisms of $\mathbb{Z}_n$ $(n > 2)$. [3]
   (d) Prove that $|\mathrm{Aut}(\mathbb{Z}_n)| = \phi(n)$. [4]
3. Universal Product Code (UPC) symbols are found on most products in retail stores. The UPC symbol is a 12 digit code identigying the manufacturer of a product and the product itself. The first 11 digits contain information about the product; the twelfth digit is used for error detection. If $d_1, d_2, \ldots, d_{12}$ is a valid UPC code, then

$$3 \cdot d_1 + d_2 + 3 \cdot d_3 + \ldots + 3 \cdot d_{11} + d_{12} \equiv 0 \mod 10$$

Figure 1: UPC Code

(a) Show that the number $0 - 50000 - 30042 - 6$ is a valid UPC number. [2]

(b) Show that if the above number is erroneously read as $0 - 05000 - 30042 - 6$, the error not detected. [2]

(c) Find a transposition error that is detected. [4]

(d) Deduce a general rule of transposition errors that will be detected. [6]

INDIAN STATISTICAL INSTITUTE

PERIODICAL EXAMINATION
M.TECH.(CS) I YEAR

PROBABILITY AND STOCHASTIC PROCESSES

Date: 14.09.2014 Maximum marks: 80 Duration: 2 and 1/2 hours

The paper contains 100 marks. Answer as much as you can, the maximum you can score is 80.

1. (a) If $X_1, \ldots, X_n$ are random variables with finite variances $\sigma_1^2, \ldots, \sigma_n^2$ respectively and $S_n = X_1 + \cdots + X_n$, then show $\mathsf{Var}(S_n) = \sum_{k=1}^{n} \sigma_k^2 + 2 \sum_{1 \le j < k \le n} \mathsf{Cov}(X_j, X_k)$.

   (b) Let $\mathcal{X} = \{1, 2, 3, 4\}$ and X and Y be random variables over $\mathcal{X}$ with the following joint distribution.

$$\begin{aligned} \Pr[X = x, Y = y] &= \tfrac{1}{2^x} \times \tfrac{1}{3} \text{ if } x = 1, 2 \text{ and } y = 1, 2; \\ &= \tfrac{1}{2^x} \times \tfrac{1}{6} \text{ if } x = 1, 2 \text{ and } y = 3, 4; \\ &= \tfrac{1}{8} \times \tfrac{1}{3} \text{ if } x = 3, 4 \text{ and } y = 1, 2; \\ &= \tfrac{1}{8} \times \tfrac{1}{6} \text{ if } x = 3, 4 \text{ and } y = 3, 4. \end{aligned}$$

   Compute the marginal distributions of X and Y. Justify whether X and Y independent. Find the means and variances of X and Y and their co-variance.

(15 + 15 = 30)

2. (a) If the probability of hitting a target is 1/5 and ten shots are fired independently, what is the probability of the target being hit at least twice?

   (b) Two dice are thrown r times. Find the probability p_r that each of the six combinations $(1,1), \ldots, (6,6)$ appears at least once.

   (c) Let X and Y be two finite non-empty sets with $|X| = n > m = |Y|$ and let $h : X \to Y$. An r-collision for h is a set of r distinct points $x_1, \ldots, x_r$ from X such that $h(x_1) = \cdots = h(x_r)$.

      i. Suppose $x_1, \ldots, x_r$ are picked independently and uniformly at random from X (uniform random sampling with replacement). What is the probability that they form an r-collision?

      ii. Let $q \ge r$ and $x_1, \ldots, x_q$ are picked independently and uniformly at random from X with replacement. Find an upper bound on the probability that there is at least one r-collision among these q points.

(6 + 6 + 13 = 25)

3. (a) Urn R contains n red balls and urn B contains n blue balls. At each stage, a ba is selected at random from each urn and they are swapped. Show that the expecte number of balls in urn R after stage k is $\frac{1}{2}(1 + (1 - 2/n)^k)$.

   (b) Let $X_1, \ldots, X_n$ be a sequence of independent random variables where each X_i take values from the set $\{0, 1, 2\}$ with probabilities p_0, p_1, p_2 respectively. Show how to extra independent and uniformly distributed random *bits* from such a sequence.

(10 + 10 = 2

4. (a) Give the formal definition of a probability space and a random variable. Properly defi all other notions which are required to arrive at the above definitions.

   (b) Let $E_1, E_2, \ldots$ be events in a (general) probability space. Show that

$$\Pr\left[\bigcup_{i\geq 1} E_i\right] \leq \sum_{i\geq 1} \Pr[E_i].$$

   (c) Define a distribution function on the real line. Define the normal density function ϕ and the normal distribution function $\Phi(x)$.

   (d) Define statistical distance $d(p, q)$ between two probability distributions on a finite n empty set Y. Show $\sum_{y\in Y} \max\{p(y), q(y)\} = d(p, q) + 1$.

   (e) Let $A_1, A_2, \ldots$ be a sequence of events. Define $\limsup_{n\to\infty} A_n$. If $\sum_n \Pr[A_n] < \infty$ show th

$$\Pr\left[\limsup_{n\to\infty} A_n\right] = 0.$$

(5 × 5 =

INDIAN STATISTICAL INSTITUTE

Mid-Semester-Examination: 2015-2016 M.Tech. (CS) First Year

Computer Organization

Date: 16.09.2015 **Maximum marks = 60** Final Credit: 30%

Time: 3 hours

Name:_____ **Roll No.:**_____

Instructions (Read carefully)

A. This is an **OPEN BOOK/OPEN NOTES** exam. Answer all questions; partial credit may be given for incomplete/incorrect answers.

B. Total points = 70; **maximum score = 60.**

---

1. **(10 points)**
(a). Represent the following 2's complement number into IEEE 754 single precision format:

1111 1111 1000 0001

(b). Consider the number N = 1 00000000 1000 0000 0000 0000 0000 000 in the IEEE 754 floating point format. Write the number N in fractional binary form.

(c). Write the number $(-3)^{1/2}$ in IEEE 754 floating point format. (3 + 4 +3)

2. **(15 points)** Let N denote a positive normalized number in the IEEE 754 floating point format, where the string E represents the exponent and M represents the significand of N as shown below:

S (1-bit)	E (8-bit)	M (23-bit)

Let *$t1* and *$t2* be two 32-bit registers as in MIPS, which hold the strings E and M respectively, each logically extended on the left (i.e., required MSB's filled with 0's). Now, given the contents of *$t1* and *$t2* as inputs, (i) suggest an algorithm for checking whether or not N can be expressed as a 32-bit 2's complement integer; (ii) Write a code in MIPS to implement this. You may use integer arithmetic instructions (signed, unsigned, immediate), logical (AND, OR, XOR, including immediate mode), branch, and shift (logical left/right, arithmetic left/right) instructions, as needed. Assume availability of general pupose registers such as *$t0, $t1,....., $s0, $s1,....., $zero,* as necessary. (8 +7)

3. **(10 points)** You are given a 32-bit integer A in 2's complement form stored in a register *$t0*. We want to compute $\lfloor 63*A/4 \rfloor *9$, where *, /, $\lfloor \rfloor$ represent multiplication, division, and floor operations, respectively. Write a MIPS code to evaluate the expression and save the result in register *$t1*. You are **not allowed** to use *mul* (multiply) or *div* (divide) instruction. Justify the rationale behind your algorithm.

4. **(10 points)** A program P contains 20% tasks which are to be sequentially executed on a machine M1. A new 80-core machine M2 has now been designed, which consists of 80 copies of M1 on a chip. Calculate the maximum achievable speed-up S_A in respect to execution time of the new machine M2 relative to machine M1, for the program P following Amdahl's argument. Next, compute the scaled speed-up S_G for the machine M2 relative to M1 following Gustavson's argument. Derive the expressions for speed-up in both the cases and justify your derivation. In your analysis, you may neglect the inter-processor transaction time. (5 + 5)

5. **(15 points)** Consider the following MIPS code running on a machine M1, whose clock cycle time (CCT) is 10 *ns*. Assume that each of the instructions can be completed within a clock cycle. How many instructions will be executed by this code? What is the total amount of CPU-time spent before the code halts.

```
        lw $t1, -50($t2)
        srl $t1, $t1, 16
        lui $t3, 7FFF (in hex)
        ori $t3, FFFF (in hex)
Loop    xor $t4, $t1, $t3
        add $t1, $t1, $t4
        sll $t1, $t1, 16
        bne $t1, $zero, Loop
        sw $t2, 1000($t3)
```

6. **(10 points)** Show the scheme of an adder that will add two (nk)-bit 2's complement integers $A = a_{nk}\, a_{nk-1} \ldots a_1$ and $B = b_{nk}\, b_{nk-1} \ldots b_1$ such that the cost of the adder = $O(nk \log k)$, and delay = $O(n \log k + k)$.

INDIAN STATISTICAL INSTITUTE
M. Tech (CS). I year : 2015–2016
Data & File Structures
Mid Semester Examination

Date: 18. 09. 2015 Marks: 50 Time: $2\frac{1}{2}$ Hours

Answer any part of any question. The question is of 55 marks. The maximum marks you can get is 50. Please write all the part answers of a question at the same place.

1. (a) Write a C program for the following task. Input the last two digits of your roll number (in case roll number is not available, put the last two digits of your year of birth), say integer r. Input a prime p in the range 30 to 50. Create a data set of 11 integers $x_0, \ldots, x_{10}$, where

$$x_i = (r + 20) * (i + 1) \bmod p.$$

   (b) Write a C program to construct a heap.

   (c) Execute your code of 1(b) on the data available from 1(a), considering that the integers are arriving in the order $x_0, x_1, \ldots x_{10}$ one after another. Explain each step with proper figure. Do not insert any duplicate data.

5+5+5 = 15

2. (a) Describe the data structure of a binary search tree and write down a C program for insertion of a new data.

   (b) Execute your insertion function (explain with proper figures of binary tree) on the data set $x_1, \ldots, x_{10}$ available from 1(b), considering that the integers are arriving in the order $x_0, x_1, \ldots x_{10}$ one after another. Do not insert any duplicate data.

(2+5) + 8 = 15

3. (a) Write a C function to multiply two integer matrices A, B and to store the result in R. The matrices A, B, R, along with the necessary dimensions, should be passed as parameters to the function.

   (b) Using the matrix multiplication function implemented above, write a C program to read the matrices A, B from a file and to write the resultant product matrix R in the same file without removing the existing contents.

7 + 8 = 15

P. T. O.

4. Explain what happens when the following codes (or pieces of codes) are executed. Give proper justification to your answers.

(a) `char *p, *q; while (*p++ = *q++);`

(b)
```
int i, k = 1, n = 5;
for (i = 0; k < n+1; i = k-i) { printf("%d\n", k); k = k+i; }
```

(c)
```
#include <stdio.h>
main(){
        int arr[10][10], i, j;
        for (i = 0; i < 10; i++)
        for (j = 0; j < 10; j++)
                arr[i][j] = i*j + 1;
        printf("%x\n", arr); printf("%x\n", arr+1);
        printf("%x\n", *arr); printf("%x\n", *arr+1);
        printf("%x\n", **arr);
}
```

The first line of output of this program is fef56970.

(2+3) + 5 = 10

INDIAN STATISTICAL INSTITUTE
M. Tech. (CS) I year : 2015–2016
Data & File Structures
Semestral Examination

Date: 16. 11. 2015 Marks: 100 Time: 3 Hours

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

1. (a) Write a function in C using iteration that can evaluate

$$f(n) = (f(n-1))^2 + (f(n-2))^3$$

with initial conditions $f(0) = 0, f(1) = 1$.

(b) Write a function in C to check whether a string is a palindrome.

(c) Write a C program that can insert and search a record inside a file. Consider that each record contains employee identification number, the name, address and salary. You should use "fread" and "fwrite" functions available in C.

$5 + 5 + 10 = 20$

2. (a) What is a heap?

(b) Explain an algorithm to construct a heap and show how your algorithm works with the data set 121, 37, 6, 121, 91, 98, 67, 85.

(c) What is the time complexity of your algorithm above?

(d) How this data structure can be efficiently used for sorting?

$2 + 8 + 6 + 4 = 20$

3. (a) Describe the data structure of a binary search tree.

(b) Write down a C program for deletion of a key in this data structure.

(c) Consider a binary search tree with the key values 25, 10, 30, 5, 11, 7 (arriving in this sequence). Show how your deletion program works when you delete the key 25 from the tree.

$4 + 10 + 6 = 20$

4. (a) Write down a non-recursive C function for inorder traversal in a binary tree.

(b) Describe the algorithm for constructing a binary search tree when only its pre-order traversal data is available.

$10 + 10 = 20$

P. T. O.

5. (a) Briefly explain the concept of a height balanced (AVL) tree.

   (b) What is the worst case time complexity to search a key in such a tree containing N nodes?

   (c) Write down the insertion algorithm for this data structure.

$3 + 7 + 10 = 20$

6. (a) What is an m-way search tree. What are the additional constraints over m-way search tree to define a B-Tree? What is a 2-3 tree?

   (b) Estimate the minimum as well as the maximum number of key values in a B-Tree of order m having height l.

   (c) How do you decide the value of m for an efficient implementation?

$(2 + 2 + 2) + 8 + 6 = 20$

7. (a) Write the algorithm for insertion of a key value in a B-Tree.

   (b) Explain your insertion algorithm for a B-Tree of order 3 with the key values 77, 55, 32, 49, 78, 79, 89, 62, 99, 53, 35 (arriving in this order),

$10 + 10 = 20$

8. (a) Explain two hashing strategies with clear description of collision resolution.

   (b) Select a specific hashing strategy among the above two and implement a function in C programming language that can manage search and insertion in a hash table.

$10 + 10 = 20$

9. (a) Explain the splay tree with suitable examples.

   (b) Briefly discuss the concepts of indexed file and relative file.

$10 + (5 + 5) = 20$

INDIAN STATISTICAL INSTITUTE

Semestral Examination

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

Discrete Mathematics

Date : Nov.. 23, 2015 Maximum Marks : 100 Duration : 3:00 Hours

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

1 Let R is reflexive relation on a set A. Show that R is an equivalence relation if and only if (a, b) and (a, c) are in R implies that (b, c) is in R. [8]

2 Let $(P, \leq)$ be a partially ordered set. Suppose the length of the longest chains in P is n. Prove that the elements in P can be partioned into n distinct antichains. [10]

3 Derive the solution for the recurrence relation $a_n - 5a_{n-1} + 6a_{n-2} = 7^n$ with $a_0 = 1$ and $a_1 = 6$. [12]

4 Use a truth table to show that $(a \rightarrow b) \wedge (a \wedge \neg b)$ is a contradiction. Also show that the nagation of a quantified statement $\neg[(\forall x)L(x)]$ is equivalent to $(\exists x)(\neg L(x))$. [6+6=12]

5 Let S be a set of n elements. $\Im$ is a family of nonempty subsets of S such that for some fixed k, $|A \cap B| = k$ for every distinct $A, B \in \Im$. Prove that $|\Im| < n$. [15]

6 Prove that the Ramsey number $R(s, t)$ is less than $\binom{s+t-2}{s-1}$ for $s, t > 1$. [10]

7 Let $G(X, Y; E)$ be a bipartite graph with bipartision X and Y. Prove that G has a matching of X into Y if and only if $|N(S)| \geq |S|$ for all $S \subseteq X$ where $N(S)$ denotes the set of vertices having a neighbor in S. Derive a formula for the number of matchings in $K_{n,n}$ that do not match x_i to y_i for any i. [10+10=20]

8 Let $G(V, E)$ be a triangle free planar graph with at least three vertices. Prove that $|E| \leq 2|V| - 4$. (Triangle free implies no cycle of length three). [12]

9 Write an algorihm that outputs a triangle free graph having chromatic number $k+1$ from a triangle free graph having chromatic number k. Prove that the algorithm you have given is correct. [10+10=20]

INDIAN STATISTICAL INSTITUTE

SEMESTRAL EXAMINATION
M.TECH.(CS) I YEAR

ELEMENTS OF ALGEBRAIC STRUCTURES

Date: Maximum marks: 100 Duration: 3 hours 26·11·15

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

1. (a) Let U, V be two ideals of a ring R. Let UV be the set of all elements that can be written as finite sums of elements of the form uv where $u \in U$ and $v \in V$. Prove that UV is an ideal of R. Moreover, Show that $UV \subset U \cap V$. [5+5]

   (b) Let V be a finite dimensional vector space over F, and T be a homomorphism of V onto V. Prove that T is an isomorphism. [10]

   (c) Write an algorithm to find the gcd of two polynomials over a field F. [10]

   (d) Construct a finite field of 2^3 elements. [10]

   (e) Let F be a field. Let $g \in F[x]$ be a polynomial of degree n. Let (g) be the ideal generated by g in $F[x]$. Prove that $F[x]/(g)$ is an n dimensional vector space over F. [10]

2. Let f be a non-zero polynomial over $\mathbb{Z}[x]$. The content of f, denoted by $c(f)$, is defined as the g.c.d. of the non-zero coefficients of f.

   (a) Let $f, g \in \mathbb{Z}[x]$ be two non-zero polynomials. Prove that $c(f.g) = c(f).c(g)$ [10]

   (b) If $f \in \mathbb{Z}[x]$ factors as $f = gh$ with $g, h \in \mathbb{Q}[x]$, then prove that there exists $a, b \in \mathbb{Q}$, such that $ag, bh \in \mathbb{Z}[x]$ and $f = (ag).(bh)$. [10]

   (c) Let $f \in \mathbb{Z}[x]$ be of positive degree whose leading coefficient is not divisible by the prime p. If $f \mod p$ is irreducible, then prove that f is irreducible in $\mathbb{Q}[x]$. [5]

   (d) Let $f \in F_q[x]$ and suppose $u \in F_q[x]$ satisfies $u^q \equiv u(\mod f)$. Prove that

$$f = \prod_{a \in \mathcal{F}_q} gcd(f, u - a).$$

   [10]

3. (a) Let F be a field. Let $f(x)$ be a polynomial of degree d in $F[x]$. Assume that, for distinct $\alpha_1, \cdots, \alpha_n$ of F with $d < n$, we have $f(\alpha_i) = \beta_i$. Prove that

$$f(x) = \sum_{i=1}^{n} \beta_i \left(\prod_{j \neq i} \frac{x - \alpha_i}{\alpha_i - \alpha_j} \right) \qquad [10]$$

(b) Prove that, for a square matrix P over a field F, the columns are linearly independent, if and only if the rows are linearly independent. [5]

(c) Let $a_0, \cdots, a_m, b_0, \cdots, b_m$ be elements of the field F with the a_is being non-zero. Prove that, the columns of the matrix

$$\begin{pmatrix} a_0 & a_1 & \cdots & a_m \\ a_0 b_0 & a_1 b_1 & \cdots & a_m b_m \\ \vdots & \vdots & \ddots & \vdots \\ a_0 b_0^m & a_1 b_1^m & \cdots & a_m b_m^m \end{pmatrix}$$

are linearly independent if and only if the b_js are distinct. [10]

INDIAN STATISTICAL INSTITUTE

First-Semestral-Examination: 2015-2016 M.Tech. (CS) First Year

. Computer Organization

Date: 28 November, 2015 Credit: 50%

Time: 3 hours

Name:_____ **Roll No.:**_____

Instructions (Read carefully)

A. Answer all questions; partial credit may be given for incomplete/incorrect answers.
B. This is an **OPEN-BOOK/OPEN-NOTES** exam.
C. Total points = 60; **maximum score =50.**

---

1. **(5 points)** For each of the statements below, write True (**T**) or False (**F**):

(i) The only way to detect overflow while performing the addition of two 2's-complement integers in the CPU hardware is to check the sign-bits of the inputs and the output.

(ii) Multi-cycle implementation of data paths for processor design improves the CPI value of the computer compared to that of the single cycle implementation.

(iii) In multi-cycle processor implementation, the clock cycle time is determined by the longest delay in executing an instruction.

(iv) A pipelined machine can be viewed as a single-cycle un-pipelined machine with an improved value of CPI.

(v) TLB (translation lookaside buffer) is a buffer that stores the updated data words waiting to be written in the secondary memory.

2. **(3 points)** Consider the following MIPS code:

```
lui  $ t1, 7FFF (in hex)
ori $t1, $t1, FFFF (in hex)
addu $t1, $t1, $t1
sll t1, t1, 2
addi $t1, $t1, 9
```

The content of the register $t1 after execution of the above code is (choose one):
(i) 0, (ii) 1, (iii) -1, (iv) a number causing overflow, (v) none of these.

3. (**4 points**) Consider the following two floating point numbers A and B in IEEE 754 single precision format:

A: 0 1111 1110 1000 0000 0000 0000 0000 000

B: 0 1111 1110 0100 0000 0000 0000 0000 000

The result of floating point addition (A + B) is (choose one):

(i) 0 1111 1111 0110 0000 0000 0000 0000 000

(ii) 0 1111 1110 0100 0000 0000 0000 0000 000

(iii) 1 1000 0000 1100 0000 0000 0000 0000 000

(iv) 0 1111 1111 0000 0000 0000 0000 0000 000

(v) none of the above.

5. (**2 points**) Higher associativity in cache organization is likely to provide (choose one):

(i) reduced miss rate, higher hit time, and unchanged miss penalty;
(ii) higher miss rate, higher hit time, and lower miss penalty;
(iii) reduced miss rate, reduced hit time, and unchanged miss penalty;
(iv) reduced miss rate, unchanged hit time, and lower miss penalty;
(v) reduced miss rate, reduced hit time, and higher miss penalty.

6. (**3 points**) The characteristics and purpose of using a victim cache are (choose one):
(i) It is an additional cache used in the hierarchy to increase the cache size;
(ii) It is an additional cache kept as a pre-fetch buffer to hold blocks which may be needed later, and hence, it reduces miss rate;
(iii) It is a fully-associative cache that is provided to bring newly requested blocks from memory when the regular cache is full;
(iv) It is an additional cache that only holds the blocks which were replaced by new incoming blocks to the main cache;
(v) None of the above.

7. (**3 points**) Consider a full binary tree with 15 nodes including 8 leaf-nodes. Each leaf node appears at depth 3 (assume the root has depth 0). The tree represents an information-search scenario as experienced while web search such as use of Wikipedia. The in-order labeling of the tree is {1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15}. For cache-oblivious efficient search, the nodes should be linearly stored in the main memory as:

(i) 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15;
(ii) 8, 4, 12, 2, 1, 3, 6, 5, 7, 10, 9, 11, 14, 13, 15;
(iii) 8, 4, 2, 1, 3, 6, 5,7, 12, 10, 9, 11, 14, 13, 15;
(iv) 8, 4, 12, 2, 6, 10, 14, 1, 3, 5, 7, 9, 11,13, 15
(v) none of the above.

8. (**4 points**) Consider a cache-main memory (MM) system, where the miss penalty is 32 clock cycles, which includes 4 cycles for sending address from the CPU to MM, 24 cycles access time of MM per word, and 4 cycles for sending a word from MM to cache. Let the cache block size be 4-word; MM is one-word organized, interleaved with 4 banks; MM-cache bandwidth = 1-word; cache miss rate is 2%; average # MM-access per instruction = 1.2. Ideal CPI (without cache misses) = 2. The CPI of the real machine will be (show your computation and choose one):

(i) 2.53, (ii) 2.86, (iii) 2.77, (iv) 3.54, (v) none of the above.

9. (**20 points**) The memory hierarchy of a computer system consists of virtual memory, main memory and cache and is organized as follows (assume 1 word = 32 bits).

(i) Virtual memory: 8 virtual pages; page size = 2K words; page replacement policy is FIFO.
(ii) Main memory (MM): 4 physical page frames; page size = 2K words; MM has 2-way interleaving; Each line per bank of MM = 1 word;
time to send the line address from CPU to MM = 5 clock cycles;
time to access a line in MM = 15 clock cycles;
time to deliver a word from MM to cache = 4 clock cycles;
width of the memory data bus = 32 bits.

(iii) Cache:
The cache is physically addressed, write-through, direct-mapped, has a capacity of 4K words, cache block size = 2 words;
Currently, virtual pages 0, 1, 4, 5 are in MM, and they were loaded in the MM in the same order in time (i.e., page 0 first, page 1 next, so on). The current page table is given by:

Virtual page	Page frame
0	3
1	1
2	not in MM
3	not in MM
4	2
5	0
6	not in MM
7	not in MM

The CPU requests access to an item X whose virtual address is 0100 1001 0010 10. Assume that currently cache is empty.

(i) (5 points) Write the sequence of events that would take place while processing on this request (e.g, cache hit/miss, address translation, page fault).
(ii) (5 points) What will be the physical address of X after being uploaded to MM (in binary)?
(iii) (5 points) Where should X be located in the cache (cache index in binary)?
(iv) (5 points) What would be the miss penalty of the cache for a read miss in clock cycles?

10. (**16 points**) Consider the following MIPS code sequence:

```
Loop: LW $t1, 0 ($t2)
      SW $t1, 48 ($t4)
      SUB $t2, $t2, $t1
      BNZ $t2, Loop
```

(a) Show the timing of the above instruction sequence for a 5-stage pipelining architecture without any forwarding or other enhancement.

(b) Assume that normal *forwarding* hardware is now provided and assume that the branch is handled by predicting it as *not taken*, and that zero-testing hardware *has not* been moved to the ID-stage.

(c) If the above loop is executed 20 times, how many clock cycles would be required to complete the program, with the above assumptions?

(d) Is it possible to improve the timing further by re-ordering of instructions? If so, indicate how it can be achieved.

INDIAN STATISTICAL INSTITUTE

SEMESTRAL EXAMINATION
M.TECH.(CS) I YEAR

PROBABILITY AND STOCHASTIC PROCESSES

Date: 20.11.2015 Maximum marks: 100 Duration: 3 hours

The paper contains 120 marks. Answer as much as you can, the maximum you can score is 100.

1. (a) A bin contains m white balls and n black balls. Balls are taken out uniformly at random without replacement. What is the expected number of black balls left when all the white balls have been taken out?

   (b) If a random variable X takes only non-negative integer values, then show that

$$E[X] = \sum_{n \geq 0} \Pr[X > n].$$

(10 + 10 = 20)

2. (a) Let $A = (a, \alpha)$ and $B = (b, \beta)$, where a, b, α, β are integers with $b > a \geq 0$, $\alpha, \beta > 0$. Show that the number of paths for a random walk from A to B which touch or cross the x-axis equals the number of all paths for a random walk from $A' = (a, -\alpha)$ to B.

   (b) Let u_n be the probability that a return to origin occurs at epoch n for a random walk that started at the origin. Find u_n.

   (c) Show that the probability that in time interval 0 to $2n$, the particle spends $2k$ time units on the positive side and $(2n - 2k)$ time units on the negative side equals $u_{2k}u_{2n-2k}$.

(6 + 6 + 8 = 20)

3. (a) State and prove the Chapman-Kolmogorov equations.

   (b) Show that a finite Markov chain has no null states.

   (c) Let $\{Y_k\}$ be a sequence of mutually independent random variables, each assuming the values ± 1 with probabilities 1/2. Let $X_n = (Y_n + Y_{n+1})/2$. Is $X_1, X_2, \ldots$ a Markov chain? Justify your answer.

   (d) Let e_1, e_2 and e_3 be the three states of a time invariant Markov chain. A transition is made from state e_i to e_j with probability $(i + j)/(6 + 3i)$. Write the state transition matrix of the Markov chain. Does the Markov chain have a stationary distribution? Justify your answer and if it has one, then find it.

(e) Define what is meant by an ergodic state of a Markov chain. Show that if a state e_j is ergodic, then for all i,

$$p_{i,j}^{(n)} \to f_{i,j}\mu_j^{-1} \text{ as } n \to \infty,$$

where μ_j is the mean recurrence time for state e_j and $f_{i,j}$ is the probability that the chain starting in state e_i will ever pass through state e_j.

$(6 + 7 + 7 + 10 + 10 = 40)$

4. (a) Define rth central moment. Show that for the normal distribution, all odd order central moments vanish. Define Kurtosis. Show that for the normal distribution, Kurtosis is 0.

(b) Define Fisher information. Suppose X follows $\mathsf{Poisson}(\lambda)$. Find the Fisher information in X about λ.

(c) Let $X_1, \ldots, X_n$ be independent random variables each following $\Phi(\mu, \sigma^2)$. Obtain the MLE for μ and an unbiased estimator for σ^2.

(d) Let $X_1, \ldots, X_N$ be independent random variables each following $\mathsf{Bernoulli}(p)$ with p unknown. It is required to test $H_0 : p = 1/4$ versus $H_1 : p = 1/2$. Design a most powerful level α test for this scenario. Show how to use the normal approximation of the binomial distribution to obtain approximate expressions for the Type-I and Type-II error probabilities.

$(8 + 7 + 10 + 15 = 40)$

Indian Statistical Institute

Semester-I 2015-2016

M.Tech.(CS) - First Year

End-semester Examination (30 November, 2015)

Subject: Introduction to Programming

Total: 60 marks Maximum marks: 50 Duration: 4 hrs.

INSTRUCTIONS

1. All programs should be written in C.
2. Please **strictly** adhere to the following conventions for naming your files. The main file for problem number i should be named `cs15xx-probi.c`, where `15xx` is your roll number. Any supporting files should be named `cs15xx-probi-aux1.c`, `cs15xx-probi-aux1.h`, `cs15xx-probi-aux2.c`, and so on.
3. Create a single shell script named `compile.sh` that contains the commands required to compile **all** your programs.
4. Please put **ALL** your files in a **single** directory named `cs15xx`. Create a tarball from this directory using the following command: `$ tar czf cs15xx.tgz ./cs15xx`.
5. Mail the tarball to `mandar.mitra@gmail.com`. Use "Programming End-sem" as your subject line.
6. You may consult your notes, manual pages or other documentation available on your machine, but you will be permitted Internet access **only** when submitting your answer.
7. For Question 4, please submit a written answer in the provided answer sheet in addition to your program. Your answer should clearly describe the data structure and algorithm that you use, using diagrams if necessary.
8. You may also submit a **brief** written answer / solution sketch for the other problems. Your programs will be tested using randomly generated test cases. Even if your programs do not work correctly on these test cases, you may get partial credit based on your written answers.

P.T.O.

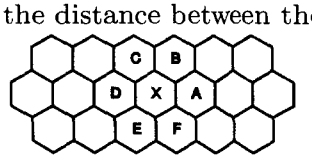
1. Write a program to print a diamond shape on the screen using asterisks (*). The "width" of the diam defined as the number of asterisks in any diagonal of the diamond, will be specified as a comman argument. You may assume the argument will be an odd number between 1 and 50.

Example:

```
$ a.out 1
*
$ a.out 5
  *
 ***
*****
 ***
  *
$
```

Marks: 8 (correctness: 4, time / space efficiency: 2, programming style: 2).

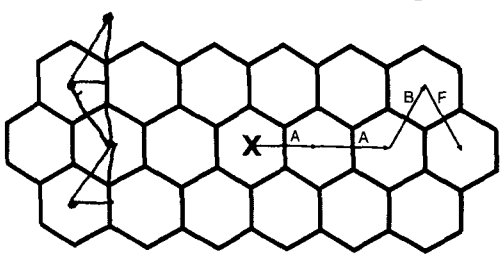
2. A hexagonal grid is shown below. An ant that is initially sitting at the centre of the cell marked X crawling. At each move, the ant moves in one of the six directions marked A, B, C, D, E, and F centre of an adjacent cell. Assume that the distance between the centres of adjacent cells is 1.0 uni



Write a program that takes a sequence of moves made by the ant as command-line arguments and the final displacement of the ant from its initial position.

Example: The figure on the right marks the path taken by the ant in the first example.

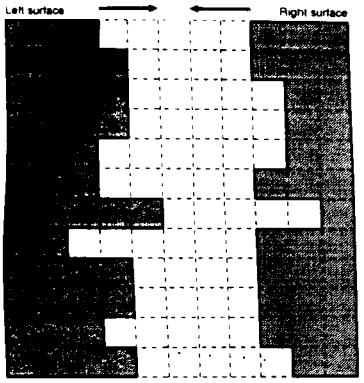
```
$ a.out AABF
3
$ a.out FEDCBA
0.0
$ a.out BCB
2.6458
```



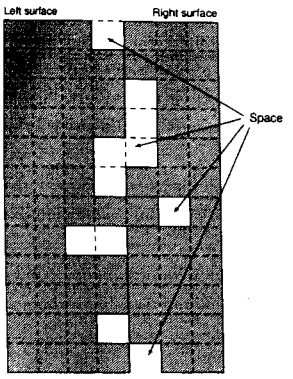
Marks: 16 (correctness: 7, time / space efficiency: 5, programming style: 4).

$3 \times \frac{\sqrt{3}}{2}$ $\frac{1}{2}$

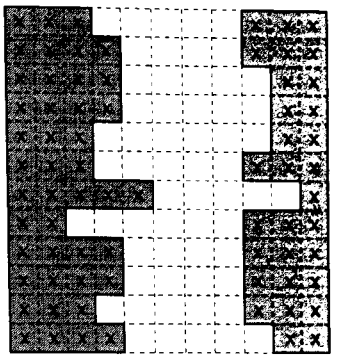
3. The following diagrams show the slightly rough surfaces of two machine parts before assembly, and after assembly when they are in contact with each other. Assume that the diagram (before assembly) is represented by an $m \times n$ matrix of characters. Each element of the matrix is either an `X` (to mark the machine parts) or a blank (indicating space). The first and last columns will always consist of `X`s only (no spaces). The matrix representation of the following diagram (superimposed on the diagram itself) is shown in the rightmost figure below.



Before assembly



After assembly


Matrix representation

Each row of the matrix contains a **single** stretch of zero or more blanks separating the left surface from the right surface. For a given diagram, you have to determine the total empty area — the total number of blanks trapped between the surfaces — that will exist after assembly, i.e., after the left surface has been brought into contact with the right surface. For the above example, your answer should be 11.

Note that the two surfaces are brought into contact by moving them strictly horizontally towards each other until a rightmost `X` of the left surface of some row is immediately to the left of the leftmost `X` of the right surface of that row.

Assume that the matrix representation of the diagram is given to you in an input file called `input.txt`. This file will consist of one or more lines of `X`s and blanks, each corresponding to one row of the image. All lines will have the same length; each line will begin and end with a series of `X`s, and will have a single sequence of zero or more blanks in between. Your program should print the total number of spaces trapped between the 2 surfaces after they have been brought into contact as described above.

Example: input.txt

```
XXXX␣␣␣␣␣XXXXX
XXX␣␣␣␣␣XXXXXXX
XXXXX␣␣␣␣␣XXXX
XX␣␣␣␣␣␣␣XXXXXX
```

```
$ a.out
4
$
```

Marks: 18 (correctness: 10, time / space efficiency: 4, programming style: 4).

P.T.O.

4. Let $p_1, p_2, \ldots, p_n$ be n positive integers representing the stock price of a company over a period consecutive days. The *span* of day i is defined as the maximum number of consecutive days (starting a including day i) over which the stock price p_i remains maximum. Write a program that takes $p_1, p_2, \ldots$ as command-line arguments and computes and prints the span of each day.

Example:

```
$ a.out 5 4 3 3 4 2 6
6 5 2 1 2 1 1
```

Marks: 18 (correctness: 6, time / space efficiency: 8, programming style: 4).

NB: For this question, you are *required* to submit a written answer.

INDIAN STATISTICAL INSTITUTE

Periodical Examination

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

Design and Analysis of Algorithms

Date : 22.2.2011 Maximum Marks : 60 Duration : 3 Hours

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

1. Let L be an array containing a sequence of n integers. Write an expected linear time in-place algorithm for finding the median of the members in L. [By an in-place algorithm we mean that you can swap the lements in the array and you are allowed to allocate $O(1)$ working storage in your algorithm. Justify the time and extra-space complexity of your algorithm.

   An element is said to be a *frequent* element if it occurs strictly more than $\frac{n}{4}$ times in the given set. Design an $O(n)$ time algorithm to find a *frequent* element in a set of n elements, if one exists.

   [(5+4+2)+9=20]

2. (a) Given a simple polygon with n vertices in $\mathbb{R}^2$, design an $O(n)$ time algorithm for reporting the edges of their convex-hull in order. Justify the time complexity of your proposed algorithm. Hints: Apply Graham-Scan algorithm appropriately to show that the time complexity is linear.

   (b) Describe an $O(n \log n)$ time algorithm for the 2-d farthest pair problem. The input to your algorithm should be an array of n points $P[1, 2, \ldots, n]$, where $P[i].x$ and $P[i].y$ are the x- and y-coordinates of the i-th point. Explain the steps of the algorithm clearly, and analyze the time complexity of each step. Also, justify the correctness of your proposed algorithm. You may assume that there are no duplicate x- or y-coordinates, if it helps to simplify your algorithm. You are also advised to avoid using complicated algorithmic tools unless it is required to achieve your targeted time complexity.

   [8+8=16]

3. Consider the UNION-FIND problem, where each set is represented by a tree, and the set name is written at the root of that tree.

   The *union by rank* heuristic for the union operation of two sets works as follows: Let U and V be two sets represented by two trees rooted at u and v respectively. If $|U| < |V|$, then u is attached with v; otherwise, v is connected to u. The name of the new set is writen at the root (v or u) of the resulting tree, and its size is set to $|U| + |V|$.

   Show that if the *union by rank* heuristic is followed, then the height of the tree of a set containing k nodes can be at most $\lfloor \log k \rfloor$.

State the method of path compression when a sequence of union and find operations are executed. What is the advantage of this method with respect to the execution time of the algorithm. (You do not need to give any formal proof.)

You also state the data structures to be used for storing the sets in the implementation of the above two methods for the UNION-FIND problem.

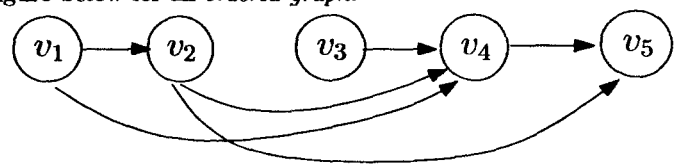
[5+5+5=15]

4. Let $G = (V, E)$ be a directed graph with nodes $V = \{v_1, v_2, \ldots, v_n\}$. We say that G is an *ordered graph* if it has the following properties:

   (i) Each edge goes from a node with lower index to a node with higher index. That is, every directed edge has the form (v_i, v_j) with $i < j$.

   (ii) Each node except v_n has at least one edge leaving it. That is for every node v_i, $i = 1, 2, \ldots, n-1$, there is at least one edge of the form (v_i, v_j).

   See the figure below for an *ordered graph*.



   The length of a path is the number of edges in that path. The goal in this problem is to identify the longest path from v_1 to v_n in the graph G.

   Show that the following algorithm does not correctly solve this problem, by giving an example of an *ordered graph* on which it does not return the correct answer.

   while there is an edge out of the node w
   choose the edge (w, v_j)
   for which j is as small as possible
   Set $w = v_j$;
   Increase L by 1;
   endwhile
   Return L as the length of the longest path.

   In your example, you tell what is the correct answer and what the algorithm will return.

   [10]

5. The transitive closure of a directed acyclic graph (DAG) $G = (V, E)$ is another DAG $G' = (V, E')$ where an edge $e = (v_i, v_j) \in E'$ implies that there is a directed path from v_i to v_j in the graph G. Design an $O(|V||E|)$ time algorithm for computing the transitive closure of a DAG $G = (V, E)$. [10]

INDIAN STATISTICAL INSTITUTE
First Semestral Examination : 2015-16

Course Name: M. TECH. (CS) I YEAR

Subject Name: Computer Networks

Date: 23.02.2016

Maximum Marks: 60 Duration: 2 hours

Answer as much as you can.

1. a) A large population of ALOHA users generate 50 requests/sec, including both originals and retransmissions. Time is slotted in units of 40 msec.

i) What is the chance of success at the first attempt?
ii) What is the probability of exactly k collisions and then a success?
iii) What is the expected number of transmission attempts needed?

b) Frames of 1000 bits are sent over a 1 Mbps satellite channel. Assume that the propagation delay over a satellite channel is 250 ms. Acknowledgements are always piggybacked onto data frames. The headers are very short. Three bit sequence numbers are used. What is the maximum achievable channel utilisation for:
a) Stop-and-wait
b) Go-back-n
c) Selective repeat

7+8 = 15

2. a) A group of N stations share a 56 kbps pure ALOHA channel. Each station outputs a 1000 bit frame on an average of once every 100 seconds, even if the previous one has not yet been sent (e.g. the stations can buffer outgoing frames). What is the maximum value of N?

b) Consider two hosts A and B, connected by a single link of rate R bps. Suppose that the two hosts are separated by m meters, and suppose the propagation speed along the link is s meters/sec. Host A is to send a packet of size L bits to Host B.

i) Express the propagation delay, d_{prop}, in terms of m and s.
ii) Determine the transmission time of the packet, d_{trans}, in terms of L and R.

c) A system uses the Stop-and-Wait ARQ Protocol. If each packet carries 1000 bits of data, how long does it take to send 1 million bits of data if the distance between the sender and receiver is 5000 km and the propagation speed is 2×10^8 m/s? Ignore transmission, waiting, and processing delays. We assume no data or control frame is lost or damaged. 5+(2+2)+6 = 15

3. a) A cyclic redundancy code has 8 message bits and uses the generator polynomial:

$$x^5 + x^4 + x + 1$$

(i) Deduce the redundant checkbits which would be appended to the following message at the sender side: 11101001
(ii) Verify your answer to (i) showing that the receiving side's check of the message extended with the redundant bits succeeds.

b) We need a dataword of at least 11 bits. Find the values of k and n in the Hamming code C(n, k) with $d_{min} = 3$.

c) Explain whether the following code is linear or not:

Dataword	*Codeword*
00	00000
01	01011
10	10111
11	11111

d) We need to use synchronous TDM and combine 20 digital sources, each of 100 Kbps. Each output slot carries 1 bit from each digital source, but one extra bit is added to each frame for synchronization. Answer the following questions:

i) What is the duration of an output frame?
ii) What is the output data rate? 5+4+2+4 = 15

4. a) A network with one primary and four secondary stations uses polling. The size of a data frame is 1000 bytes. The size of the poll, ACK, and NAK frames are 32 bytes each. Each station has 5 frames to send. How many total bytes are exchanged if there is no limitation on the number of frames a station can send in response to a poll?

b) Prove that for a non-persistent CSMA, the throughput S is given by:

$$S = \frac{Ge^{-aG}}{G(1+2a)+e^{-aG}},$$

where a is the ratio of propagation delay to packet transmission time and G is the offered traffic rate. 7+8 = 15

5. a) i) Suppose you have the following 2 bytes: 01011100 and 01100101. What is the 1s complement of the sum of these 2 bytes?
ii) For the bytes in part (a), give an example where one bit is flipped in each of the 2 bytes and yet the 1s complement doesn't change.

b) If an Ethernet destination address is 07:01:02:03:04:05, what is the type of the address (unicast, multicast, or broadcast)? Explain.

c) Start with the unit vector $w = [+1]$ and do the following:

i) Use the Walsh table to develop codes for a network system with 4 stations.

ii) Show that the receiver will receive the exact bits by the CDMA multiple access method. Assume that the four stations that share the channel have the following data for transmission during a 1 bit interval:

Station 1: none, station 2: bit 1, station 3: bit 0, and station 4: bit 1.

(2+3)+2+(2+6) = 15

6. Suppose there is a CSMA/CD-like multiple access protocol. In this protocol, time is slotted and all adapters are synchronized to the slots. Unlike slotted ALOHA, however, the length of a slot (in seconds) is much less than a frame time (the time to transmit a frame). Let S be the length of a

slot. Suppose all frames are of constant length $L = kRS$, where R is the transmission rate of the channel and k is a large integer. Suppose there are N nodes, each with an infinite number of frames to send. We also assume that the propagation delay $d_{prop} < S$, so that all nodes can detect a collision before the end of a slot time. The protocol is as follows:

• If, for a given slot, no node has possession of the channel, all nodes contend for the channel; in particular, each node transmits in the slot with probability p. If exactly one node transmits in the slot, that node takes possession of the channel for the subsequent $k - 1$ slots and transmits its entire frame.

• If some node has possession of the channel, all other nodes refrain from transmitting until the node that possesses the channel has finished transmitting its frame. Once this node has transmitted its frame, all nodes contend for the channel.

Note that the channel alternates between two states: the productive state, which lasts exactly k slots, and the non-productive state, which lasts for a random number of slots. Clearly, the channel efficiency is the ratio of $k/(k + x)$, where x is the expected number of consecutive unproductive slots.

a) For fixed N and p, determine the efficiency of this protocol.

b) For fixed N, determine the p that maximizes the efficiency.

c) Using the p (which is a function of N) found in (b), determine the efficiency as N approaches infinity.

d) Show that this efficiency approaches 1 as the frame length becomes large.

4+4+4+3 = 15

INDIAN STATISTICAL INSTITUTE

Semestral Examination: 2015-16

Course Name: M. Tech. I Year

Subject Name: Database Management Systems

Date: 24/02/2016 **Maximum Marks: 100** **Duration: 3hours**

Note: Answer all questions

1. Define all the variations of the join operation. Express them in terms of cross product, selection and projection, if possible. Why is the join operation given special attention for query optimization?

12

2. a) Give a set of FDs for the relation schema R(W, X, Y, Z) with primary key {W, X} under which R is in 2NF, 3NF but not in BCNF.
   b) Consider the following set of FDs:
   - $X \rightarrow YZ$
   - $Y \rightarrow Z$
   - $X \rightarrow Y$
   - $XY \rightarrow Z$
   - $XC \rightarrow W$

   Find a irreducible set.

12

3. Consider the following relations:
   Suppliers(supplier_no: int, supplier_name: char(50), status: int, city: char(40))
   Parts(part_no: int, part_name: char(100), color: char(10), weight: real)
   Projetcs(project_no: int, project_name: char(20), city: char(40))
   Shipment(supplier_no: int, part_no: int, project_no: int, quantity: int)
   Give SQL formulation or relational algebra or tuple / domain relational calculus expression for the following queries:
   a) Find the maximum value of part_no
   b) All parts with same name, color and weight.
   c) All the projects for which there are no shipments.
   d) Find the Suppliers who supply all available colored parts except Yellow Color parts.

12

4. A socio-economy survey has been done in a set of remote villages in Eastern India to study their agricultural practices. A village is identified by a name, which may be considered unique. Each village has certain number of houses indentified by a house number. Each house belongs to a certain caste. In each house, the number of persons involved in agriculture may vary. Some house may also have some non-agricultural come. Villages have three types of lands – highly fertile, moderately fertile and un-fertile. Un-fertile lands are used for raising fruit trees. In each village, each plot of land is of specific land type and is identified by a plot number and cropping practices (mono-cropping and multi-cropping). Each plot is owned by any one of the houses in that village. Village authority, for each village, keeps an account of the crops cultivated in that village, in a calendar year and the average yield of each of them in tons/acre.
   Draw a suitable ER/EER Diagram.

12

5. 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.

12

Examination: Mid-Semester II(2015-16)
M.Tech I Year
Automata, Languages and Computations.

Date:25-02-16 Maximum Marks:80 Duration:2hours 40 mins.

Note: Answer as many as you can. Maximum score is 80
Unless otherwise stated, notation used is as defined in the class.

1. Define a non-deterministic finite automaton(NFA).
   Construct an NFA that accepts all strings over $\{a, b\}$ containing an even number of occurences of ab. Explain(without proof) your construction. Find its equivalent DFA. [3+8]

2. Construct a DFA that accepts all binary strings which are binary representations of (non-negative) integers which are multiples of 3.(No proof) [6]

3. Let $\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3$ be 3 DFAs accepting $\mathcal{L}_1, \mathcal{L}_2, \mathcal{L}_3$ respectively over the same alphabet. Construct a single DFA that accepts $(\mathcal{L}_1 - \mathcal{L}_2) \bigcup \mathcal{L}_3$. Explain your construction [8]

4. State Kleene's Theorem.
   Define regular expressions over an alphabet Σ.
   Let $\mathbf{r}$ and $\mathbf{s}$ be two regular expsessions. Prove the following identity

   $$(\mathbf{r} + \mathbf{s})^* = (\mathbf{r}^* + \mathbf{s}^*)^*.$$

   Consider the following equation in X

   $$X = \mathbf{r} + X.\mathbf{s},$$

   where $\mathbf{r}$ and $\mathbf{s}$ are regular expressions. Show that the equation has a solution. [2+3+7+7]

5. Write an algorithm to test, given a DFA $\mathcal{M}$, whether $\mathcal{L}(\mathcal{M}) = \phi$. Prove its correctness. [9]

6. Find a CFG that generates the set of all palindromes over the alphabet $\{a, b\}$. [6].

7. Show that if $\mathbf{G}$ is a positive CFG and $w \in \mathcal{L}(\mathbf{G})$ then there is a derivation tree for w in $\mathbf{G}$. [8]

8. When is a grammar said to be in Chomsky Normal Form(CNF)? Convert the following grammar into an equivalent grammar in CNF.

$$S \to aB/Ab;\ A \to aAB/a$$
$$B \to b/ABb$$

[2+6]

9. Given a regular language $\mathcal{L}$ write down a regular grammar that generates $\mathcal{L}$ (No proof required) [5]

10. State Bar-Hillel's Pumping Lemma.
Use it to show that $\{a^p : p \text{ is prime}\}$ is not context-free. [3+7]

Indian Statistical Institute

Semester-II 2015-2016

M.Tech.(CS) - First Year

Mid-term Examination (27 February, 2016)

Subject: Operating Systems

Total: 70 marks Maximum marks: 60 Duration: 3 hrs.

Please keep your answers brief and to the point.

1. (a) Consider the following sequence of steps used by the kernel to effect a context-switch from process P to P'.

> 1. Run the scheduling algorithm. Let P' be the selected process.
> 2. If $P == P'$, goto 5.
> 3. Save context of P.
> 4. Restore context of P'.
> 5. Return to user mode.

Now consider a context switch from process P'' to P. What problem arises if, in step 4, the context of P is restored faithfully using exactly the values saved in step 3 during the switch from P to P'? Your answer should clearly explain what happens at the individual machine instruction level.

(b) Explain what possible values can be returned by the `fork`, `execve`, and `exit` system calls. Your answer should include information about what process(es) can access these return values.

(c) Suppose a user starts the Firefox browser by running `$ firefox` from the command prompt in a terminal. While browsing, she clicks on a link to a PDF file, which is opened for viewing in a document viewer program.

(i) Draw a diagram showing the parent-child relationship between the proc structures corresponding to the terminal, browser and document viewer processes.

(ii) Explain how the diagram will change if the user now exits the browser.

[6+6+(2+4)=18]

2. Consider 3 processes, P_0, P_1 and P_2, that are created in that order at almost the same time. Assume that each of these processes has exactly 1 CPU burst (of duration d_0, d_1 and d_2 resp.), and no I/O bursts. Construct scenarios by choosing appropriate values of d_0, d_1 and d_2 ($\leq$ 10ms), and any other parameter if necessary, such that

(a) $T_F = T_R = T_S$;

(b) $T_F > T_R > T_S$;

(c) $T_R > T_F > T_S$;

(d) $T_R > T_F = T_S$;

where T_F, T_S and T_R are the average turnaround times for the 3 processes if they are scheduled using the First Come First Served, Shortest Job First, and Round Robin scheduling algorithms. Justify your answer in each case. [2+6+6+4=18]

3. The following questions all pertain to version 2.6 of the Linux kernel.

   (a) Explain clearly how the scheduler uses the static priority of a process.

   (b) What are the ranges of static priority values assigned to conventional and real-time processes?

   (c) Suppose you compile and execute the following program as an ordinary user. Assume the `get_static_prio()` is a function that returns the static priority of the calling process. Explain what output the program will produce.

```
int main() {
   nice(-5);
   printf("%d\n", get_static_prio());
   nice(10);
   printf("%d\n", get_static_prio());
   return 0;
}
```

[5+2+5=12]

4. (a) Consider the following proposed solution to the 2-process Critical Section Problem (CSP).

```
1  shared char want[2] = {0,0};
2  shared int turn = 0;
3  P_i() // i == 0 or 1
4  { while (1) {
5      want[i] = 1;
6      while (want[j] && turn!=i);
7      critical_section();
8      want[i] = 0;
9      turn = j;
10     remainder_section();
11   }
12 }
```

   Show that the above scheme does not guarantee mutual exclusion by constructing an appropriate interleaved sequence of instructions executed by two processes P_0 and P_1.

   (b) Suppose a processor provides a `Swap(x,y)` operation that interchanges the contents of two variables atomically. Show how you would use this to implement a solution to the n-process CSP that satisfies the mutual exclusion, progress and bounded waiting properties.

[10+12=22]

Examination: Semester II(2015-16)
M.Tech I Year
Automata, Languages and Computations.

Date: 18.04.16 Maximum Marks:100 Duration:3 hours

Note: Answer as many as you can. Maximum score is 100
Unless otherwise stated, notation used is as defined in the class.

1. Consider the following grammar **G**:

$$S \rightarrow aB/bA$$

$$A \rightarrow a/aS/bAA$$

$$B \rightarrow b/bS/aBB$$

Is the grammar ambiguous? Justify. [6]

2. (a) Explain in detail when a pushdown automaton (PDA) accepts a string by empty stack.

   (b) Show that if $\mathcal{L}$ is a context-free language then there is a PDA that that accepts $\mathcal{L}$.

   (c) When is a PDA said to be deterministic? Show that if $\mathcal{L}$ is a language accepted by a deterministic PDA by empty stack, then there exists a deterministic PDA $\mathcal{M}'$ that accepts $\mathcal{L}$ by final states and $\mathcal{L}$ has the prefix property, *i.e.*, no proper prefix of a string in $\mathcal{L}$ is in $\mathcal{L}$.

   (d) Show that if $\mathcal{L}$ is a context-free language and $\mathcal{R}$ is regular, then $\mathcal{L} \bigcap \mathcal{R}$ is context-free by constructing a suitable PDA.(Show only the major steps.)

   [5+7+9+6]

3. (a) Design (i)single-tape (ii) multitape Turing machines(TMs) that accept the set of all strings over $\{a, b\}$ containing an equal numbers of a's and b's. Compare the number of moves made by the respective TMs on accepting a string of length n.
   Also, construct a TM that uses only $\lceil \log n \rceil$ cells (not counting cells on the input tape) that accepts all binary strings in which the number of 0's is twice the number of 1's.

(b) A language $\mathcal{L}$ over Σ is said to be recursive if both $\mathcal{L}$ and $\Sigma^* - \mathcal{L}$ are recursively enumerable. Show that $\mathcal{L}$ is recursive iff there is a TM accepting $\mathcal{L}$ that halts on all inputs. [(10+7)+8]

4. (a) When is a function $f : \mathbb{N}^k \to \mathbb{N}$ said to be (Turing) computable? Show that the following functions and predicates are recursive.

i. $f(n,m) = [n/m] = \begin{cases} 0 & \text{if } m = 0 \\ \text{the largest integer} \leq n/m & \text{if } m > 0 \end{cases}$

ii. $Prime(n) \longleftrightarrow n$ is prime.

iii. $Divide(n,m) \longleftrightarrow m$ divides n.

iv. $p(n) =$ the nth prime number.

v. $lcm(n,m) = \begin{cases} 0 & \text{if } n = 0 \text{ or m} = 0 \\ \text{the least common multiple of } n,m & \text{otherwise} \end{cases}$

vi. $gcd(n,m) = \begin{cases} 0 & \text{if } n = 0 \text{ or m} = \\ \text{the greatest common divisor of } n,m & \text{otherwise} \end{cases}$

(b) Consider the following function.

$$B(x,y) = \begin{cases} 0 & \text{if } x = 0 \text{ or } y = 0 \\ x \dot{-} (\mu z \leq x(\exists w \leq x(w.(x \dot{-} z) = x) & \\ \quad \text{and } \exists u \leq y(u.(x \dot{-} z) = y))) & \text{otherwise} \end{cases}$$

Here "'$\mu z \leq xP(x)$"' means "'the smallest $z \leq x$ such that $P(x)$ holds"'. Describe in simple English the function B. [24+4]

5. (a) Define the classes $\mathcal{P}$ and $\mathcal{NP}$. When is a language said to be $\mathcal{NP}$-complete?

(b) Show that SAT is in $\mathcal{NP}$.

(c) Describe the CHROMATIC NUMBER problem. Assuming that 3-SAT is $\mathcal{NP}$- complete, prove that the CHROMATIC NUMBER problem is also $\mathcal{NP}$-complete. [6+4+10]

INDIAN STATISTICAL INSTITUTE

Semestral Examination: 2015-16

Course Name: M. Tech. I Year

Subject Name: Database Management Systems

Date: 19/04/2016 **Maximum Marks: 100** **Duration: 3hours**

Note: Answer all questions

---

1. a) Compare B Trees and B+ Trees.
   b) Create a B Tree of Order 5 with the sample keys:
   10, 30, 29, 13, 77, 16, 55, 79, 67, 21, 84, 56, 83, 99, 15, 78, 80, 86, 87, 82
   c) Now the following nodes are deleted from the above tree in the following order:
   21, 29
   Show the updated tree in each case.

7+7+6=20

2. a) Explain the difference of data file structure between Oracle and file based database.
   b) Answer the following (write the correct option):
   i) The minimal set of super key is called
   A. Primary key
   B. Secondary key
   C. Candidate key
   D. Foreign key
   E. All of the above
   ii) A relation that has no partial dependencies is in which normal form ?
   A. First
   B. Second
   C. Third
   D. BCNF
   E. All of the above
   iii) A functional dependency between two or more non-key attributes is called
   A. Transitive dependency
   B. Partial transitive dependency
   C. Functional dependency
   D. Partial functional dependency
   E. All of the above
   iv) Which of the following is true regarding Referential Integrity?
   A. Every primary-key value must match a primary-key value in an associated table
   B. Every primary-key value must match a foreign-key value in an associated table
   C. Every foreign-key value must match a primary-key value in an associated table
   D. Every foreign-key value must match a foreign-key value in an associated table
   E. All of the above
   v) Logical design of database may be called
   A. Database Instance
   B. Database Snapshot
   C. Database Schema
   D. Database file
   E. All of the above

15+5=20

INDIAN STATISTICAL INSTITUTE
First Semestral Examination : 2015-16

Course Name: M. TECH. (CS) I YEAR

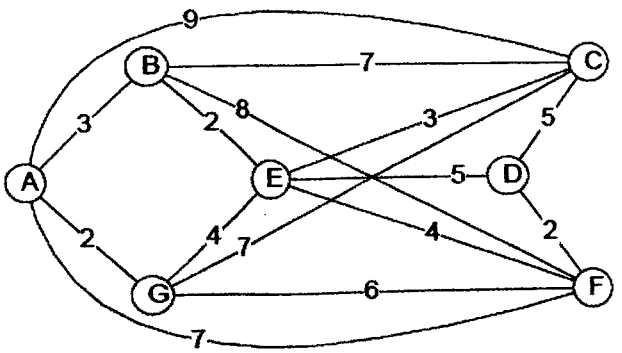
Subject Name: Computer Networks

Date: 22/04/2016

Maximum Marks: 100 Duration: 3 hours

Answer as much as you can.

1. a) Assume the network shown below uses a distance vector routing protocol and all nodes are in their initial state. Assume C sends an update to all its neighbours and then E sends an update to all its neighbours. Show the complete distance vector table of B in the state after it has received both updates. (7)



b) Dijkstra's algorithm gives you the shortest path from a source to a destination. Assume you want to get a second path (a backup path) that does not have any link in common with the shortest path (it can have routers in common). How can you extend Dijkstra's algorithm to give you this second path (if it exists)? (6)

c) An ISP is granted a block of addresses starting with 120.60.4.0/22. The ISP wants to distribute these blocks to 100 organizations with each organization receiving just eight addresses. Design the sub-blocks and give the slash notation for each sub-block. Find out how many addresses are still available after these allocations. (7)

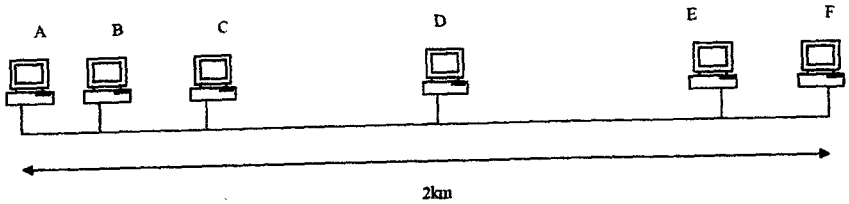
2. a) A system has n layer protocol hierarchy. User applications generate messages with M bytes. At each of the layers, an h-byte header is added. What fraction of the network bandwidth is wasted with headers for such message transmission? (2)

b) Consider a slotted ALOHA system having four stations. If the respective offer-loads for these stations are $G_1 = 0.1$, $G_2 = 0.5$, $G_3 = 0.2$, and $G_4 = 0.2$ packets per second, find the individual throughput rate for each station and the total throughput. (5)

c) A network with one primary and four secondary stations uses polling. The size of a data frame is 1000 bytes. The size of the poll, ACK, and NAK frames are 32 bytes each. Each station has a total of 5 frames to send. How many bytes, in total, are exchanged if there is no limitation on the number of frames a station can send in response to a poll? (5)

P.T.O

d) Consider the following linear network that is 2 km long.



Assume that propagation speed of electromagnetic waves sent over the medium is 2×10^8 m/s and the transmission rate of the network is 10 Mbps. What is the minimum frame size (in bits) necessary to ensure that CSMA/CD will work properly for this network? Explain your reasoning. (5)

e) If a binary signal is sent over a 5 kHz bandwidth channel whose signal to noise ratio is 25dB, what is the maximum achievable data rate? (3)

3. a) A supermarket has two girls ringing up sales at the customers. If the service time for each customer is exponential with mean 6 minutes and if the people arrive in a Poisson fashion at the rate of 12 per hour,
   i) What is the probability for having to wait for service?
   ii) What is the expected percentage of idle time for each girl? (3+3 = 6)

b) Suppose N packets arrive simultaneously to a link at which no packets are currently being transmitted or queued. Each packet is of length L and the link has transmission rate R. What is the average queuing delay for the N packets? (4)

c) Consider a router buffer preceding an outbound link. Let N denote the average number of packets in the buffer plus the packet being transmitted. Let a denote the rate of packets arriving at the link. Let d denote the average total delay (i.e., the queuing delay plus the transmission delay) experienced by a packet. Suppose that on average, the buffer contains 10 packets, and the average packet queuing delay is 10 msec. The link's transmission rate is 100 packets/sec. What is the average packet arrival rate, assuming there is no packet loss? (6)

d) Consider the delay in sending an x-bit message over a k-hop path in a circuit-switched network and in a (lightly loaded) packet switch network. The circuit set-up time is s seconds, the propagation delay is d seconds per hop, the packet size is p-bits, and the data rate is b bps. Derive the condition under which the packet network have a lower delay. (4)

4. a) The bit stream 10011101 is transmitted using the standard CRC method. The generator polynomial is x^3+1. Show the actual bit string transmitted. Suppose the third bit from the left is inverted during transmission. Show that this error is detected as the receiver's end. (3+3 = 6)

b) Consider an 8 bit block cipher. If we view each mapping induced by this cipher as a key, then how many possible keys does this cipher have? (2)

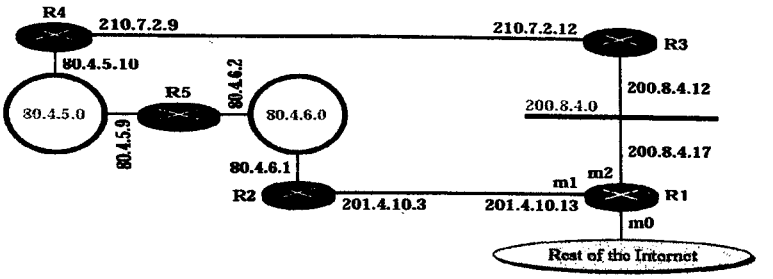
c) Suppose N people want to communicate with each of $N-1$ other people using symmetric key encryption. All communication between any two people, i and j, is visible to all other people in this group of N, and no other person in this group should be able to decode their communication. How many keys are required in the system as a whole? Now suppose that public key encryption is used. How many keys are required in this case? (4)

d) Consider RSA with $p = 5$ and $q = 11$ (symbols bear their usual meanings).

P.T.O

i. What are n and z?
ii. Let e be 3. Why is this an acceptable choice for e?
iii. Encrypt the message $m = 8$ using the key (n, e). Let c denote the corresponding ciphertext. Show all work. (2+2+4 = 8)

5. a) Prepare the routing table for Router R1 in the network shown below: (4)



b) Given the address 23.56.7.91 and the default class A mask, find the beginning address (network address). (3)

c) The address 43:7B:6C:DE: 10:00 has been shown as the source address in an Ethernet frame. The receiver has discarded the frame. Why? (3)

d) Suppose the length of a 10Base5 cable is 2500 m. If the speed of propagation in a thick coaxial cable is 200,000,000 m/s, how long does it take for a bit to travel from the beginning to the end of the network? Assume there are 10 μs delay in the equipment. (4)

e) Consider a network with address 150.46.0.0
i. What class of network does this address represent?
ii. It is required to have a total of 90 subnets in order to accommodate the WAN links and LAN segments at each site. The busiest LAN segment does not have more than 200 hosts on it. The IP routing protocol is RIP Version 1. What subnet mask should you use? (2+4 = 6)

6. a) What do you mean by a loopback address? What range of addresses in IPv4 is usually reserved for this purpose?. (2+1 = 3)

b) What are the differences between a master device in a Bluetooth network and a base station in an 802.11 network? (3)

c) If the bit string 0111101111101111110 is subjected to bit stuffing, what is the output string? (3)

d) A 3000 km long T1 trunk is used to transmit 64-byte frames using a Go-back-n protocol. If the propagation speed is 6 μs/km, how many bits should the sequence number be? (6)

e) A channel has a bit rate of 5 kbps and a propagation delay of 25 msec. For what range of frame sizes does stop-and-wait give an efficiency of at least 60%? (5)

INDIAN STATISTICAL INSTITUTE

Semestral Examination - 2016

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

Design and Analysis of Algorithms

Date : April 26, 2016 Maximum Marks : 100 Duration : 4 Hours

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

1. Given three sets of points A, B and C, each of size n, lying on three horizontal lines $y = 0$, $y = 1$ and $y = 2$ respectively. Design an $O(n^2)$ time algorithm to test whether there exists at least one triple (a, b, c), $a \in A$, $b \in B$ and $c \in C$ such that the points a, b and c are collinear. [10]

2. Suppose you are going on a trip to Manas-Sarovar starting from the base-camp along a path (there is no alternate sub-path from one point to another point along that path). You have a list of $n - 1$ possible camp-sites $\{1, 2, \ldots, n-1\}$ where you can camp in along the way (0 and n are considered to be the base-camp and Manas-Sarovar respectively). You want to do this trip in exactly K days, stopping K-1 nights to camp. The goal is to plan this trip so that you can minimize the maximum amount of walking done in a single day, say d. As an example, if the trip involves 3 days of walking, and you walk 11, 14, 12 miles on each day respectively, then the cost is 14. Another schedule that involves walking 11, 13, 13 miles on each day has cost 13. Thus, if only this two trips are feasible, then $d = 13$.

   The location of the camp-sites are specified in advance, and you can only camp at a camp-site. Thus, given a problem instance $(c_1, c_2, \ldots, c_n; K)$, where c_i is the distance of i-th camp-site from $i-1$-th camp-site, and K as defined earlier, design an $O(n^2K)$ time algorithm for outputing the distance d, such that on each day we walk at most d miles. Also, d should be the smallest possible value with this property. [15]

3.(a) State the Ford-Fulkerson's algorithm for computing maximum flow from the source vertex s to the sink vertex t in a directed graph where each edge is attached with a capacity.

(b) Show that the Ford-Fulkerson's algorithm may take infinite time to terminate based on the nature of its edge capacities.

(c) Show that if the Ford-Fulkerson's algorithm terminates, it will produce the maximum flow value from s to t. [10+7+8=25]

4.(a) In a *simple network* each node has either indegree 1 or outdegree 1.

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 and f is the value of the maximum flow.

(b) In the context of designing an algorithm for computing the maximum matching for a non-bipartite graph, the following result is important: *If at some stage of augmenting the matching, there is no augmenting path from a node u, then there will never be an augmenting path from u in the subsequent stages.* Prove this statement, and state why it is important.

[10+10 = 20]

5.(a) Consider two 2×2 matrices $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ and $B = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}$. The elements of the matrices A and B are real numbers. C is a 2×2 matrix obtained by multiplying A and B. Show that computation of C can be done using at most 7 multiplications. You may use as many addition/subtraction as required.

(b) Next, write a divide and conquer algorithm for multiplying two matrices of size $n \times n$ (n can be any arbitrary integer, not necessarily of the form 2^k) such that the time complexity of your algorithm is $O(n^{\log_2 7})$.

(c) Justify the time and extra space (apart from the space required for storing the input matrices) complexities of your algorithm. [6+6+(4+4)=20]

6.(a) 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.

6.(b) Show that if the edge costs of a complete graph G satisfy triangle inequality, then a traveling salesman tour of cost at most $1.5 \times opt$ can be found in polynomial time, where *opt* is the cost of the optimum tour of the given problem. [8+12=20]

7. Consider a special case of set cover optimization problem. The input is a universal set $S = \{x_1, x_2, \ldots, x_n\}$, and a set $\Sigma = \{S_1, S_2, \ldots, S_m\}$ of subsets of S, such that $\cup_{i=1}^{m} S_i = S$. Each element in S appears in at most two subsets of Σ. The objective is to choose the minimum number of subsets from Σ such that their union covers the entire set S.

Design a polynomial time 2-factor approximation algorithm for this version of set cover problem.

[10]

Indian Statistical Institute
Semester-II 2015-2016
M.Tech.(CS) - First Year
Semestral Examination (29 April, 2016)
Subject: Operating Systems

There are 4 questions in this paper, carrying a total of 110 marks. You may answer as many questions as you like. The maximum marks that you can score is 100.

Duration: 3.5 hrs.

Please keep your answers brief and to the point.

1. (a) Consider a multi-threaded application. For each of the following, state whether it is shared by all threads or not: (i) global variables, (ii) local variables, (iii) open files.

   (b) Suppose 2 threads within a process execute the following code. Assume that `A` and `i` are variables of the appropriate types.

   ```
   for (i=0; i < 100; i++) putchar(A[i])
   ```

   Each of `A` and `i` may be a local variable, or a global variable. This gives rise to 4 possibilities. For each of these possibilities, compute the minimum and maximum number of characters printed on the screen when the process executes the above code. Clearly justify your answer.

   If the same explanation applies in multiple cases, simply state this. You do not have to repeat your explanation.

   (c) Explain how asynchronous I/O can be useful for an implementation of a user-level threads library.

   [3 + 17 + 8 = 28]

2. (a) In a demand-paged system, it takes 100 nanoseconds to access memory. The page table has 8 entries and is held in registers. It takes 10 milliseconds to service a page fault if an empty frame is available or if the victim frame is not dirty. If the victim frame is dirty, the page fault service time is 20ms. Assuming that the victim frame is dirty 80% of the time, find the maximum page fault rate for which the effective memory access time remains within 200 nanoseconds.

   (b) Consider a demand-paged system where the degree of multiprogramming is currently fixed at four. The system is periodically monitored for utilization, and the following readings are obtained:

   - CPU utilization: 13%, disk utilization: 95%;
   - CPU utilization: 13%, disk utilization: 3%;
   - CPU utilization: 87%, disk utilization: 3%;

   Explain what is most probably happening in each situation. Also suggest a way to increase CPU utilization in the first two cases.

   (c) Consider the i386 memory management hardware.

   (i) How many segments can be present in a process' logical address space? How many of these can be used at a time?

   (ii) How many bits of each page table entry are needed to store the frame address for that page?

   [12 + 12 + (4+2) = 30]

3. (a) Consider an SVR2 filesystem with the following parameters:
   - the block size for this filesystem is 512 bytes;
   - each block number can be stored in 32 bits.

   (i) What is the theoretical upper limit on the size of a file in such a system?

   (ii) Compute the minimum and maximum number of disk blocks occupied by a file of size (A) 4,000 bytes (B) 16,000 bytes.

   (b) When a file is "created" using the `creat` system call, what permissions are checked in each of the following situations and why? (i) the file already exists; (ii) the file does not exist (i.e., it is *actually* created).

   (c) In the *umount* system call, how does the kernel determine if any file/directory on the mounted filesystem is currently in use?

[(4+(6+6)) + (4+4) + 4 = 28]

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

$$Total = [\ 5\ 6\ 7\ 6\ 5\ ]$$

$$Max = \begin{bmatrix} 5 & 4 & 3 & 2 & 1 \\ 1 & 3 & 0 & 3 & 2 \\ 0 & 1 & 2 & 1 & 3 \\ 2 & 0 & X & 0 & 4 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} \quad Alloc = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 \\ 2 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} \quad Req = \begin{bmatrix} 4 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 3 & 2 \\ 0 & 0 & 2 & 1 & 1 \\ 0 & 0 & 0 & 0 & 3 \\ 1 & 1 & 1 & 0 & 0 \end{bmatrix}$$

$Total[i]$ specifies the total number of instances of resource i that exists 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.

(a) Compute the minimum value of X for which the system will be in an unsafe state.

(b) Show that the system is not in a deadlocked state.

[14 + 10 = 24]

INDIAN STATISTICAL INSTITUTE

SEMESTRAL-I EXAMINATION
M.TECH.(CS) I YEAR

PROBABILITY AND STOCHASTIC PROCESSES
(BACK PAPER)

Date: 10.06.16 Maximum marks: 100 Duration: 3 hours

Answer all questions. Each question carries 10 marks.

1. Let A, B, C be three arbitrary events. Find expressions for the following events.
   (a) At least one of A, B, C occurs.
   (b) None of the events A, B, C occur.
   (c) Not more than two of the events occur.

2. Let X_1 and X_2 be independent random variables with Poisson distributions $p(k; \lambda_1)$ and $p(k; \lambda_2)$. Show that $X_1 + X_2$ has the Poisson distribution $p(k; \lambda_1 + \lambda_2)$.

3. There are three persons and you want to pick one such that the probability of any one of them getting chosen is 1/3. You are given a fair coin. Describe a method for doing this.

4. State the weak law of large numbers for a sequence of i.i.d. random variables having a common distribution. Assuming that the variance exists, prove the law.

5. Suppose a Markov chain has two states 0 and 1. With probability p_i (with $0 < p_i < 1$) the chain moves from state i to state $(1 - i)$ and with probability $(1 - p_i)$ it remains in state i. Find the stationary distribution of this Markov chain.

6. Suppose that in an election, a candidate P get p votes and another candidate Q gets q votes, with $p > q$. Show that the probability that throughout the counting there are always more votes for P than for Q equals $(p - q)/(p + q)$.

7. Let $\mathcal{E}$ be a recurrent event in a Markov chain and

$$\begin{aligned} u_n &= \mathsf{P}[\mathcal{E} \text{ occurs at the } n\text{th trial }]; \\ f_n &= \mathsf{P}[\mathcal{E} \text{ occurs for the first time at the } n\text{th trial }]. \end{aligned}$$

   Let $U(s)$ and $F(s)$ be the generating functions for $\{u_n\}$ and $\{f_n\}$. Show that

$$U(s) = 1/(1 - F(s)).$$

8. Define mean deviation from a measure of central tendency. Show that the mean deviation is minimum when taken about the median.

9. The frequency distribution of family-size for 250 families is given below.

family-size	frequency
1	4
2	22
3	25
4	45
5	52
6	41
7	36
8	15
9	7
10	3

Find the mean, median and the mode.

10. When is a statistic called sufficient for a parameter? Let $X_1, \ldots, X_n$ be i.i.d. Geometric(p with common p.m.f. $f(x;p) = p(1-p)^x$, $x = 0, 1, 2, \ldots$. Show that $\sum_{i=1}^{n} X_i$ is sufficient for p