INDIAN STATISTICAL INSTITUTE

Mid-Semester Examination: 2010-2011

M. Tech. (CS) 2^{nd} Year

Artificial Intelligence and Expert Systems

Date: 20.09.2010 Maximum Marks: 60 Duration: 2 hours

Answer all questions in brief.

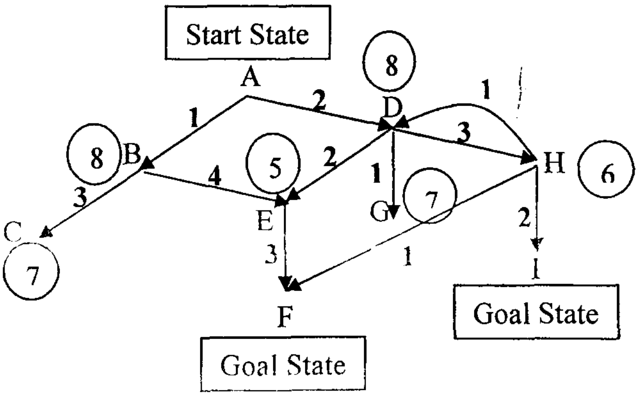
1. In *farmer-fox-goose-grain* puzzle, a farmer wishes to cross a river taking his fox, goose, and grain with him. He can use a boat which will accommodate only the farmer and one possession. If the fox is left alone with the goose, the goose will be eaten. If the goose is left alone with the grain it will be eaten. Draw a state space search tree for this puzzle using left-bank and right-bank to denote left and right river banks, respectively. [10]

2. What do you mean by *ridge* and *plateau*? Explain *simulated annealing* approach. What is the difference between *simulated annealing* and *steepest ascent hill climbing* approach? [1 + 1 + 6 + 2 = 10]

3. Prove that the set of states expanded by algorithm $A*$ is a subset of those examined by breadth first search. [5]

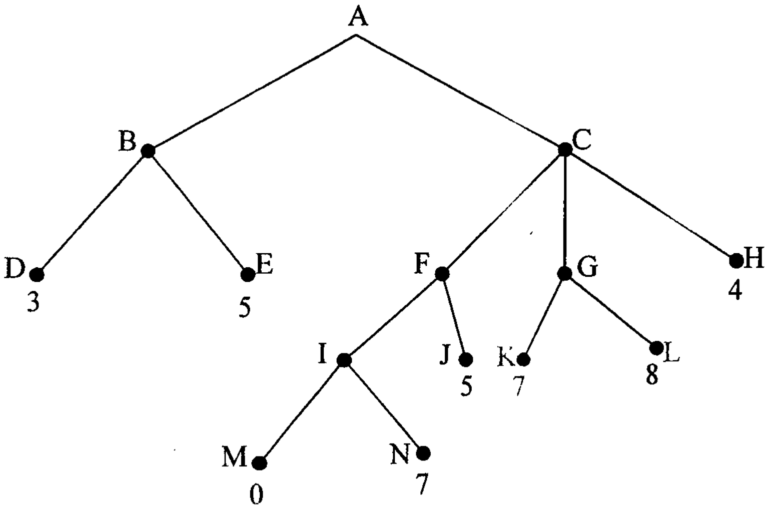
4. Define monotone property of a heuristic. Prove that any monotonic heuristic is admissible. [5 + 5 = 10]

5. Execute the *uniform cost search* and *best first search* algorithms on the following search graph, and show the solution path, along with its cost and list the expanded nodes for each case (each node of the graph is represented by a letter and the encircled value is the heuristic evaluation of the corresponding node, while the bolded numerical value represents the actual length of the path between two nodes). [5 + 5 = 10]



6. Perform the *minimax* search procedure on the game tree shown below in which static scores are all from the first player's point of view and MAX is allowed to move first. Perform the left-to-

right and right-to-left α-β pruning procedure on this tree and show how many nodes can b pruned away. Discuss why a different pruning occurs. [3 + 5 + 5 + 2 = 15



Indian Statistical Institute

M. TECH. (CS) 2 Year : 2010–2011
Mid-semester Examination
Subject: Cryptology

Date: 20/09/2010 Time: 2 hours Maximum Marks:40

Note: The paper carries 50 marks. Maximum you can score is 40.

1. State and prove Piling-up lemma. [10]

2. Prove that $|\mathsf{Prob}(l_u = f) - \frac{1}{2}| = \frac{|W_f(u)|}{2^{n+1}}$, where symbols have their usual meaning. [10]

3. Suppose a cryptosystem achieves perfect secrecy for a particular plaintext probability distribution. Prove that perfect secrecy is maintained for any plaintext probability distribution. [10]

4. suppose a sequence of plaintext blocks, $x_1, \cdots, x_n$ yields the ciphertext sequence $y_1, \cdots, y_n$. Suppose that one ciphertext block, say y_i, is transmitted incorrectly. Show that number of blocks that will be decrypted incorrectly is equal to one if ECB or OFB modes are used for encryption; and equal to two if CBC or CFB modes are used. [10]

5. Suppose we construct a keystream in a synchronous stream cipher using the following method. Let $K \in \mathcal{K}$ be the key, let $\mathcal{L}$ be the key stream alphabet, and let Σ be a finite set of states. First, an initial state $\sigma_0 \in \Sigma$ is determined from K by some method. For all $i \geq 1$, the state σ_i is computed from the previous state σ_{i-1} according to the following rule:

$$\sigma_i = f(\sigma_{i-1}, K),$$

where $f : \Sigma \times \mathcal{K} \rightarrow \Sigma$. Also, for all $i \geq 1$, the keystream element z_i is computed using the following rule:

$$z_i = g(\sigma_i, K),$$

where $g : \Sigma \times \mathcal{K} \rightarrow \mathcal{L}$. Prove that any keystream produced by this method has period at most $|\Sigma|$.

[10]

INDIAN STATISTICAL INSTITUTE

M. Tech (Computer Science) II year, 2010 – 11
Mid-Semester Examination
Pattern Recognition and Image Processing

Date: 22/09/10 **Maximum marks: 50** **Duration: 150 minutes** (3)

Note: Answer all the questions

1. State the Bayes decision rule for three-class classification problem and show that it minimizes the probability of misclassification. [2+10=12]

2. Let there be two classes C_1 and C_2 with prior probabilities P_1 and P_2, and class conditional density functions p_1 and p_2. Let λ_{ij} denote the cost of classification of an observation from class *i* to class j, *i,j=1,2.* Let $\lambda_{ij} > \lambda_{ii} > 0 \; \forall i = 1,2$. Derive the Bayes decision rule for minimum risk in this set up. [10]

3. (i) State the minimum within cluster distance criterion.
   (ii) Describe the k-means algorithm for clustering.
   (iii) Give an example of a data set and two of its initial partitions for which the resultant clustering would be different when k-means algorithm is applied.
   [4+4+3=11]

4. Write short notes on the following.
   (i) Training and test sets.
   (ii) Estimation of parameters for normal distribution and estimation of error probability for a decision rule.
   (iii) Measures of dissimilarity between sets for agglomerative clustering procedures.
   [4+4+4=12]

5. (i) State the K-nearest neighbor decision rule for pattern classification.
   (ii) Describe the minimum distance classifier for pattern classification.
   [3+2=5]

Indian Statistical Institute

Semester-1 2010-2011

M.Tech.(CS) - Second Year

Mid-semester Examination (27 September, 2010)

Subject: Compiler Construction

Maximum marks: 40 Duration 3 hrs.

Please keep your answers brief and to the point.

1. (a) Function headers in C consist of a return type, a function name, and an argument list (possibly empty) contained in parentheses. An example function header is given below:

```
int somename(char c, int n)
```

Assuming that the only permitted return types are `char`, `int`, and `void`, and the only permitted argument types are `char`, `float`, and `int`, write a regular definition for function headers in C. (If your definition is incorrect, you will get partial credit if you can provide regular definitions for the various "components" of a function header.)

(b) A *mailbox* file consists of several email messages one after another. Assume the following.

- Each email message consists of a header and a body.
- The header and body are each followed by a single blank line.
- The header consists of *From*, *To*, *Subject*, *Date* and other fields. Each field in the header occupies a single line, and starts with the name of the field, immediately followed by a colon (:), a single space, and the contents of that field. Some example header fields are shown below:

```
From: pcm@isical.ac.in
From: "S.N. Bose" <satyen@gmail.com>
Subject: IMPORTANT: mid-sem exam cancelled!
Subject: notice
Date: Fri Sep 24 17:00:00 IST 2010
```

Write a Lex programme that takes a mailbox file as input, and prints a summary of the mailbox contents on the screen. The summary consists of the From, Subject and Date fields for each message, along with the number of lines in the message's body. The summary for a mailbox containing two messages is given below:

```
From: pcm@isical.ac.in
Subject: important
Date: Fri Jan  1 09:10:00 IST 2010
Lines: 10
From: "S.N. Bose" <satyen@gmail.com>
Subject: Re: E=mc^2
Date: Thu Oct  5 20:10:00 IST 2000
Lines: 982
```

Hint: In Lex, . is a regular expression that matches any single character other than the newline.

Note: You may make reasonable assumptions where necessary. **Clearly state any assumptions you make.**

Warning: Lex always tries to find the longest possible match for any given pattern.

[4+10=14]

P.T.O.

2. Consider the following grammar G:

$$\begin{array}{lllllllllllllll} E & \rightarrow & L(E) & \quad E & \rightarrow & \&L & \quad E & \rightarrow & *E & \quad E & \rightarrow & L.L & \quad E & \rightarrow & L \\ L & \rightarrow & L[E] & \quad L & \rightarrow & \mathbf{id} & & & & & & & & & \end{array}$$

E is the start symbol; L is a non-terminal; all other symbols are terminals.

(a) Eliminate left-recursion from the grammar if necessary.

(b) Perform left-factoring if necessary.

Let the grammar obtained after the above steps be G'.

(c) Compute the *FIRST* set for the right hand side of each rule of G'.

(d) Compute the *FOLLOW* set for each non-terminal of G'. Show your rough work.

(e) Construct the $LL(1)$ parsing table for G'.

(f) Is G an $LL(1)$ grammar? Is G' an $LL(1)$ grammar? Justify your answer.

(g) Construct the canonical collection of $LR(0)$ (i.e. *SLR*) items for the original grammar G.

[2+2+4+5+4+2+12=31]

INDIAN STATISTICAL INSTITUTE

Backpaper Examination

M. Tech. (CS) II year: 2009–2010

Information and Coding Theory

Date: 27. 09. 2010 Marks: 100 Time: 3 Hours

Note: Part A and Part B should be answered in separate answerscripts.

Part A

1. Define linear and nonlinear code. Give example of each. [4+6=10]

2. State and prove the following bounds

   (a) Sphere-packing bound.

   (b) Plotkin's bound.

   [10+10=20]

3. Construct a 2-error correcting BCH code for n=15. How would you decode it? [15+5=20]

Part B

1. Let X and Y be random variables where:

$$X = \begin{cases} 0 & \text{with probability } \frac{3}{4} \\ 1 & \text{with probability } \frac{1}{4} \end{cases}$$

$$Y = \begin{cases} 0 & \text{with probability } \frac{1}{2} \\ 1 & \text{with probability } \frac{1}{2} \end{cases}$$

   Find $H(X)$ and $H(Y)$. [5+5=10]

2. State and prove the following

   (a) Chain Rule of entropy.

   (b) Shearer's Lemma.

   [10+10=20]

3. Write encoding and decoding algorithm of Huffman Code. [15]

4. Define Binary Symmetric Channel. [5]

INDIAN STATISTICAL INSTITUTE

Mid-Semester Examination : (2010-2011)

M.Tech.(CS) II Year

Database Management Systems

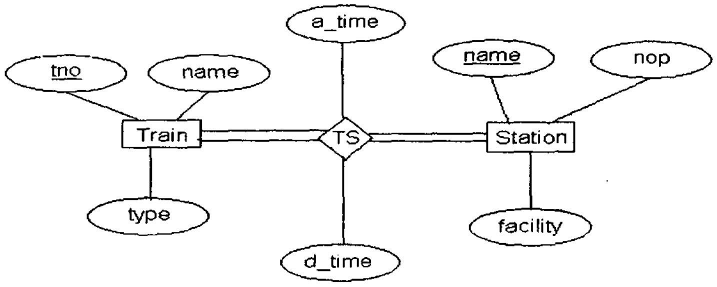
Date: 29.09.2010 **Maximum Marks: 40** **Duration: 2 Hrs.**

1. A relation R(A, B, C, D, E, F, G) has been decomposed as R1(A, B, C, D, E) and R2 (C, D, E, F, G) against the following Functional Dependencies :
{A→BC, CD→E, B→EF, CDE→F}

    a) Explain whether the relations R1 and R2 are dependency preserving.
    b) Justify that the above two relations are not in BCNF.
    c) Derive the possible candidate keys for R.
    d) Decompose the relation R to a set of normalized relations under BCNF using the given set of functional dependencies. If it is not possible to derive the relations in BCNF, derive the same using 3NF.
    e) Now, if a multivalued dependency A→→G is introduced, suggest the new set of normalized relations.

(3+3+5+6+3=20)

2. Find the errors in the ER diagram shown below and rectify them.



The key attributes are underlined. 'tno' signifies 'train number', ' a_time' and 'd_time' are the arrival and departure time of a train, 'nop' is the 'number of platforms' in a station. Each station may have different facilities like, refreshment room, retiring room, cloak room etc, shown by the attribute 'facility'. All stations may not have all the facilities. So, the attribute 'facility' is multivalued. Station name is unique, but the train name may not be unique.
Arrival time of the originating station of a train and the departure time of the final destination of a train are assigned with the value 9999. In all other cases, value of time is represented in terms of 24 hours.

a) Draw the corrected ER diagram.
b) To design a relational schema, derive a set of relations using the standard mapping rules from the corrected ER diagram.
c) Using any relational calculus, find the name of the originating station and the final destination of the train number 1054.
d) Using relational algebra, find the list of stations (provide the station name only) that do not have any retiring room facility.

(4+5+5+6=20)

-x-

ER to Relational Mapping

1. For each strong entity set E create a relation R transferring all attributes of E as the attributes of R. Primary key of E becomes the primary key of R. Each composite attribute in E is broken into its component atomic attributes before transferring to R.

2. For each weak entity set W create a relation R transferring all attributes of W as the attributes of R. Since W does not have a primary key, the primary key of its owner strong entity is borrowed as foreign key and added to R.

3. For each binary 1:1 relationship between two strong entity sets E1 and E2 mapped to relations R1 and R2 respectively, add the primary key of anyone of the two relations R1 or R2 as a foreign key to the attribute set of the other relation.

4. For each binary 1:N relationship between two strong entity sets E1 and E2 mapped to relations R1 and R2 respectively, if E2 is on the N-side of the relationship, the primary key of R1 is added as the foreign key to the attribute set of R2.

5. For each binary M:N relationship between two strong entity sets E1 and E2 mapped to relations R1 and R2 respectively, add the primary key of each relation to the attribute set of the other. The primary keys of both the entity sets E1 and E2 will jointly form the primary key of both R1 and R2.

6. Map each n-ary relationship X among n participating entity sets to a separate relation R, adding the primary keys of all the participating entity sets as the attributes of R. They jointly form the primary key of R. Even in case of a binary relationship Y that has some attributes of its own, create a separate relation R' with the attributes associated with Y and add the primary keys of the participating entity sets to R'. These primary keys jointly form the primary key of R'.

7. For each multivalued attribute A present in the entity set E, create a separate relation R with A as an attribute and also add the primary key K of E to R. K and A together will form the primary key of R.

INDIAN STATISTICAL INSTITUTE

Periodical Examination

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

Advanced Algorithms for Graph and Combinatorial Optimization Problems

Date: 1.10.2010 Maximum Marks: 60 Duration: 3 Hours

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

1. For any simple planar graph G with n vertices and e edges, prove the followings:

   (a) There is a vertex of G that has degree at most 5.

   (b) If G does not contain a cycle of length 3 then $e \leq 2n - 4$.

   [5+3=8]

2. Let $G = (V, E)$ be a comparability graph, and a transitive orientation of G is given. Define layering of the comparability graph as follows:

   The vertex v with indegree 0 is assigned in layer $\ell(v) = 1$. A vertex $u \neq v$ is assigned in layer $\ell(u) = 1 + max_{w \in adj(u)} \ell(w)$, where $adj(u)$ is the set of predecessors of u

   (a) If m layers are needed for layering all the vertices of the graph, then what is the size of the maximum clique of the graph G ?

   (b) What is the size of the maximum independent set of the graph G ?

   (c) How will you compute all maximal cliques of the graph G ? [4+6+5=15]

3. Define *tree width* of a graph $G = (V, E)$. (You may need to define *tree decomposition* of G, and its width.)

   If w is the tree width of G, then show that the vertex cover of G can be computed in $O(2^w \times w \times |V|)$ time. [5+8=13]

4.(a) Define *simplicial vertex* and *perfect elimination order.*

(b) Show that every triangulated graph $G = (V, E)$ has a simplicial vertex. Also show that, if G is not complete, then it has at least two non-adjacent simplicial vertices.

(c) Show that the perfect elimination order of the vertices of a triangulated graph can be found in time linear in the number of edges of the graph. [4+7+5=17]

5. 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 can not exceed $\frac{|V|}{f}$, where V is the set of vertices in the network and f is the value of the maximum flow.

   Show that this result helps in computing the maximum matching in a bipartite graph $G = (V, E)$ in time $O(\sqrt{|V|} \times |E|)$. (Except the desired results, you need not have to prove the intermediate results, but explicit statement of the required intermediate results are necessary.) [10+12 = 22]

INDIAN STATISTICAL INSTITUTE

Periodical Examination, Semester I, 2010

M. Tech. (CS) - II

VLSI Design and Algorithms

Date: Oct. 1, 2010 **Time : 2.5 hours** **Maximum Marks : 60**

Answer as many as you can. The maximum you can score is 60.

1. (a) State the circuit bi-partitioning problem formally along with typical constraints and objectives.
(b) Classify the various methods to solve this problem and comment on their effectiveness.
(c) Let *P* be a path graph of *n* nodes where v_1 and v_n have degree *1* and v_i is connected to v_{i-1} and v_{i+1} for $i = 2, 3, n-1$. Trace the steps of Kernighan-Lin algorithm applied to *P*, where the initial partition consists of all the nodes with odd indices in one partition and those with even indices in the other.
[4 + 4 + 6 = 14]

2. (a) What are the various layout design styles? Give a comprehensive comparison among them.
(b) Present four different wire estimation models and demonstrate with a small grid-based placement instance whether the wirelength calculation varies.
c) Name a type of placement method for which the step of terminal propagation required. Why?
[4 + 8 + 4 = 16]

3. (a) What are the necessary conditions under which an adjacency graph admits a rectangular dual?
(b) For an inherently nonslicible adjacency graph of 9 vertices,
    1. illustrate the steps of rectangular dualization on.
    2. what is the time complexity?
    3. give a corner block list representation for the floorplan obtained.
[3 + (8 + 2 + 5) = 18]

4. For the problem of global routing,
    e. give formal definition
    f. justify with an example or an counterexample as the case may be for the following statements:
        v. Lee's maze router always produces the shortest path.
        vi. Mikami-Tabuchi's line-probe router may not find a path even if one exists.
[4 + (4+4) = 12]

5. Sketch how to solve the problem of optimal floorplan sizing for a slicible floorplan topology. Argue its worst case time complexity.
[10]

INDIAN STATISTICAL INSTITUTE

Mid-Semester (Supplementary) Examination : (2010-2011)

M.Tech.(CS) II Year

Database Management Systems

Date: 12.10.2010 **Maximum Marks: 40** **Duration: 2 Hrs.**

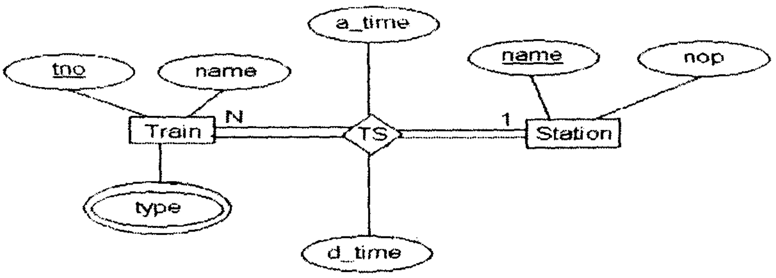
1. Consider the following schema and form the required queries:
Suppliers (sid, sname, address)
Parts (pid, pname, color)
Catalog (sid, pid, price)
The key fields are underlined. Attributes sid and pid indicate the unique ids for a supplier and a part respectively.
The queries are:
a) List the name and address of such suppliers who can supply red carpets. (USE ANY RELATIONAL CALCULUS).
b) List the name of such parts supplied by "XYZ Co" where the price is greater than Rs.500 but less than Rs.1000. (USE RELATIONAL ALGEBRA).

(5x2=10)

2. A relation R(A,B,C,D,E) has the following functional dependencies.
{A →B, BC→E, E→A}
Considering the given set of dependencies only,
i) Derive all the possible candidate keys not more than three attributes long.
ii) Justify that R cannot be decomposed into BCNF.
iii) Decompose R into 3NF.
iv) If a new multivalued dependency (E->->D) is now added, derive the new set of normalized relations.

(6+3+3+5=17)

3. Find the errors in the ER diagram shown below and rectify them.



The key attributes are underlined. 'tno' signifies 'train number', ' a_time' and 'd_time' are the arrival and departure time of a train, 'nop' is the 'number of platforms' in a station. Station name is unique, but the train name may not be unique. There are different types of trains like, LOCAL, PASSENGER, EXPRESS, MAIL etc. However, one train can be of only one type. A train will travel through many stations and many trains may stop in one station.

a) Draw the corrected ER diagram.
b) To design a relational schema, derive a set of relations using the standard mapping rules from the corrected ER diagram.

(5+8=13)

-X-

ER to Relational Mapping

1. For each strong entity set E create a relation R transferring all attributes of E as the attributes of R. Primary key of E becomes the primary key of R. Each composite attribute in E is broken into its component atomic attributes before transferring to R.

2. For each weak entity set W create a relation R transferring all attributes of W as the attributes of R. Since W does not have a primary key, the primary key of its owner strong entity is borrowed as foreign key and added to R.

3. For each binary 1:1 relationship between two strong entity sets E1 and E2 mapped to relations R1 and R2 respectively, add the primary key of anyone of the two relations R1 or R2 as a foreign key to the attribute set of the other relation.

4. For each binary 1:N relationship between two strong entity sets E1 and E2 mapped to relations R1 and R2 respectively, if E2 is on the N-side of the relationship, the primary key of R1 is added as the foreign key to the attribute set of R2.

5. For each binary M:N relationship between two strong entity sets E1 and E2 mapped to relations R1 and R2 respectively, add the primary key of each relation to the attribute set of the other. The primary keys of both the entity sets E1 and E2 will jointly form the primary key of both R1 and R2.

6. Map each n-ary relationship X among n participating entity sets to a separate relation R, adding the primary keys of all the participating entity sets as the attributes of R. They jointly form the primary key of R. Even in case of a binary relationship Y that has some attributes of its own, create a separate relation R' with the attributes associated with Y and add the primary keys of the participating entity sets to R'. These primary keys jointly form the primary key of R'.

7. For each multivalued attribute A present in the entity set E, create a separate relation R with A as an attribute and also add the primary key K of E to R. K and A together will form the primary key of R.

Indian Statistical Institute

M. TECH. (CS) 2nd Year : 2010–2011
Semester Examination
Subject: Cryptology

Date: 04/12/2010 Time: 3 hours Maximum Marks:100

Note: The paper carries 110 marks. Maximum you can score is 100.

1. Give the definition of random oracle model. Prove that a collision resistant hash function is second preimage resistant. [5 + 10 = 15]

2. Give the algorithm for Merkle-Damgard construction. Prove that Merlkle-Damgard construction gives collision resistant hash function when compression function used in it is collision resistant. [7 + 13 = 20]

3. Define RSA cryptosystem. Explain clearly how the different parameters in RSA are selected. Prove that RSA is not semantically secure. [5 + 10 + 5 = 20]

4. Describe Shank's algorithm to find discrete log. Define a group where discrete log problem is easy. [12 + 3 = 15]

5. Let E be a nonsingular elliptic curve. Define addition on the set E. Show that $(E, +)$ is an abelian group assuming addition is associative. [8 + 12 = 20]

6. How RSA can be used as signature scheme? Describe attack models and goal of adversaries on a signature scheme. Prove that there is a selective forgery using a chosen message attack on RSA signature scheme. [5 + 10 + 5 = 20]

INDIAN STATISTICAL INSTITUTE

First-Semester Examination: 2010-2011

M. Tech. (CS) 2nd Year

Artificial Intelligence

Date: 04.12.2010 Maximum Marks: 100 Duration: 3 hours

Answer all questions in brief.

1. Answer the following:
   a) How do you convert a clause of first order predicate logic into clausal representation of logic programming?
   b) Explain the difference between red cut and green cut in Prolog.
   c) Describe the difference between the following two codes written in Prolog when the goal query is "grandfather(james, X)":

grandfather(X,Y):- father(Z, Y), father(X, Z). father(james, robert). father(mike, william). father(william, james). father(robert, hency).	grandfather(X,Y):-father(X, Z), father(Z, Y). father(james, robert). father(mike, william). father(william, james). Father(robert, hency).

   d) Write a program in Prolog for post-order traversal of a binary tree. The traversal method stores the elements of the tree in a list. [4 + 4 + 5 + 7 = 20]

2. Prove that any monotonic heuristic is admissible. Describe the crossover and mutation operators of genetic algorithm. [4 + 6 = 10]

3. Consider the following set of sentences: "Mary will get her degree only if she registers as a student and passes her examination. She has registered herself as a student. She has passed her examination." Prove using both semantic tableaux approach and resolution refutation method that "she will get a degree". [5 + 5 = 10]

4. Answer the following:
   a) Prove that a clause C is a logical consequence of a set of clauses S if and only if the set $S' = S\ U\ \{\sim C\}$ is unsatisfiable.
   b) Prove using semantic tableaux that the following sentences are mutually inconsistent: "All musicians are singers. A teacher is not a singer. Mary is a teacher. Mary is a musician."
   c) What are the main features of an expert system? Explain with an example the backward chaining approach in a rule based expert system. [6 + 6 + (4 + 4) = 20]

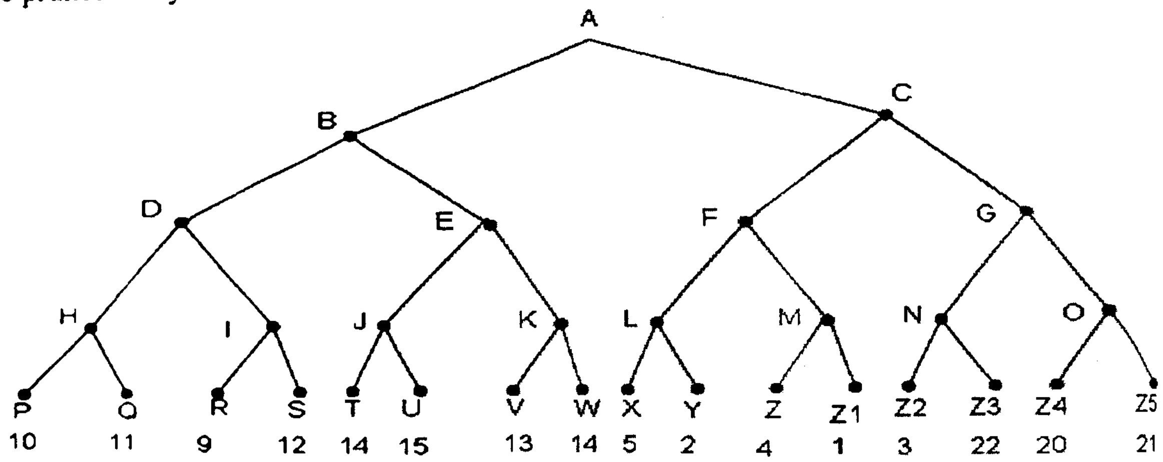
5. Answer *any two* from the following: [2 X 10 = 20]
   a) Discuss the Bayes' theorem for the probabilistic reasoning. Suppose an initial observation S_1 confirms some hypothesis h with the belief MB = 0.3. The second observation S_2 performs the same hypothesis h with the belief MB = 0.5. Find the certainty factor regarding the hypothesis h by the two observations S_1 and S_2. [6 + 4 = 10]
   b) Describe the following with suitable examples:
      (i) Semantic networks;
      (ii) Means-ends analysis. [5 + 5 = 10]

P. T.O.

c) Explain with suitable examples the differences between:
(i) Justification-based and assumption-based truth maintenance systems;
(ii) Uniform cost search and best first search. [5 + 5 = 10]

6. Answer *any two* from the following: [2 X 10 = 20]

a) Perform the minimax search procedure on the game tree shown below in which the static scores are all from the first player's point of view and MAX is allowed to move first. Perform the left-to-right α-β pruning procedure on this tree and show how many nodes can be pruned away. [4 + 6 = 10]



b) Define the constraint satisfaction problem and solve the following cryptarithmetic problem:

```
  S E N D
 +M O R E
---------------------------
 M O N E Y
---------------------------
```

[2 + 8 = 10]

c) Consider a sliding block puzzle with the following initial configuration:

W	W	W	B	B	B	E

There are three white tiles (W), three black tiles (B), and an empty cell (E). The puzzle has the following moves:

(i) A tile may move into an adjacent empty cell with unit cost.

(ii) A tile may hop over at most two other tiles into an empty cell with a cost equal to the number of tiles hopped over.

The goal of the puzzle is to have all of the black tiles to the left of all of the white tiles without regard for the position of the empty cell. Define the problem as a state space graph problem and find a sequence of moves that will transform the initial configuration to a goal configuration. What is the cost of the solution? [4 + 6 = 10]

INDIAN STATISTICAL INSTITUTE

Semestral Examination, Semester I, 2010
M. Tech. (CS) - II
VLSI Design and Algorithms

Date: Dec. 6, 2010 **Time : 3 hours** **Maximum Marks : 100**

<u>*PART I :*</u> *Answer all questions. For each question, the correct choice from the four options along with either an example or a very brief justification carries 2 marks*

1. What view of a VLSI system corresponds to the phase of high level synthesis?

A. Structural
B. Geometric
C. Behavioral
D. None of the above

2. For a given circuit, which layout design style can produce the fastest chip speed?

A. Full custom
B. Gate arrays
C. Standard cells
D. FPGAs

3. The problem of finding a rectangular dual of a given planar triangulated graph can be solved in

A. $O(n)$ time
B. $O(n \log n)$ time
C. $O(n^2)$ time
D. $O(2^n)$ time

4. The basic layout design rule for the minimum separation between two wires on diffusion region is

A. 2λ
B. 3λ
C. 1.5λ
D. 1λ

5. Which of the following problems is solvable in $O(n \log n)$ time ?

A. Left-edge algorithm
B. Balanced min-cut bi-partitioning
C. Static list-based resource-constrained scheduling
D. Concurrent global routing for all nets by integer linear programming

P.T.O.

PART II: Answer as many as you can but the maximum you can score is 90.

6. Consider the following channel with 12 columns:
   TOP = < 3 4 0 1 2 4 3 5 2 1 0 5 >
   BOT = < 1 0 3 0 4 0 5 2 1 5 4 0>

   (a) Draw its (i) horizontal constraint graph, and (ii) vertical constraint graph.
   (b) Compute the lower bound on the number of tracks required.
   (c) Trace the steps of Yoshimura-Kuh net-merging algorithm to obtain its detailed routing.
   (d) If over-the-cell routing is permitted, show the steps to solve the multi-terminal single-layer one-sided routing on the TOP row of the channel.

   [4 + 2 + 8 + 8 = 22]

7. (a) State the clock routing problem in ASICs, and sketch an algorithm to solve it.
   (b) Explain why the floorplanning algorithms for ASICs are not applicable for floorplanning in modern FPGAs with heterogeneous resources.
   (c) Give the formulations for (i) global routing and (ii) detailed routing problems in the case of FPGAs.

   [(3 + 4) + 4 + (5+5) = 21]

8. (a) Draw the DFG for the following computation:
   w := d - e; x := a * b * (c-d); y := c + x*a; z := w + e/y
   (b) What is the minimum number of time steps required? State any assumptions made.
   (c) Give its ILP formulation for time-constrained scheduling
   (d) For the above schedule, derive the minimum number of registers required.

   [6 + (2+2) + 7 + 6 = 23]

9. (a) Show two possible layouts for an 8:1 nMOS inverter and comment on the respective power dissipation when the pull-down transistor is ON.
   (b) Give a CMOS domino logic transistor level circuit diagram for a 3-input NOR gate and explain briefly how it works.
   (c) Draw an arrangement of pass transistors to realize a 2 x 2 crossbar switch.

   [(4+4) + 5 + 5 = 18]

10. (a) Illustrate with a small example the notion of fault dominance for stuck-at fault model.
    (b) Given a combinational circuit, under what conditions do each of the following hold:
    (i) *any* complete test set for single stuck-at faults also detects all multiple stuck-at faults;
    (ii) *any* complete test set for single stuck-at faults also detects all double and triple faults.
    Justify your answers briefly.
    (c) What is an ROBDD? Why do you have the requirement of ordering of variables?
    State the 3 reduction rules that convert an OBDD to a ROBDD and demonstrate their application on an example Boolean function.
    (d) Provide the *APPLY* algorithm for Boolean manipulation of two ROBDDs. Explain the working of your algorithm by showing the stepwise execution on an example.

    [4 + (3 + 3) + (1 + 1 + 3) + (2 + 3) = 20]

INDIAN STATISTICAL INSTITUTE

Semestral Examination

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

Advanced Algorithms for Graph and Combinatorial Optimization Problems

Date: 6.12.2010 Maximum Marks: 100 Duration: 3.5 Hours

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

1. Let n numbers be stored in a *read-only* array. You need to find the *median* of those numbers.

   (a) How fast you can find the median if constant amount of workspace is allowed?

   (b) If $O(k)$ space is given, show that the time complexity can be reduced to $O(n^{1+\frac{1}{k+1}})$.

   [10+12=22]

2. The *integer knapsack* problem is defined as follows:

   Let $(c_1, c_2, \ldots, c_n, K)$ be an instance of the integer knapsack problem. The objective is to identify a sequence of integers $x_1, x_2, \ldots, x_n$ such that $\sum_{i=1}^{n} c_i x_i = K$.

   Show that, any instance $(c_1, c_2, \ldots, c_n, K)$ of the *integer knapsack* problem can be solved in $O(nK)$ time. Why this algorithm is called a pseudo-polynomial time algorithm?

   [10+2=12]

3.(a) Design a 2-factor approximation algorithm for the minimum steiner tree problem of an edge-weighted undirected graph some of whose vertices are marked as terminal vertices.

(b) When a problem is said to admit a *fully polynomial time approximation scheme*?

[8+6=14]

4.(a) What do you mean by a perfect hashing ? Describe a method of perfect hashing using $O(N)$ space, where N is the number of key values present in the table.

(b) Use the hashing technique to design an expected $O(n)$ time randomized algorithm for finding the closest pair of points among a set of n points present on a 2D plane.

[10+10=20]

5.(a) State the *perfect graph* theorem. You must explicitly define all the parameters involved in the theorem.

(b) Let G be a connected undirected graph, S be a vertex separator in G, and $G_{A_1}, G_{A_2}, \ldots, G_{A_t}$ be the connected components of G_{V-S}. If S is a clique then prove that

(i) $\chi(G) = \max_{i=1}^{t} \chi(G_{S+A_i})$,

(ii) $w(G) = \max_{i=1}^{t} w(G_{S+A_i})$,

(iii) If each subgraph G_{S+A_i} is perfect, then G is perfect.

[5+(6+4+10)=25]

P.T.O.

6.(a) Consider a Monte Carlo algorithm $\mathcal{A}$ for a problem Π whose expected running time is at most $T(n)$ on any instance of size n, and produces a correct answer with probability $\gamma(n)$. Suppose further that, given a solution of Π, the checking of its correctness can be done in $t(n)$ time. Show how to obtain a Las Vegas algorithm for the problem Π that always produces a correct answer in expected time $O(\frac{T(n)+t(n)}{\gamma(n)})$.

(b) The edge contraction method for computing the mincut of an weighted undirected graph $G = (V, E)$ is as follows.

Let $(x, y) \in E$ be contracted to get the graph $G \setminus xy$. The vertices of the resulting graph is $V \setminus \{x, y\} \cup xy$, where xy is a single vertex. All the edges incident to x and y in G are incident to the vertex xy in the graph $G \setminus xy$. More than one edge incident to a vertex is replaced by a single edge with sum of weights of the original edges as the weight of the new edge. Self loop is not removed.

Write an algorithm for finding a cut of the graph G using the aforesaid contraction procedure.

Show that if G is unweighted and in each stage (when needed in your algorithm) the edge to be contracted is chosen randomly, then you can find a mincut with constant probability (You may choose your own constant value < 0.5, but have to justify your answer). State the time complexity of your algorithm.

[5+12=22]

Indian Statistical Institute

Semester II Examination 2010

M. Tech. (CS) - Second year

Subject: Internet & Multimedia Technologies

Date: 7.12.10 Full Marks: 50 Duration: 2:00 hrs.

GROUP - A: INTERNET TECHNOLOGIES (Max. Marks: 25)

1.What are the distinguishing features of E-Commerce over the traditional commerce? What is the main issue towards E-Commerce risks? [2 + 1]

2. What is data mining? How does data mining techniques help in medical sciences? Name two fields of study, which frequently play important role in the formulation of algorithms and architectures of algorithms for data mining. [1 + 1 + 2]

3. Write down the basic differences between predictive and descriptive models of data mining. What factor does help a neural network to generate nonlinear boundaries of decision regions in the decision space? [2 + 2]

4. Probability of a fraudulent credit card transaction in Kolkata is 0.005. A security algorithm predicts fraudulent transaction correctly in 99.5% of the cases. It also predicts correct 99% of the time when it decides no fraud. If for some random transaction, the algorithm decides that the transaction is a fraud one, what is the probability that the decision is wrong? [5]

5. What are the functions of a web filter? What is the Robot Exclusion Standard? What are the factors determining architecture of a search engine? [2 + 2 + 2]

6. What do you know about "Content-Transfer-Encoding"? [2]

7. Write down the physical structure of a .java source file. What is required to run an applet? What is a runnable applet? [2 + 1 + 1]

GROUP - B: Multimedia (Full Marks: 25)

1. Write short answer:
   (a) Give expression for Signal to quantization noise ratio
   (b) Give expression for Sampling interval according to Nyquist theorem
   (c) Define I-frame and P-frame in video coding
   (d) Name two major standards of video format
   (e) Write down the rows and columns in each frame of video in CIF.
   [1x5]

2. In the context of audio coding
   (a) Distinguish between 'lossless predictive coding' and 'Differential Pulse Code modulation'.
   (b) Describe the basic steps of lossless predictive coding and decoding.
   (c) Given a sequence of integer data: 5, 4, 2, 8, 6, 5, 7. Calculate the data sequence to be transmitted and the actual data generated at the receiver end using lossless predictive coding.
   [2+4+4]

3. In the context of video coding and processing
   (a) What are different types of redundancy utilized in video compression?
   (b) Suggest a simple way of reducing these redundancies.
   (c) Describe 2D logarithmic or 4-2-1 search algorithm for motion vector.
   [1+4+5]

INDIAN STATISTICAL INSTITUTE

Semester Examination : (2010-2011)

M.Tech.(CS) II Year

Database Management Systems

Date: 9.12.2010 Maximum Marks: 50 Duration: 3 Hours

Answer all questions

1. A company wants to design a database covering its employees, departments and projects. An employee is identified by a unique employee_no. For employees, the database also stores the name, address, date of birth, date of joining, designation and the salary. Each department has a unique department name. Against a department, the database also stores the department address and department budget. Besides a unique project_no, each project has a name and a project budget. The starting date and the expected date of completion are also stored for each project. A project can either be internal when it is associated to a department or external. An external project is associated with the name and address of the organization that has ordered for the project. Employees from different departments may participate in an external project and it may not be associated with any specific department. An employee works for only one department but a department has many employees. An employee, however, can work in more than one project and a project has many employees working for it. When an employee works for an external project his/her department may not be associated with that project. Moreover, an employee works for a project with certain responsibility and for a specific period of time.
   a) From the above description draw an appropriate ER/EER diagram.
   b) From the ER/EER diagram, create a set of relations using the usual mapping rules.

(6+6=12)

2. For the relation R=(A,B,C,D,E), the following dependencies are specified :
   A -> BC, CD -> E, B -> D, E -> A
   a) Derive the single attribute candidate keys of R.
   b) Explain whether the decomposition R_1(A,B,C) and R_2(C,D,E) of R is lossless.
   c) Justify that relation R cannot be decomposed to a set of normalized relations under BCNF using the given set of functional dependencies.

(4+3+6=13)

3. T1 and T2 are two transactions where T2 is nested within T1. The two transactions operate on three data items a, b and c. If crash occurs at any of the four places as shown below, what actions would be taken by the transaction manager for recovery in each case when the system incorporates,
   i) immediate update policy,
   ii) deferred update policy.

   Consider 'redo' and 'undo' are two procedures used in the recovery process where 'redo' forces the new values to the involved data items signifying as if no crash had occurred; 'undo', on the other hand, sets the old values signifying as if no execution of transactions was ever done. Show the content of the log-file in each case.

```
T1 :    read a;
        modify a;
                <- 1
        write a;
T2 :    read b;
        modify b;
        write b;
                <- 2
        commit;
```

P.T.O.

```
T1 :    read c;
                    <- 3
        modify c;
        write c;
        commit.
                    <- 4
```

(4x2=8)

4. Three relations R1(X, Y, Z), R2(M, N, P) and R3(N,X) are to be joined. The primary keys of the relations have been underlined. The relations have 100, 30 and 400 tuples respectively. The variety of values attribute N has in R2 and R3 are, V(N,R2)=15 and V(N,R3)=300.
   a) Considering number of expected tuples in the intermediate relation as the cost of join (ignoring the size of tuple produced), suggest the order of join for minimum cost.
   b) Considering the total space occupancy in the intermediate relation as the cost of join, suggest the order of join for minimum cost. The space requirements for different attributes are,
   X=30 bytes, Y=10 bytes, Z=10 bytes, M=20 bytes, N=20 bytes and P=10 bytes.
   c) If R1 and R3 are to be joined using block-oriented loop algorithm, calculate the number of disk accesses necessary in both cases when either R1 or R3 is placed in the outer loop. For each relation, consider 10 tuples occupy a block.
   d) What would be the number of disk accesses in both cases above, when relation R1 (the smaller of the two) can be totally accommodated in the main memory during execution of the join.

(6+3+4+4=17)

-x-

INDIAN STATISTICAL INSTITUTE

M. Tech.(Computer Science) II Year, 2010-11

Semestral Examination

Pattern Recognition and Image Processing

Date: 8.12.10 Maximum Marks: 100 Duration: 195 minutes

Note: This paper carries 104 marks. Answer as much as you can. Maximum that you can score is 100.

1. Draw the solution tree for branch and bound feature selection algorithm if 3 features are to be selected from 7 features. [6]

2. Let $\underline{X}' = (X_1, X_2, X_3, X_4)$ be a random vector with dispersion matrix Σ,

where $\Sigma = \begin{pmatrix} 3 & 1 & 0 & 0 \\ 1 & 3 & 0 & 0 \\ 0 & 0 & 4 & -1 \\ 0 & 0 & -1 & 4 \end{pmatrix}$. Find two principal components of $\underline{X}$. [10]

3. Describe any two feature selection algorithms. [8]

4. Write short notes on Probabilistic separability based criteria for feature selection. [5]

5. Describe the single linkage clustering algorithm. [5]

6. The following table provides frequency distribution of gray values. Find the frequency distribution of gray values after equalization of histogram. [10]

Gray value	Frequency
0	0
1	5
2	15
3	5
4	0
5	25
6	30
7	0

7. Describe Canny edge detection method for gray level images. [15]

8. Describe a method for finding line segments in a binary image using Hough Transform. [12]

(P.T.O.)

9. Describe a region based segmentation method for gray level images using quad tree. [8]

10. (a) Define skeleton of a region.
(b) Describe a thinning algorithm for an object in a binary image. [5+10]

11. Describe a method for introducing salt and pepper noise in a gray level image. [5]

12. Describe median filtering method for gray level images. [5]

Indian Statistical Institute

Semester-I 2010-2011

M.Tech.(CS) - Second Year

End-semester Examination (11 December, 2010)

Subject: Compiler Construction

Total 70 marks   Maximum marks: 60   Duration 3.5 hrs.

Please keep your answers brief and to the point.

1. Consider the following syntax-directed definition (SDD):

$$\begin{array}{ll}
S \rightarrow \textbf{id} = \textbf{expr} & \{\ S.s = \textbf{expr.size};\ \} \\
S \rightarrow \textbf{if expr then } S_1 \textbf{ else } S_2 & \{\ S_1.t = S.t + 2;\ \ S_2.t = S.t + 2; \\
 & \ \ S.s = \textbf{expr.size} + S_1.s + S_2.s + 1;\ \} \\
S \rightarrow \textbf{while expr do } S_1 & \{\ S_1.t = S.t + 4;\ \ S.s = \textbf{expr.size} + S_1.s + 2;\ \} \\
S \rightarrow S_1\ ;\ S_2 & \{\ S_1.t = S_2.t = S.t;\ \ S.s = S_1.s + S_2.s;\ \}
\end{array}$$

S is a non-terminal, all other symbols are terminals.

(a) Explain in 2-3 lines whether $S.s, S.t$ are inherited or synthesized attributes. Is the above SDD S-attributed? L-attributed? Justify your answer.

(b) Convert the SDD into a form suitable for bottom-up parsing by using marker non-terminals. Avoid introducing unnecessary marker non-terminals.

(c) Re-write the semantic actions in (b) in terms of elements of the value stack. Assume that attributes are stored in the usual positions in the value stack during bottom-up parsing.

(d) Construct a translation scheme that is equivalent to the SDD given above, but which uses the non-left-recursive version of the grammar. Make sure attributes are handled correctly.

HINT: Draw the parse tree for a suitable example string using the original grammar. Annotate the nodes with appropriate attribute values. Then construct the parse tree for the same string using the non-left-recursive version of the grammar. Use this parse tree to determine what semantic actions are needed. If your answer is incorrect, you will get partial credit if you show these steps clearly.

[3+5+10+10=28]

2. Consider the following grammar for statements (S) in a C-like language.

$S \rightarrow S\ ;\ S$
$S \rightarrow \textbf{id} = E$
$S \rightarrow \textbf{if}\ (C)\ \textbf{then}\ \{\ S\ \}\ \text{else}\ \{\ S\ \}$
$S \rightarrow \textbf{while}\ (C)\ \text{do}\ \{\ S\ \}$
$S \rightarrow \textbf{break}$
$S \rightarrow \textbf{continue}$

E and C represent arithmetic expressions and Boolean expressions, respectively. The other symbols have their usual meanings.

NOTE: The **break** and **continue** statements only have an effect if they occur inside a loop; otherwise, these statements have no effect. If a **break** statement occurs within a **while** loop, it causes execution to jump out of the loop that most closely surrounds the **break** statement. If a **continue** statement occurs

P. T. O

within a **while** loop, it causes execution to jump directly to the code that tests the condition (C) for the loop that most closely surrounds the **continue** statment.

(a) Manually translate the following snippet into 3-address code:

```
while (i < N) {
   x = x + 1;
   if (x > M) then { break } else { y = y - 1 };
   if (y < 0) then { break } else { z = z * 2 }
}
```

(b) Using your translation in (a) as a guide, write suitable semantic actions corresponding to the above grammar rules to automatically translate statements into equivalent 3-address code **using backpatching**. You may use the attributes ***truelist***, ***falselist*** and *nextlist* with their usual meanings. You may also create additional similar attributes if necessary. You need not write syntax/semantic rules corresponding to E or C.

[3+12=15]

3. Consider the following procedure in C:

```
int unique(int *a, int n)
{ int i, j;
  for (j = 0, i = 1; i < n; i++) {
      if (a[j] != a[i])
      a[++j] = a[i];
  }
  return (j + 1);
}
```

(a) Convert the body of the procedure into 3-address code. Assume that "**return** x" is a valid 3-address statement. Each time a temporary variable is needed, use a new temporary. **Do not perform any optimization at this stage.**

(b) Assuming that integers, pointers and temporaries each occupy 4 bytes, and that it takes 128 bytes to store the saved machine status, calculate the size of the activation record (AR) for unique.

(c) Draw a suitable layout for the AR (including byte offsets). Briefly justify your design.

(d) Using the above layout, write the machine code for the calling sequence and return sequence corresponding to a call to unique from **main**. Assume that (i) the stack grows from low addresses to high addresses; (ii) the stack pointer points to the beginning (i.e. lowest address) of an AR; and (iii) the AR for main takes 320 bytes.

(e) Identify the leaders in your 3-address code (part (a)) and draw the flow-graph.

(f) Optimize your intermediate code by using whichever of the following techniques are applicable: global common sub-expression elimination, copy propagation, dead code elimination, code motion, induction variable elimination.

[6+4+2+5+(3+4)+3=27]

INDIAN STATISTICAL INSTITUTE

Back Paper Examination : (2010-2011)

M.Tech.(CS) II Year

Database Management Systems

Date: 12.1.11 Maximum Marks: 50 Duration: 2 Hours

Answer all questions

1. To design a database for its employees using relational model, an organization generates a single relation with the following attributes:
(employee_no, name, address, designation, salary, department_name, department_address, department_budget)
Every employee has a unique employee_no. Each department also has a unique department_name. While an employee works in only one department, a department may have many employees. Two employees having same designation get same salary.
   a) Using the above properties, decompose the single relation into a set of normalized relations under BCNF.
   b) Now, if the organization wishes to add two hobbies of each employee to the relation under one attribute "hobby", what would be the new set of normalized relations.

   (6+5=11)

2. Two relations R_1 (a,b,c,d,x) and R_2 (a,b,c,d,y) have four common attributes. The fifth attribute x and y of R_1 and R_2 respectively also belong to the same domain. Find the intersection of the two relations without using 'union', 'intersection' or 'set difference' operators.

   (7)

3. Let R_1 (A,B,C) and R (A,D,E) are two relations where A is the common attribute. Attributes B and E are in the same domain. Attributes C and D are also in the same domain. However, the domain of B and E is different from the domain of C and D. Out of the two equivalent algebraic expressions given below, which one you would prefer and why?

   i) $\Pi_A (R_1 \cup R_2)$

   ii) $(\Pi_A (R_1)) \cup (\Pi_A (R_2))$

   (7)

4. For a relation R(A,B,C,D,E,F), a query wants to execute:

   $$\pi_{A,B,C} (\sigma_{B>500} (R))$$

   System maintains a B+ tree index for (A,B,C) on R. However, the index is unclustered. In your opinion, what strategy the query optimizer should take for executing the above query? The entire relation occupies 100 pages while the index structure needs 5 pages only.

   (7)

5. Two transactions T_o and T_1 are executed sequentially as shown below. Both the transactions are manipulating the data item A. If crash occurs in one of the four places (1 to 4) as indicated in the schedule, explain the recovery action the system would undertake if it follows a deferred update log maintenance strategy with standard 'redo' and 'undo' routines. If the log contains a check point after the commit of T_o, would there be any change in the recovery process after the crash points 3 and 4 ?

   Schedule:
   T_o: read (A)
   A=A-1

P.T.O.

```
       write(A)
       ----------------------(1)
       commit
       ---------------------(2)
T1:  read (A)
        A=A+2
        write (A)
        --------------------(3)
        commit
        -------------------(4)
```

(3x6=18)

INDIAN STATISTICAL INSTITUTE
Mid-Semester Examination : 2010-2011
M.Tech (CS) II Year
Advanced Database Theory and Applications

Date : 21.02.2011 **Max. Marks – 40** **Time – 2 Hours**

1.

Time Slot	T1	T2	T3	T4
1	read(A)			
2			read(B)	
3		write(A)		
4				read(C)
5	write(B)			
6	write(C)			
7		write(B)		
8				write(A)
9			write(C)	

The above concurrent schedule involves four transactions T1 to T4 using three data items A, B and C for read and write.

a) Drawing a precedence graph show whether the above schedule is conflict serializable?

b) To examine view serializability, draw the labeled precedence graph for three data items separately to show whether they are individually view serializable?
Also draw the composite labeled precedence graph to examine whether the schedule is view serializable considering all the three data items together?

(4+(3x3)+4=17)

2. Let us consider that timestamp based protocol is used for the execution of the concurrent schedule given in Question 1. According to the schedule, each read or write operation needs one time slot. In case of any time conflict, a transaction is allowed to rollback and restart with a new timestamp higher than all the existing timestamps. No such conflict is ignored. i.e. Thomas' Write Rule is not considered. It is assumed that a rolled back transaction is rescheduled immediately. The unused time slots obtained due to the rollback of a transaction may be utilized by a rescheduled transaction. A transaction may have any number of rollback and restart. Find the total number of time slots required to execute all the four transactions without changing the given schedule. The timestamps of the four transactions are related as,

$$TS(T1) < TS(T2) < TS(T3) < TS(T4)$$

While the execution is in progress, show the changes in the values of the read and write timestamps of different data items. Initial values of all the read and write timestamps are less than TS(T1).

(15)

3. Keeping all other conditions same in the problem described in Question 2, any transaction once rolled back is rescheduled immediately and it joins at the end of the current schedule with its timestamp unaltered. Count the number of rollbacks that would occur before all the four transactions are completed, if the system employs, wound-wait scheme for deadlock prevention.

(8)

-x-

Indian Statistical Institute

Mid-Semester Examination 2010-2011
M. TECH.(CS) II Year
Subject: Document Processing and Retrieval
Full Marks: 50 Duration: 2 hrs.

Date: 22nd February 2010

(Answer all questions)

1. Write down the block diagram of a document analysis system. What is the skew of a document image? Describe Hough transform based approach for skew detection of a document. [6+2+8]

2. What do you mean by off-line and on-line recognition? Which of these two approaches, in your opinion, is simpler in recognition point of view and why? [3 +4]

3. Generate a feature based tree classifier to recognize following printed alphanumeric characters (O, D, M, W, P, R, S, Z, X, 7, 9, 4). [12]

4. Discuss two rotation invariant feature extraction methods for character recognition. [10]

5. What do you mean by a run of black pixels? How the run information of a character image can be used to detect the stroke width of a character? [2+3]

INDIAN STATISTICAL INSTITUTE

Periodical Examination: (2010 – 2011)
M.Tech. (CS) II Year

Parallel Processing: Architectures and Algorithms

Date: 22/02/2011 Maximum Marks: 60 Duration: 2 hrs

1. (a) According to the multiplicity of instruction and data streams, classify the parallel machine architectures with a schematic diagram for each.
   (b) Show the schematic diagram of a Parallel Random Access Machine (PRAM), and mention the underlying assumptions of the model.
   (c) Show that the CRCW model of PRAM with N processors can always be simulated on the EREW model having N processors with an $O(\log N)$-fold increase in the processing time.

[2+3+3=8]

2. Given *m* sets of numbers, each containing *n* elements: $S_j = \{xj1, xj2\ ,\dots \dots, xjn\}, 1 \leq j \leq m$, and *n* is a power of 2, it is required to find the sum of each set: $SUM_j = \sum_k xjk, 1 \leq j \leq m, 1 \leq k \leq n$. Develop a parallel architecture that may produce the results in $(\ \log n + m - 1)$ steps. How many processors are required and what will be the interconnection network? Mention the input data distribution and the delivery of output. Calculate the speed-up and utilization.

[4+2+2+2 = 10]

3. Given the choices: i) 2-D mesh, ii) binary hypercube, and iii) cube-connected cycle, you are asked to select the best interconnection network for interconnecting 64 nodes of a multicomputer. Let *d* be the maximum node degree, *k* the network diameter, and *B* the bisection width of a network. Rank the three architectures according to the parameter $(B.\ (d.k.)^{-1})$.

[10]

4. (a) Given two processes P_i and P_j with corresponding sets of inputs and outputs as I_i, I_j and O_i, O_j respectively, state the conditions to be satisfied for *Pi* and *Pj* to be executable in parallel, i.e., *parallelizable*. Is the relation of '*parallelizable*' an *equivalence relation*?

   (b) Consider the following program segment with six instructions:

$P_1 : A = B + C$

$P_2 : D = B \times A$

$P_3 : E = B \times C$

$P_4 : F = D + C$

$P_5 : B = E + A$

$P_6 : C = F + A$

Draw the data dependence graph considering each statement as a process.
Show a possible scheduling of the processes exploiting the maximum parallelism existing among the processes. Ignore the resource dependences. Assuming that addition takes 10 time units, multiplication

100 time units, and inter-processor communication 300 time units, calculate the speed-up achieved by your scheduling.

[(2+1)+(4+3+2) = 12]

5. (a) Show the timing diagrams for memory access of the *S-access* and *C-access* interleaved memory organizations and mention their merits and demerits.
(b) What is an embedding? Prove that a *(2N-1)*-node completely-filled binary tree is not contained in a *2N*-node hypercube *($N=2^n$)*.

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

6. (a) Prove that if each node of an $N \times N$ mesh contains one packet to be routed to a unique destination, following farthest-first strategy, routing can be completed in (2N-2) steps.

(b) In an $N \times N$ mesh network, how many distinct shortest paths exist between a pair of nodes (x_1, y_1) and (x_2, y_2), $0 \leq x_1, y_1, x_2, y_2 \leq N-1$, where (x, y) denotes the position of a node in row x and column y. Two paths are distinct if they differ at least in one edge. How many paths of these are node disjoint (no common node, except the source and destination)?

[5+(3+2)= 10]

INDIAN STATISTICAL INSTITUTE
Mid-Semester Examination
M. Tech. (CS) II year
Information and Coding Theory

Date: 23. 02. 2011 Marks: 30 Time: 3 Hours

Instructions (Read carefully):

1. There are six questions, each carrying five marks.
2. Answer each question clearly and precisely.

Questions:

1. Construct a *binary linear code* with parameters $[n, \log_2(n+1), \frac{n+1}{2}]_2$. Provide a full proof that your code is indeed a binary linear code and it satisfies the above parameters. From this code, construct another *binary linear code* with parameters $[n+1, \log_2(n+1), \frac{n+1}{2}]_2$.

2. Let C_i be an $[n_i, k_i, d_i]_q$ linear code over $\mathbb{F}_q$, for $i = 1, 2$. Define

$$C = \{(a+c, b+c, a+b+c) \mid a, b \in C_1, c \in C_2\}$$

   (a) Is C a linear code? If yes, then what are its parameters?

   (b) Let G_1 and G_2 be generator matrix of C_1 and C_2 respectively. Then what is the generator matrix of C ?

   (c) What can you say about distance of C ?

3. Show that if there exists a linear code C with parameters $[n, k, d]_q$ where d is *even*, then there exists a linear code C' with parameters $[n, k, d]_q$ such that every codeword in C' has even weight.

4. Let C be an *binary linear code* and let $\overline{C}$ be the code derived by taking *complement* of all codewords of C.

   (a) Show that if $(1, 1, \ldots, 1) \in C$ then $C = \overline{C}$.

   (b) Prove or refute: $\overline{C}$ is a linear code.

   (c) Prove of refute: $C \cup \overline{C}$ is a binary linear code.

5. The *heaviest codeword* problem is defined as follows: Upon receiving a parity check matrix H that fully defines a *binary linear code* C, find the codeword $c \in C$ with the *maximum weight* (i.e., find c such that $wt(c) \geq wt(c')$, for all $c' \in C$). Give an efficient (polynomial-time) algorithm for this problem or show that it is NP-complete. You can assume that finding the distance of C is NP-complete.

6. Prove the following:

   (a) There exists no binary linear code with parameters $[2^m, 2^m - m, 3]$, for any $m \geq 3$.

   (b) Let C be a binary linear code with parameters $[2^m, k, 4]$ for some $m \geq 3$. Then $k \leq 2^m - m - 1$.

Indian Statistical Institute

M. Tech Computer Science

Mid-Semester Examination -2011

GROUP - B Computer Vision

Duration – 1hour 15 min.
Full Marks – 25

Answer any two questions

Q1. (i) write down the differences between the perspective projection and orthographic projection. 4

(ii) What kind of projection will you have for a wide angle lens and a telephoto lens ? Give reasons for your answer. 4

(iii) What do you mean by radiance and irradiance ? 4

Q2. (i) Show that for a telephoto lens, image irradiance is directly proportional to the scene radiance and inversely proportional to the square of the f-number of the lens. 7

(ii) What is BRDF ? How is it constrained by the Helmoltz reciprocity condition ? 5

Q3. (i) What is convolution between two functions ? Show that convolution is commutative and associative. 7

(ii) What do you mean by the point spread function ? 5

eatness - 1

INDIAN STATISTICAL INSTITUTE
Mid-Semestral Examination
M. Tech (CS) - II Year (Semester - II)
Multi-dimensional Search & Computational Geometry

Date : 25 February, 2011 Maximum Marks : 50 Duration : 3 Hours

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

1. Any triangulation of any n-sided simple polygon has exactly n - 2 triangles. Suppose that the polygon has h polygonal holes each having k sides. As a function of n, h and k, how many triangles will such a triangulation have? [5]

2. Given an arrangement A of n lines in the plane and given an arbitrary line l, what is the maximum number of edges of the cells of A containing the portion of line l? [5]

3. The convex hull is a somewhat non-robust shape descriptor, since if there are any distant outlying points, they will tend to dominate the shape of the hull. A more robust method is based on the following iterative approach. First compute the convex hull of all the points, remove the vertices of the hull. Then compute the convex hull of the remaining points, and again remove the vertices of the hull. Repeat this until no more points remain. The final result is a collection of nested convex polygons (where the last one may degenerate to a single line segment or single point).
Given a set P of n points in the plane, devise an $O(n^2)$ time algorithm to compute this iterated sequence of hulls. [8]

4. (a) You are given a convex polygon in the plane having n_c sides and an x-monotone polygon having n_m sides. What is the maximum number of intersections that might occur between the edges of these two polygons? (You may use O-notation.)
(b) You are given two x-monotone polygonal chains P and Q with a total of n vertices between then. Prove that P and Q can intersect at most $O(n)$ times. [10]

5. An arrangement of n lines in the plane has exactly n^2 edges. How many edges are there in an arrangement of n planes in 3-dimensional space? Explain briefly. [7]

6. In Fortune's algorithm, let $\{p_1, p_2, \ldots p_n\}$ denote the sites that have been processed up to this point in time. Show that the number of parabolic arcs along the beach line is at most $2n$-1. [7]

7. Let $\{p_1, p_2, \ldots p_n\}$ be a set of n points with weights $\{w_1, w_2, \ldots w_n\}$ respectively. For any query rectangle $[x : x'] [y : y']$ our objective is to report the top k largest weighted points inside the query rectangle $[x : x'] [y : y']$. If the number of points inside the query rectangle $[x : x'] [y : y']$ is less than k, then we will report all the points. Built an efficient data structure so that for any query rectangle one can report the top k largest weighted points in $O(log^2n + k)$ time. Write the storage space and construction time of the data structure. [15]

INDIAN STATISTICAL INSTITUTE

✓

Periodical Examination

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

Topics in Algorithm and Complexity

Date : 28.2.2011 Maximum Marks : 60 Duration : 3 Hours

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

1.(a) Let L be a decision problem, and let Π be a minimization problem, and let $\alpha > 1$. Write down the conditions for α-gap-introducing reduction from L to Π. Introduce conditions for gap preserving reduction from problem $|Pi_1$ to problem $|Pi_2$.

(b) Given a graph G, to decide whether G have a Hamiltonian cycle is NP-hard problem. From this information, using some gap introducing reduction, show that for any polynomial time computable function $\alpha(n)$, Traveling Salesman Problem cannot be approximated within a factor of $\alpha(n)$, unless P = NP.

(c) State the PCP theorem. [6 + 8 + 6 = 20]

2. Describe the **LazySelect** algorithm for finding the k-th order statistics $x_{(k)}$ among a set of n univariate data $x_1, x_2, \ldots, x_n$ $(n >> k)$. Show that the probabaility of finding $x_{(k)}$ in the first pass of the **LazySelect** is $1 - O(n^{-\frac{1}{4}})$. Hence show that the expected time complexity of the **LazySelect** algorithm is $O(n)$. [6+6+6=18]

3.(a) A simple polygon P is star-shaped if there exists a point $q \in P$ such that for every point $p \in P$, the line segment $[p, q]$ completely lies inside P. Give a linear time algorithm for testing whether an input polygon P is star-shaped.

(b) For a two variate linear programming problem, if the constraints are considered one by one, then write an algorithm for computing the optimum solution of the problem. [10+10=20]

4.(a) Let $G = (V, E)$ be an undirected graph with n vertices. We label the vertices in V randomly with the set of integers $\{1, 2, \ldots, n\}$. Now, we obtain a directed graph G' by assigning direction of an edge (i, j) as $i \rightarrow j$ if $i < j$; otherwise the direction is $j \rightarrow i$.

Prove that if G has a path P of length k, then P is also a path in G' with probability $\alpha = \frac{2}{(k+1)!}$.

P.T.

24

(b) Now consider the following algorithm:

```
repeat steps 1 and 2 $t = \frac{1}{\alpha}$ times
1.      Assign directions to the edges of $G$ randoly to get $G'$.
2.      Find longest path in $G'$.
end (repeat)
If the path length is less than $k$ in each time,
                    then report failure to get a path of length $k$
                    else  report success
end.
```

Prove that the above algorithm finds a path of length $k = O(\frac{\log n}{\log\log n})$ (if one exists) in randomized polynomial time.

[8+8=16]

INDIAN STATISTICAL INSTITUTE

Mid Semester Examination: (2010 - 2011)

Course Name: M.Tech (CS) **Year: II year**

Subject Name: Neural Networks & Applications

Date: February 28, 2011 **Maximum Marks: 50** **Duration: 2 hrs**

Answer all questions.

1. Describe various transfer functions that are used in the design of artificial neural networks. (10)

2. Consider a set of 10 dimensional patterns being distributed in 3 classes. You need to design a multilayer perceptron-based classifier. Assume that the model consists of a single hidden layer with 12 nodes. Derive the equations for modifying the weights of the neural network, under batch-mode learning. (30)

3. Show how the momentum term helps in backpropagation learning. (10)

INDIAN STATISTICAL INSTITUTE

Mid-Semester Examination: 2010-2011

M. Tech. (CS) 2nd Year

Advanced Image Processing

Date: 01.03.2011 Maximum Marks: 50 Duration: 2 hours

Answer any five questions in brief.

1. Describe the Fast Fourier Transform algorithm and discuss its computational complexity. [10]

2. (a) Obtain the Fourier transform of the Laplacian of a two-variable function $f(x, y)$. Assume that x and y are continuous variables.
   (b) State and prove the *convolution* theorem. [4+6]

3. Show that the Fourier transform of the autocorrelation function of $f(x)$ is its power spectrum. Prove that the origin of the Fourier transform of a two-variable function $f(x, y)$ can be moved to the center of its corresponding $N X N$ frequency square by multiplying $f(x, y)$ by $(-1)^{x+y}$. [5+5]

4. Assuming pin-hole camera model for perspective projection, prove that
   (a) Distant objects appear small.
   (b) A set of parallel lines not perpendicular to z-axis in 3D, is mapped to a set of concurrent lines in 2D. [5+5]

5. State three basic principles of photometric transformation and then derive

$$g(x,y)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(\alpha,\beta)\,h(x-\alpha,y-\beta)\,d\alpha\,d\beta+\eta(x,y)$$

All terms carry their usual meaning. [3+7]

6. Derive the Wiener filter for image restoration. Mention one limitation of this approach. [9+1]

Indian Statistical Institute

M. TECH. (CS) Second Year
Mid-semester Examination
Subject: Advanced Cryptology

Date: 02/02/2011 Time: 2 hours Maximum Marks:40

Note: The paper carries 50 marks. Maximum you can score is 40.

1. Explain Shamir's (t, w)-threshold scheme. Prove that it is a perfect secret sharing scheme. [3 + 7]

2. Take RSA cryptosystem. Assume there exists an adversary $\mathcal{A}$ running in time t for which

$$\Pr[\mathcal{A}([x^e \bmod n]) = x] = 0.01,$$

where the probability is taken over random choice of $x \leftarrow \mathbb{Z}_n^*$. Show that it is possible to construct an adversary $\mathcal{A}'$ for which

$$\Pr[\mathcal{A}'([x^e \bmod n]) = x] = 0.99,$$

The running time t' of $\mathcal{A}'$ should satisfy $t' = \text{poly}(||n||, t)$. [10]

3. Give an example of an Encryption Scheme (different from ElGamal) which is IND-CPA secure but not IND-CCA2 secure. Justify your answer. [10 + 2]

4. Define EUF-CMA and SUF-CMA security of Signature Scheme. [2 + 2]

5. Consider the following Encryption Scheme

- **Setup**(1^k)
  (a) Choose two prime numbers "p" and "q" such that $2^{k-1} < p < q < 2^k$.
  (b) Compute $n = pq$
  (c) Compute $\phi(n) = (p - 1)(q - 1)$
  (d) Choose "e" such that $\gcd(e, \phi(n)) = 1$.
  (e) Compute "d" such that $ed \equiv 1 \bmod \phi(n)$.
  (f) Message Space $(\mathcal{M}) = \mathbb{Z}_n^*$
  (g) Ciphertext Space $(C) = \mathbb{Z}_n^* \times \mathbb{Z}_n^*$

  Public Key $(PK) \equiv (n, e)$
  Secret Key $(SK) \equiv (p, q, d)$

- **Encryption**(m, PK)
  (a) Choose $m_1 \overset{R}{\leftarrow} \mathbb{Z}_n^*$
  (b) Compute $m_2 = m_1^{-1} m \bmod n$ (Note: $m_1 m_2 \equiv m \bmod n$)

P.T.O.

(c) Compute $c_1 \equiv (1+m_1)^e \bmod n$

(d) Compute $c_2 \equiv (1+m_2)^{m_1} \bmod n$

Ciphertext $= C$

- **Decryption**$(C = (c_1, c_2), SK)$

  (a) Compute $m_1' \equiv c_1^d \bmod n$

  (b) Compute $l \equiv m_1^{-1} \bmod \phi(n)$

  (c) Compute $m_2' \equiv c_2^l - 1 \bmod n$

  (d) Compute $m' \equiv m_1' m_2' \bmod n$

  Return m'

Questions

(a) What is the condition when Decryption may fail?

(b) Show that above scheme is not IND-CCA2 secure. [4+

✓

INDIAN STATISTICAL INSTITUTE

M. Tech. (CS) II Year (2010-11): II Semester

Periodical Examination

ADVANCED PATTERN RECOGNITION

Date: 03.03.2011 Duration: 2 Hours Marks: 60

Note: Answer all the questions

1. (a) Distinguish between probability and fuzzy membership.

   (b) Write a short note on max-min composition rule of inference. **[5+5=10]**

2. (a) Suppose X is a data set and $A_1, A_2, \ldots, A_c$ are subsets of X. Write the criteria for the family $\{ A_1, A_2, \ldots, A_c \}$ to be a valid fuzzy c-partition of X.

   (b) State when fuzzy c-means algorithm converges to hard c-means algorithm.

   (c) Give the mathematical justification of the rule for updating the membership values of a sample point to various clusters in each iteration of the fuzzy c-means algorithm. **[2+2+8=12]**

3. (a) Write the basic steps of an algorithm for decomposing a feature space.

   (b) Describe the classification strategy of this method.

   (c) Give sketches of three typical classification problems (with two features and two classes) where this technique can be effective but other conventional classification approaches may not provide satisfactory performance. **[10+5+3=18]**

4. (a) Write a short note on activation function in connection with the perceptron algorithm.

   (b) Sketch the general structure of the multilayer perceptron (MLP).

   (b) State the back propagation learning algorithm for MLP. **[5+3+12=20]**

INDIAN STATISTICAL INSTITUTE

Mid-Semester Examination: 2010-11(Second Semester)

Course Name: M.Tech. (CS) 2nd Year

Subject Name: Natural Language Processing

Date: 03.03.2011 Maximum Marks: 50 Duration: 2 hours

Note: Each question carries equal marks.

Q1. Consider an English example:

Which book did she review without reading?

Here *reading* has a "gap" as shown below with underscores:

Which book did she review without reading____?

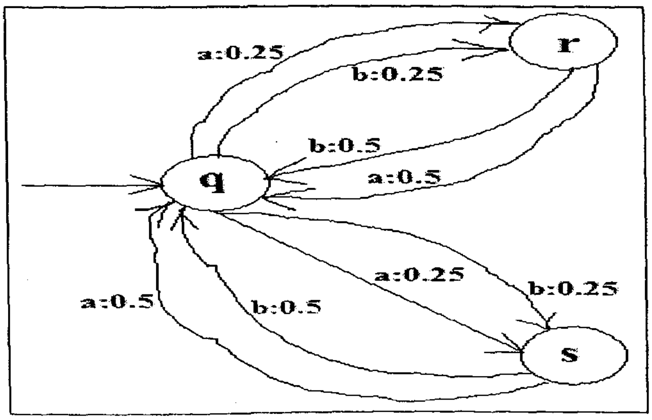
We don't know "without reading what?". This gap is considered as "parasitic" gap (parasitic on "review what?") and cannot easily stand on its own.

Assume that "parasitic gap" is really a rare syntactic construction and this occurs on average once in 100,000 sentences. You have developed an approach to identify sentences with parasitic gaps. Your approach is pretty good, but not perfect. If a sentence has a parasitic gap, your method will say so with probability 0.95 and if it doesn't, it will wrongly say it does with probability 0.005. Suppose the test says that a sentence that contains a parasitic gap. What is the probability that this is true?

[Hints: If $A \subseteq \cup_{i=1}^{n} B_i$, P(A) > 0, and $B_i \cap B_j = \phi$ for i ≠ j then according to Bayes' theorem:

$$P(B_j \mid A) = \frac{P(A \mid B_j)P(B_j)}{P(A)} = \frac{P(A \mid B_j)P(B_j)}{\sum_{i=1}^{n} P(A \mid B_i)P(B_i)} .]$$

Q2. Consider the given HMM (as shown in the figure) as a guessed HMM for a training sequence **aabb**. Show the output probabilities (i.e., the new HMM) at the end of the first iteration of the training algorithm.



What are the probabilities of observing **aabb** using (i) the guessed HMM and (ii) the HMM you get after the first iteration. Are you satisfied with the change in this probability? Explain your answer.

Please turn over

✓

Q3. Assume there are four languages (L_1, L_2, L_3, and L_4) and you are given sufficiently large sample for each language. Formulate a method to identify the language (one out of the given four) in which a short segment of text is written. Your identification method should be based on training itself on the sample text given for each known language. For example, each of the following lines is text in a different language:

doen is ondubbelzinnig uit
pretender a un emploi
uscirono fuori solo alcune
look into any little problem

You method will take one line as input and identify the language in which the line is written. Explain your answer by providing pseudocode of your proposed method.

[Hint: This is a classification task, in which you may use some of the language modeling techniques taught in your class.]

Q4. Automatic detection of sentence boundary is a non-trivial problem in NLP. This is so because the question, "what is a sentence?" has no definite answer. We normally say that a sentence is "something ending with a '.', '?' or '!'." The problem is that periods in text can be used either to mark an abbreviation or to mark the end of sentence. Sometimes what is on one or the other or even both sides of the punctuation marks colon, semicolon, and dash (':', ';', and '—') might best be thought of as a sentence by itself, as ':' in this example: I have two goals in my life: First one is to do Ph.D. in NLP and the second one is to teach at MIT.

Assume you have developed of an HMM based POS tagger. Could you formulate a sentence-boundary detection method using this HMM? Explain your answer by providing a mathematical foundation and suitable illustrations.

INDIAN STATISTICAL INSTITUTE
Mid-Semester Examination: 2010-11(Second Semester)

Course Name: M.Tech. in Computer Science
Subject Name: Distributed Systems

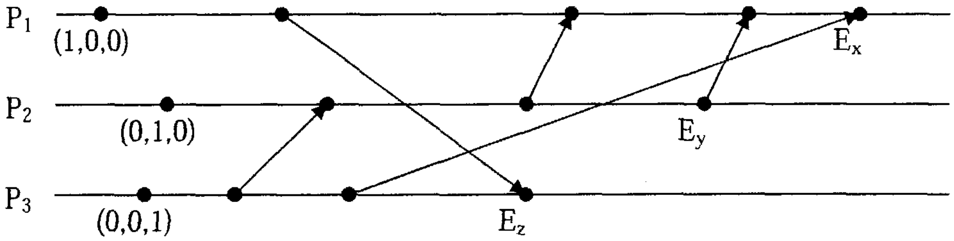
Date: 04. 03. 2011 *Maximum Marks: 40* *Duration: 2 hours*

Instructions:

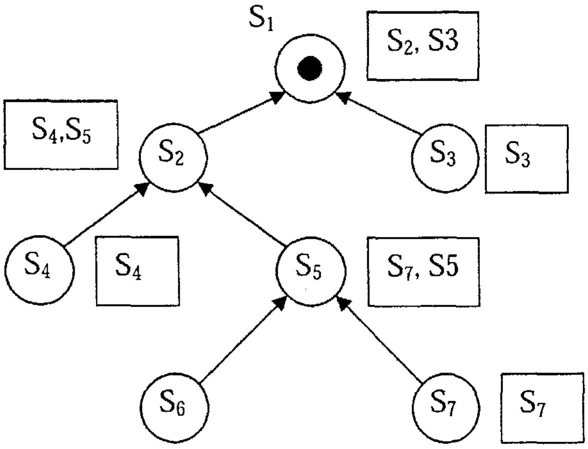
Answer all questions. All parts of a question must be answered in the same place. This is an open book examination.

1. In the following space-time diagram, vector clocks are being used to timestamp the events. The initial timestamps are given. Write the timestamps of the events E_x, E_y, and E_z.

[3 marks]



2. The figure shows a state of Raymond's tree based algorithm for mutual exclusion. S_1 currently holds the token. The request queue at each node is shown beside the node. For example, at S_5, there is a pending request from S_7 followed by one of its own.



(a) Which sites are requesting?

(b) In what sequence will the requesting sites enter the critical section?

(c) Which of these represent possible sequences in which the requests were made? Answer Yes/No for each of the following.

(i) S_7, S_3, S_4, S_5
(ii) S_4, S_3, S_7, S_5
(iii) S_4, S_3, S_5, S_7

[2 + 4 + 3 = 9 marks]

P.T

24

3. Which of the following algorithms assume FIFO channels? For the ones which require FIFO channels, you must indicate what may happen if the channels are not actually FIFO.
   (a) Ricart-Agrawala Algorithm for Mutual Exclusion
   (b) Chandy-Lamport's Algorithm for Global State Recording
   (c) Birman-Schiper-Stephenson protocol for Causal Ordering of messages

[4 X 3 = 12 marks]

4. (a) The k-exclusion problem is similar to the mutual exclusion problem, but here at most k nodes can be in the critical section at the same time, for some predefined constant k. Design an algorithm to achieve k-exclusion in a distributed system. Argue in brief why no more than k nodes can be in the critical section at the same time in your algorithm. Analyze the message complexity (per critical section entry) of your algorithm.

   (b) Explain the types of failures (such as deadlock, starvation, and others) that can happen under the following situations:
   i. The channels do not satisfy FIFO delivery in Lamport's mutual exclusion algorithm
   ii. A control message is lost in Huang's termination detection algorithm
   iii. A node S_i crashes (shuts off) in Maekawa's algorithm

[4 + 4 = 8 marks]

5. A distributed database transaction, T_1, transfers Rs 10,000 from account A at site S_1 to account B at site S_2. The server process P_1 (at S_1) deducts Rs 10,000 from the account A and sends a message, M, to the server process P_2 (at S_2) instructing it to add Rs 10,000 to the account B at S_2. At about the same time a second transaction, T_2, attempts to find the total balance of the accounts. For this purpose T_2 needs to request P_1 to record the state of account A and to request P_2 to record the state of account B.

   a. Draw a space-time diagram for P_1 and P_2. Suppose the state of A is recorded after P_1 sends the message M. Show the valid region in your space-time diagram in which the state of B must be recorded to get the correct total balance.

   b. Re-draw the space-time diagram showing the new messages when Chandy-Lamport's global state recording algorithm is used by T_2. Assume that the algorithm is initiated at S_1 after P_1 sends the message M. Explain how the algorithm guarantees that a consistent global state is detected. [4 + 4 = 8 marks]

✓

INDIAN STATISTICAL INSTITUTE

Semestral Examination: (2010 - 2011)

Course Name: M.Tech (CS) **Year: 2nd year**

Subject Name: Neural Networks & Applications

Date: April 30, 2011 **Maximum Marks: 100** **Duration: 3 hrs**

Answer all the questions.

1. State and prove fixed-increment convergence theorem for a single-layer perceptron. (5+20 = 25)

2. Show how Oja's neural network model can extract the first principal component of a data set. Mention all the assumptions that you have made. (20)

3. a) Consider a ten-dimensional data set containing 100 points. The data is known to have 4 groups. Design a Self Organizing Feature Map for clustering this data. Explain its learning rule.

   b) Why is the learning rule called competitive learning? ((6+2) + 2 = 10)

4. Describe the operation of a Radial Basis Function neural network. (15)

5. (a) Consider the patterns (1,1,1) and (-1,-1,-1). Show how they can be stored in a Hopfield network.

   (b) Assume that the activation of a node of the above Hopfield network at its stable state is either +1 or –1. Which of the above fundamental memories will be retrieved in response to the following input patterns: (1,1,-1), (1,-1,-1), (1,-1,1), (-1,-1,1), (-1,1,1) and (-1,1,-1). (8+12 = 20)

6. Write short notes on the following: (10+10 = 20)

   a) Spatio-temporal model of a neuron

   b) Self organization

INDIAN STATISTICAL INSTITUTE

Semestral Examination

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

Topics in Algorithm and Complexity

Date : 30.4.2011 Maximum Marks : 100 Duration : 3.5 Hours

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

1. Formulate the set cover problem as an integer program.

   Show that the dual fitting based analysis for the greedy set cover actually establishes an approximation gurantee of H_k for the greedy set cover algorithm.

   Propose a rounding scheme in LP relaxation for set cover problem to achieve an approximation gurantee of f, where f is the frequency of the most frequent element. Justify the approximation factor. [5+13+12=30]

2. Scheduling on unrelated parallel mechines: Given a set J of jobs, a set M of machines, and for each $j \in J$ and $i \in M$, a positive integer p_{ij}, the time taken to process job j on machine i, the problem is to schedule the jobs on the machines such that the maximum processing time of any machine is minimized.

   Formulate the Scheduling on unrelated parallel machines in an LP setting.

   Propose an algorithm that achieves a constant factor approximation guarantee for the Scheduling on unrelated parallel machines problem.

   Show that the proposed algorithm guarantees an approximation factor of 2. [4+8+8=20]

3. Show that the decision version of the Independent Set problem for an undirected graph $G = (V, E)$ is $W[1]$-complete, but if G is planar then it admits a fixed parameter tractable algorithm. [6+6=12]

4. The *cluster editing problem* is defined as follows:

   A graph $G = (V, E)$ and a nonnegative integer k are given. Find out whether we can transform G by deleting and adding at most k edges, into a graph that consists of a disjoint union of cliques.

   Give a simple $O(3^k \times n)$ time algorithm for this problem.

   Indicate whether it is possible to improve the time complexity to $O(\alpha^k n)$, where $\alpha < 3$. Justify your answer. [8+12=20]

5. Given a set P of n points in 2D, the closest pair in P is defined as $CP(P) = \min_{p,q \in P} dist(p.q)$.

   (a) Show that given P and a distance r, checking whether $CP(P) \leq r$ or $CP(P) > r$ can be done in $O(|P|)$ time.

   (b) Use this result to propose an algorithm for computing $CP(P)$ for the point set P that runs in expected $O(|P|)$ time.

[8+10=18]

6. (i) Show that the class PSPACE is closed under complementation.

   (ii) Consider the following two-player game GRAPHGAME on an undirected graph $G = (V, E)$ where player A and B alternate moves. Player A makes the first move. A move consists of choosing a vertex in the graph. When v is chosen, v and all its neighboring vertices $N(v)$ are removed. Also, all edges having one end-point in $v \cup N(v)$ are removed. Player A wins if and only if player B is the first player left with no vertices to choose.

   Show that GRAPHGAME is in PSPACE.

[4+10=14]

INDIAN STATISTICAL INSTITUTE
Semester Examination : (2010-2011)
M.Tech.(CS) II Year

Advanced Database Theory & Applications

Date: 02.05.2011 **Maximum Marks: 50** **Duration: 2 Hours**

Students will have to answer all questions

1) In a distributed database environment an organization maintains a database of its suppliers for supply of office and project related items. The projects are executed at different locations. The relations maintained for this purpose are:

Supplier (sid, sname, slocation, turnover)
Item (ino, itype, iname, make)
Catalog (sid, ino, unit_price)
Project (pno, pname, budget, plocation, fagency)

Primary keys of the relations are underlined. Besides the unique id, Supplier relation maintains the name and location of operation (slocation) of each supplier. Average annual turnover of each supplier is also kept. While ino is a unique item number against each item, item type (itype) can be either office-item or project-item. Item name (iname) and manufacturer (make) of each item are also kept in Item relation. Catalog provides the unit_price of each item against each supplier. Each project is identified by a unique project number (pno). Name of the project (pname), its budget, location where it is executed (plocation) and the name of the funding agency (fragency) are also kept in the Project relation.

In the present status of the database, the projects are getting executed at Delhi, Bangalore and Kolkata where Kolkata is the Head-quarter. The entire database is maintained at Kolkata. Other project sites maintain only the horizontal fragments of the relations relevant to their activities. The organization has adopted the following guideline for placing orders to its suppliers:

- Orders for office-items are always placed to the local suppliers from the project location.
- Orders for project-items with unit_price less than or equal to Rs.5 lakhs can also be placed locally to the local suppliers.
- Order for any project-item costing more than Rs.5 lakhs is placed from the Head-quarter.

Following the above guidelines, find the horizontal fragments (both primary and derived) that need to be stored at different project sites.

(15)

2) Two organizations of the type described in Question 1, maintain centralized databases for its suppliers. While one follows the schema of Question 1, the other organization differs only in the way project related information is maintained. While Supplier, Item and Catalog relations are maintained exactly in th same way, the second organization designs its project data as:

Project (pno, name, pbudget, location, agency-name)
Fagency (agency-name, address, alocation)

It may be assumed that a funding agency always prefers to have its projects executed at its own location (alocation=plocation).

Now, the two organizations wish to design a global database to maintain all the information available in both the schemas while data is stored at individual databases. Derive a global conceptual schema considering the final schema to be relational as well.

Show different steps of merging in detail, including identification of conflicts, processes of resolution and the different tables required for the purpose of merging.

(20)

3) Let A be an object belonging to the security class X and B is a subject belonging to the security class Y. Explain in each of the following cases whether B will be permitted to execute the desired operations on A when the concerned DBMS implements both mandatory and discretionary access control mechanisms. The mandatory system follows the Bell-Lapadula model. Discretionary system can be used only if both the subject and object belong to the same security class.

   i) The owner of A has granted explicit read and write privileges to B and B wants to execute them when X>Y.

   ii) Under the same set of discretionary access rights, B wants to read and write on A when X=Y.

   iii) B wants to read and write on A when X<Y, but the owner of A has not granted any privilege to B. B, however, has received the required privileges from another subject C who in-turn has received them from the owner of A with grant option. C belongs to security class X.

(15)

-x-

INDIAN STATISTICAL INSTITUTE
Semestral Examination
M. Tech. (CS) II year: 2010–2011
Information and Coding Theory

Date: May 4, 2011 Maximum Marks: 60 Time: 210 minutes

Use **separate** answerscript for each group.

Group A

Maximum Marks 40

1. A discrete memoryless source emits a sequence of statistically independent binary digits with probabilities $P(1) = 0.005$ and $P(0) = 0.995$. Packets are created taking 100 digits at a time and a binary codeword is assigned for every string containing three or fewer ones.
   (a) Assuming that all codewords are of the same length, find the minimum length required to provide codewords for all sequences with three or fewer ones. [6]
   (b) Calculate the probability of observing a source sequence for which no codeword has been assigned. [6]

2. Let X be a random variable over $\{0,1\}^n$ such that every element in $Sup(X)$ occurs with the same probability, where $Sup(X)$ denotes the support of X. Let, D be another random variable over $\{0,1\}^n$ such that at least ϵ fraction of $Sup(D)$ lies in $Sup(X)$. Prove that
$$H(D) \leq H(X) + \epsilon n.$$
[6]

3. Let X_n be a source over $\{0,1\}^n$ such that every element in its support, $Sup(X)$, occurs with the same probability. Let $H(X_n) \leq \kappa(n)$. Assume that there exists a Probabilistic Polynomial Time Turing Machine T such that for all $z \in \{0,1\}^n$, $T(z) = 1 \Rightarrow z \in Sup(X_n)$. Prove that for $s = \Omega(n)$, s-pseudoentropy of X_n is at most $\kappa(n) + \mathtt{neg}(n)$, where $\mathtt{neg(n)}$ is a negligible function. [6]

4. Using entropy, prove that finding the minimum of n numbers requires at least $n-1$ binary operations. [6]

5. Let $\mathcal{F}$ and $\mathcal{A} = \{A_i\}_{i \in I}$ be a collection of subsets of $[n]$, such that each element of $[n]$ appears in at least r members of $\mathcal{A}$. For $i \in I$, define $F_i = \{F \cap A_i : F \in \mathcal{F}\}$. Prove that
$$\prod_{i \in I} |F_i| \geq |\mathcal{F}|^r.$$
[6]

6. (a) Write algorithms for encoding and decoding of the Huffman Code. [4+4]

(b) Let $C(X)$ be the expected codeword length of the Huffman Code. Prove or disprove,

$$H(X) < C(X) < H(X) + 1.$$

[6]

Group B

Maximum Marks 20

1. Let C be a t-error correcting $[n, k, d]$ Reed-Solomon code, where $n \geq 2t + k$.

   (a) Describe a simple, exponential time decoding algorithm for C. [3]

   (b) Describe a polynomial time decoding algorithm for C. [4].

2. Let C be a binary code, consisting of 2^k codewords. Then how many codewords are required to specify C if:

   (a) C is simply a code. [1]

   (b) C is a linear code. [1]

   (c) C is a cyclic code. [1]

3. Let C be an $[n, k, d]$ Reed-Solomon code. Then, either prove or disprove the following: [5]

   C is also a cyclic code.

4. Let C be an $[n, k, d]$ linear code. Show that the problem of *the nearest neighbor decoding* for C is NP-complete. [5]

✓

INDIAN STATISTICAL INSTITUTE

Semestral Examination

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

Multi-dimensional Search & Computational Geometry

Date : 6 May, 2011 Maximum Marks : 100 Duration : 3:30 Hours

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

1. Let P be a non-convex polygon. Describe an algorithm that computes the convex hull of P in $O(n)$ time. [10]

2. Let S_1 be a set of n disjoint horizontal line segments and let S_2 be a set of m disjoint vertical line segments. Give an $O((n+m)log(n+m))$ time algorithm to count how many intersections there are in $S_1 \cup S_2$. [12]

3 A simple polygon P is star-shaped if there is a point q in the interior of P such that for each point p on the boundary of P, the open line segment qp lies entirely within the interior of P. Suppose that P is given as a counterclockwise sequence of its vertices $<v_1, v_2, ..., v_n>$. Write an efficient algorithm to determine whether P is star-shaped polygon. (Note: You are not given the point q.) Prove the correctness of your algorithm. [15]

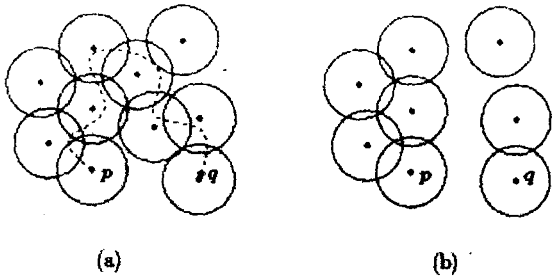
4. Given a set of n points in the plane, write an efficient algorithm for computing the smallest enclosing disc of the given set of points. [15]

5. (a) Define a strip to be the region bounded by two (non-vertical) parallel lines. The width of a strip is the vertical distance between the two lines.
Given a set of n points in the plane, present an efficient algorithm which finds the non- vertical strip of minimum width that encloses all of these points.

   (b) In the primal plane, there is a triangle whose vertices are the three points p, q, and r and there is a line l that intersects this triangle. What can you infer about the relationship among the corresponding dual lines p^*, q^*, r^*, and the dual point l^*? Explain. [15 +15=30]

6. Let *P* be a set of *n* points in the plane. Give an *O(nlog n)* time algorithm to find for each point *p* in *P* another point in *P* that is closest to it. 8

7. You are given a set of *n* sites *P* in the plane. Each site of *P* is the center of a circular disk of radius 1. The points within each disk are said to be safe. We say that *P* is *safely connected* if, given any *p, q* $\in$ *P*, it is possible to travel from *p* to *q* by a path that travels only in the safe region. (For example, the disks of Fig. 1(a) are safely connected, and the disks of Fig. 1(b) are not.) Present an *O(n log n)* time algorithm to determine whether such a set of sites *P* is safely connected. Justify the correctness of your algorithm and derive its running time. [20]



(a) (b)

Fig. 1

Second semestral Examination 2010-11
M. Tech (CS) II Year
Computer Vision

Answer Group A and Group B in separate answer scripts

GROUP A

Date: 06.05.2011 Full Marks: 50 Duration: 1 hour 45 mins

Attempt all questions:

1. What is generic about edge detection by Hough transform? Explain Hough space in the context of straight line edges. Explain why Canny's method is sometimes referred to as optimal edge detection stating clearly what those parameters are, that are actually optimized. 2+4+4=10

   OR

   Derive the Marr-Hildreth operator. In what way is it related to visual receptive field? How is the Laplacian operator connected to the diffusion equation? How can the scale space theory be proposed from this equation? 3+2+2+3=10

2. What are the mathematical models of Weber and Fechner in perception? Do you agree with the following statement that 'if stimulus increases multiplicatively, perception increases additively'? Can Mach's mathematical model be called the pioneering contrast theory in vision? What is the basic difference between the Gestalt approach and the contrast approach in vision? 4+2+3+4=13

   OR

   Explain the importance of the following three terms in the context of visual mechanism:

   a) illumination b) reflectance c) luminance.

   How are these related to each other according to Helmholtz? What is the Young-Helmholtz theory of colour vision? What are the two primary pathways of vision? What is raw primal sketch? 6+2+2+1+2=13

3. Explain what you mean by Fourier transform of an image and spatial frequencies. How can you explain brightness assimilation and brightness in terms of these concepts? In what way can binocular disparity assist in depth perception? 4+4+4=12

4. What are the basic components of any vision system? Make a clear demarcation between animal vision and machine vision. How is logarithmic compression at photoreceptor level related to light and dark adaptation? Do you agree with the statement that 'all vision systems necessarily depend on electromagnetic (em) sources'? If not, describe any non-em based vision system, or else justify your answer. 4+5+2+4=15

Indian Statistical Institute

M. Tech Computer Science, II Year

Semester Examination -2011

GROUP – B

Subject – Computer Vision Date: 06- 05-2011

Duration – 1hour 30 min.
Full Marks – 50

Answer any four questions.

Q1. What is meant by the reflectance map of Lambertian objects? In the gradient space how is this reflectance map depicted? What is the nature of the reflectance map for the material of the Moon? Give reasons for your answer. 3+3+3+3

Q2. Using the shape from shading technique, how can you estimate the shape of an object for a rotationally symmetric light source? 12

Q3. What is photometric stereo? Describe a technique, using photometric stereo, to find the shape of an object. Assume the reflectance map to be perfectly linear. 4+8

Q4. What are the error terms to be minimized in the computation of optical flow for the continuous case? Minimize the total error and get your solution. Give an example to show that the optical flow field always does not correspond to motion field of an object.

3+6+3

Q5. Distinguish between gnomonic and stereographic projections. Discuss how you will get the shape of an object using stereographic projection co-ordinates. 6+6

For neatness - 2

Indian Statistical Institute

Advanced Image Processing

M.Tech.(CS)-II year 2010-11

Full marks:100 Time:3 Hours

Date: 10.05.2011

Answer any **ten** questions. All questions carry equal marks.

1. (a) Define: Dilation and Erosion. State and prove the duality property between these two operators. [5]

   (b) Define: Opening and Closing. Prove that closing is an idempotent operator. [5]

2. (a) Derive the expression for gradient magnitude in mean-squares sense using best plane fit approach. [8]

   (b) How is this related to Robert's gradient? [2]

3. (a) Define 'bpp' and 'PSNR'. What is the objective of image compression in terms of 'bpp' and 'PSNR'? [4]

   (b) Write down the basic steps of 'vector quantization'? [3]

   (c) Give an algorithm for codebook generation for VQ image compression. [3]

4. Consider the following block of gray levels:

9	8	2	1
7	6	2	3
8	4	3	6
4	2	7	8

   (a) Calculate the compressed and reconstructed representation of the block using BTC (Block Truncation Coding). [8]

   (b) Calculate the compression ratio and MSE. [2]

5. (a) Define: *Medial Axis.* What do you mean by medial axis transform? [4]

   (b) Describe a two-pass algorithm for Distance Transform and then find the medial axis for 8-connectivity. [6]

6. (a) Define: Euler number or Genus. How a component labeling algorithm can be used to compute Genus? [4]

(b) Describe a morphological algorithm to compute Euler number or Genus? [6]

7. (a) Define: *Principal Axis* of an image. Calculate the moment of inertia of an image about the principal axis. [2+4]

(b) If $\overline{m}_{ij}$ denotes the (i, j) central moments of an image f and θ presents the slope of the principal axis, then prove that $\tan 2\theta = \frac{2\overline{m}_{11}}{\overline{m}_{20}-\overline{m}_{02}}$. [4]

8. (a) Define: *Mutual Information.* How do you compute mutual information between two images. [5]

(b) Suppose that the origin of the $(\acute{x}, \acute{y})$ system is at position (α, β) relative to the (x, y) system; that the scale factors of $\acute{x}$, $\acute{y}$ are λ and ν times those of x and y, respectively; and that the $\acute{x}$-axis makes counterclockwise angle θ with the x-axis. If (x_p, y_p) and $(\acute{x}_p, \acute{y}_p)$ are the coordinates of a point P in the (x, y) and $(\acute{x}, \acute{y})$ systems, respectively, then prove that

$$\acute{x}_p = \lambda(x_p - \alpha)\cos\theta + \nu(y_p - \beta)\sin\theta$$

$$\acute{y}_p = -\lambda(x_p - \alpha)\sin\theta + \nu(y_p - \beta)\cos\theta$$

[5]

9. Describe the thresholding method proposed by N. Otsu. How do you extend this method to obtain multiple thresholds. [7+3]

10. State and prove the correlation theorem. Show that the Fourier transform of the autocorrelation function of $f(x)$ is its power spectrum. [7+3]

11. Write short notes on:

(a) Image registration

(b) Fast Fourier transform [5x2]

Indian Statistical Institute

M. TECH. (CS) Second Year
End-semester Examination
Subject: Advanced Cryptology

Date: 11/05/2011 Time: 3.5 hours Maximum Marks:100

Note: The paper carries 110 marks. Maximum you can score is 100.
Answer Group A and Group B in seperate answer sheets.
Notations used are as defined in the class

Group A

1. Describe Fiat-Shamir Signature scheme. Using Forking Lemma, show that this scheme is EUF-NMA secure in random oracle model assuming factorization problem is hard. [8 + 8]

2. Consider the following Signature Scheme

- **Setup(1^k)**
  (a) Choose a prime number p, such that $2^{k-1} < p < 2^k$
  (b) Choose g such that $< g > = \mathbb{Z}_p^*$
  (c) Choose $\alpha \xleftarrow{\mathcal{R}} \mathbb{Z}_{p-1}$
  (d) Compute $y = g^\alpha \pmod{p}$
  (e) Message Space $(\mathcal{M}) = \mathbb{Z}_{p-1}$
  (f) Signature Space $(\sigma) = \mathbb{Z}_p^* \times \mathbb{Z}_{p-1}$

  Public Key $(PK) \equiv (p, g, y)$
  Secret Key $(SK) \equiv (\alpha)$

- **Sign(m, SK)**
  (a) Choose $r \xleftarrow{\mathcal{R}} \mathbb{Z}_{p-1}$
  (b) Compute $\beta = g^r \pmod{p}$
  (c) Compute $s = r + \alpha m \pmod{(p-1)}$

  Message-Signature pair$= (m, (\beta, s))$

- **Verify($(m, (\beta, s), PK)$)**
  (a) Check $g^s \overset{?}{=} \beta y^m \pmod{p}$

  If above step is correctly verified, return *true* else *false*.

Question

(a) Show that this scheme is not EUF-CMA secure. [8]

3. Define dm-IND-iCCA and dm-EUF-iCMA security of signcryption scheme. Show that Encrypt and Sign paradigm is not fm-IND-iCCA secure. [5 + 5 + 4]

4. Consider the following Signcryption Scheme

- **Setup(1^k)**

  (a) Choose a prime number p, such that $2^{k-1} < p < 2^k$

  (b) Choose g such that $< g >= \mathbb{Z}_p^*$

  (c) Choose $\alpha \overset{\mathcal{R}}{\leftarrow} \mathbb{Z}_{p-1}$

  (d) Compute $y = g^\alpha \pmod p$

  (e) Choose two secure hash functions $H_1 : \mathbb{Z}_p^* \to \{0,1\}^{k/2}$ and $H_2 : \{0,1\}^{k/2} \to \{0,1\}^{k/2-1}$

  (f) Choose a secure Symmetric Encryption algorithm $\mathcal{E}$ and its corresponding Decryption algorithm $\mathcal{D}$

  (g) Message Space $(\mathcal{M}) = \{0,1\}^*$

  (h) Ciphertext Space $(\mathcal{C}) = \{0,1\}^* \times \{0,1\}^{k/2-1} \times \mathbb{Z}_{p-1}$

  Public Parameters $(PP) = (p, g, H_1, H_2, \mathcal{E}, \mathcal{D})$
  Public Key $(PK) \equiv (y)$
  Secret Key $(SK) \equiv (\alpha)$

  (Notation: "α_R" and "y_R" denote the secret key and public key of receiver respectively.)
  ("α_S" and "y_S" denote the secret key and public key of sender respectively.)
  ("$\mathcal{E}_{key}(m)$" denotes encryption of message "m" using symmetric key encryption algorithm "$\mathcal{E}$" with the key "key")
  ("$\mathcal{D}_{key}(c)$" denotes decryption of ciphertext "c" using corresponding symmetric key decryption algorithm "$\mathcal{D}$" with the key "key")

- **Signcryption(m, PK_R, SK_S, PP)**

  (a) Choose $x \overset{\mathcal{R}}{\leftarrow} \mathbb{Z}_{p-1}$

  (b) Compute $h = H_1(y_R^x \pmod p))$

  (c) Compute $r = H_2(h)$

  (d) Compute $s = \frac{x}{r+\alpha_S} \pmod{(p-1)})$

  (e) Compute $c = \mathcal{E}_h(m)$

  Ciphertext $C = (c, r, s)$

- **Designcryption($C = (c, r, s), SK_R, PK_S, PP$))**

  (a) Compute $h' = H_1(((y_S . g^r)^s)^{\alpha_R} \pmod p))$

  (b) Check $r \overset{?}{=} H_2(h')$

  (c) If above step is correctly verified, then return $\mathcal{D}_{h'}(c) = m'$ else $\perp$

Question

(a) Show that this scheme is not fm-EUF-iCMA secure. [8]

5. Describe Boneh-Franklin Identity Based Encryption Scheme. Show that this scheme is IND-ID-CPA secure under Bilinear Diffie-Hellman assumption. [8 + 8]

6. Consider the following Identity Based Encryption Scheme
($\mathbb{Z}_n^{odd} = \{x | 1 \leq x \leq n-1$ and x is odd $\}$)

- **Setup(1^k)**
  (a) Choose two prime numbers "p'"and "q'" such that $2^{k-1} < p' < q' < 2^k$ and $p = 2p' + 1$ and $q = 2q' + 1$ are also prime numbers.
  (b) Compute $n = pq$
  (c) Choose a secure hash function $H : \{0,1\}^* \rightarrow \mathbb{Z}_n^{odd}$
  (d) Message Space $(\mathcal{M}) = \mathbb{Z}_n$
  (e) Ciphertext Space $(\mathcal{C}) = \mathbb{Z}_n$
  Public Parameters $(PP) = (n, H)$
  Master Secret Key $(MSK) = (p, q)$
- **Key Extraction(ID, MSK, PP)**
  (a) Compute $e = H(ID)$
  (b) Compute d such that $ed \equiv 1 \pmod{\phi(n)}$
  Return $SK_{ID} = d$
- **Encryption(m, ID_R, PP)**
  (a) Compute $e_R = H(ID_R)$
  (b) Compute $c \equiv m^{e_R} \pmod{n}$
  Ciphertext $C = c$
- **Decryption($C = c, SK_{ID_R}, PP$)**
  (a) Compute $m' = c^{SK_{ID_R}} \pmod{n}$
  Return m'

Question

(a) Show that above Identity Based Encryption can be totally broken if adversary is provided Key Extraction Oracle only.
(No need to show it rigorously. Just give the sketch of the attack.) [8]

Group B

1. According to the result by Dolev et al., any single round PSMT protocol requires $n \geq 3t+1$ channels. On the other hand, according to the result by Srinathan et al., any single round PSMT protocol over $n \geq 3t+1$ channels has a lower bound of $\Omega\left(\frac{n\ell}{n-3t}\right)$ field elements (from a finite field $\mathbb{F}$) on the communication complexity, to securely send a message $m^{\mathbf{S}}$ containing ℓ field elements from $\mathbb{F}$. That is, any single round PSMT protocol must have a communication complexity of $\Omega\left(\frac{n\ell}{n-3t}\right)$ field elements in order to securely send a message $m^{\mathbf{S}}$ containing ℓ field elements. Now design a single round PSMT protocol which sends a message containing ℓ field elements and have a total communication complexity of $\mathcal{O}\left(\frac{n\ell}{n-3t}\right)$ field elements. You should take $n \geq 3t+1$ and not $n = 3t+1$. Similarly, the message size ℓ should be general and you are not supposed to design a protocol for a specific value of ℓ and n. Your single round protocol should have a communication complexity of $\mathcal{O}\left(\frac{n\ell}{n-3t}\right)$ field elements for any n satisfying the condition $n \geq 3t+1$ and any value of ℓ. You have to prove the secrecy and communication complexity of your protocol. 10 Marks.

   Hint: The special case of $n = 3t+1$ and $\ell = 1$ was discussed in the class. Try to build upon that idea.

2. Suppose there are $n = 2t+1$ channels between **S** and **R**, t of which can be under the control of adversary $\mathcal{A}_t$. **S** has a message $m^{\mathbf{S}}$, which is a sequence of ℓ elements from a finite field $\mathbb{F}$. Suppose **S** wants to *reliably* communicate $m^{\mathbf{S}}$ to **R**, then a very simple solution is *broadcast*: simply send $m^{\mathbf{S}}$ over all the n channels. This requires a communication complexity of $\mathcal{O}(n\ell)$ field elements.

   Now suppose **R** somehow knows the *exact identity* of k out of the possible t channels which are under the control of the adversary where $0 < k < t$. On the other hand, **S** only knows the value k, but not the exact identity of those k corrupted channels. Moreover, there is no way by which **R** can let **S** know the identity of k corrupted channels which are known to **R**. Now in this scenario, both **S** and **R** have some partial information about the corrupted channels: **S** knows the value k and **R** knows the identity of k corrupted channels. Now design a single round protocol, using which **S** can reliably communicate $m^{\mathbf{S}}$ to **R**, such that the communication complexity of the protocol is $\mathcal{O}\left(\frac{\ell n}{k}\right)$ instead of $\mathcal{O}(n\ell)$. 10 Marks.

   Hint: You have to use the property of RS code and the relationship between the number of errors that can be corrected by RS codes, the degree of the polynomial used for encoding in the RS code and the length of the RS code.

3. During our discussion on pseudo-basis, we assumed that $n = 2t+1$. Discuss the problem that will arise if instead we assume $n = 2t$. In another words, state the property which will no longer hold if we assume $n = 2t$. You have to argue mathematically. 10 Marks.

4. Show that there does not exist any single round VSS protocol (i.e., only one round in the sharing phase) with $n = 3t$ parties. Argue mathematically. 10 Marks.

   Hint: Try to relate with another problem which you know is un-solvable in one round with $n = 3t$ entities, of which at most t can be corrupted.

✓

INDIAN STATISTICAL INSTITUTE

M. Tech. (CS) II Year (2010-11): II semester

Semestral Examination

ADVANCED PATTERN RECOGNITION

Date: 12.05.2011 Duration: 195 minutes Marks: 100

Note: This paper carries 115 marks. Answer as much as you can. Maximum marks you can get is 100.

1. (a) Distinguish between probability and fuzzy membership.
   (b) Discuss about the parameters of fuzzy c-means cluster seeking algorithm.
   (c) Give sketches of three typical recognition problems (with two features and two classes) where a fuzzy classification technique can be effective but other conventional classification approaches may not provide satisfactory performance. **[5+3+3]**

2. Describe the back propagation learning algorithm for the multilayer perceptron (MLP). **[8]**

3. (a) Consider a two-class (C_1 and C_2) problem with the following training samples, each having two features (x_1 and x_2),

   $C_1 = \{(1,1)', (1,3)', (3,1)'\}$ and $C_2 = \{(2,1)', (2,3)', (3,2)'\}$.

   Use the Gini impurity index to create a full grown decision tree for this data.
   (b) Explain the concept of surrogate splits in a decision tree. **[12+4]**

4. (a) Describe k nearest density estimation technique.
   (b) Derive k- nearest decision rule based on the above density estimation procedure. **[5+5]**

5. (a) Define the maximal information compression index.
   (b) State a feature selection procedure based on the index.
   (c) Describe the perceptron learning algorithm. **[3+7+8]**

6. Write short notes on the following concepts:
   (i) k-fold cross validation,
   (ii) VC dimension,
   (iii) Support vectors,
   (iv)Karush Kuhn Tucker (KKT) conditions,
   (v) Multi-objective genetic algorithms (MOGA). **[5x4]**

7. (a) Describe selection operation in genetic algorithms.
   (b) Describe elitist model of genetic algorithms. **[6+6]**

P.T O.

8. Let $I = <X, A>$ be a decision table, where $X = \{x_1, ..., x_7\}$ is a nonempty set of finite objects (the universe) and $A = C \cup D$ is a nonempty finite set of attributes. Here, $C = \{A_1, A_2\}$ and $D = \{Walk\}$ are the set of condition and decision attributes, respectively.

X	A_1	A_2	***Walk***
x_1	16-30	50	yes
x_2	16-30	0	no
x_3	31-45	1-25	no
x_4	31-45	1-25	yes
x_5	46-60	26-49	no
x_6	16-30	26-49	yes
x_7	46-60	26-49	no

Explain lower and upper approximations, boundary region, degree of dependency and significance of an attribute with the above example data. **[5x2=10]**

9. (a) Distinguish between decision-theoretic and syntactic approaches to pattern recognition.

(b) Describe the basic operations of a syntactic classifier with the help of a block diagram. **[4+6]**

INDIAN STATISTICAL INSTITUTE
End-Semester Examination: 2010-11(Second Semester)

Course Name: M.Tech. in Computer Science
Subject Name: Distributed Systems

Date: 13. 05. 2011 *Maximum Marks: 100* *Duration: 3 hours*

Instructions:

Answer ***all*** *questions from Part A and any* **four** *from Part B. All parts of a question must be answered in the same place.*

Part-A: *Answer all questions*

1. Indicate whether each statement below is true or false. No justification is necessary.
   (a) In Lamport's algorithm for distributed mutual exclusion, a process can enter the critical section only when its request is at the top of the request queues of all processes.
   (b) In order to break a deadlock in the OR-request model, we must abort at least one process from each cycle of the wait-for dependency graph.
   (c) There can be no deterministic algorithm for leader election in anonymous networks.
   (d) Traversal algorithms are not wave algorithms.

[4 X 2 = 8]

2. The clockwise order of process ids in a ring is 8 – 2 – 3 – 6 – 11 – 5 – 7 – 1 – 9 – 10 – 12 – 4. Indicate the set of processes remaining in contention at the end of each round if the Hirschberg Sinclair algorithm is executed to elect the process having maximum id as the leader. Assume all the processes start together and the system is synchronous.

ROUND-1	
ROUND-2	
ROUND-3	
ROUND-4	

[4]

3. The vector timestamps of some events are given below.

TS (e_1) = ⟨3, 9, 2⟩ TS (e_2) = ⟨2, 1, 2⟩ TS (e_3) = ⟨5, 7, 5⟩

TS (e_4) = ⟨6, 7, 7⟩ TS (e_5) = ⟨5, 3, 9⟩ TS (e_6) = ⟨6, 3, 9⟩

Which events are concurrent with e_3? [4]

P.T.O

/

4. In an implementation of Birman-Schiper-Stephenson's protocol to run on a completely connected network with FIFO links, checking for the second condition ($\mathbf{C_j\,[k] \geq VT_m[k]}$ **for all** $\mathbf{k \neq i}$) got left out by mistake. Show clearly with an example why the protocol will not work correctly in this system. [4]

Part-B: *Answer any four questions*

1. **[Self-Stabilization and Deadlock Detection]**

   (a) Consider a variant of Dijkstra's self stabilizing token ring algorithm in which there are two exceptional machines (as opposed to only one in Dijkstra's solution) – both having the same definition. Will the processes stabilize to states where exactly one process has the privilege? Will they deadlock? Justify your answer.

   (b) Analyze the truth of this claim: *In the OR-request model, a deadlocked process must belong to one or more directed cycles in the wait-for graph.* Your answer must begin with True/False and then provide the justification/counter-example.

   (c) What will be the impact of a message loss in Chandy et al.'s diffusion computation based deadlock detection algorithm for the OR-request model. Will it miss a real deadlock? Will it report a phantom deadlock? Justify your answer.

   [8 + 6 + 6 = 20]

2. **[P2P and Cloud]**

   (a) A word's anagram is a different word that is written with the same letters (e.g. sail and lisa). The goal is to discover all the anagrams in a large volume of text. The final output must group all words that are anagrams of each other and must not include words that have no anagrams.

   Write the pseudo codes for the map & the reduce functions that implement a solution to the above problem.

   (b) We have a network composed of 4 peers which are completely connected. With this network we want to build a 2-Dimensional CAN Network, with the constraint that each peer will manage approximately the same address space area. Present a network joining algorithm to achieve this goal. Your algorithm should consider the fact that some parts of the address space may be queried more frequently than others. [10 + 10 = 20]

P.T.O

3. **[Agreement Protocols and Leader Election]**

a) Explain how a solution to the Byzantine agreement problem can be used for a solution to the problems of consensus and interactive consistency.

b) Consider the leader election algorithm which uses extinction on waves. For each of the following cases present a brief (one/two line) justification indicating whether the case is possible if a message is lost in the algorithm?

i. A wrong leader is elected (that is, not the one with the best id)

ii. No leader is elected

iii. Two groups of processes elect two different leaders

[8 + 12 = 20]

4. **[Routing Algorithms and Spanning Trees]**

(a) Enumerate the main differences between the Chandy-Misra shortest path routing algorithm and the distributed extension of the Floyd-Warshall algorithm. Your answer should be short and to-the-point.

b) Consider an asynchronous system with reliable links and a given node r. Design an algorithm to create a BFS spanning tree of the network rooted at r. At the termination of the algorithm, every node should know its parent in the BFS tree and the root r should know that the BFS tree has been built completely. Can you do it with polynomial message complexity? [10 + 10 = 20]

5. **[Balanced Sliding Window Protocol, Wave and Traversal Algorithms]**

(a) Prove that the Balanced Sliding Window Protocol satisfies the liveness requirement.

(b) What is the main modification done by Awerbuch on the classical depth-first search algorithm? Analyze the effect of this modification on the time complexity and message complexity of the algorithm.

[10 + 10 = 20]

INDIAN STATISTICAL INSTITUTE

Semestral Examination: (2010 – 2011)

M.Tech. (CS) II Year

Parallel Processing: Architectures and Algorithms

Date: 16 /05/2011 Maximum Marks: 100 Duration: 3 hrs

Answer any five questions

1. a) State Bernstein's conditions for parallelism.

b) Given two 2 × 2 matrices A and B, it is required to compute the product matrix P = A× B, and finally the sum S of the four elements of P.

(i) Show the program graph assuming the operations 'multiply' and 'addition' as fine grains.

(ii) For an arbitrary processor, it is given that the addition operation needs 20 CPU cycles, and the multiply operation takes 100 CPU cycles. The interprocessor communication delay is 200 cycles. Assume that data are initially loaded in respective processors.

Show the scheduling of the fine grain program using maximum number of identical processors required to utilize the software parallelism existing in the program. Find the speed-up and utilization.

(iii) Use grain packing to optimize the number of processors to improve the speed-up and utilization. Compare the results with those obtained in (ii) and justify your answer.

[3+(4+7+6)=20]

2. a) Show the block diagram of an 8 × 8 omega network using 2 × 2 switches. Find the number of permutations realizable in a single pass by the network and justify your answer.

b) A permutation P: 0 →6, 1 →1, 2 →5, 3 →2, 4 →4, 5 →0, 6 →7, 7 →3, is to be realized on an 8 × 8 omega network with 2 × 2 switches. Show the path matrix, and hence draw the conflict graph. What is the minimum number of passes required to route P in the network?

c) Draw the block diagram of a 3-stage non-blocking N × N Clos' network using (n × m) switches at the input stage. Prove that (2n-1) middle-stage switches are sufficient for non-blocking operation of the network.

[(3+3)+(3+3+2)+(3+3)=20]

3. Draw a network for sorting 8 elements using Batcher's odd-even merge sort technique. Write down the parallel algorithm for sorting **n** elements following this method.

Derive an expression for the time required to sort **n** elements. Hence justify if the algorithm is cost-optimal or not.

[(4+6+7+3) =20]

P.T.O

4. Consider an $\mathbf{n} \times \mathbf{n}$ mesh of processors where the boundary processors, i.e., the processors in row 1 and column 1, are only capable of handling input-output operations. Design an ***O(n)*** algorithm for multiplying two $\mathbf{n} \times \mathbf{n}$ matrices A and B on this architecture. Show necessary diagrams to explain the operations. Justify that this is the fastest possible algorithm for the given architecture and find the cost.

[12 + 3 + 5 = 20]

5. a) Consider a CRCW SM SIMD computer with $\mathbf{n}^3$ processors. Given the $\mathbf{n} \times \mathbf{n}$ adjacency matrix of an **n**-node graph *G*, design an algorithm to compute the connectivity matrix *C* of *G* on this computer in ***O(log n)*** time where

c_{jk} = 1, if a path exists from node v_j to v_K, or k = j,

= 0, otherwise.

Mention how the simultaneous write operations to the same memory location are to be resolved.

b) Given an array {a_1, a_2, ..., a_n}, write an algorithm for computing the prefix sums S_i =(a_1 + a_2 +..... + a_i), for $\mathbf{1 \leq i \leq n}$, in ***O(log n)*** time on an EREW SM SIMD machine.

[(10+2)+8=20]

6. Answer in brief:

a) Prove that H_n, a hypercube of order n, is Hamiltonian, for n ≥ 2.

b) Find an expression for the Moore bound on the number of nodes in a regular graph with degree **d** and diameter **k**.

c) Show an embedding of a **5** × **3** mesh on a **16**-node hypercube with dilation 2.

d) Show that the bisection width of a pyramid of size $\mathbf{P = 4^k}$ is $\mathbf{2\sqrt{P}}$.

[5 × 4 = 20]

Indian Statistical Institute

Semester Examination 2010-2011
M. TECH. (CS) II Year
Subject: Document Processing and Retrieval
Full Marks: 100 Duration: 3 hrs.

16·05·11

(Answer all questions)

1. Discuss a method to generate 400 dimensional gradient based features for character recognition. [10]
2. What is the tilt of a word image? How does tilt differ from skew? Describe a method to correct the tilt of an English word image. [2+2+5]

3. Generate a feature based tree classifier to recognize the following printed alphanumeric characters **(D,O,Q,6,9,4, S,Z,X,Y,K).** [10]

4. What do you mean by a multi-script document? Discuss the OCR technology for multi-script documents. Discuss two robust features to separate printed degraded Devnagari script from Roman script. [2+4+6]

5. What are the differences between holistic recognition and segmentation based recognition? What do you mean by real-word errors and non-word errors? Discuss about different types of touching that may appear in the text of a document. [3+3+6]

6. Let $x_1,x_2, x_3 \ldots\ldots x_n$ be n consecutive contour pixels of a character. Discuss a rotation invariant feature extraction method based on $x_1,x_2, x_3 \ldots\ldots x_n$. [6]

7. In the context of information retrieval, describe the binary independence model for probabilistic document relevance ranking. Mention clearly the assumptions associated with it.
Given a query and a document, define the odds of relevance of the document. Using suitable assumptions, simplify the odds of relevance in terms of the probabilities of a term appearing in relevant and non-relevant documents. Describe a method to iteratively estimate these probabilities using relevance feedback. [5 + 3 + 7 + 5 = 20]

8. Write short notes on (a) Word-spotting (b) Profile features (c) Signature verification (d) Hidden Markov model [4x5=20]