PhD Qualifier Examination

Department of Computer Science and Engineering

Date: 03-Nov-2016 **Part A, SIT Syllabus** **Maximum Marks:** 50

Answer any five questions.

A.1 Write only the answers (no need of any explanation). **($2\frac{1}{2}\times 4$)**

(a) In each case, state whether $f(n) = O(g(n))$, or $f(n) = \Omega(g(n))$, or both (that is, $f(n) = \Theta(g(n))$).

i) $f(n) = 5\log(n)$, and $g(n) = \log(n^3)$.

ii) $f(n) = n2^n$, and $g(n) = 3^n$.

(b) Given a sorted array A and an interval $[a,b]$, we have to find how many elements of A lie in this interval. What is the worst-case time complexity of the best known algorithm for this?

(c) Let S and T be two binary search trees with m and n nodes, respectively. All the elements in S are distinct. Similarly, all the elements in T are distinct. We have to find the common elements of S and T. What is the worst-case time complexity of the best known algorithm for this?

(d) Which of the following operations is/are asymptotically more efficient in a height-balanced binary search tree compared to a binary max heap?

i) searching a key element
ii) insertion
iii) deletion of maximum element
iv) deletion of minimum element

A.2 You are given an array $A[0\ldots n-1]$ storing exactly n of the $n+1$ integers $0,1,\ldots,n$. This means that exactly one integer x in the range $0,1,\ldots,n$ is missing in A. Your task is to determine x.

(a) If A is unsorted, propose an $O(n)$-time $O(1)$-space algorithm to find x. **(5)**

(b) If A is sorted (in ascending order), propose an $O(\log n)$-time $O(1)$-space algorithm to find x. **(5)**

A.3 Consider a rooted tree T with each node storing an integer key. Each node may have an arbitrary number of children. Let us impose a left-to-right ordering on the children of each node. We then index the nodes as $0,1,2,3,\ldots$ level by level starting from the root, and in each level from left to right (see Figure 1 for an example). We use an array A of integer pairs to store the tree as follows. Consider the node at index i in the tree. Let k be the key stored at this node, and let the index of the parent of this node be p. Then, the i-th entry in the array A is the pair (k,p). The index of the parent of the root is -1. Let n be the number of nodes in T. Assume that A is read-only and cannot be modified.

(a) Propose an algorithm that, given the array A and its size n, prints the keys stored in all the leaf nodes of the tree. Your algorithm should run in $O(n)$ time and use $O(1)$ extra space. **(5)**

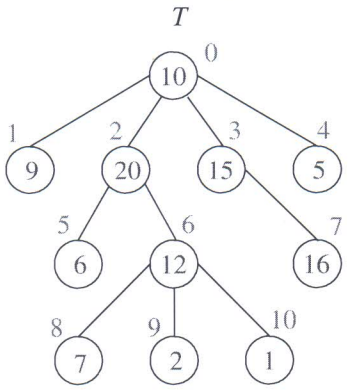
(b) Propose an algorithm that, given A, its size n, and an index $i \in \{0,1,2,\ldots,n-1\}$, prints the keys stored in all the child nodes of the node stored at index i. Your algorithm should run in $O(t+\log n)$ time and use $O(1)$ extra space, where t is the number of children of the node at index i. **(5)**

A.4 You are doing some stress-testing on a particular model of glass jars to determine the height from which they can be dropped without being broken. The setup for this experiment is as follows. You have k jars of the particular model, and a ladder with n rungs. You want to find the highest rung from which you can drop a jar and not have it broken. Call this the *highest safe rung*. The problem is to determine the highest safe rung in as few jar-drops as possible. You can reuse a jar as long as it is not broken.

(a) Suppose $k = 1$. Devise a strategy to determine the highest safe rung with at most n jar-drops. **(3)**

(b) Suppose $k = 2$. Devise a strategy to determine the highest safe rung with at most $f(n)$ jar-drops, where $f(n)$ is $o(n)$ (that is, $\lim_{n\to\infty} \frac{f(n)}{n} = 0$). **(7)**

Figure 1: Example for Question A.3



	0	1	2	3	4	5	6	7	8	9	10
A	10	9	20	15	5	6	12	16	7	2	1
	−1	0	0	0	0	2	2	3	6	6	6

A.5 (a) Define a weighted graph, and explain its adjacency-list representation with a suitable example. **(1+2)**

(b) Let $G = (V, E)$ be a weighted, undirected, connected graph. Each of its edges has weight either 1 or 2. What can be the minimum possible weight of a/the Minimum Spanning Tree (MST) of such a graph? Suggest an efficient algorithm to decide whether the given graph G has an MST with this minimum possible weight. Explain its worst-case time complexity. Assume that the input graph G is specified by an adjacency-list representation. **(1+4+2)**

A.6 You are given an unsorted array $A = (a_0, a_1, \ldots, a_{n-1})$ of n integers (positive, negative, or zero). A *run* in A is an interval $[i, j]$ of indices (not necessarily maximal) such that $a_i \leqslant a_{i+1} \leqslant \cdots \leqslant a_{j-1} \leqslant a_j$. Write an efficient C/C++/Java function that, given A and n, computes a run in A with the largest possible sum. Notice that the sum of an empty run is zero. **(10)**

A.7 Let $A = (a_0, a_1, \ldots, a_{n-1})$ be an array of n positive integers. We call A *super-increasing* if $a_i > a_0 + a_1 + \cdots + a_{i-1}$ for all $i = 1, 2, 3, \ldots, n-1$. For example, the array $(2, 3, 6, 14, 28, 60, 125)$ is super-increasing, whereas the array $(2, 3, 6, 11, 20, 50, 100)$ is not super-increasing.

(a) Write an efficient C/C++/Java function that, given A and n, determines whether A is super-increasing. **(4)**

(b) Write an efficient C/C++/Java function that, given a super-increasing array A, its size n, and a positive integer x, determines whether x can be written as

$$x = a_{i_1} + a_{i_2} + \cdots + a_{i_k}$$

for some $k \geqslant 1$, and for array indices satisfying $0 \leqslant i_1 < i_2 < \cdots < i_k \leqslant n-1$. **(6)**

A.8 Let $A = (a_0, a_1, \ldots, a_{n-1})$ be a sorted array of n distinct positive integers, B a positive integer, and $C = (c_0, c_1, \ldots, c_{n-1})$ a second array of n positive integers (not sorted, and may contain duplicate entries). A set $(i_1, i_2, \ldots, i_k)$ of indices satisfying $0 \leqslant i_1 < i_2 < \cdots < i_k \leqslant n-1$ for some $k \geqslant 1$ is called *feasible* if $a_{i_j} - a_{i_{j-1}} \geqslant B$ for all $j = 2, 3, \ldots, k$. The *cost* of a feasible set $(i_1, i_2, \ldots, i_k)$ of indices is $c_{i_1} + c_{i_2} + \cdots + c_{i_k}$. Your task is to maximize the cost over all feasible sets of indices. Call this maximum cost $M(n)$.

(a) Derive a recurrence relation for $M(n)$, that is, express $M(n)$ in terms of $M(i)$ for $i = 1, 2, \ldots, n-1$, where $M(i)$ is the maximum cost for the first i items $(A[0 \ldots i-1], B, C[0 \ldots i-1])$. Also supply the required initial condition(s). (**Hint:** For deriving the recurrence, consider the two cases: $i_k < n-1$ and $i_k = n-1$, where i_k is the last index in a feasible set. If $i_k < n-1$, then all the indices in the feasible set are chosen from $0, 1, 2, \ldots, n-2$. If $i_k = n-1$, consider the constraint $a_{i_k} - a_{i_{k-1}} \geqslant B$.) **(4)**

(b) Write an efficient C/C++/Java function to compute $M(n)$ using the recurrence of Part (a). **(6)**

Software Engineering

Total Marks =25

Question Nos. 1 to 10 carry one mark each. Indicate exactly one option that you feel is the most appropriate. [1x10=10]

1. The purpose of error seeding is which one of the following?
   a) Determine the origin of the bugs
   b) Plant Trojans
   c) Determine the number of latent bugs
   d) Plant bugs in the code to check the efficacy of maintenance

2. Which one of the following types of testing is usually not performed during system testing?
   a) Stress testing
   b) Functionality testing
   c) Recovery testing
   d) White box testing

3. During unit testing of a software, why is it important to test the boundary values?
   a) It reduces testing costs as boundary values are easily computed by hand.
   b) Debugging is easier when testing boundary values.
   c) The correct execution of a function on all boundary values proves that the function is correct.
   d) Programming the boundary conditions is usually error-prone in practice.

4. If branch coverage has been achieved on a unit under test, which one of the following test coverage is implicitly implied?
   a) Path coverage
   b) Multiple condition coverage
   c) Statement coverage
   d) Data flow coverage

5. Which one of the following class relations is indicated among the class(es) in the following statement: "A course is the prerequisite of many courses and has many courses as its prerequisite."
   a) N-ary association
   b) Unary association
   c) Binary association
   d) Aggregation

6. Consider the following requirement for a certain software to be developed: "The output of the program shall be produced within 10secs of any input being given by the user 90% of the time." Which one of the following can be said about this requirement?
   a) Ambiguous and Verifiable
   b) Ambiguous and Nonverifiable
   c) Non-ambiguous and Non-verifiable
   d) Non-ambiguous and Verifiable

7. Consider the following partial description of a security software. "An infrared sensor would detect any intrusion and alert the security officer." Which one of the following types of requirements is this?
   a) Functional
   b) Non-Functional
   c) Design Requirement
   d) Design constraint

8. Consider the statement: "An employee is either a worker or a manager." Assuming that Employee and Manager to be two classes, what can be said about the relationship between these two classes?
   a) Association
   b) Generalization-specialization
   c) Containment
   d) Polymorphism

9. If a software product of size S takes m months to develop, then according to the COCOMO II early design stage estimation model, how long (in months) will it take to develop a product of size 2×S?
   a) Greater than 2 × m months
   b) Greater than 3 × m months
   c) Less than 2 × m months
   d) Greater than 4 × m months

10. What is the correct order in which a software project manager estimates various project parameters while using COCOMO?
   a) cost, effort, duration, size
   b) cost, duration, effort, size
   c) size, effort, duration, cost
   d) size, cost, effort, duration

11. Represent the following program using a DFD (Data Flow Diagram). **[5]**

```
main(){
   int a[100],b[100];
   for(i=0; i<100;i++)  f(a[i],b[i]);
}
```

```
f(int a,int b){
 if (a>b) f1();
      else f2();
      }
```

```
f1(){
     return;
 }
```

```
f2(){
     return;
 }
```

12. Develop the use case diagram for the following web-based Inventory Management System (IMS). You need not provide a text description for your use case diagram. **[5]**

    Using IMS, the store keeper should be able to view the specific products that are in stock. IMS should allow the store keeper to generate an indent for the goods whose stock he observes to be below a threshold. To generate an invoice, he must first log in. The sales personnel use a separate sales management software (SMS) for carrying out sales transactions. When a sales transaction on sales management software (SMS) completes, a reduce inventory function of the IMS is automatically invoked by SMS. When supply for any goods arrives, the store keeper should be able to update the inventory. The store keeper should be able to generate a report summarizing the daily transactions.

13. Consider the following description about the academic set up of a certain Institute. This description forms a part of the requirements specification document that has been prepared to for automating the clerical activities associated with the academic activities of the Institute. Develop the class diagram that can be inferred from the description. **[5]**

    Some instructors are professors, while others have job title adjunct professor. Departments offer many courses, but a course may be offered either by a single Department or at most jointly by two departments. Courses are taught by instructors, who may teach up to three courses. Instructors are assigned to one department. But, at times instructors are given joint appointment in two Departments. One of the instructors of a Department also serves as the Head of the department.

Comprehensive Exam Autumn 2016-17

Database Management System (Full Marks: 25)

1. A university has a large number of students and several departments. A student can belong to multiple departments (i.e., there is possibility of joint enrolment). Faculty members can also belong to multiple departments. Each department offers many subjects. However, each subject is offered by only one department. The following additional points may be noted.
   I. Each department has a department name and is uniquely identified by a department code.
   II. A student is uniquely identified by a roll no. Other attributes of student are name, date of birth and address. A student can have one or more email addresses.
   III. A subject has a subject code, a subject name, no. of credits and the semester (Spring/Autumn) in which it is offered.
   IV. A student has to register for several courses. For each course registered by a student, the year (e.g., 2016, 2017, etc.) needs to be maintained and also the grade obtained.

   Answer the following questions based on the above data requirement specifications:
   a. Draw an Entity-relationship (E-R) diagram for capturing the above-mentioned information, clearly identifying all the entities, relationships, cardinalities and attributes (including the type of attribute like key attribute, multi-valued attribute, derived attribute, etc.).
   b. Derive the relational schema from the above E-R model. The schema should not have any redundant relation and it should be in 3NF. For each relation, clearly identify all the columns, primary and foreign keys.

   [4+4=8]

2. Construct a B+ tree for the following set of key values: 5, 7, 10, 17, 8, 20, 16, 11, 18, 9. Consider that the keys are inserted in the above specified order. The maximum number of key values that can fit in any node (leaf as well as non-leaf) is 3. Show the tree at the end of each insert. Once the tree construction is complete, perform the following operations: Delete 17, Delete 16, Insert 17, Delete 18. Show the tree at the end of each operation.

   [5+4=9]

3. Consider a relational schema with the following relations P(AB), Q(ABD) and R(AE):

P(AB)

A1	B2
A1	B1
A2	B3
A3	B1
A3	B3
A4	B2

Q(ABD)

A1	B1	D1
A1	B2	D2
A2	B3	D1
A1	B1	D3
A3	B3	D1
A3	B1	D1
A4	B2	D2

R(AE)

A1	E1
A2	E1
A3	E1
A4	E3
A5	E4
A6	E4

Show the intermediate results and the final result of evaluating the following relational algebra query on the above schema:

$$\Pi_{AD}[(\sigma_{(B="B1")}[P \bowtie R]) \bowtie (\sigma_{(D="D1")}[P \bowtie Q])] \quad \textbf{[4]}$$

4. Consider a schedule of 3 transactions T1, T2 and T3 as shown below where r1(x) denotes "transaction T1 reads data item x", w3(m) denotes "transaction T3 writes data item m", etc.

   S: r2(x), r1(m), w3(x), w2(m), r3(x), w1(y), w2(y), w3(m), r1(x)

   Is the schedule S given above (a) view serializable (b) conflict serializable? **[2+2=4]**