Ph.D. Comprehensive Examination

Computer Science Department
University of Miami

August 19, 2022

Student Name: Jia Deng

Student Number:

Problem number	Points (10 max)
1	5.625
2	
3	0
4	
5	
6	5
7	2
8	
9	
10	5
Total:	

point gained: 5.625

point gained: 4.375

1. **Data organization; Algorithms and complexity**

   Each item x in a set S has a unique key $key[x]$. We need to implement the following operations.

   (a) Search (S, key)
   (b) Insert (S, x)
   (c) Successor (S, x)
   (d) Predecessor (S, x)

   Give the 4 running times as an O() for the following implementations:

   (a) Ordered (sorted) array,
   (b) Ordered doubly linked list,
   (c) Min-Heap, and
   (d) Hash table

(a) ordered array
- search $O(\log(n))$
- wrong Insert $O(\log(n))$
- wrong Successor $O(\log(n))$
- wrong Predecessor $O(\log(n))$

(b) Ordered doubly linked list
- Search $O(n)$
- wrong Insert $O(n)$
- wrong Successor $O(n)$
- wrong Predecessor $O(n)$

(c) Min-Heap
- Search $O(\log(n))$
- wrong Insert $O(\log(n))$
- wrong Successor $O(\log(n))$
- wrong Predecessor $O(\log(n))$

(d) Hash table
- Search $O(1)$
- Insert $O(1)$
- Successor $O(1)$
- Predecessor $O(1)$

2. **Program control and structure; Programming language and notations**

Suppose that procedure *swap* is declared as follows:

```
procedure swap( x, y: integer);
    procedure f(): integer;
        var z: integer;
        begin // f
            z = x; x = y; return z;
        end // f
    begin // swap
        y = f();
    end // swap
```

Describe the effect of the procedure call *swap*(i, A[i]) under each of the following parameter passing methods:

(a) Call-by-value

(b) Call-by-reference

(c) Call-by-value-result

(a) Procedure swap in effect assign the value of i to A[i]. Since f() change the value of z to the value of x, statement. y = f() change the value of A[i] to the result of f(), i.

(b) Call-by-reference pass the memory address of the parameters. Swap in effect switch the address between i and A[i]

(c) call-by-value-result in effect switch the value between variables i and A[i]

no mathmatical content

point gained:

0

3. **Software engineering**

From the software engineering point of view, any software development process can be divided into several sub-disciplines:

(a) Requirement Analysis
(b) Functional Specification
(c) Architectural Design
(d) Implementation
(e) Testing and Evaluation
(f) Maintenance

Choose three sub-disciplines or tasks within these sub-disciplines that involve a mathematical approach, and illustrative them with examples.

(a) Requirement Analysis may involve mathematical approach. If we require a system that produces an accuary above certain level, or works properly with certain confidence level. We should analyse the goal mathematically.

(d) Implementation involve mathematical approach, complex algorithms often come with maths theoritical background. We need to implement the methods in accordance with the theories.

(e) Testing and evaluation should involve mathematical approach. In order to evaluate the performance, we need to quantify the results and analyse them.

4. **Systems**

(a) dynamic linked libraries can support shared library code, allowing one copy of a library routine to be used by several different processes.

absolute **relative** **static** **dynamic** **none of these is correct**

(b) When it is not known at compile time where a process will reside in memory, relocatable code must be generated.

logical **physical** **absolute** **relocatable**

(c) A UNIX process calls *fork()* to create a child process as shown: *pid = fork()*;

i. What value will be assigned to pid in the parent process by the call to *fork()*?

the parent's process id **the child's process id** **zero** **none of these**

ii. What value will be assigned to pid in the child process by the call to *fork()*?

the parent's process id **the child's process id** **zero** **none of these**

(d) The Banker's algorithm is used for deadlock prevention.

denial **prevention** **avoidance** **recovery**

(e) Belady's anomaly can affect the performance of the FIFO page replacement algorithm.

FIFO **LRU** **optimal** **SJF**

(f) direct access files are made of fixed length records that allow programs to read and write records in no particular order.

sequential **direct** **logical** **none of these is correct**

(g) When an I/O request is being handled for a user's process, which term refers to the policy of returning control to the user process before the I/O is completed?

synchronous I/O **asynchronous I/O** **delayed I/O** **none of these**

(h) Which multithreading model requires that a new kernel thread be created for each new user thread?

many-to-one **one-to-one** **many-to-many** **none of these is correct**

(i) A process that does not affect, and is not affected by, another process is referred to as:

static **independent** **cooperating** **dynamic** **unbounded**

5. **Software, Programming Techniques**

Given that

$B(x)$ means "x is a bear"
$F(x)$ means "x is a fish", and
$E(x, y)$ means "x eats y",

what is the best English translation of

$\forall x[F(x) \rightarrow \forall y(E(y, x) \rightarrow B(y))]$?

(a) All fish eat bears.

(b) Every fish is eaten by some bear.

(c) Bears only eat fish.

(d) Every bear eats fish.

(e) Only bears eat fish.

6. **Networking and Communications**

Point gained: 2

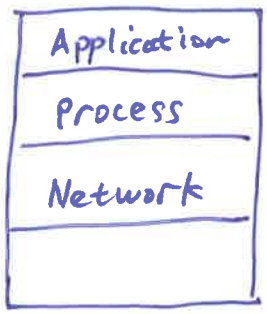
(a) Draw a diagram showing layers of the Internet Protocol Stack and briefly discuss role of each layer.

2

(b) Describe functions of each layer when a file is transferred from a source to destination using (*file transfer protocol* (FTP).

0

(a)



Application layer handles the programs for transfering data packets.

Process layer manages every processes.

Network layer controls the network infrastructure.

2

Type text here

point gained: 10

7. **Algorithms and complexity**

Describe an algorithm that takes two input lists of integers $A = a_1, ..., a_n$ and $B = b_1, ..., b_m$ and delivers the list of all the elements that belong to A but not to B. A and B do not contain redundant elements, however, the elements of A and B might have a large range.

The algorithm should run in $O(n \log m + m \log m)$ time.

Sort list B by using MergeSort, this should run in $O(m \log m)$ time. Create a empty list Result.

Then check each input item whether it's in list A and list B. If the integer is in list A but not in B, append it to Result.

Output the result list.

Search a item in sorted list B runs in $O(\log m)$. Assume input size n, checking every input runs in $O(n \log m)$ time. Then the algorithm has the run time of $O(n \log m + m \log m)$.

8. **Automata and language theory**

Consider the following grammar:

$G \longrightarrow S\ \$\$$
$S \longrightarrow A\ M$
$M \longrightarrow S \mid \epsilon$
$A \longrightarrow a\ E \mid b\ A\ A$
$E \longrightarrow a\ B \mid b\ A \mid \epsilon$
$B \longrightarrow b\ E \mid a\ B\ B$

S
Am
AS | A ε
A* ε
aE | bAA
aaB
a
aabE
aaBB
baE aE
| baaB aaB

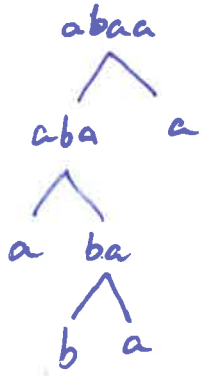
(a) Describe the language that the grammar generates in English.

(b) Show a parse tree for the string `a b a a`.

(c) Is the grammar LL(1)? If so, show the parse table; if not, identify a prediction conflict.

(a) The language generates strings that contain {a, b}

(b)



(c) LL(1)

9. **Discrete Structures**

Recall that the Hamiltonian Cycle Problem is the problem of deciding, on input graph G, whether G has a cycle that visits all the nodes exactly once. Show that this problem is polynomial time decidable if the input is restricted to the graphs with the property that each node has at most two neighbors (i.e., at most two adjacent nodes).

The input graphs with nodes that have at most two neighbors.

We can write a program to traverse the input graph. It starts with a random node, select an node that connects to the it which hasn't been visited and doesn't form a cycle. Otherwise it backtrack to the neighbor which hasn't been selected.

If it traverses all nodes and comes back to the starting node, there is a Hamiltonian cycle. This algorithm runs in polynomial time. Therefore the problem is polynomial time decidable.

lacking technical operations
point gained: 5 points

10. **Other Topics**

Give a detailed explanation of any one approach to machine learning. Give a substantial example that illustrates the technical operation of the approach, and demonstrates interesting knowledge learned.

Random Forest is a machine learning approach that belongs to ensemble learning family.

The purpose of ensemble learning is to combine the predective power of many weak learners through some voting mechanism. In the case of RF, it usually consists of large number of decision trees. Each tree is trained on a subset of the feature space, i.e. randomly selected features from the training data. The input data is also randomly sampled by using bootstrap. Therefore each tree in the forest learns some properties of the training dataset.

when predicting new data samples, each decision tree makes a prediction on the labels. The class label which gets the majority vote is the final result for classification.

RF is very effective in terms of classification accuracy. It's also very efficient and fast. The randomness in data sampling and feature space partition boost the performance of RF. The method of ensemble is also very powerful since the majority vote significantly improve accuracy compared to individual trees.