# Ph.D. Comprehensive Examination

Computer Science Department
University of Miami

August 14, 2020

Student Name: Jiaxong Chen Student Number:

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

1. **I.A Data organization; III.A 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, $O(n)$, $O(1)$, $O(n)$, $O(1)$
   (b) Ordered doubly linked list, $O(n)$, [illegible], $O(1)$, $O(1)$
   (c) Min-Heap, and [illegible], $O(\log(n))$, $O(\log(n))$, $O(\log(n))$, $O(\log(n))$
   (d) Hash table $O(1)$, $O(1)$, $O(1)$, $O(1)$

2. **I.B Program control and structure; I.C 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) Call by value [illegible]

Z = X, [illegible], [illegible].

[illegible] Z is assigned by a specific value X, not change.

(b) Call by reference:

If a variable value changes, [illegible] the variables points to that [illegible] also changes.

(c) Call by value-result:

A variable is pointed to a variable which is a result of a function.

3. **I.D 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) communicate with the clients and understand the requirment of their needs. Then, translate the requirments as needed functions. For example, a bank needs a system (data) to store all transactions.

(c) Based on the list of the system functions, design an architecture of the system. Needs to split the functions into small pieces. For bank system, the choice of computer language, different parts including security etc..

(d) Based on the architecture of the system, implement the functions into codes., for a bank system, write the codes to generate the data system.

4. **I.E Systems**

(a) _____ 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, _____ 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 _____.

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

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

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

(f) _____ 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:

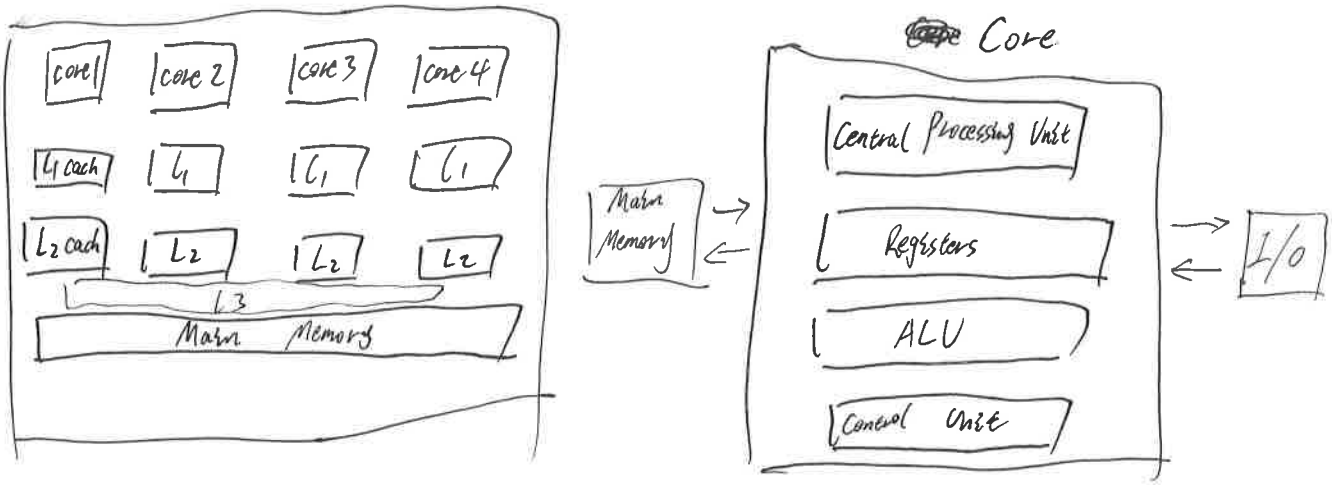
**static** **independent** **cooperating** **dynamic** **unbounded**

5. **II Computer Organization**

(a) Draw an architecture of a quad-core processor and discuss the role of each module in your diagram.

(b) Find a binary representation of the decimal number 0.1.

(a)




- Main Memory stores the data and instructions.
- Core executes the instructions and contains a control Unit which interprets the instructions.
- I/O is controlled by the control unit in core.

(b) please check back side.

$0.1 \times 2 = 0.2$ → 0
$0.2 \times 2 = 0.4$ → 0
$0.4 \times 2 = 0.8$ → 0
$0.8 \times 2 = 1.6$ → 1
$0.6 \times 2 = 1.2$ → 1
$0.2 \times 2 = 0.4$ → 0
$0.4 \times 2 = 0.8$ → 0
$0.8 \times 2 = 1.6$ → 1
$0.6 \times 2 = 1.2$ → 1
$0.2 \times 2 = 0.4$ → 0
$0.4 \times 2 = 0.8$ → 0
0
1
1
0
0
⋮

$123 / 2 = 61$ → 1
$61 / 2 = 30$ → 1
$30 / 2 = 15$ → 0
$15 / 2 = 7$ → 1
$7 / 2 = 3$ → 1
$3 / 2 = 1$ → 1
$1 / 2 = 0$ → 1
$0 / 2 = 0$ → 0

$0001100110011001100\cdots \rightarrow 1.1001100110011\cdots \times 10^{-4}$

$\Rightarrow \quad -4 + 127 = 123 \rightarrow 1101111$

$\because 0.1 > 0 \quad \therefore 0$

$0.1 \Rightarrow$

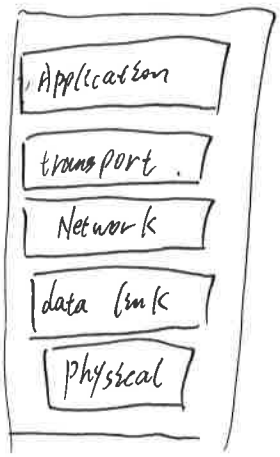
$0 \quad 1101111 \quad 1001100110011001100110011\cdots$

6. **II.D Networking and Communications**

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

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

(a)




physical : the physical [redacted] offeres for carrying data, like fiber optic cable

data link : implements on [redacted] NIC which contains a controler that provides framing, link access, reliable access and error detection.

Network : Router receives the packet from a network which shares the same first 24 bits in their IPv4 address then it check the [redacted] forwarding table and find the destination where the packet send to.

Transport : TCP provides secure transportation. It could make sure [redacted] the requested data will arrive and no repeated data.

Application : the client machine sends requests via socket, [redacted] the [redacted] transpat protocol and destination address, port number.

(b) [redacted] Address : the source and destination address, [redacted] port number are stored in the packet.

control line sends requests and login credential.

the data line transfer the actual data via HTTP.

7. **III.A 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.

1. sort A via merge sort $O(n \log n)$

2. sort B via merge sort $O(m \log m)$

3. iterate A, starting $i = 0$, $j = 0$ $O(n)$

i in index of A; j in index of B

if $A[i] = B[j]$, then add this number into result.list.

elif $A[i] < B[j]$, then $i = i + 1$

elif $A[i] > B[j]$, then $j = j + 1$ until find a value in B which is no smaller than $A[i]$

4. return result.list

the algorithm run in $O(n \log n + m \log m)$

8. **Automata and language theory**

Consider the following grammar:

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

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

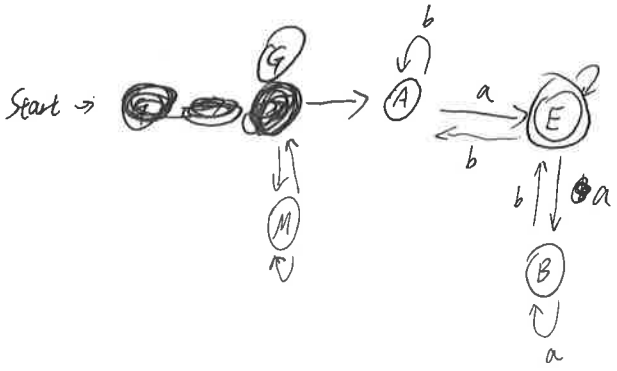
(b) Show a parse tree for the string `a b a a`.

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

(a)

$G \rightarrow S$

$G \rightarrow AM = \varepsilon \mid aaba$




(b)

A E A E

abaa → A: a E

b A A

a E a E

ε ε

AM.

αE|bAA S|ε

9. **III.C 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).

$G = (V, E)$

```
u, v ⊆ V.
For each pairs (u, v):
    res = True
    if there is a edge between u and v:
        res = True.
    else:
        res = False.
        return res
return res.
```

Since we defines each node has are most two neighbors, so we could finish the above algorithm in $O(n^2)$. by checking if there is a edge between every pairs of vertices.

10. **IV 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.

K-mean.

1. Decide the number of cluster (K).
2. Randomly choose K points in data as cluster points.
3. For each data point, calculate the distances between the data point and all K cluster points. Then, we set that data point to the cluster which has the smallest distance.
4. For each cluster, calculate the mean of all points belong to that cluster. and set it as cluster point.
5. Redo the step 3 and 4. until converge.

There are many interesting thing for K-mean including how to choose K, how to initially select K cluster points. These techniques are experimental. K-mean usually has good performance.