Ph.D. Comprehensive Examination

Computer Science Department
University of Miami

August 14, 2020

Student Name: Chenguang Zhao Student Number: C12120763

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) ①
(b) Insert (S, x) ②
(c) Successor (S, x) ③
(d) Predecessor (S, x) ④

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) ① $O(\log n)$ binary search
② $O(n)$ need move one by one along array
③ $O(1)$ index +1
④ $O(1)$ index -1

(b) ① $O(n)$ can't binary search
② $O(n)$ $O(n)$ find + $O(1)$ insert
③ $O(1)$ build-in feature
④ $O(1)$ same as ③

(c) ① $O(n)$ not garantee left node < right node no binary search
② $O(\log n)$ add from right ~~bottom~~, height is $\log n$, do swap if small
③ $O(n)$ same as ①
④ $O(n)$ same as ①

(d) ① $O(1)$ from key to value.
② $O(1)$ add key → value
③ $O(n)$ find one by one in table list
④ $O(n)$ same as ③

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) For call-by-value, parameters' value are passed to x, y, x=i, y=A[i]
y=i, after calling f(), the values of i, A[i] are not changed.

(b) For call-by-reference, parameters are references of i, A[i].
x=*i, y=*A[i]. After calling f(), i and A[i] references are swapped. Thus, their values are swapped.

(c) For call-by-value-result, At first, x, y make a copy of values of i and A[i]. In f(), x, y swap their values At the end of swap, x, y copy values back to i and A[i]. Thus, their values are changed.

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) At requirement analysis, we need to figure out the scope of problems. Therefore, we can estimate the needs of end users and the complexity of the system. Based on these requirements, we have to control the time for project and the cost of project with in the budget.

Time Budget ≥ time for (a) (b) (c) (d) (e), in total.

Cost Budget ≥ Total cost for (a) (b) (c) (d) (e).

(c) At architectural design, we need to design a system to fit the requirements of expected system. For example, system has to allow thousands of users to submit request at the same time. In order to reduce the waiting time, we may have to run jobs parallel. We have to take speed ① into consideration.

time eclipse of a job ≥ $\frac{\text{job runs on one cpu}}{\text{number of cpus}}$ (best cases: close to it)

(f) At maintence, we have to fine tune the system, add new features, and boost performance. If we boost performance, time complexity of data structure is important. If a system has a lots of search and insert operations and a few predecessor and successor operations. We can use a Hash table, which has O(1) for search and insert. This largely improve the performance

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)

There are 4 cores, each core have the full components of a cpu core. Each one have two level private caches for fast memory access. Four cores have a L3 share memory for working together. L3 caches is connected to memory controller for I/O. L3 caches is connected to Interconnection, which is a system bus ~~have~~ helping 4 cores to share and work playing a role ~~of~~ Interface.

(b)

$0.1 \times 2 = 0.2$
$0.2 \times 2 = 0.4$
$0.4 \times 2 = 0.8$
$0.8 \times 2 = 1.6$
$0.6 \times 2 = 1.2$
$0.2 \times 2 = 0.4$
$0.4 \times 2 = 0.8$
$0.8 \times 2 = 1.6$
$0.6 \times 2 = 1.2$
$0.2 \times 2 = 0.4$
$0.4 \times 2 = 0.8$
$0.8 \times 2 = 1.6$
$0.6 \times 2 = 1.2$
$0.2 \times 2 = 0.4$
$0.4 \times 2 = 0.8$
$0.8 \times 2 = 1.6$

binary representation of decimal 0.1 at 16 precision is:

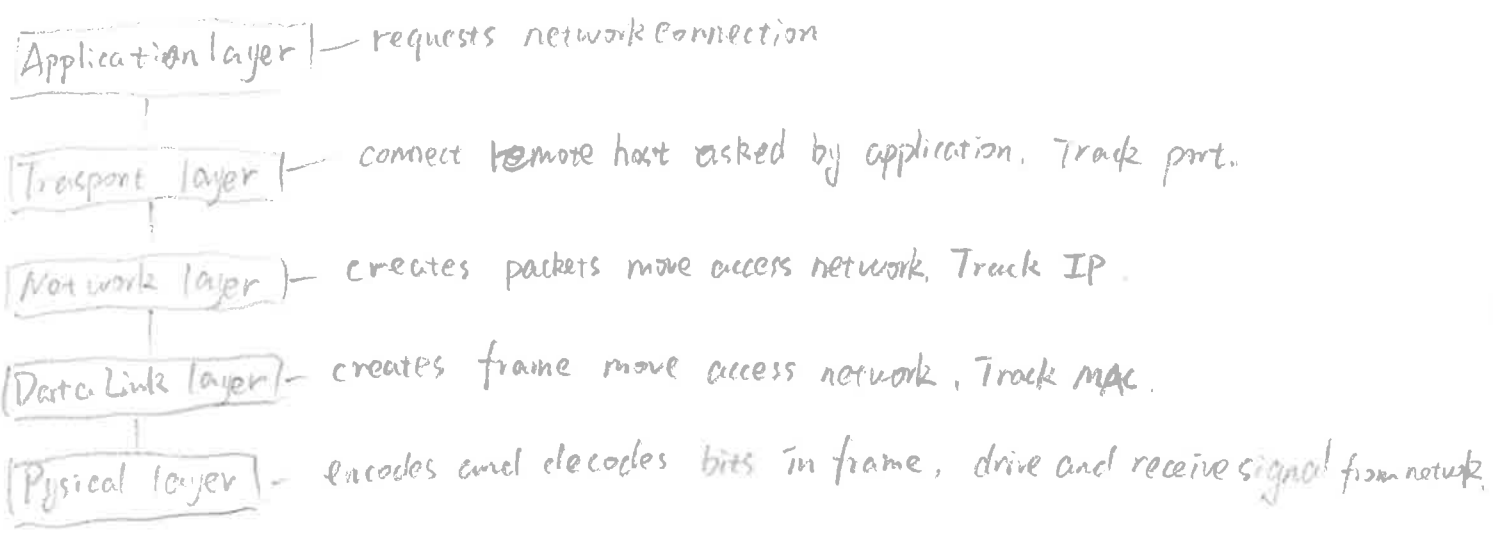
0.0001100110011001

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

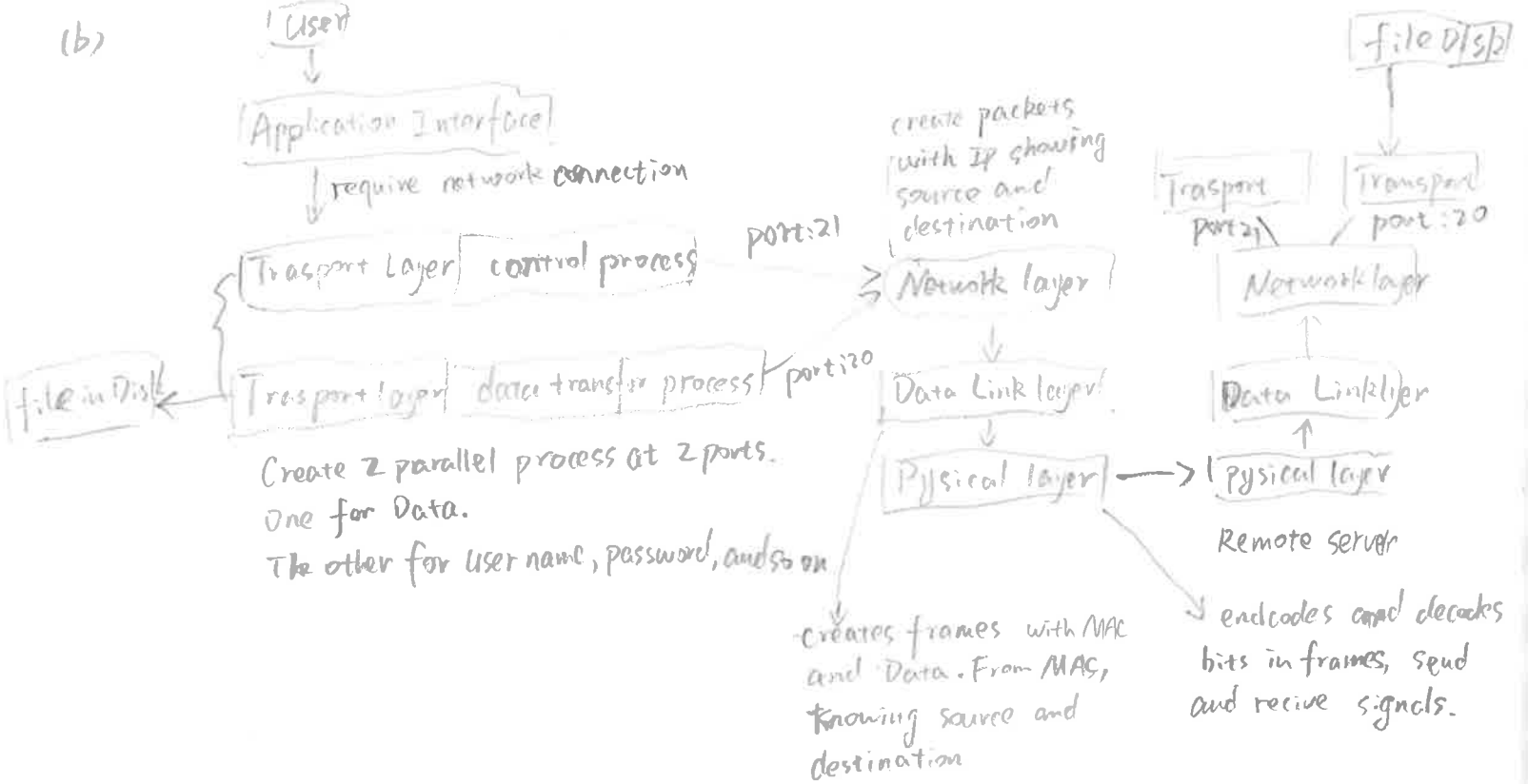
   (a) Draw a diagram showing layers of the Internet Protocol Stack and briefly discuss role of each layer.

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

(a).



(b)



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.

Step 1:

Sort A into increasing order $O(n \cdot \log n)$

Sort B into Increasing order $O(m \log m)$

Step 2: set j=0 ; C is a list

for i, a in enumerate(A): $O(n+m)$

if A[i] == B[j] : Continue

elif A[i] < B[j]:

Append A[i] to C ; i++;

else: # A[i] > B[j]

j++

If j reach the last element in B, add A[i:] to C.

else loop ends automately.

The list C has elements in A but not in B.

This algorithm runs in $O(n \cdot \log n + m \cdot \log m + n + m) = O(n \log n + m \log m)$

8. **Automata and language theory**

Consider the following grammar:

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

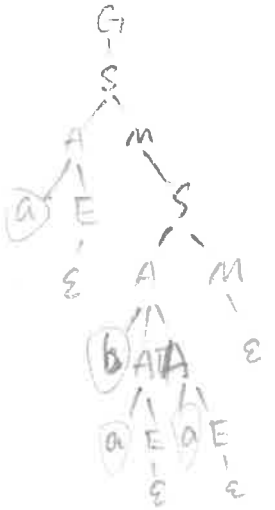
(a) Describe the language that the grammar generates in English.

(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) Strings belong to the language may start with a or b, ~~consist as loose~~ ~~one a or~~ the length of string is at least 1. Continous a no more than 3. every a few 'a' is followed by a few 'b' and then a few 'a'.

(b)



abaa

(c) LL(1) is context free grammer

This gramar is not context free.

Assume it is context free, for a string s ∈ language, there is a length p. and s can be cut into 5 parts $uvxyz$, where $|vxy| \leq p$, $|vy| > 0$, for $i \geq 0$ $uv^ixy^iz \in$ the language.

But we know that there is no long repeat 'a' or 'b' exists in language. when $i \geq 4$, uv^ixy^iz must $\notin$ language. Contradiction

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).

Give graph G (V, E) , Each v has at most two edges.

If a node has 0 or 1 edge, it can't build a cycle.

~~the~~ build unvisited, visited set. keep current_node.

Step 1:

Choose a node as start node; Perform DFS expand its neighbors.

If nodes has <2 neighbors :
    return false

If node is visited and unvisited is not empty:
    backtrack;

If node is visited and unvisited is empty :
    return true

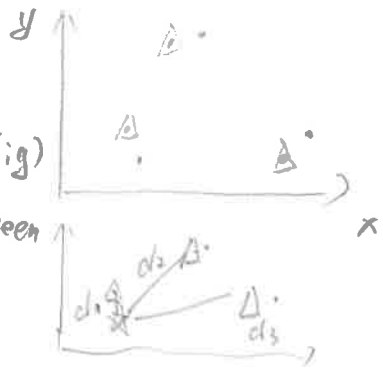
return false.

runs at $O(V+E) \leq O(V+2V) = O(3V) = O(V)$

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-means is a clustering ~~algorithm~~ method. It cluster data into k presetted classes. No overlapping classes.



1. ~~set~~ the number of class. K=3
2. Random choose 3 nodes as centroids. (Δ in fig)
3. Choose a node and calculate distance between the node and centroids (☆ in fig)
   Assign node to the nearest class.
4. Repeat 3 for all nodes in fig
5. Based on updated classes. calculate means of classes as updated centroids
6. Repeat 3, 4, 5, until classes and centroids don't change.

k-means is very efficient and ~~easy~~ machine learing ~~app~~ approach. It's good a unsupervised problems, especially you know how many classes are expected. Obviously, the backwards of k-means is that you have to give a super parameter k. There are some ways to solve it. Thus, I will use k-means when I met unsupervisied ~~the~~ known classes problem.