

Ph.D. Comprehensive Examination

Computer Science Department
University of Miami

August 14, 2020

Student Name: Xu Pan 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,
- (b) Ordered doubly linked list,
- (c) Min-Heap, and
- (d) Hash table

	search	Insert	Successor	Predecessor
Ordered array	$O(\log n)$	$O(n)$	$O(1)$	$O(1)$
ordered doubly linked list	$O(n)$	$O(1) / O(n) \text{ (plus search)}$	$O(1) / O(n) \text{ (plus search)}$	$O(1) / O(n) \text{ (plus search)}$
Min-Heap	$O(n)$	$O(\log n)$	$O(n)$	$O(n)$
Hash table	$O(1)$	$O(1)$	$O(n)$	$O(n)$

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

```

$x = 1 \quad y = A[1]$
 $z = 1 \quad x = A[1]$
 $y =$

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) change nothing outside swap. Only values are passed into swap. And swap only don't print or do anything rather than switch some values.

(b) $i = A[i]$
 $A[i] = i$

Because variables changed in swap remained.

(c) same as (b).

But will be different if in multi-thread programs. call-by-value-result behaves like call-by-value when the function is running, but will give new values back to arguments when "end".

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.

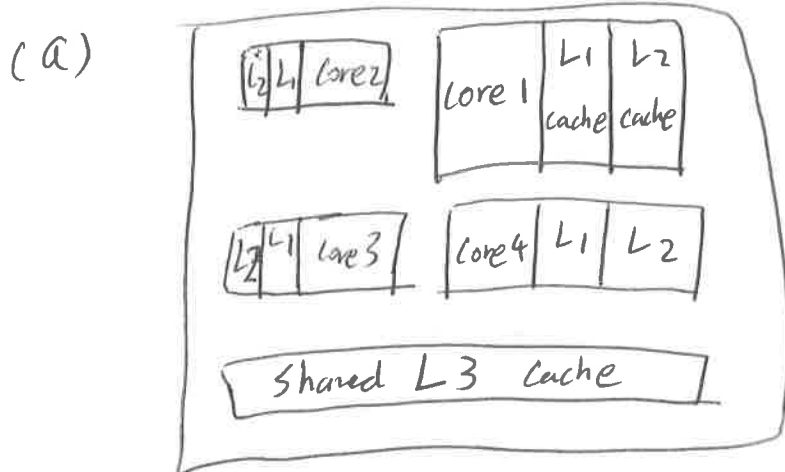
- (a) Requirement Analysis is about to specify the problems that are going to be solved by the new system. And how feasible is that, which includes the development budget, time frame, resource.
- (b) Functional Specification. According to the requirement, specifications need to be determined. For example which platform, ~~and~~ framework, package, interface, protocol. ~~At~~ choosing them properly is crucial for meeting the requirement.
- (d) Implementation. Coders and hardware designers will turn designs into runnable programs and products.

4. I.E Systems

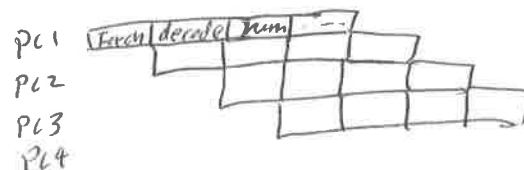
- (a) _____ linked libraries can support shared library code, allowing one copy of a library routine to be used by several different processes.
absolute relative static **dynamic** none of these is correct
- (b) When it is not known at compile time where a process will reside in memory, _____ code must 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** *should be child's process id*
- (d) The Banker's algorithm is used for deadlock *detection, and scheduling*
denial **prevention** **avoidance** recovery *proper*
- (e) Belady's anomaly can affect the performance of the _____ page replacement algorithm. *Least-Running-Time*
FIFO LRU optimal **SJF** *FIFO, LRU are scheduling algorithms.*
First-In-First-out
- (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 *when using user level thread implementation, one kernel thread can support many user threads.*
- (i) A process that does not affect, and is not affected by, another process is referred to as:
static **independent** cooperating dynamic unbounded
multi-programming, which was a fancy name when most computers can only run one process at a time.

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.



Core is to do the computation (states switching)
 It has ALU, registers, control flow, data flow.
 It usually (always) uses pipeline ~~pipeline~~. Fetch PC code,
 decode, run, write back.



L1 is to store ^{cache} most recently used addresses and values.

L2 is to store recently used addresses and values.

L3 can be read by all cores, hopefully can reduce fetch time when multi-thread or multi-process are used.

(b)

$$\begin{array}{r}
 \times 0.1 \\
 \hline
 \times 2 \\
 \hline
 0.2 \\
 \times 2 \\
 \hline
 0.4 \\
 \times 2 \\
 \hline
 0.8 \\
 \times 2 \\
 \hline
 1.6
 \end{array}$$

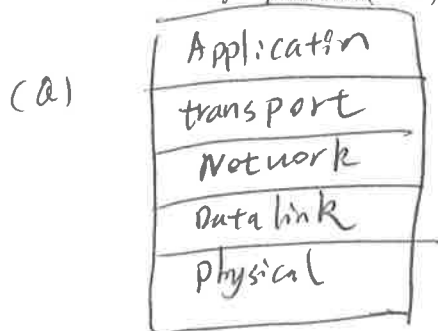
$$\begin{array}{r}
 \times 0.2 \\
 \hline
 \times 2 \\
 \hline
 1.2 \\
 \times 2 \\
 \hline
 0.4
 \end{array}$$

loop

$$(0.2)_{10} = (0.00011)_2$$

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)).



- (1) Physical layer is about how signal is passing through air or cable for example, what inter-pulse-interval is used, high-low voltage? frequency band? FM, AM?
- (2) Data link defines how packets are pass from one local machine to another local machine. It uses MAC address to identify machines. And if using hub, the sender broadcast packet to all neighbors. If using switch, it is point-to-point.
- (3) Network layer. IP is the only popular protocol. It defines how to find a remote machine. Usual a packet must go through many routers to find the receiver. Each end-point has a unique IP address. Routers used forward table to determine where to send the packet. Each packet, in its header, has a max life time counter. If it decrements to zero, router drops it.
- (4) Transport layer defines how user want to send the packets. lossless, or as fast as possible? TCP is a lossless protocol. The sender will know the receiver get the message correctly or not.

UDP, otherwise, doesn't care about if the ~~use~~ receiver get the message correctly.

TCP uses three way hand shake to initialize the channel.

There are algorithms about how to ensure lossless.

Practically, authorization is required to initiate TCP connection.

(5) Application layer. There are many kinds of application.

For example, http and web browser, FTP, E-mail, online game, streaming....

(6) FTP is a application layer protocol.

Transport layer, it uses two TCP connections, one for control, one for data.

Network layer gives it a header about sender and receiver information.

Packets ~~reach~~ reaches receivers local network

bounce in many local networks, finally.

physical layer = it may go through underwater ~~cables~~ cables and wifi.

7. III.A Algorithms and complexity

Describe an algorithm that takes two input lists of integers $A = a_1, \dots, a_n$ and $B = b_1, \dots, b_m$ 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 $O(n \log n + m \log m)$ time.

(1) use heap sort or merge sort to sort A & B

They are known to be $O(n \log n + m \log m)$

(2) $i = 0$ // for indexing A
 $j = 0$ // for indexing B

while $i < n$ or $j < m$

if $A[i] = B[j]$

$i++$

if $A[i] < B[j]$

$C.append(A[i])$

$i++$

if $A[i] > B[j]$

$C.append(B[j])$

$j++$

it is $O(m+n) \ll O(n \log n + m \log m)$

final time complexity is $O(n \log n + m \log m)$

8. Automata and language theory

Consider the following grammar:

$$G \rightarrow S \$$$

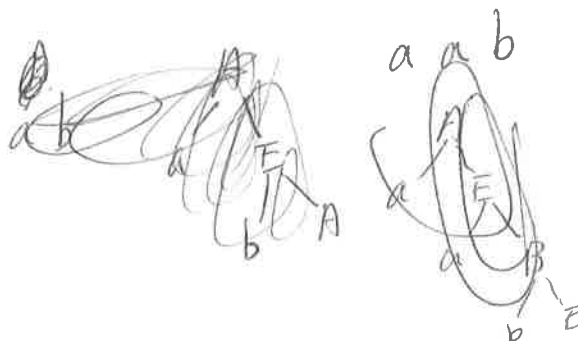
$$S \rightarrow A M$$

$$M \rightarrow S | \epsilon$$

$$A \rightarrow a E | b A A$$

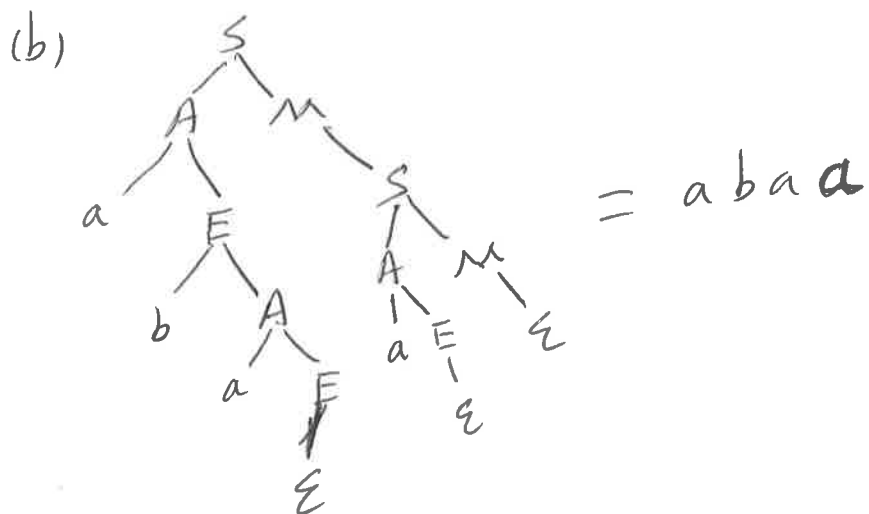
$$E \rightarrow a B | b A | \epsilon$$

$$B \rightarrow b E | a B B$$



- (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) it is a string with a and b, ~~maybe it is not~~



(c) It is LL(1)

	a	b
A	aE	bAA
E	aB	bA
B	aBB	bE

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).

Firstly, only when all vertexes have exactly two edges, there may exist a Hamiltonian Cycle.

It converts to a depth first search problem.

when all vertexes are visited ^{in one branch}, and the last node visited has neighbor the first node, that means we find a Hamiltonian cycle.

If we finished depth first search, and no finish condition occurred, that means there is no Hamiltonian cycle.

The only modification from DFS is we need a array to track the nodes visited in one branch.

The DFS time complexity is $O(V+E)$ ~~$O(V)$~~ $= O(3V) = O(V)$ is polynomial.

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.

~~Naive~~ Naive Bayesian Classifier.

we have observation with n features $\vec{x} = (x_1, x_2, \dots, x_n)$

$$P(C_i | \vec{x}) = \frac{P(C_i) P(\vec{x} | C_i)}{P(\vec{x})}$$

in Naive Bayes $P(\vec{x} | C_i) = P(x_1 | C_i) \cdot P(x_2 | C_i) \cdot \dots = \prod_k P(x_k | C_i)$

the chosen class is

$$i = \arg \max_i P(C_i) \prod_k P(x_k | C_i)$$

$P(x_k | C_i)$ is learned from training set, with a assumed distribution.

for example $P(x_k | C_i) \sim N(\mu_{ki}, \sigma_{ki})$

$$P(C_i) = \frac{\text{no. } C_i \text{ in training set}}{\text{no. training set.}}$$

bass and trout are two classes, each has feature weight and length.

first learn $P(\text{weight} | \text{bass}) \overset{\sim N}{P(\text{length} | \text{bass})} P(\text{weight} | \text{trout}) \overset{\sim N}{P(\text{length} | \text{trout})}$
 $P(\text{bass}) \overset{\sim N}{P(\text{trout})}$

then classify if $P(\text{bass}) \cdot P(x_{\text{weight}} | \text{bass}) P(x_{\text{length}} | \text{bass})$
 $> P(\text{trout}) P(x_{\text{weight}} | \text{trout}) P(x_{\text{length}} | \text{trout})$

then x is bass, otherwise x is trout.

