

Name: _____

ALGORITHMS QUALIFYING EXAM

This exam is open:

- books
- notes

and *closed*:

- neighbors
- calculators

The *upper bound* on exam time is 3 hours.

Please write your name at the top of each page.

Please put all your work on the exam paper.

Partial credit will only be given if your work is shown.

Good luck!

Name: _____

Part 1: Choose 30 points from this part**1: Function Asymptotic Order of Growth (10 points)**

- a) (4 points) **Arrange the four functions below in order of increasing asymptotic size. That is, find an arrangement g_1, g_2, g_3, g_4 of the functions satisfying g_1 in $O(g_2)$, g_2 in $O(g_3)$, g_3 in $O(g_4)$.**

Justify your answer mathematically.

$$8n^2 \lg^4 n \quad 5^{3n} \quad \left(\frac{5}{6}\right)^{4n} \quad 2^{9 \log_8 n} \lg \lg n$$

Name: _____

b) (6 points) Given the following facts about functions $f_1(n), f_2(n), f_3(n), f_4(n) :$

$$f_1(n) \in \Omega(8n^2 \lg^4 n)$$

$$f_2(n) \in \Theta\left(\left(\frac{5}{6}\right)^{4n}\right)$$

$$f_3(n) \in \Theta\left(2^{9 \log_8 n} \lg \lg n\right)$$

$$f_4(n) \in O\left(5^{3n}\right)$$

**Can we conclude from these 4 facts that $f_3(n) \in \Omega(f_1(n))$?
Why or why not?**

**Can we conclude from these 4 facts that $f_1(n) \in \Omega(f_2(n))$?
Why or why not?**

Name: _____

2: Solving a Recurrence (10 points)

Consider the following recurrence:

$$T(n) = \begin{cases} 8T\left(\frac{n}{2}\right) - n^3 + 7n^4 + 5 & n > 1 \\ 1 & n = 1 \end{cases}$$

- a)** (5 points) Find a tight bound on the closed form solution to the recurrence $T(n)$ by building a recursion tree and solving the resulting summation.

Name: _____

b) (5 points) Can the recurrence be solved using the Master Theorem?

If yes, explain why and show how to use the Master Theorem to obtain the result you obtained in (a).

If not, explain why not.

Name: _____

3. (5 points) Discrete Fourier Transform

Compute the Discrete Fourier Transform of the vector: (3,2)

Name: _____

4. (5 points) Number Theory

a) (3 points) List all the elements of the set Z_{21}^* for the *finite multiplicative group modulo 21*: (Z_{21}^*, \cdot_{21}) .

b) (2 points) Calculate the number of elements of Z_{21}^* using Euler's phi function.

Name: _____

5. (10 points) Longest Common Subsequence

Given a sequence $X = \langle x_1, x_2, \dots, x_m \rangle$, another sequence $Z = \langle z_1, z_2, \dots, z_k \rangle$ is a subsequence of X if and only if there exists a strictly increasing sequence $\langle i_1, i_2, \dots, i_k \rangle$ of indices of X such that, for all $j = 1, 2, \dots, k$, we have:

$$x_{i_j} = z_j$$

Using this definition of subsequence, find the longest sequence Z of characters that is a subsequence of the string

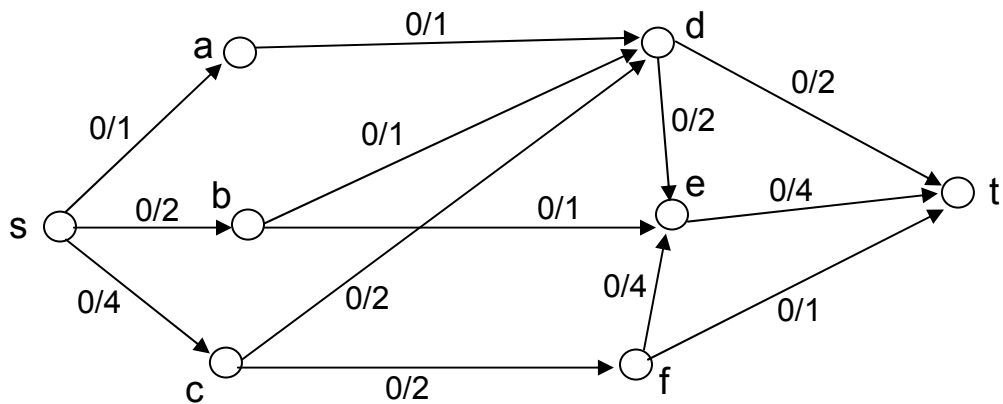
$X = \text{"skullandbones"}$

and is also a subsequence of the string

$Y = \text{"lullabybabies"}$.

Name: _____

6. (10 points) **Flow Networks:** Show the result of executing the Ford-Fulkerson algorithm on the flow network below, where node s is the source and node t is the sink.



What is the value of the maximum flow from s to t ?

Name: _____

Part 2: Choose 35 points from this part**1. (15 points) NP-Hardness:**

The NP-hard PARTITION problem is stated as follows [Garey & Johnson]:

INSTANCE : Finite set A and a size $s(a) \in \mathbb{Z}^+$ for each $a \in A$.

QUESTION : Is there a subset $A' \subseteq A$ such that $\sum_{a \in A'} s(a) = \sum_{a \in A - A'} s(a)$?

The KRR problem is stated as follows:

INSTANCE : A collection of 2D rectangles B and a 2D rectangular container C .

QUESTION : Does there exist a set of translations

of the rectangles in B that place them into C without overlapping each other?

(Assume that all rectangles are axis-parallel; that is, their sides are either vertical or horizontal. Furthermore, rectangles of B are not allowed to overlap the edges of the container C .)

Use PARTITION to prove that KRR is NP-hard.

Name: _____

2. (15 points) The 1D BIN-PACKING problem is stated as follows [Garey & Johnson]:

INSTANCE : A finite set U of items, a size $s(u) \in \mathbb{Z}^+$ for each $u \in U$, a positive integer bin capacity B and a positive integer K .

QUESTION : Is there a partition of U into disjoint sets U_1, U_2, \dots, U_K such that the sum of the sizes of the items in each U_i is B or less?

Consider the following pseudocode:

(Assume that INSERTION-SORT sorts the items in U by nonincreasing size. Also assume that item u "fits" in U_j implies that the sum of sizes of items in U_j + the size of u does not exceed B .)

BIN - PACK(U, B)

$U \leftarrow$ INSERTION - SORT(U)

$j \leftarrow 1$

for each $u \in U$

do if u does not fit into U_j

then $j \leftarrow j + 1$

place u into U_j

(see next page)

Name: _____

- a) (5 points) Provide a tight upper bound on the worst-case running time of BIN-PACK.

Name: _____

- b) (5 points) Describe an input to BIN-PACK that causes it to achieve the worst-case running time derived in (a).

Name: _____

- c) (5 points) Is BIN-PACK guaranteed to always minimize j ? Provide a proof to justify your answer.

Name: _____

3. (10 points) Consider 2 problems A and B.

Suppose that:

- a general instance A' of problem A is transformed into an instance B' of problem B;
- the transformation takes $\Omega(n \lg n)$ time;
- it is shown that a solution to B' corresponds to a solution to A' ;
- problem A takes $\Omega(n \lg n)$ time.

Can this reduction prove that problem B requires $\Omega(n \lg n)$ time?

Explain carefully why or why not.

Name: _____

4. (10 points) Consider the Minimum Spanning Tree (MST) problem.

Suppose that:

- the vertices of the input graph correspond to 2D points;
- the weight of an edge between 2 vertices u and v is equal to the Euclidean distance between the u and v .

This creates the Euclidean MST problem.

Now, suppose that you are given a procedure CLOSEST-PAIR that accepts the input graph from the Euclidean MST problem and returns a pair of vertices whose Euclidean distance from each other is the smallest over all pairs of vertices in the graph. Assume that the worst-case asymptotic running time of CLOSEST-PAIR is in $O(|V|^2)$, where $|V|$ is the number of vertices in the graph.

Can CLOSEST-PAIR be used within the MST algorithm of Kruskal or the MST algorithm of Prim to reduce its worst-case asymptotic running time?

Explain carefully why or why not.

Name: _____

Part 3: Designing an Algorithm: Answer one of these 2 questions
(35 points)

1. Given 2 collections R and B of 1D intervals on the x-axis of the form:

- $R = \{r_1, r_2, \dots, r_n\}$ where $r_i = (r_{si}, r_{fi})$ describes a single 1D interval for which r_{si} is the x-coordinate of the starting point and r_{fi} is the x-coordinate of the finishing point.
- $B = \{b_1, b_2, \dots, b_m\}$ is defined similarly in terms of $b_j = (b_{sj}, b_{fj})$.

Design an efficient algorithm that reports the starting and finishing coordinates of each interval of the x-axis that is both in an interval of R and in an interval of B.

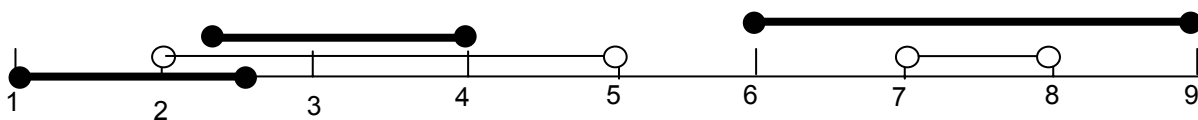
For example, for the figure below, the following results would be reported in the order given:

An interval that is in both R and B starts at x=2.

An interval that is in both R and B ends at x=4.

An interval that is in both R and B starts at x=7.

An interval that is in both R and B ends at x=8.



Parts (a), (b) and (c) on the next 3 pages ask you for pseudocode, correctness justification, and efficiency analysis.

Name: _____

(a) (15 points) PseudoCode:

Name: _____

(b) (10 points) Correctness:

Name: _____

(c) (10 points) Upper Bound on Asymptotic Worst-Case Running Time:

Name: _____

2. Sorting: This problem involves a sorting scenario related to exploring large datasets over the Web. In this scenario, metadata (data about data) is gathered that describes a set of raw data. Suppose that you have a list of n metadata records, where $n \gg 50,000$, and that all of these records can fit together into the main memory of the computer that is acting as a Web server. Suppose also that you need to write an application to process a list of metadata on the Web server. Your application will be reused often.

You may assume that the metadata list is represented as a (1-dimensional) array named **A** and that each element of **A** is a pointer to a metadata record. Each metadata record contains information about a *single* database containing raw data has been analyzed in order to produce that metadata record:

- **D**: a unique database identifier corresponding to the database that contains the associated raw data. You have no information about the range of values for **D**.
- **M**: an integer value that is equal to the minimum age over all people in the database. You may assume that no age in any database has a value > 100 .
- **G**: a text string containing the name of the geographic region associated with the database. You may assume that **G** has no more than 10 characters.

Describe a sorting strategy that will *efficiently* sort **A** so that the metadata records appear in nondecreasing **G** (lexicographic) order. If there is more than one metadata record with the same **G** value, then these records appear in nondecreasing **M** order. Your sorting strategy need not be stable (i.e. the relative order of two records with the same **G** and **M** values need not be preserved).

Your description should contain:

- a) pseudocode
- b) correctness justification
- c) upper bound on asymptotic worst-case running time
- d) classify your sort: is it comparison-based, non-comparison-based, or hybrid?

Name: _____

(a) (10 points) Pseudocode:

Name: _____

(b) (10 points) Correctness:

Name: _____

(c) (10 points) Upper Bound on Asymptotic Worst-Case Running Time:

Name: _____

d) (10 points) Classify your sort: is it comparison-based, non-comparison-based, or hybrid?