THE AUSTRALIAN NATIONAL UNIVERSITY

Second Semester 2008

COMP3600/COMP6466
(Algorithms)

Writing Period: 3 hours duration

Study Period: 15 minutes duration

Permitted Materials: None

Answer ALL Questions

All your answers must be written in the boxes provided in this booklet. You may be provided with scrap paper for working, but it must **not** *be used to write final answers.*

There is additional space at the end of the booklet in case the boxes provided are insufficient. Label such overflow boxes with the question number.

Do not remove this booklet from the examination room.

Name (family name first):

Student Number:

Official use only:

Q1 (25)	Q2 (20)	Q3 (12)	Q4 (15)	Q5 (12)	Q6 (16)	Q7 (10)	Total (100)

QUESTION 1 [25 marks]

(a) Order the following complexity classes from slowest-growing to fastest growing. If any of the classes are the same, indicate this fact and justify it.

$$\Theta(n^{4.5}),\ \Theta(\log^2 n),\ \Theta(3^n),\ \Theta(n^2 \log^{3/2} n),\ \Theta(\log_2 n^{100}),\ \Theta(n^4/(6n+5))$$

QUESTION 1(a) (i)	**[3 marks]**

(b) Let $f(n)$, $g(n)$ and $h(n)$ be three positive functions. Either prove that the following statement is true, or provide a counterexample:

(i) $\max\{f(n), g(n)\} = O(f(n) + g(n))$.

QUESTION 1(b)(i) **[4 marks]**

(ii) $f(n) = O(g(n))$ implies $2^{(f(n)} = O(2^{g(n)})$.

QUESTION 1(b)(ii) **[4 marks]**

(c) Using the $\Theta(.)$ notation, give asymptotic upper bounds for $T(n)$ in the following recurrences. Note that you are **not** allowed to use the Master theorem.

(i) $T(n) = 5T(n/7) + n^2$, assume that n is power of 7 and $T(n)$ is constant for $n \leq 7$.

QUESTION 1(c)(i) **[5 marks]**

(ii) $T(n) = \sqrt{n}T(\sqrt{n}) + n$, assume that n is power of 2 and $T(n)$ is constant for $n \leq 4$.

QUESTION 1(c)(ii) **[4 marks]**

(d) Given n elements in a set S and k with $1 \leq k \leq n$, the following algorithm $\mathcal{A}$ similar to the given linear algorithm `SELECT` in the textbook can be used to find the first k largest elements in S, which proceeds as follows.

Step 1. The elements in S are grouped with size 27 each, except the last group which contains $n - \lfloor n/27 \rfloor * 27$.

Step 2, Find the median of each group. As a result, there will be a sequence of medians consisting of the medians of the groups. Let x be the median of the median sequence.

Step 3. The set S is now partitioned into three disjoint sets R_1, R_2 and R_3 which contain elements less, equal to or greater than x.

Step 4. Call algorithm $\mathcal{A}$ for finding the first $(|R_1| - (n - k + 1) + 1)$ largest elements on R_1 recursively if $|R_1| \geq n - k + 1$. Otherwise, if $|R_1 \cup R_2| \geq n - k + 1$, return the elements in R_3 and $k - |R_3|$ of xs; otherwise, call algorithm $\mathcal{A}$ for finding the first k largest elements on set R_3 recursively.

Write a recurrence for the running time of algorithm $\mathcal{A}$ and show the algorithm works within the linear time.

QUESTION 1(d) (more room on next page)	**[5 marks]**

Answer box continued over page.

QUESTION 1(d) continued. **[5 marks]**

QUESTION 2 [20 marks]

(a) Describe three common probing methods for hash tables with open addressing. What is the number of different probing sequences derived from each of them? assume that m is the size of the hash table.

QUESTION 2(a) **[6 marks]**

(b) There are two types of data structures used to represent disjoint sets. What are they? To speed-up the `UNION` operation on the disjoint sets, heuristics for them have been introduced. Describe how to implement the `UNION` operation on the data structures, using the heuristics briefly.

QUESTION 2(b)	**[4 marks]**

(c) Assume that there are n elements with keys, to build a dynamic directory for these elements, either a binary search tree or a red-black tree will be employed for such a purpose, can you detail the advantage of using the red-black tree rather than the binary search tree?

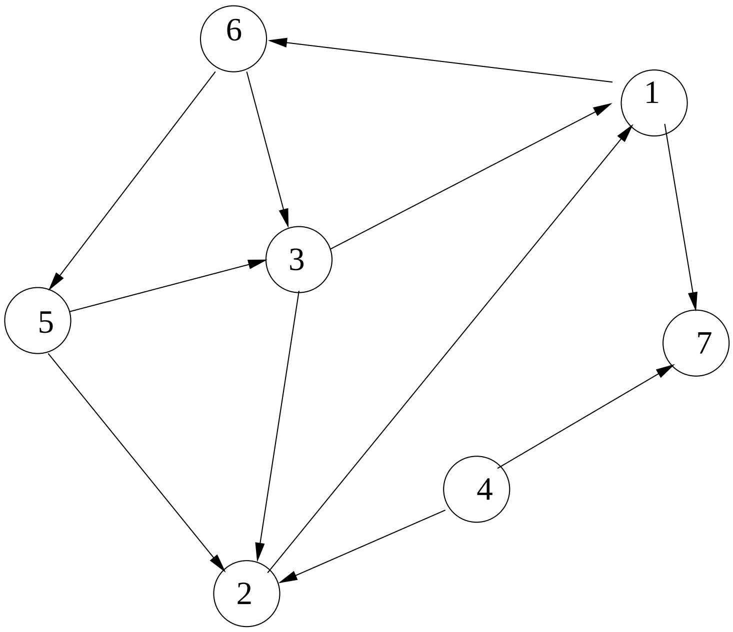
QUESTION 2(c) **[4 marks]**

(d) Given an integer sequence $10, 4, 19, 23, 22, 15, 28, 27, 12, 9, 11, 16$, which is stored in an array $A[1..12]$. Construct a binary min-heap for the numbers in the sequence.

QUESTION 2(d) **[6 marks]**

QUESTION 3 [12 marks]

(a) Perform a Depth-First Search on the directed acyclic graph (DAG) below, starting at vertex 1 and exploring neighboring vertices **in order of their labels**. Give your results as a diagram that shows the order of visited vertex and the start and finish times of each vertex. Classify the edges in the DAG as tree edges, forward edges, back edges and cross edges if they exist.



QUESTION 3(a)	**[6 marks]**

(b) Using the results of the part (a), list the vertices in topological order.

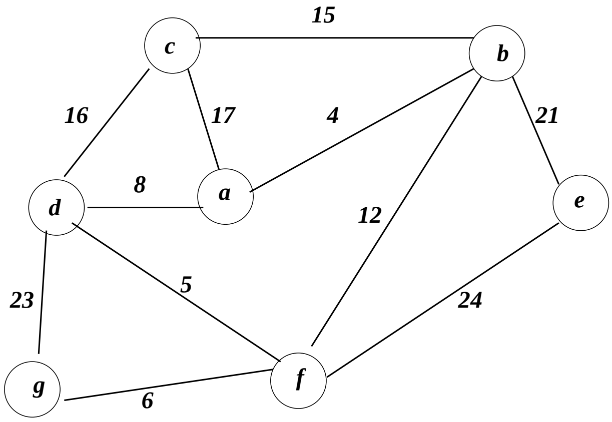
QUESTION 3(b) **[2 marks]**

(c) Assume that each edge in the DAG is assigned a weight, there is an algorithm for finding single-source shortest paths in the DAG. What is the running time of that algorithm if the DAG contains m directed edges and n vertices?

QUESTION 3(c) **[4 marks]**

QUESTION 4 [15 marks]

(a) Use Kruskal's algorithm to find a minimum spanning tree on the following graph. Indicate the order in which the edges are chosen.



QUESTION 4(a) (more room on next page) **[10 marks]**

QUESTION 4(a) (continued)

(b) Suppose you are given a connected graph G, with edge costs that are all distinct. Prove that G has a unique minimum spanning tree.

QUESTION 4(b) **[5 marks]**

QUESTION 5 [12 marks]

(a) Consider the sequence alignment problem over a four-letter alphabet $\{z_1, z_2, z_3, z_4\}$, with a given gap cost and given mismatch costs. Assume that each of these parameters is a positive integer.

(i) Suppose you are given two strings $A = a_1 a_2 \ldots a_n$ and $B = b_1 b_2 \ldots b_m$ and a proposed alignment between them. Give an $O(mn)$ dynamic programming algorithm to decide whether this alignment is the `unique` minimum-cost alignment between A and B.

QUESTION 5(a)(i) (more space on next page) **[8 marks]**

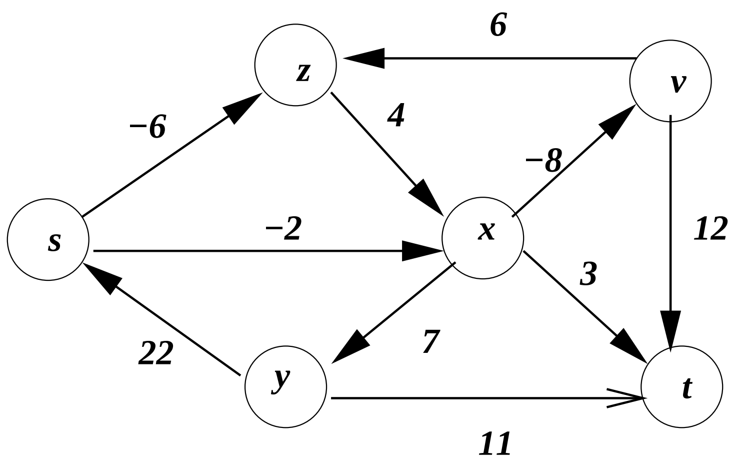
QUESTION 5(a)(i) (continued)

(ii) Briefly indicate what extra needs to be done to find the unique alignment example rather than just its cost.

QUESTION 5(a)(ii)	**[4 marks]**

QUESTION 6 [16 marks]

(a) Apply either the Dijkstra algorithm or the Bellman-Ford algorithm to find the distance from vertex s to each of the other vertices in the following graph. Explain your choice of which of the two algorithms you are using. Give enough intermediate information to demonstrate that you are applying the algorithm correctly (don't just give the final result).



QUESTION 6(a) (more room on next page)	**[12 marks]**

Answer box continued over page.

QUESTION 6(a) continued. **[12 marks]**

(b) How to modify the Floyd-Warshall algorithm to find the transitive closure of a directed graph $G(V, E)$? What is the running time of the proposed algorithm? if $|V| = n$ and $|E| = m$.

QUESTION 6(b) **[4 marks]**

QUESTION 7 [10 marks]

This question is only for COMP6466/PhD/Honours students.

Given a directed telecommunication network $G(V, E)$, associated with each link $e \in E$, there are two weights: one is the communication cost $c(e)$, and another is the communication delay $d(e)$ on the link e which is an integer.

The question is: given a pair of nodes u and v and an integer D, to find a shortest path in G from node u to node v such that (i) the cost sum of links in the path is minimised; and (ii) the total communication delay in the path from node u to node v is no more than the given threshold D.

Devise an algorithm for the problem and analyze its running time complexity. (**Hint:** You may use Dijkstra's algorithm and dynamic programming technique to solve this problem).

QUESTION 7 (more room on next page)	**[10 marks]**

Answer box continued over page.

QUESTION 7 continued. **[10 marks]**

Additional answers. Clearly indicate the corresponding question and part.

Additional answers. Clearly indicate the corresponding question and part.

Additional answers. Clearly indicate the corresponding question and part.

Additional answers. Clearly indicate the corresponding question and part.

Additional answers. Clearly indicate the corresponding question and part.