Candidate Number

G6017

THE UNIVERSITY OF SUSSEX

BSc and MComp SECOND YEAR EXAMINATION
January 2022 (A1)

Program Analysis

Assessment Period: January 2022 (A1)

DO NOT TURN OVER UNTIL INSTRUCTED TO BY THE LEAD INVIGILATOR

Candidates should answer TWO questions out of THREE.
If all three questions are attempted only the first two answers will be marked.

The time allowed is TWO hours.

Each question is worth 50 marks.

At the end of the examination the question paper and/or answer book, used or unused, will be collected from you before you leave the examination room.

1.

(a) Precisely specify the conditions under which the following algorithm returns true, and then discuss, in detail, the running time of the algorithm. If you think it has different best- and worst-case running times then these should be considered separately, and you should explain the conditions under which best and worst-cases arise.

You must fully explain your answer and use O, Ω and θ appropriately to receive full marks.

Algorithm Ex1 $\big((a_1, \ldots a_n), (b_1, \ldots, b_m)\big)$

```
k ← 0
for i ← 1 to n do
        j ← 1
        while j ≤ m do
                If a_i == b_j
                        k ← k + 1
                j ← j + 1

return k > 0
```

[10 marks]

(b) Precisely specify the conditions under which the following algorithm returns true, and then discuss, in detail, the running time of the algorithm. If you think it has different best- and worst-case running times then these should be considered separately, and you should explain the conditions under which best and worst-cases arise.

You must fully explain your answer and use O, Ω and θ appropriately to receive full marks.

Algorithm Ex2 $\big((a_1, \ldots a_n), (b_1, \ldots, b_n)\big)$

```
q ← true
for i ← 1 to n  do
        j ← n
        while j > 0 and q  == true do
                If a_i == b_j
                        q ← false
                j ← j − 1

return q
```

[10 marks]

(c) Precisely specify the conditions under which the following algorithm returns true, and then discuss, in detail, the running time of the algorithm. If you think it has different best- and worst-case running times then these should be considered separately, and you should explain the conditions under which best and worst-cases arise.

You must fully explain your answer and use O, Ω and θ appropriately to receive full marks.

Algorithm Ex3 $\big((a_1, \ldots a_n), b\big)$

$q \leftarrow 0$
$w \leftarrow false$
for $i \leftarrow 1$ to $n - 1$ do
    $z \leftarrow a_i + a_{i+1}$
    If $z < b$
        $q \leftarrow q + 1$
    else
        $q \leftarrow q - 1$
If $q < 0$
    $w \leftarrow true$

return w

[10 marks]

(d) A data pattern analyser is to be built that can detect and count up the number of occurrences of two and three letter same letter sequences in a sequence (e.g. (a,a) or (b,b,b)). The analyser should stop if it encounters ***** in the sequence and return the number of occurrences found up to that point in the form of a 2-tuple $(\#2LetterSequences, \#3LetterSequences)$. A 3 letter sequence should only count as a 3 letter sequence, not an occurrence of two 2 letter sequences. No letter in the input sequence ever occurs more than 3 times in a row.

So, for example:

Input string	**2 letter sequences found**	**3 letter sequences found**
(a,b,a,b,b,a,a,*****)	2	0
(x,x,x,y,y)	1	1
(p,q,p,z,z,*****)	1	0
(*****,a,a)	0	0
(x,x,a,d,*****, s,s,s,l)	1	0

Produce a formal statement of this problem, and then write an algorithm to solve the problem using a pseudo code style similar to the one shown in parts (a) to (c). State the bounds on the best and worst case performance of your algorithms using O, Ω and θ appropriately to receive full marks.

[10 marks]

(e) A file is protected by a random password consisting of n binary bits. All password combinations are equally probable. To access the file we need the correct password. The process of applying the password to the file takes 10ms regardless of the value of n. Brute force attack is always a viable basic strategy for guessing a password.

To ensure that the file remains sufficiently secure, we need to ensure that there is no more than a 1% chance over 30 days that the password is guessed by a hacker program utilizing brute force working 24 hours a day, 7 days a week. How many bits should be specified for the password?

[10 marks]

2.

(a) A student has been asked to put some parcels on a shelf. The parcels all weigh different amounts, and the shelf has a maximum safe loading weight capacity of 100 Kg. The weight of parcels are as follows (in Kg):

$\boldsymbol{parcel}$	$\boldsymbol{weight\ (Kg)}$
1	8
2	50
3	2
4	15
5	4
6	5
7	20

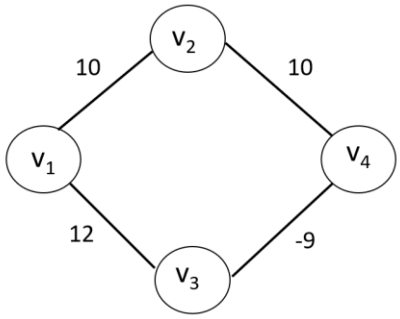
The student has been asked to load the maximum weight possible parcels on the shelf subject to the maximum safe loading weight.

State two possible approaches for a greedy algorithm solution to solve this problem. In each case, state clearly the result you would get from applying that approach to this problem, stating whether the solution is optimal or not. If

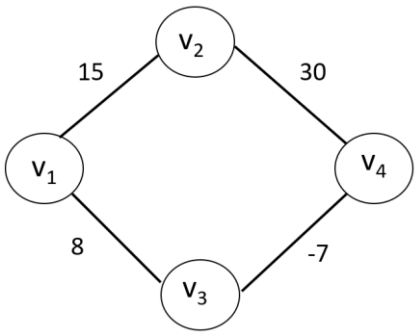
your answer does not produce an optimal solution, what algorithm could be employed to find one?

[10 marks]

(b) One example of a greedy algorithm is the Dijkstra algorithm for finding the lowest cost path through a weighted graph. The diagram below shows two weighted graphs that a student wants to investigate using Dijkstra's algorithm. In each case the task it to find the lowest cost of reaching every node from v_1. Each graph has a single negative weight in it.



Graph (a)



Graph (b)

One of the graphs will yield a correct analysis of the lowest cost for all vertices, and the other will produce an incorrect analysis. Which of the two graphs will produce the incorrect analysis, and explain why the greedy nature of Dijkstra's algorithm is responsible for the incorrect analysis. Your answer should include the key concept of an invariant.

[5 marks]

(c) The priority queue is a widely used data structure. Priority queues may be implemented using binary heaps and simple linear arrays. For the basic priority queue operations of:

- Building an initial queue
- Taking the highest priority item off the queue
- Adding a new item to the queue

Compare and contrast the running time complexities (best and worst cases) associated with implementations using binary heaps and simple linear arrays. You may find it helpful to use diagrams to support your answer.

[10 marks]

(d) A recursive algorithm is applied to some data $A = (a_1, \dots, a_m)$ where $m \geq 2$ and m is even. The running time T is characterised using the following recurrence equations:

$T(2) = c$ when the size of A is 2
$T(m) = T(m-2) + c$ otherwise

Determine the running time complexity of this algorithm. Note that m is even and the problem size reduces by 2 for each recursion.

[10 marks]

(e) Another recursive algorithm is applied to some data $A = (a_1, \dots, a_m)$ where $m = 2^x$ (i.e. 2, 4, 8,16 …) where x is an integer ≥ 1. The running time T is characterised using the following recurrence equations:

$T(1) = c$ when the size of A is 1, and c is a constant
$T(m) = 2T\left(\frac{m}{2}\right) + m$ otherwise

Determine the running time complexity of this algorithm. You will find it helpful to recall that:

$$\sum_{x=1}^{x=\infty} \frac{1}{2^k} \rightarrow 1$$

And

$$2^{log_2(x)} = x$$

[15 marks]

3.

(a) The subset sum problem can be reliably solved optimally using the dynamic programming algorithm shown below:

```
SubsetSum(n, W)

Let B(0, w) = 0 for each w ∈ {0, ..., W}
  for i ← 1 to n
    for w ← 0 to W
      if w < w_i then
        B(i, w) ← B(i − 1, w)
      else
        B(i, w) ← max (w_i + B(i − 1, w − w_i), B(i − 1, w))
```

Where n is the number of requests, W is the maximum weight constraint, w_i is the weight associated with request i, and B is the solution space.

You are given a set of requests and their corresponding weights. The maximum weight constraint W is 12.

i	w_i
1	1
2	7
3	10
4	6
5	3
6	2

Copy the following solution space table to your answer book (do not write your answer on the question paper) and complete the table to determine the optimal subset sum.

		w												
		0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
i	**6**													
	5													
	4													
	3													
	2													
	1													
	0													

[10 marks]

(b) The sequence alignment problem may be solved by the following dynamic programming algorithm:

SequenceAlignment(X, Y):

Let $B(i, 0) \leftarrow i \times \gamma$ for each $1 \leq i \leq n$
Let $B(0, j) \leftarrow j \times \gamma$ for each $1 \leq j \leq m$
For $i \leftarrow 1\ to\ n$
    For $j \leftarrow 1\ to\ m$
        $B(i,j) \leftarrow \min\ [\delta(x_i, y_j) + B(i-1, j-1),$
        $\gamma + B(i, j-1),$
        $\gamma + B(i-1), j\ ]$

Where $X = (x_1, \ldots, x_n)$ and $Y = (y_1, \ldots, y_m)$ are two sequences to be aligned, $\delta(p, q)$ is a penalty associated with matching symbol p to q, and γ is a gap penalty.

The sequence alignment algorithm is applied using the following data:

$X = (a, b, c)$

$Y = (a, b, a, b, b)$

$\gamma = 4$

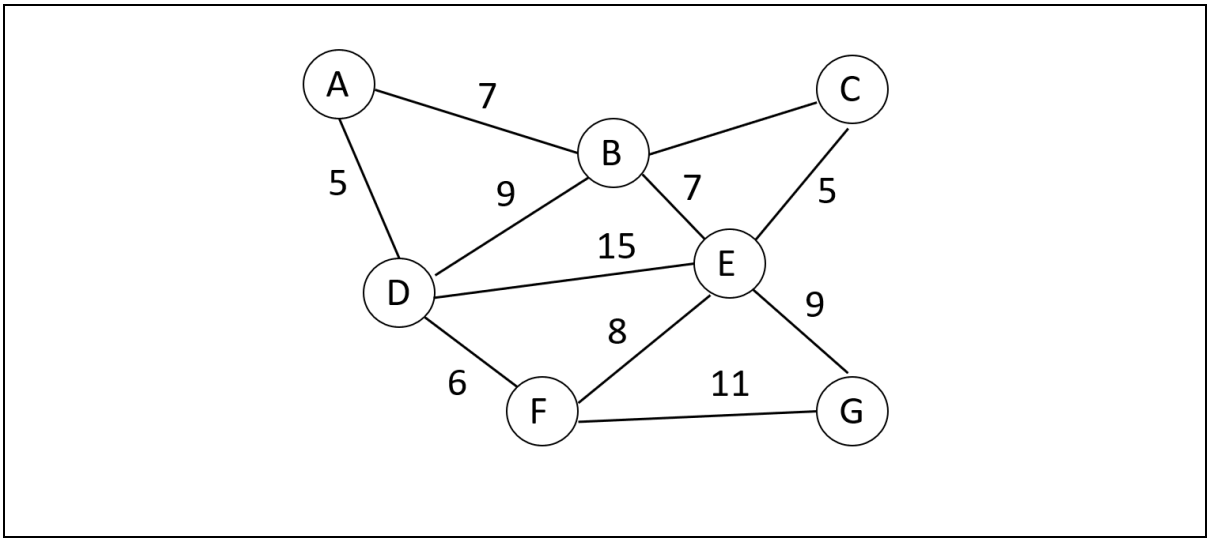
The delta function is defined for symbols in the alphabet $\{a, b, c\}$:

	a	**B**	**c**
a	0	4	5
b	4	0	7
c	5	7	0

Generate the problem space matrix B and thus determine the optimal alignment between X and Y.

[15 marks]

(c) Draw a Minimum Spanning Tree derived from the graph shown below.



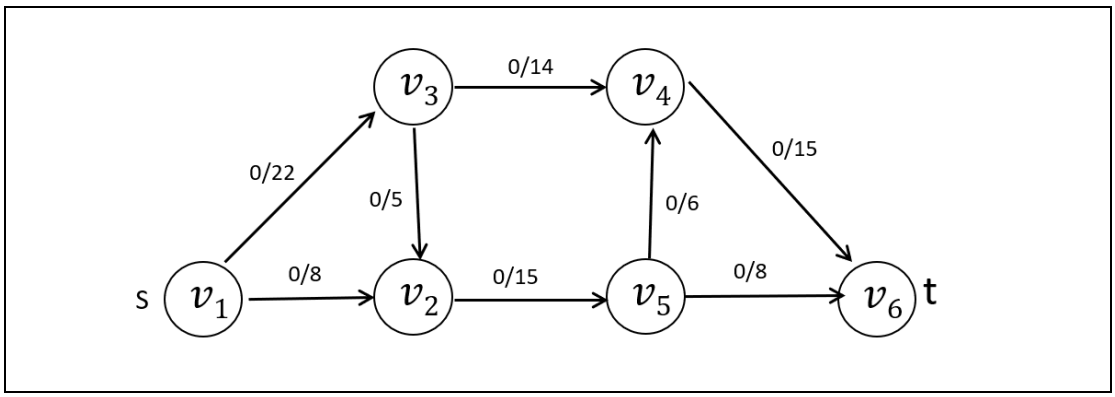
[5 marks]

(d) The Ford-Fulkerson algorithm is used to determine network flow. The diagram below represents a data network that connects a Data Service Provider (DSP)

connected to $v_1(s)$ to a customer connected to $v_6(t)$. Each edge represents a single data transmission link.

The notation p/q indicates a current actual forwards flow p measured in Gb/s in a pipe with a maximum capacity of q also measured in Gb/s.

At the outset no data is being sent by the DSP to the customer.



i. Show the residual graph that will be created from the initial empty flow. When drawing the residual graph, show a forward edge with capacity x and a backward edge with flow y by annotating the edge $\vec{x}; \overleftarrow{y}$.

[2 marks]

ii. What is the bottleneck edge of the path (s, v_3, v_4, t) in the residual graph you have given in answer to part (a) ?

[2 marks]

iii. Show the residual graph after incorporating the simple path (s, v_3, v_4, t) that results from augmenting the flow based on the residual graph you have given in answer to part (a).

[4 marks]

iv. Repeat the process outlined above incorporating additionally the simple paths (s, v_3, v_2, v_5, t) , (s, v_2, v_5, v_4, t) and (s, v_2, v_5, t) showing each residual graph, to determine the maximum flow between s and t, and thus the maximum data bandwidth that can be achieved between the DSP and the customer.

[12 marks]

End of paper