THIS PAPER IS NOT TO BE REMOVED FROM THE EXAMINATION HALL

UNIVERSITY OF LONDON

CO3325 ZA

BSc Examination

COMPUTING AND INFORMATION SYSTEMS, CREATIVE COMPUTING and COMBINED DEGREE SCHEME

Data Compression

Friday 3 May 2019: 14:30 – 16:45

Time allowed: 2 hours and 15 minutes

DO NOT TURN OVER UNTIL TOLD TO BEGIN

There are **FOUR** questions on this paper. Candidates should answer **THREE** questions. All questions carry equal marks, and full marks can be obtained for complete answers to a total of **THREE** questions. The marks for each part of a question are indicated at the end of the part in [.] brackets.

Only your first **THREE** answers, in the order that they appear in your answer book, will be marked.

There are 75 marks available on this paper.

A handheld calculator may be used when answering questions on this paper but it must not be pre-programmed or able to display graphics text or algebraic equations. The make and type of machine must be stated clearly on the front cover of the answer book.

Question 1

Parts (a)–(f) require true or false answers with justifications. You should read each statement carefully and then determine whether the statement is TRUE or FALSE. If the statement is true, explain why it is true or give a supportive example. If it is false, explain what should be the correct answer or provide a counterexample. Additional instructions for some parts are given as 'Hints' in square brackets after each statement.

(a) The code (0, 01, 011, 0111) for (A,B,C,D) is uniquely decodable.

[Hint: use 001001100000111 as an example.] [4]

(b) It is possible to find a shorter prefix code than the binary code (0, 10, 001, 011).

[Hint: give your reasons.] [4]

(c) In the context of video compression, *P pictures* mean *predictive* pictures, and they are frames to be compressed by spatial compression. [4]

(d) The static Huffman encoding method starts from the bottom and builds the linked list taking a 'bottom-up' approach, while Shannon-Fano method is similar but starts from the top. [4]

(e) If 129 distinct colours are required for an application in a RGB colour model, the *colour depth* value required should be at least 8 bits. [4]

(f) Given the sine wave signal $s(t) = 2sin(880\pi t)$, the maximum sample rate should be 880Hz.

[Hint: show your calculation.] [5]

Question 2

Consider the binary source (A,B), where the probability of A is 0.2. Answer each part of the questions below and show all your work. The following logarithm values may be useful for your work.

x	0.1	0.2	0.3	0.4	0.6	0.7	0.8	0.9
$\log_2 x$	-3.32	-2.32	-1.74	-1.32	-0.74	-0.51	-0.32	-0.15

(a) Compute the entropy of the source. [3]

(b) Demonstrate how the canonical static Huffman code can be devised for the source. [3]

(c) What is the average codeword length? [2]

(d) Let the *compression efficiency* be defined as $\eta = \frac{H}{L}$, where H is the entropy of the source and L is the average codeword length. What is the value of the *compression efficiency* of the Huffman coding in (b)? [2]

(e) Demonstrate how the *compression efficiency* of the Huffman coding in (b) may be improved. [10]

(f) Demonstrate (in a table as follows) how the Huffman tree in part (e) may be represented in an array. [5]

i	1	2	3	4	5	6	7	...
X[i]								...

Question 3

(a) Decode the following string using the HDC algorithm. Explain the meaning of each control symbol that you use. [10]

`r4n1Ar2n6BB5522r31n30ABr3Cn2BC`

(b) What is the compression ratio? [3]

(c) Copy the following table into your answer book, and fill it in to show the probability distribution of the source based on the symbol frequencies: [8]

character	B	5	2	1	0	A	C	␣
frequency								
probability (in fraction)								

(d) What is the entropy of the source? [4]

For decimal precision, assume a scale of 2, *i.e.* two digits after the decimal point. The following logarithm values may be useful for your work (note all the decimals in the table have a scale of 2).

x	6	12	24
$\log_2 x$	2.58	3.58	4.58

Question 4

(a) The following flawed pseudocode is meant to outline the Arithmetic algorithm. Identify the flaws and offer your corrections. [8]

```
1. L <- 0 and d <- 1
2. If x is within [L,L+d*p1)
3.     then output s1, leave L unchanged, and
4.          set d<-d*p1
5.     else if x is within [L+d*p1, L+d)
6.          then output s2, set L<- L*d+p2 and d<-d*p2
7. If the_number_of_decoded_symbols
8.             > the_required_number_of_symbols
9.     then go to step 2.
```

(b) A binary sequence of length 5 (symbols) was encoded on the binary alphabet (A,B) using the Arithmetic encoding algorithm. Suppose that the probability Pr(A)=0.25 and the encoded output is 0.13. Demonstrate in a table such as the one below how the Arithmetic decoding algorithm would derive the original sequence of symbols step by step. Assume that in your table the values of i column represent iterative steps; the values of the [L, L+d) column represent the values upon completion of each iteration; and the d*p1 and d*p2 columns show intermediate partial results. [17]

i	L	d	d*p1	d*p2	[L, L+d)	Output
0	0	1				
1						
2						
3						
4						
5						

END OF PAPER