Examiners' commentary 2018–2019

CO3325 Data compression – Zone B

General remarks

Data compression is a technical subject and the examination questions are often precise and specific. Each question may consist of a number of parts and each part may contain a number of sub-questions explicitly or implicitly. Questions are often labelled as Question 1, Question 2, and so on.

Parts are often labelled as (a), (b), etc. Explicit sub-questions are often labelled as i. ii., etc. To achieve a good grade, you must understand the examination questions. For example, you would need to know at least how many sub-questions are required, and whether there are any further implicit tasks involved in each part or sub-question. The examination can be viewed as a piece of **written communication between you and the examiner**. As such communication gives no second chance for clarification and is highly constrained by time, you would need to demonstrate your knowledge in the most efficient manner in the examination. For example, you may need to decide promptly if it would be better to draw a diagram with a short notation as your answer, or rather write a long paragraph of descriptions.

As Data compression is a Level 6 examination, you need to demonstrate your competence in problem solving. This includes sometimes sharing your interpretation of given specifics in questions. For example, if necessary, you may leave notes to the examiner on your script to clarify certain issues to avoid any potential confusion, or to add assumptions to simplify your solutions.

The questions in the examination do not necessarily have an equal level of complexity. You will want to potentially secure as many marks as possible and as early as possible. For example, you might choose to attempt the questions you find easier first, or to give only itemised answers first, adding details later if time permits.

If you follow these guidelines as well as those in the subject guide and textbooks, your examination should be an enjoyable experience. It is a good opportunity for you to check the level of your knowledge and to celebrate your academic achievements.

The performance of the candidates in 2018–19 was very good, though a range of marks – from the high first class to the low failure – was seen.

Comments on specific questions

Question 1

This question tested candidates' basic knowledge learnt from the course. A good approach to answer such a question was to provide solutions following precise instructions, e.g., to write down on the script a TRUE or FALSE, followed by justification and an example (or counterexample). By doing this, you offer an insight into the truth base of each part of the question.

a. It is the way the tree is built (not the linked list) in the Shannon-Fano and Huffman encoding algorithms that differs. Some candidates did not spot

the wrong object 'linked list', which should be tree. Some others spent too much time on the 4-mark sub-question.

b. A variable-length code can be uniquely decodable if one codeword serves as a separator. Some mistakes made this year included failing to realise that '0' serves as a separator, and being vague about that fact.

c. The 'I' pictures mean intra pictures and they are the frames to be compressed by spatial compression. Some candidates showed lack of knowledge in this area.

d. It is impossible to find a shorter prefix code due to the Kraft-McMillan inequality. Some candidates gave a correct conclusion but did not provide the important justification. Some other candidates did not seem to understand the question.

e. The colour depth is represented by the number of bits used to hold a colour value in an RGB colour model. This was a bookwork question and most candidates answered correctly.

f. The Nyquist theorem can be applied to solve the problem. Some candidates said their calculator did not work, but the calculation could really have been done by hand. If not, a formula should be given as an answer.

Question 2

This question was about Huffman coding.

a. This part of the question tested understanding of the entropy. Good answers included the mathematical expression and calculation. Note the unit of the entropy is bits.

b. Good answers included the Huffman tree and the code.

c. This part of the question requires you to show the calculation. The result unit 'bits' should be included.

d. This part required, again, the mathematical formulae. The calculation result should have been presented as a percentage.

e. A good answer to this part would consist of four sections, including the extended alphabet and its probability distribution, the Huffman tree and code, the average length of the codewords, and finally the efficiency of the coding. A big problem with candidates' answers seems to be a lack of knowledge about the approach of alphabet extension. Some other candidates showed incomplete knowledge of this.

f. The answer to this part should have been the node contents of the Huffman tree. This part was about the implementation knowledge using the efficient array data structure, but few candidates managed to answer correctly.

Question 3

a. This was a bookwork question consisting of two sub-questions. An easy way to answer this part of the question was to first answer the first sub-question and then explain all the symbols used to answer the second sub-question. Marks were lost due to candidates giving: no explanation; a lack of detail in the explanation; or an incomplete list of symbols.

b. This part was straightforward and tested knowledge about one of the compression measures. Some candidates gave the mathematical formulae without calculation. Some others confused compression ratio with rate.

c. There were two explicit requirements in this part of the question. Good answers were presented in two sections. The first section gave the character frequency. Assuming it is a closed system, the probability distribution of the source can be calculated.

d. A good answer to this part of the question would simply be the calculation of the entropy. Most candidates answered this part well.

Question 4

This question tested candidates' knowledge of the Arithmetic decoding algorithm.

a. There are two implied sub-questions and good answers included explicit answer sections for both. In your answer, the first section should give the errors identified. The second section should be your suggested corrections.

b. This part of the question required a filled table of values to demonstrate understanding of how the decoding algorithm works. Some candidates did not manage to give the correct answer.