University of St Andrews

MARTINMAS 2022-23 EXAMINATION DIET
SCHOOL OF COMPUTER SCIENCE

MODULE CODE:	**CS3302**
MODULE TITLE:	Data Encoding
EXAM DURATION:	3 hours
EXAM INSTRUCTIONS:	a. Answer all three questions b. Each question carries 20 marks

This assessment consists of exam-style questions and you should answer as you would in an exam. You cannot copy or paraphrase text or material from other sources and present this as your own work. Your exam answers should be entirely your own work without unacknowledged input from others. If you are in any doubt, you should clearly acknowledge the origin of any material, text passages or ideas presented (e.g. through references). You must not co-operate with any other person when completing the exam, which must be entirely your own work. You must not share any information about the exam with another person (e.g. another student) or act on any such information you may receive. Any attempt to do so will be dealt with under the University's Policy for Good Academic Practice and may result in severe sanctions. You must submit your completed assessment on MMS within 3 hours of you downloading the exam. Assuming you have revised the module contents beforehand, answering the questions should take no more than three hours.

1. Encryption

(a) Alice and Bob are establishing a shared key using Diffie-Hellman key exchange, with prime number $p = 353$ and $g = 3$ (the roles of p and g are as in the lecture notes). Alice chooses random number 6 and Bob chooses random number 11. What values do Alice and Bob send each other? What is the shared key? Show your working (you can use a calculator). [3 marks]

(b) Verify that the S-box of AES maps 0x65 to 0x4D. Show all intermediate steps in the calculation, including long divisions and matrix and vector multiplication and addition. [11 marks]

(c) Consider a mode of operation whose encryption is characterised by the formulas:

$$\begin{aligned} c_1 &= E(x_1) \\ c_i &= E(x_i \oplus c_{i-1} \oplus x_{i-1}), \quad \text{for } i > 1 \end{aligned}$$

where:

- x_i is the i-th block of plaintext
- c_i is the i-th block of ciphertext
- E is a symmetric-key encryption function (e.g. DES or AES with a fixed key)
- D is the corresponding decryption function (with the same key)
- $\oplus$ stands for bitwise XOR of two bit strings of the same length

Give the matching formulas for decryption. Justify your answer. [2 marks]

(d) For the mode of operation as in question (c), can we decrypt the rest of a message if one block is lost? Justify your answer. [2 marks]

(e) For the mode of operation as in question (c), can we decrypt the rest of a message if two consecutive blocks are received in the wrong order? Justify your answer. [2 marks]

[Total marks 20]

2. Compression

(a) Assume a fixed code alphabet B. A set K of code words with the Fano condition is called complete if there is no $w \in B^*$ such that $w \notin K$ and $K \cup \{w\}$ has the Fano condition; in other words, we cannot add any further code words without violating the Fano condition.

A set K of code words with the Fano condition is called safe if every string in B^* is a prefix of at least one string that is a concatenation of code words from K.

Show that a set K of code words with the Fano condition is complete if and only if it is safe. [4 marks]

(b) Show that a set K of code words with the Fano condition is complete if and only if $\sum_{w \in K} \frac{1}{k^{|w|}} = 1$, where k denotes the size of the code alphabet, and $|w|$ denotes the length of a code word w. Your answer can refer to lemmas and theorems from the lecture notes. [5 marks]

(c) Consider the information source with source alphabet $\{a, b, c, d, e\}$, with probabilities $P(a) = 0.11, P(b) = 0.06, P(c) = 0.64, P(d) = 0.1$, and $P(e) = 0.09$. Construct a Huffman coding, with code alphabet $\{0, 1\}$. What is the average length of this coding? [4 marks]

(d) A petshop only sells cats and dogs, and each pet is either black or white. Every time a pet is sold, the petshop owners send a message to their accountant saying whether a cat or a dog has been sold, and they call their supplier saying whether a black pet or a white pet has been sold.

On average, the number of cats sold is twice the number of dogs sold. A quarter of the cats sold is white and a fifth of the dogs sold is white.

What is the entropy of messages sent to the accountant?

What is the entropy of calls to the supplier?

What is the mutual information of messages sent to the accountant and calls to the supplier?

(You can use a calculator, and round to 3 decimal places.) [7 marks]

[Total marks 20]

3. Error correction

(a) Let L be a binary linear code with basis

$$B = \{0000111, 1001101, 0111011, 0011001\}.$$

Find:

(i) A generator matrix G_L of L in normal form.

(ii) A generator matrix $G_{L'}$ of a code L' in standard form, where L' is a code equivalent to L, stating clearly why they are equivalent.

[6 marks]

(b) We say that L is a (n, k, d)-code. Determine n, k and d. How many errors could L detect and how many errors could L correct? Is L perfect? [5 marks]

(c) Let K be a binary code with generator matrix

$$G = \begin{pmatrix} 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}.$$

Suppose that we have a message $m = 110$, and we wish to encode m using K before feeding it through a noisy information channel. What is the code word c that is passed into the channel?

Assume that the channel is a (zero-memory) binary symmetric channel with error probability $p = 0.015$. What is the probability of receiving $x = 010101$ on the other end of the channel? Give your answer to 3 decimal places. [3 marks]

(d) Find a parity check matrix H of K and use it to decode the received word $x = 010101$. Include in your solution all precomputation needed. [6 marks]

[Total marks 20]

***** END OF PAPER *****