TOTAL: 50 MARKS Answer **all** questions.

This assignment contributes 5% of your overall course mark. Submit your assignment as a single PDF file to Assignment Drop Box. Include all workings and explanations. Marks will be deducted for ambiguous solutions. Zero marks are awarded if the answers contain no explanation. Also, refer to the Departmental policy on cheating and plagiarism. Cut-and-paste without acknowledgment of the source is not acceptable.

Assignment Drop Box (https://adb.ec.auckland.ac.nz/adb/).

Departmental Policy on Cheating on Assignments: see Assignments page of the course web site.

Due date: 11:59 pm, Friday 17 August, 2012

1. Consider the following assignments:

$$\begin{pmatrix} & code_1 & code_2 & code_3 & code_4 \\ A & 0 & 0 & 1 & 1 \\ B & 100 & 1 & 01 & 01 \\ C & 10 & 00 & 001 & 001 \\ D & 11 & 11 & 0001 & 000 \end{pmatrix}$$

(a) Which of the above assignments are codes? Justify your answer.

[2 marks]

(b) Which of the above codes are prefix free? Justify your answer.

[2 marks]

(c) Which of the above codes are uniquely decodable? Justify your answer.

[2 marks]

(d) For those codes that are uniquely decodable, give the encoding of ABBCCCDDDD.

[2 marks]

(e) For each of the following strings

(i) 010011

(ii) 1000

(iii) 11110000001

indicate whether they are encodings in each of the uniquely decodable codes above of some string over the alphabet $\{A, B, C, D\}$; for each affirmative answer give the string encoded.

[2 marks]

2. Is $\{1, 011, 01110, 1110, 10011\}$ uniquely decodable? Justify your answer; in case of negative answer, find a string with two encodings.

[10 marks]

3. For each set of natural numbers:

i) 133333333333, 2, 1.

[3 marks]

ii) 32, 16, 8, 4, 2

[2 marks]

iii) 3, 1, 4, 3.

[2 marks]

iv) 1, 2, 50, 2.

[3 marks]

check whether there is a prefix binary code whose codewords lengths are exactly the provided numbers. In each case justify your yes or no answer; in case of affirmative answer construct a code with the specified requirements.

4. Devise two correct Huffman trees and their corresponding codewords for the letters A, B, C, D, E, F with frequencies given in the following table:

Letter	Frequency
A	15%
B	15%
C	10%
D	10%
E	30%
F	20%

[10 marks]

5. A file F consisting of n characters, formed with lower-case letters (26), upper-case letters, and extra m characters, has to be efficiently stored.

(a) How many bits do you need to store F using the ASCII code?

[2 marks]

(b) What is the maximum valued for m to be able to store F using a 6-bit code? Present your solution and calculate the size of the compressed file.

[6 marks]

(c) How much size reduction (percentage) have you obtained?

[2 marks]

Solution 1.

- All assignments are one-to-one, so they are codes.
- $Code_1$ is not prefix free because 10 is a proper prefix of 100; $Code_2$ is not prefix free because 1 is a proper prefix of 11;$Code_3$ and $Code_4$ are prefix-free.
- $Code_3$ and $Code_4$ are prefix-free, so uniquely decodable. $Code_1$ is not uniquely decodable because 100 = 10-0. $Code_2$ is not uniquely decodable because 11=1-1.
- The encoding in $Code_3$ of ABBCCCDDDD is 1010100100100100010001001. The encoding in $Code_3$ of ABBCCCDDDD is 1010100100100100000000000.
- 010011 is the encoding of BCA for $Code_3$ and $Code_4$
- 1000 is not a valid encoding for $Code_3$; 1000 is the encoding of AD for $Code_4$
- 11110000001 is not a valid encoding for $Code_3$ but is a valid code for $Code_4$ and the encoding is AAAADDA.

Solution 2.
Let $C = \{1, 011, 01110, 1110, 10011\}$ and apply the Sardinas-Patterson algorithm. We obtain: $C_1 = \{110, 0011, 10\}, C_2 = \{10, 0, 011\}$, so 011 is a codeword common to both C and C_2, hence C is not uniquely decodable. We have:

$$01110 - 1110 - 011 = 011 - 1 - 011 - 10011 = 011101110011.$$

Solution 3.
We use Kraft's theorem to decide wether there is a prefix binary code for the given codewords lengths.

i) Because $2^{-333333333333} + 2^{-2} + 2^{-1} = 3/4 + 2^{-333333333333} < 1$ because $2^{-33333} < 1/4$, so by Kraft's theorem there is prefix binary code for which the codewords lengths are exactly: 333333333333, 2, 1, for example, the code $0^{333333333333}, 01, 1$.

ii) Because $2^{-32} + 2^{-16} + 2^{-8} + 2^{-4} + 2^{-2} < 1$ by Kraft's theorem there is a prefix code whose codewords lengths are: 32, 16, 8, 4, 2 for example, the code $00, 10^3, 110^6, 1110^{13}, 11110^{28}$.

iii) Because $2^{-3} + 2^{-1} + 2^{-2} + 2^{-3} = 1$ by Kraft's theorem there is a prefix code whose codewords lengths are 3, 1, 2, 3, for example, $000, 1, 01, 001$.

iv) Because $2^{-1} + 2^{-2} + 2^{-50} + 2^{-2} = 1 + 2^{-50} > 1$, by Kraft's theorem there is no prefix code whose codewords lengths are 1, 2, 50, 2.

Solution 4.
Write the letters in the order F, C, D, A, B, E and use Huffman's procedure we get the code, and finally change 0 into 1 and 1 into 0:

Letter	Code 1	Code 2
A	100	011
B	101	010
C	010	101
D	011	100
E	11	00
F	00	11

Solution 5.
1) If the file has n characters and each is stored as a 7-bit code, then one needs $7n$ bits.

2) The file uses $26 + 26 = 52$ characters, so with a 6-bit code one can code $2^6 = 64 > 52$ characters, hence the maximum value for m is 64-52=12.

3) A file with n characters will be coded by $6n$ bits, so the size reduction is (from $7n$ to $6n$) is about 14.3%.