

M.Sc. Examination 2018
Semester-I
Computer Science
Course : MCSC-13
(Information and Coding Theory)

Time : 3 Hours

Full Marks : 40

Questions are of value as indicated in the margin

Answer **any four** questions.

1. An information source has a 5 symbol alphabet, Symbols are named as A, U, I, O and E with individual probabilities of occurrence 0.4, 0.05, 0.1, 0.05 and 0.4 respectively. Each symbol requires 8 bits to be transferred in an uncompressed form.
 - a) Find the theoretical maximum compression ratio achievable by lossless compression. 2
 - b) Verify how closely that limit can be reached by Shannon-Fano encoding. 3
 - c) Find the probabilities of occurrences for which the result of (a) is the lowest. 2
 - d) Find the compression ratio achieved through static Huffman encoding for the probabilities found in c. 3
2. An information source has a 5 symbol alphabet, Symbols are named as A, U, I, O and E with individual probabilities of occurrence 0.4, 0.05, 0.1, 0.05 and 0.4 respectively. Each symbol requires 8 bits to be transferred in an uncompressed form. Arithmetic coding is used to encode the symbols of the source using 16-bit floating point representation with 12 bit mantissa and 4 bit exponent (e.g 0.1256 is represented as [1245,4]).
 - a) Check if the sequence IEEEEIU can be fully encoded using the representation avoiding overflow. 6
 - b) Find the codewords for AAAAA and AA. 2
 - c) What is the maximum length sequence which can be encoded using the representation? 2
3.
 - a) Show the dynamic Huffman tree and transmitted code sequence for an input sequence AIEEIAE. 6
 - b) Illustrate the decoding operation for the same example. 4
4. The string "THIS IS THE THESIS" is to be encoded and transmitted using dictionary-based encoding schemes. Each letter and requires 8 bit in an uncompressed form. Ignore " " and " " characters.
 - a) Find the compressed size using LZ77 with 8 character search and 4 character lookahead buffer. 5
 - b) Compare the result of (a) with the compressed size achieved by LZW with dictionary size 32 and 6 basic symbols T, H, E, I, S and '\b']. 5
5. A discrete memoryless channel has input $X = \{X_0, X_1, X_2, X_3\}$ and output $Y = \{Y_0, Y_1, Y_2, Y_3\}$. Input probabilities of set X are $\{1/4, 3/8, 1/4, 1/8\}$ respectively. X and Y has the following joint distribution matrix [i, j]th element is $\text{Prob}(X_i, Y_j)$:

P.T.O.

(2)

	X_0	X_1	X_2	X_3
Y_0	1/32	1/16	1/8	1/32
Y_1	1/16	1/8	1/32	1/32
Y_2	1/16	1/16	1/16	1/16
Y_3	1/4	0	0	0

- a) Find the measure of predictability of X upon observing Y. 5
- b) Without calculating, comment on possible change in the result of (a) if the last row is changed to $\{1/16, 1/16, 1/16, 1/16\}$ and justify. 2
- c) Find the channel capacity for this DMC. 3
6. While testing a system it is observed that, average bit error probability over a channel is 0.01. An error correction code for correcting single bit error is to be deployed. The maximum codeword length is chosen such that the frame error rate of the codeword frame < 0.1 .
- a) Choose an appropriate coding scheme and encoding length. 2
- b) Show with an example that all single errors are correctable. 5
- c) Justify (b) from minimum free distance consideration. 3
- [Hint : If frame length = L bit, bit error probability = p, then frame error probability = probability that any single bit among the L bits is in error.]
7. a) Find the maximum number of bit errors which are always corrected by a rate -1/3 4-state convolution encoder with generators 111, 101 and 011. 4
- b) Illustrate that error correction performance with a suitable example. 6
-