Version AGiF/3

EGT3
ENGINEERING TRIPOS PART IIB

Thursday 4 May 2023 9.30 to 11.10

Module 4F5

ADVANCED INFORMATION THEORY & CODING

*Answer not more than **three** questions.*

All questions carry the same number of marks.

*The **approximate** percentage of marks allocated to each part of a question is indicated in the right margin.*

*Write your candidate number **not** your name on the cover sheet.*

STATIONERY REQUIREMENTS
Single-sided script paper

SPECIAL REQUIREMENTS TO BE SUPPLIED FOR THIS EXAM
CUED approved calculator allowed
Engineering Data Book

10 minutes reading time is allowed for this paper at the start of the exam.

You may not start to read the questions printed on the subsequent pages of this question paper until instructed to do so.

You may not remove any stationery from the Examination Room.

1 (a) Recover the missing digits i_1 and i_2 in the sum

$$8674534 \; + \; 19i_18342 = 10672i_276$$

given that the second operand is divisible by 9. [10%]

(b) Find an integer n such that $R_{144}(100n) = 12$. [10%]

(c) What is the inverse of 27 in the ring $\langle \mathbb{Z}_{143}, \oplus, \odot \rangle$? [10%]

(d) What is $X^5 \cdot X^4$ in GF(8)? [10%]

(e) Determine a parity-check matrix of the code over GF(5) defined by its generator matrix

$$\mathbf{G} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix}.$$

[10%]

(f) In a toy example of the Diffie-Hellman public key distribution system, we operate in $\langle \mathbb{Z}_p, \odot \rangle$ where $p = 59$ is a prime for which $p - 1 = 2 \cdot 29$ has a "large" prime factor. The generator is $\alpha = 2$. Your secret key is $x_A = 27$ and you wish to generate a common secret key to communicate with Bob whose public key is $y_B = 43$ which he derived from his secret key x_B.

(i) What is your public key y_A? [15%]

(ii) Use the Diffie-Hellman protocol to generate a common key and convert it to a six-digit binary key that can be used in a conventional secret key cryptosystem. [15%]

(iii) Eve suggests that, in order to improve the security of Diffie-Hellman, one might want to double-exponentiate secret keys: instead of using x_A as your secret key in Diffie-Hellman, use $x'_A = 2^{x_A}$ and similarly Bob uses $x'_B = 2^{x_B}$. Explain why the resulting system still determines a common secret key but actually weakens Diffie-Hellman instead of strengthening it. [20%]

2 A coded Redundant Array of Independent Disks (RAID) consists of 5 hard disks of 1 tebibyte (1 TiB) each (2^{40} bytes or 2^{43} bits) each. The aim of a RAID is to ensure data recovery in the event that some of the disks malfunction.

(a) Describe the simplest possible coding system that permits storage of 4 TiB on the RAID so that all data can be recovered in the event that a single hard disk becomes unavailable. [10%]

(b) Describe a coding system that permits storage of 3 TiB on the RAID so that all data can be recovered in the event that up to two hard disks become unavailable. [10%]

(c) Consider a Reed-Solomon (RS) code over GF(16) defined via the irreducible (not primitive) polynomial $\pi(X) = 1 + X + X^2 + X^3 + X^4$ and the root of unity $\beta = X$. What is the length of the code? [10%]

(d) The RS code is defined so that the first digits of a codeword's Discrete Fourier Transform (DFT) are zero. Determine the 3-dimensional code's parity-check matrix. [10%]

(e) Using the logarithmic table of GF(16) below, determine the code's systematic encoder matrix. [20%]

(f) The RAID has a 12 bit control string that specifies the encryption standard applied to the data. Hard disks 0, 1, and 2 store 4 bits of this control string each. Hard disks 3 and 4 store parity-check symbols obtained from the 12 bit string using the systematic RS encoder matrix you determined in Part (e). The 4 bits stored on hard disk 0 are $[1, 1, 0, 1]$, corresponding to the element $1 + X + X^3$ of GF(16). Hard disk 1 and 2 have malfunctioned and any information stored on those hard disks has been erased. Hard disks 3 and 4 contain the bits $[0, 1, 0, 0]$ and $[0, 0, 0, 0]$, respectively. Use RS decoding to recover the 12 bit control string. [40%]

α^0	1	α^4	$X + X^2 + X^3$	α^8	$1 + X^3$	α^{12}	X
α^1	$1 + X$	α^5	$1 + X^2 + X^3$	α^9	X^2	α^{13}	$X + X^2$
α^2	$1 + X^2$	α^6	X^3	α^{10}	$X^2 + X^3$	α^{14}	$X + X^3$
α^3	$1 + X + X^2 + X^3$	α^7	$1 + X + X^2$	α^{11}	$1 + X + X^3$	α^{15}	1

3 (a) Consider a channel code $\mathcal{C}_n$ of length n with M codewords constructed over alphabet $\mathcal{X}$. The codewords are generated at random. Let $p_{e,m}(\mathcal{C}_n)$ be the error probability of the m-th codeword of a given codebook $\mathcal{C}_n$ for $m = 1, \ldots, M$ and $\mathbb{E}[p_{e,m}(\mathsf{C}_n)]$ its average over the random-coding ensemble.

(i) Show that for any $s > 0$ it holds that

$$\mathbb{P}\big[p_{e,m}(\mathcal{C}_n) \geq 2^{1/s} \cdot \mathbb{E}[p_{e,m}(\mathsf{C}_n)^s]^{1/s}\big] \leq \frac{1}{2}. \qquad [15\%]$$

(ii) Define the random variable $Z_m = \mathbb{1}\Big\{p_{e,m}(\mathcal{C}_n) < 2^{1/s} \cdot \mathbb{E}[p_{e,m}(\mathsf{C}_n)^s]^{1/s}\Big\}$ and explain the meaning of the quantity

$$\mathbb{E}\Big[\sum_{m=1}^{M} Z_m\Big]. \qquad [10\%]$$

(iii) Show that

$$\mathbb{E}\Big[\sum_{m=1}^{M} Z_m\Big] \geq \frac{M}{2}. \qquad [20\%]$$

(b) Consider now a code $\mathcal{C}'_n$ with $M' = 2M - 1$ codewords of length n constructed over alphabet $\mathcal{X}$. The codewords are independently generated with distribution Q_{X^n}. Let $p_{e,m'}(\mathcal{C}'_n)$ be the error probability of the m'-th codeword of a given codebook $\mathcal{C}'_n$ for $m' = 1, \ldots, M'$ and $\mathbb{E}[p_{e,m'}(\mathsf{C}'_n)]$ its average over the random-coding ensemble.

(i) Compare the rate of $\mathcal{C}'_n$ with that of $\mathcal{C}_n$ in Part (a). [10%]

(ii) Show that there exists a code $\tilde{\mathcal{C}}_n$ with M codewords obtained by removing codewords from $\mathcal{C}'_n$ such that $p_{e,m}(\tilde{\mathcal{C}}_n) < 2^{1/s} \cdot \mathbb{E}[p_{e,m'}(\mathsf{C}'_n)^s]^{1/s}$ for $m = 1, \ldots, M$. [15%]

(iii) Using that $\big(\sum_i a_i\big)^s \leq \sum_i a_i^s$ for $0 \leq s \leq 1$ and that for a given code

$$p_{e,m'}(\mathcal{C}'_n) \leq \sum_{\bar{m} \neq m'} \sum_{y^n} \sqrt{W^n(y^n|x^n(\bar{m}))W^n(y^n|x^n(m'))},$$

where $W^n(y^n|x^n)$ is the n-use channel transition probability, show that

$$\mathbb{E}[p_{e,m'}(\mathsf{C}'_n)^s] \leq (M'-1) \sum_{x^n} \sum_{\bar{x}^n} Q_{X^n}(x^n) Q_{X^n}(\bar{x}^n) \Big(\sum_{y^n} \sqrt{W^n(y^n|x^n)W^n(y^n|\bar{x}^n)} \Big)^s.$$

[20%]

(iv) Finally, show that there exists a code with M codewords such that for $0 \leq s \leq 1$,

$$p_{e,m}(\tilde{\mathcal{C}}_n) < (4(M-1))^{\frac{1}{s}} \Bigg(\sum_{x^n} \sum_{\bar{x}^n} Q_{X^n}(x^n) Q_{X^n}(\bar{x}^n) \Big(\sum_{y^n} \sqrt{W^n(y^n|x^n)W^n(y^n|\bar{x}^n)} \Big)^s \Bigg)^{\frac{1}{s}}.$$

[10%]

4 (a) Let P_V be the probability distribution of a discrete memoryless source V over alphabet $\mathcal{V}$. For $0 \leq \rho \leq 1$, consider the function

$$E_s^V(\rho) = \log \left(\sum_{v \in \mathcal{V}} P_V(v)^{\frac{1}{1+\rho}} \right)^{1+\rho} \tag{1}$$

(i) Sketch the function $E_s^V(\rho)$. Explain the connection of the function $E_s^V(\rho)$ with the random-coding average probability of error $\bar{p}_e$. Explain how it can be used to show that there exist fixed-length source codes of rate $R = \frac{1}{n} \log_2 M$ such that for $R > H(V)$, their probability of error tends to zero as $n \to \infty$, where M is the number of indices that are used for compression. [15%]

(ii) Let P_Z be the probability distribution of a discrete memoryless source Z over alphabet $\mathcal{V}$. Assume that $P_Z(z) \neq P_V(v)$ for some pairs $(z, v) \in \mathcal{V} \times \mathcal{V}$ and that $H(Z) = H(V)$. Sketch the functions $E_s^V(\rho), E_s^Z(\rho)$ and compare them. [15%]

(iii) Consider the sources described Part (a)(ii). Show that, when compressing the two sources in parallel, there exist codes of rate $R > 2H(V)$ such that the error probability decays exponentially with exponent given by

$$\max_{0 \leq \rho \leq 1} \rho R - \left(E_s^V(\rho) + E_s^Z(\rho) \right).$$

[20%]

(b) Consider a Bayesian multiple hypothesis testing problem between distributions with $|\mathcal{V}|$ hypotheses $P_{Y|V}(y|v)$ for $v \in \mathcal{V}$, defined over the same alphabet $\mathcal{Y}$.

(i) Define the hypothesis testing problem and the corresponding average error probability $\bar{\varepsilon}$. [20%]

(ii) For $\gamma > 0$, define the set

$$\mathcal{S}(\gamma) = \left\{ (v, y) \in \mathcal{V} \times \mathcal{Y} \, : \, \frac{P_{VY}(v, y)}{P_Y(y)} \leq \gamma \right\}.$$

Show that the error probability of a given test is such that

$$\bar{\varepsilon} \geq \mathbb{P}\left[\frac{P_{VY}(V, Y)}{P_Y(Y)} \leq \gamma \right] - \gamma.$$

[30%]

END OF PAPER

THIS PAGE IS BLANK