EE5848 Topics in Information Theory & Coding (2020-21)
Programming Assignment - 1 (30 marks)

Due at 6am, 25 Mar. 2021

Submission Instructions: You may form teams of size 1-2 students (only among students crediting the course). Exactly one of the team members must upload the simulation result online through Google Forms. You must upload

1. the simulation result as a figure in '.pdf' format.
2. the program/script that generates this output (in Matlab/python).

Plagiarism, in any form, will be taken seriously. Please write the program on your own. You could be asked to present the script file and explain the algorithm orally at short notice. You are allowed to discuss with other teams, but the program must be written on your own.

Programming Language: You can use Matlab, Python or any other tool for this programming assignment.

The Problem: You must plot the complementary cumulative distribution functions of the overall completion time of the following three distributed matrix-vector multiplication schemes:

1. **Code A**: a $(32, 16)$ code which is related to Reed-Muller code and whose generator matrix is described further below. This curve must be obtained using Monte-Carlo simulation.
2. **Code B**: using a $(32, 16)$ MDS code. This curve must also be obtained using Monte-Carlo simulation.
3. the uncoded scheme for $n = 32$ workers. You can either use the expression derived in lectures to directly plot this curve or use Monte-Carlo simulation.

You must use the following mother runtime distribution to conduct the simulation:

$$F(t) = P[T_0 \leq t] = 0 \text{ if } t \leq 1 \quad \text{and } F(t) = 1 - \exp\big(-(t-1)\big) \text{ for } t > 1.$$

There are $n = 32$ workers in the system. For the two coded schemes above, the individual runtimes $T_1^{\text{coded}}, \ldots, T_{32}^{\text{coded}}$ are independent and identically distributed random variables with CDF $F(kt)$. For the uncoded scheme the individual runtimes have CDF $F(nt)$.

For all the schemes, you must find the probability $P(\text{overall runtime} > t)$ for $t = 0.10, 0.11, 0.12, 0.13, 0.14$. Plot all three curves (with t in the x-axis and the probability on the y-axis) in the same figure. The y-axis must be in the logarithmic scale, i.e., the y-axis must be $\log_{10}[\text{probability}]$. You can achieve this in Matlab by using the command `semilogy`, and in python by using `matplotlib.pyplot.semilogy`. Please use horizontal and vertical grid lines in your plot.

Reference value: For $t = 0.115$, these probabilities are approximately equal to 0.26 and 0.16 for the Reed-Muller coded scheme and the MDS coded scheme, respectively.

Generator Matrix for Code A:

Start with the 2×2 matrix $H_2 = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$. Given an $n \times n$ matrix H_n, define

$$H_{2n} = \begin{bmatrix} H_n & 0 \\ H_n & H_n \end{bmatrix}.$$

Sort the rows of H_{32} in the descending order of number of '+1's in each row, i.e., with the first row containing most number of $+1$. Define the generator matrix G of Code A as the submatrix of this sorted matrix consisting of its first 16 rows.

Simulation Idea:

One possible way to conduct the simulation is as follows. For each of the two coded schemes, you must generate several independent samples (about 10^5 or more) of $T_{\text{overall}}^{\text{coded}}$, and store them in an array. The fraction of elements in the array with values exceeding t is an estimate for $P(T_{\text{overall}}^{\text{coded}} > t)$. To generate one sample of $T_{\text{overall}}^{\text{coded}}$ you will generate $n = 32$ iid random variables (each with CDF $F(kt)$) representing the individual runtimes of the workers. For these values of the worker runtimes, find the overall runtime of the two coded scheme, and store it in the array.