Technical University of Munich
Coding for Communications and Data Storage
Prof. Dr.-Ing. Antonia Wachter-Zeh

Coding Theory for Storage and Networks

Lab 1: Regenerating Codes

Solve the tasks in *RegeneratingCodes.sage*. Complete the lines with ? markers in comment or the noted blocks.

In this lab, consider a distributed storage system with n nodes, each nodes stores α symbols, such that a data collector can reconstruct the whole storage by contacting any k nodes. The repair degree of the system is d and the repair bandwidth of each node if β.

Problem 1: MSR Codes by ZigZag Codes

In this task, we use an $(n, k = n-2, d = n-1)$ ZigZag code ([Tamo, Wang and Bruck, '11]) to construct an $(n, k, d) = (5, 3, 4)$, $\alpha = 4$, $\beta = 2$ Regenerating Code.

The $(5, 3, 4)$ ZigZag code for the message $\boldsymbol{u} = (u_1, \ldots, u_{12})$ over $\mathbb{F}_3$ is given below:

Node 1	u_1	u_2	u_3	u_4
Node 2	u_5	u_6	u_7	u_8
Node 3	u_9	u_{10}	u_{11}	u_{12}
Node 4	$u_1 + u_5 + u_9$	$u_2 + u_6 + u_{10}$	$u_3 + u_7 + u_{11}$	$u_4 + u_8 + u_{12}$
Node 5	$u_1 + 2u_7 + 2u_{10}$	$u_2 + 2u_8 + u_9$	$u_3 + u_5 + u_{12}$	$u_4 + u_6 + 2u_{11}$

Denote the content of storage at each node $i \in \{1, \ldots, 5\}$ by $\boldsymbol{s}_i$, e.g., $\boldsymbol{s}_1 = (u_1, u_2, u_3, u_4)$.

1. Determine the encoding matrix $\boldsymbol{A}_i$, $i = 1, \ldots, 5$ of each user such that $\boldsymbol{s}_i = \boldsymbol{A}_i \boldsymbol{u}$.
   Hint: Every $\boldsymbol{A}_i$ is a 4×12 matrix over $\mathbb{F}_3$.

2. Measure the time spent to encode the data to be stored on the $n = 5$ nodes.

   Hint: Use the funtion time.time().

3. Assume node 1 has failed. Repair it with minimum bandwidth.

   Hint: First determine which symbols from which nodes should be downloaded.

   How much time does it take to get the node up and running? Assume downloading one symbol cost 1 second. Calculate the time spent on downloading and measure the time spent on decoding the downloaded symbol to repair Node 1.

4. Calculate the minimum time spent to reconstruct the file by contacting the first $k = 3$ nodes. Assume downloading one symbol cost 1 second.

5. Compare the time spent in task 3 and task 4. Is it more efficient to regenerate the node with minimum bandwidth than reconstruct the file? Justify your answer.

Problem 2: MSR Codes by Product Matrix Codes

In this task, we use an $(n, k, d = 2k-2)$ Product Matrix Code ([Rashmi, Shah and Kumar, '11]) to construct an $(n, k, d) = (5, 3, 4)$ Regenerating Code over $\mathbb{F}_{13}$. Let the information stored be $\boldsymbol{u} = (2, 2, 3, 5, 6, 10)$.

1. Construct the Vandermonde matrix Ψ and the information matrix $\boldsymbol{M}$.

   Measure the time spent to encode the data to be stored on the $n = 5$ nodes, denoted the encode data matrix by $\boldsymbol{S}$.

2. Assume Node 4 has failed, i.e., the 4-th row of $\boldsymbol{S}$ is erased. How do you repair it by connecting $d = 4$ nodes with bandwidth $\beta = 1$?

   Hint: Determine the coding vector at the helper nodes side and the coding matrix at the repair node side.

   How much time does it take to get the node up and running? Here we ask you to measure the time spent on downloading and repairing Node 4.

3. Reconstruct the file by contacting Node 1, 2, 3.

   Hint: Set up and solve the linear system of equations.

   Calculate the time spent on downloading and measuring the time spent on decoding (assuming download one symbol cost 1 second).

4. Compare the time consumption for repair and reconstruct. Is it more efficient to regenerate the node than reconstruct the file? Justify your answer.

5. Compare the time consumption for repair and reconstruct in these two problems. Which code is better? Justify your answer.