On Optimal Heterogeneous Regenerating Codes

Krishna Gopal Benerjee and Manish K. Gupta, Senior Member, IEEE
Dhirubhai Ambani Institute of Information and Communication Technology Gandhinagar,
Gujarat, 382007 India
Email: krishna_gopal@daiict.ac.in, mankg@computer.org

Abstract

Heterogeneous Distributed Storage Systems (DSSs) are close to the real world applications for data storage. Each node of the considered DSS, may store different number of packets and each having different repair bandwidth with uniform repair traffic. For such heterogeneous DSS, a failed node can be repaired with the help of some specific nodes. In this work, a family of codes based on graph theory, is constructed which achieves the fundamental bound on file size for the particular heterogeneous DSS.

Index Terms

Heterogeneous DSS, graphical construction, heterogeneous regenerating codes, repair bandwidth.

I. INTRODUCTION

DATA storage is a big challenge for mankind since ancient times. Recently emerged Cloud computing provides an excellent way to store the data in a Distributed Storage Systems (DSSs). In such a DSS, data file is stored on \( n \) distinct nodes in such a smart way that the complete file can be retrieved by connecting certain number of nodes. In the case of node failure, system has to repair the failed node by either generating functional equivalent of the data loss or by generating the exact data that was lost on that node. In order to provide reliability systems use either simple replication or MDS (maximum distance separable) erasure codes. Simple replication uses more space (so it is bad for storage minimization) and erasure MDS code approach is not efficient for bandwidth minimization in a node repair process. To optimize these conflicting parameters data storage and bandwidth, in a seminal paper Dimakis et. al [1] introduced family of codes called regenerating codes. These regenerating codes received attention in several papers [1–4].

Consider a DSS of total \( n \) distinct nodes. Homogeneous Regenerating codes are specified by the parameters \( [n, k, d, \alpha, \beta, B] \), where \( B \) is the file size and \( \alpha \) is the number of packets on each node. In
order to get a file, user has to connect $k$ ($< n$) nodes out of total $n$ nodes. The particular $\alpha$ and $k$ are known as node storage capacity and reconstruction degree for the DSS [5]. In case of a node failure, data can be recovered by contacting $d$ nodes and downloading $\beta$ packets from each node. Thus total bandwidth for a repairing a node is $d\beta$, where $d$ and $\beta$ are known as repair degree and repair traffic respectively. By optimizing both $\alpha$ and $\beta$ in different order, we get two kind of regenerating codes called Minimum Storage Regenerating (MSR) codes and Minimum Bandwidth Regenerating (MBR) codes. Many researchers constructed MBR and MSR codes using combinatorial designs, graphs, sequences and matrices etc [6, 7].

For $(n, k)$ heterogeneous DSS with dynamic node storage capacity, repair traffic, repair degree and reconstruction degree, a fundamental bound is established in [8]. The computational complexity to calculate parameters, for the fundamental bound achieving codes, is very high. Hence, in this work, we consider a special case (considered in [9]) by choosing constant repair traffic and reconstruction degree. Further, we calculate relations on parameters of codes which achieves the bound. For the particular parameters, the optimal codes (the fundamental bound achieving codes) are constructed using graphs.

Organization: The paper is organized as follows. Section 2 describes the model of heterogeneous DSS and collects the necessary background including fundamental bound on file size for the heterogeneous DSS. Conditions for the optimal codes are established in Section 3. Graphical construction of the family of such optimal codes is given in Section 4. Final section concludes the paper with general remarks.

II. MODEL

In heterogeneous DSS, a file is divided into encoded packets and the encoded packets are distributed among $n$ distinct nodes $U_i$ ($i = 1, 2, \ldots, n$) such that each node has storage capacity $\alpha_i$ and repair degree $d_i$. An user can reconstruct the file by downloading data from any $k$ ($< n$) nodes. If a node $U_i$ fails then data collector will download $\beta$ packets from specific $d_i$ nodes out of remaining $n - 1$ nodes. The particular $d_i$ nodes are called helper nodes for the failed node $U_i$. In such a case, repair bandwidth for a node $U_i$ is $\gamma_i = d_i\beta$.

An example of such heterogeneous DSS is illustrated in Figure 1. In this example, a file with size 3 ($= B$) is stored in $(n = 6, k = 2)$ heterogeneous DSS with repair traffic $\beta$ is 1. In the particular DSS, node storage capacity $\alpha_i$ is 2, 2, 2, 3, 2 and 2 for $i = 1, 2, 3, 4, 5, 6$ (see Figure 1). Note that $\alpha_i = \gamma_i = d_i$ for each $i = 1, 2, 3, 4, 5, 6$.

Again, a set of $d_i$ helper nodes which are used for repairing the failed node $U_i$, is called as surviving set. Formally, surviving sets are defined as follows.
Fig. 1. A file is divided into 3 (= B) distinct coded packets \( x_1, x_2 \) and \( x_3 \) on field \( \mathbb{F}_q \). These three packets are encoded into thirteen distinct packets and distributed in \((6,2)\) heterogeneous DSS.

**Definition 1.** (Surviving Set): In a \((n,k)\) heterogeneous DSS, surviving set of a node \( U_i \) \((1 \leq i \leq n)\) is a set of \( d_i \) nodes which are used for repairing the node \( U_i \). Note that there could be several surviving sets for a given node \( U_i \). Indexing all the distinct surviving sets by a positive integer \( \ell_i \), let us denote them by \( S_i^{(\ell_i)} \) \((\ell_i = 1, 2, \ldots)\). For a particular node \( U_i \), number of distinct surviving sets are finite say \( \tau_i \) then \( \ell_i = 1, 2, \ldots, \tau_i \).

For \((6,2)\) heterogeneous DSS (as shown in Figure 1), the surviving sets of each node are listed in Table I.

**TABLE I**

SURVIVING SETS FOR NODES OF \((6,2)\) HETEROGENEOUS DSS CONSIDERED IN FIGURE 1.

| Nodes \( U_i \) | All possible surviving sets \( S_i^{(\ell_i)} \) | \# surviving sets \( \tau_i \) |
|-----------------|----------------------------------------|---|
| \( U_1 \)    | \( S_1^{(1)} = \{U_4, U_6\} \), \( S_1^{(2)} = \{U_2, U_6\} \). | 2 |
| \( U_2 \)    | \( S_2^{(1)} = \{U_1, U_3\} \), \( S_2^{(2)} = \{U_1, U_5\} \), \( S_2^{(3)} = \{U_4, U_3\} \), \( S_2^{(4)} = \{U_4, U_5\} \). | 4 |
| \( U_3 \)    | \( S_3^{(1)} = \{U_2, U_4\} \), \( S_3^{(2)} = \{U_2, U_5\} \), \( S_3^{(3)} = \{U_5, U_4\} \), \( S_3^{(4)} = \{U_5, U_6\} \). | 4 |
| \( U_4 \)    | \( S_4^{(1)} = \{U_1, U_3, U_5\} \), \( S_4^{(2)} = \{U_2, U_3, U_5\} \), \( S_4^{(3)} = \{U_1, U_6, U_5\} \), \( S_4^{(4)} = \{U_2, U_6, U_5\} \). | 4 |
| \( U_5 \)    | \( S_5^{(1)} = \{U_2, U_4\} \), \( S_5^{(2)} = \{U_3, U_5\} \). | 2 |
| \( U_6 \)    | \( S_6^{(1)} = \{U_3, U_1\} \), \( S_6^{(2)} = \{U_4, U_1\} \). | 2 |
Similar to the parameters of the regenerating codes for homogeneous systems [5], we provide the parameters of Heterogeneous Regenerating codes in the next remark.

**Remark 2.** For a \((n, k)\) heterogeneous DSS, regenerating codes over a field \(\mathbb{F}_q\) are described by the parameters \([n, k, d, \bar{d}, \beta, B]\), where \(B\) is the file size, \(\beta\) is the repair traffic, \(\bar{d} = [d_{i,j}]_{1 \times n}\) and \(\bar{d} = [\alpha_{i,j}]_{1 \times n}\) are one dimensional arrays of repair degree \(d_i\) and node storage capacity \(\alpha_i\) for node \(U_i\) indexed with \(i = 1, 2, \ldots, n\).

Note that multi-node failure can be assumed as a sequence of single node failure within a small time interval. So the sequence of surviving sets are needed to repair such multi node failure. Formally, the surviving sequence can be defined as follows.

**Definition 3.** (Surviving Sequence): For a \((n, k)\) heterogeneous DSS, surviving sequence \(\left< S_{f_j}^{(\ell_{f_j})} \right>_{j=1}^n\) is a sequence of surviving sets picked up randomly one for each node \(U_{f_j}\), where \(f_j\) is some permutation on set \(\{1, 2, \ldots, n\}\) and \(\ell_{f_j} \in \{1, 2, \ldots, \tau_{f_j}\}\).

For some particular surviving sets, a possible surviving sequence
\[
\left< S_{f_j}^{(\ell_{f_j})} \right>_{j=1}^6 = \left< S_{f_5}^{(1)}, S_{f_3}^{(1)}, S_{f_4}^{(2)}, S_{f_6}^{(2)}, S_{f_1}^{(1)}, S_{f_2}^{(2)} \right>
\]
associated with failed nodes \(U_5, U_3, U_4, U_6, U_1\) and \(U_2\) in \((6, 2)\) heterogeneous DSS (see Table I).

In [9], heterogeneous DSS is mapped with acyclic directed graph called information flow graph. Analyzing min-cut of the information flow graph, a fundamental bound on file size \(B\) is computed for such \((n, k)\) heterogeneous DSS. The bound is described in following theorem.

**Theorem 4 (Fundamental Bound).** For a \((n, k)\) heterogeneous DSS, the file size \(B\) must satisfy the following inequality

\[
B \leq \min_{\left< S_{f_j}^{(\ell_{f_j})} \right>_{j=1}^k} \left\{ \sum_{j=1}^k \min \left\{ \alpha_{f_j}, \left( \min_{\lambda=0} \left\{ \left. \beta \right| \left( \bigcup_{\lambda=0} \left\{ U_{f_\lambda} \right\} \right) \right. \right) \right\} \right\},
\]

where \(\{U_{f_\lambda}\} = \phi, 0 \leq \lambda < j \leq k, S_{f_j}^{(\ell_{f_j})} \in \left< S_j^{(\ell_{f_j})} \right>_{j=1}^n, \mathcal{F} \) is the set of all surviving sequences \(\left< S_{f_j}^{(\ell_{f_j})} \right>_{j=1}^n\) with length \(n\) and \(\ell_{f_j} \in \{1, 2, \ldots, \tau_{f_j}\}\).

In [9], it is shown that there exist code which achieves the fundamental bound for such \((n, k)\) heterogeneous DSS. Hence, one can get the optimal codes by reducing parameters which meets the fundamental bound. In the next section, parameters for the optimal codes are computed by minimizing node storage capacity and repair bandwidth.
III. CONDITIONS FOR OPTIMALITY

Consider a \((n, k)\) heterogeneous DSS with \(\tau_i\) number of surviving sets \(S_i^{(\ell_i)}\) and repair degree \(|S_i^{(\ell_i)}| = d_i\) (\(\ell_i = 1, 2, \ldots, \tau_i\) and \(i = 1, 2, \ldots, n\)). If \(\alpha_i > |S_i^{(\ell_i)}|\beta\) then the failed node \(U_i\) can not be repaired so \(\alpha_i \leq |S_i^{(\ell_i)}|\beta\) for each \(i\) and \(\ell_i\). For optimality, \(\alpha_i = d_i\beta\). Hence for constant repair traffic \(\beta\), node storage capacity \(\alpha_i\) and repair degree \(d_i\) are proportional to each other. Consider \(c_i \in \{0, 1\} \subset \mathbb{R}\) such that \(\sum_{i=1}^{n} c_i = 1\) and \(c_i / \alpha_i = c_j / \alpha_j\) for \(1 \leq i < j \leq n\). Hence, \(\alpha_i = c_i \sum_{i=1}^{n} \alpha_i = c_i \alpha^* (i = 1, 2, \ldots, n)\).

So, the parameters \(\alpha_i\) and \(c_i\) are proportional to each other. Again, \(k\) is reconstruct degree so, \(B \leq \sum_{i \in \mathcal{K}} \alpha_i = \sum_{i \in \mathcal{K}} c_i \alpha^*\) for any arbitrary set \(\mathcal{K} \subset \{1, 2, \ldots, n\}\) such that \(|\mathcal{K}| = k\). Hence, \(B \leq \sum_{i=1}^{k} c_i \alpha^*\) for \(c_1 \leq c_2 \leq \ldots \leq c_n\). For optimum case, one can reduce \(\alpha^*\) up to \(\alpha_{min}^*\) such that

\[
B = \sum_{i=1}^{k} c_i \alpha_{min}^* \Rightarrow \alpha_{min}^* = B \left(\sum_{j=1}^{k} c_j\right)^{-1},
\]

Similarly for a fixed proportional factor \(\alpha_{min}^*\), one can minimize the repair traffic \(\beta\) such that Bound 4 holds with equality. For a specific surviving sequence \(\left\langle S_{j}^{(\ell_j)}\right\rangle_{j=1}^{n}\) with sufficient large repair traffic \(\beta\), the inequality \(\alpha_{f_m} \leq \left|S_{j}^{(\ell_j)}\right\rangle_{j=1}^{n} \left(\bigcup_{\lambda=0}^{m-1} \{U_{f_\lambda}\}\right)\beta\) holds for each \(m = 1, 2, \ldots, k\). If we choose \(\beta = \beta_{min}\)

\[
\beta_{min} = \max_{1 \leq m \leq k} \left\{ \max_{j=1}^{l} \left\{ \alpha_{f_m} \left|S_{j}^{(\ell_j)}\right\rangle_{j=1}^{n} \left(\bigcup_{\lambda=0}^{m-1} \{U_{f_\lambda}\}\right)^{-1}\right\}\right\}
\]

then \(\beta_{min}\) is the minimum value of repair traffic \(\beta\) which ensures \(\left|S_{j}^{(\ell_j)}\right\rangle_{j=1}^{n} \left(\bigcup_{\lambda=0}^{m-1} \{U_{f_\lambda}\}\right)\beta_{min} \geq \alpha_{f_m}\) for each \(f_m\) of an arbitrary surviving sequence.

Formally the results can be summarized by the following theorem.

**Theorem 5.** Consider a \((n, k)\) heterogeneous DSS with given surviving sets \(S_i^{(\ell_i)}\) (\(i = 1, 2, \ldots, n\); \(\ell_i = 1, 2, \ldots, \tau_i\) for some \(\tau_i \in \mathbb{Z}\)). A family of codes with \(\alpha_i = c_i \alpha_{min} = d_i \beta_{min}\) and \(\beta = \beta_{min}\), achieves the Fundamental Bound 4, where \(\alpha_{min}\) and \(\beta_{min}\) can be calculated by Equations (1) and (2).

Next section presents a construction of an optimal family of regenerating code based on graph.

IV. FAMILY OF OPTIMAL CODES

Graph \(\mathcal{H}(\mathcal{V}, \mathcal{E})\) is representation of vertex set \(\mathcal{V}\) and edge set \(\mathcal{E}\) such that \(\mathcal{E} = \{E : |E| = 2\} \subset \mathcal{V}\). For a graph \(\mathcal{H}(\mathcal{V}, \mathcal{E})\), degree of vertex \(v_i \in \mathcal{V}\) (denoted by \(\text{deg}(v_i)\)) is the total number of edges \(E \in \mathcal{E}\) such that \(v_i \in E\). An edge \(E \in \mathcal{E}\) is called loop if \(E = \{v_i, v_i\}\) for some \(i \in \{1, 2, \ldots, n\}\). Similarly, two edges \(E_l, E_m \in \mathcal{E}\) are called parallel edges if \(E_l = E_m\) (\(l \neq m\)). A graph \(\mathcal{H}(\mathcal{V}, \mathcal{E})\) is called simple if the graph does not have loop or parallel edges. Two vertices \(v_i\) and \(v_j\) are adjacent if
For the respective $(4,2)$ heterogeneous DSS, packets $y_j$ ($j = 1, 2, ..., 10$) are distributed among $4$ $(= n)$ nodes (see Figure 2). Observe that $S_1^{(1)} = \{U_3, U_4\}$, $S_2^{(1)} = \{U_3, U_4\}$, $S_3^{(1)} = \{U_1, U_2, U_4\}$ and $S_4^{(1)}$
= \{U_1, U_2, U_3\} and \beta_{\text{min}} = 1. Hence, the graphical construction can be summarized by the following theorem.

**Theorem 6.** Consider a simple connected graph \( \mathcal{H}(V, E) \) with \( 2 \leq \text{deg}(v_m) \leq \text{deg}(v_j) \) (for \( 1 \leq m < j \leq n \)) such that two arbitrary vertices from \( \{v_1, v_2, \ldots, v_k\} \) are not adjacent through an edge. A \((n = |V|, k = |\{v_1, v_2, \ldots, v_k\}|)\) heterogeneous DSS associated with the graph \( \mathcal{H}(V, E) \), achieves the Fundamental Bound 4, where the node storage capacity \( \alpha_i \) and repair degree \( d_i \) are each \( \text{deg}(v_i) \) and repair traffic \( \beta \) is 1.

**V. CONCLUSION**

Motivated by the real world applications we considered heterogeneous DSSs with dynamic repair degree and node storage capacity. A graphical construction framework is used to construct a family of optimal regenerating codes meeting heterogeneous DSS fundamental bound on file size.

**REFERENCES**

[1] A. Dimakis, P. Godfrey, M. Wainwright, and K. Ramchandran, “Network coding for distributed storage systems,” in INFOCOM 2007. 26th IEEE International Conference on Computer Communications. IEEE, May 2007, pp. 2000 –2008.

[2] A. Dimakis, K. Ramchandran, Y. Wu, and C. Suh, “A survey on network codes for distributed storage,” Proceedings of the IEEE, vol. 99, no. 3, pp. 476 –489, march 2011.

[3] S. Pawar, N. Noorshams, S. El Rouayheb, and K. Ramchandran, “Dress codes for the storage cloud: Simple randomized constructions,” in Information Theory Proceedings (ISIT), 2011 IEEE International Symposium on, 31 2011-aug. 5 2011, pp. 2338 –2342.

[4] M. K. Gupta, A. Agrawal, and D. Yadav, “On weak dress codes for cloud storage,” CoRR, vol. abs/1302.3681, 2013.

[5] A. Dimakis, P. Godfrey, Y. Wu, M. Wainwright, and K. Ramchandran, “Network coding for distributed storage systems,” Information Theory, IEEE Transactions on, vol. 56, no. 9, pp. 4539–4551, Sept 2010.

[6] N. Silberstein, A. Rawat, and S. Vishwanath, “Error-correcting regenerating and locally repairable codes via rank-metric codes,” Information Theory, IEEE Transactions on, vol. 61, no. 11, pp. 5765–5778, Nov 2015.

[7] K. G. Benerjee and M. K. Gupta, “On dress codes with flowers,” Signal Design and Its Applications in Communications, The Seventh International Workshop on, pp. 108–112, Sept. 2015.

[8] ——, “Tradeoff for heterogeneous distributed storage systems between storage and repair cost,” CoRR, vol. abs/1503.02276, 2015.

[9] ——, “On heterogeneous regenerating codes and capacity of distributed storage systems,” CoRR, vol. abs/1402.3801, 2014.