Repairing Schemes for Tamo-Barg Codes

Han Cai*, Ying Miao†, Moshe Schwartz‡§, and Xiaohu Tang*
*School of Information Science and Technology, Southwest Jiaotong University,
Chengdu, 610031, China, (e-mail: hancai@aliyun.com;xhutang@swjtu.edu.cn)
†Faculty of Engineering, Information and Systems, University of Tsukuba,
Tennodai 1-1-1, Tsukuba 305-8573, Japan (e-mail: miao@sk.tsukuba.ac.jp)
‡Department of Electrical and Computer Engineering, McMaster University,
Hamilton, Ontario L8S 4K1, Canada (e-mail: schwartz.moshe@mcmaster.ca)
§School of Electrical and Computer Engineering, Ben-Gurion University of the Negev,
Beer Sheva 8410501, Israel (on a leave of absence)

***Abstract*—We study the problem of repairing erasures in locally repairable codes beyond the code locality under the rack-aware model. We devise two repair schemes to reduce the repair bandwidth for Tamo-Barg codes under the rack-aware model, by setting each repair set as a rack. The first repair scheme provides optimal repair bandwidth for one rack erasure. We then establish a cut-set bound for locally repairable codes under the rack-aware model. Using this bound we show that our second repair scheme is optimal. Furthermore, we consider the partial-repair problem for locally repairable codes under the rack-aware model, and introduce both repair schemes and bounds for this scenario.**

I. Introduction

WITH the expanding volume of data in large-scale cloud storage and distributed file systems, like Windows Azure Storage and Google File System (GoogleFS), disk failures have become a norm rather than an exception. To protect data from such failures, the simplest solution is to replicate data packets across different disks. However, this approach suffers from large storage overhead. Consequently, coding techniques have been developed as an alternative solution.

Regenerating codes and locally repairable codes are two different methods to improve the repair efficiency of failed nodes. Over the past decade, many results have been obtained in this area, e.g., see [7], [9], [17], [19], [22], [28]–[30], [32], [33] for regenerating codes, and see [2], [3], [5], [6], [10], [14], [16], [20], [25]–[27], [31] for optimal locally repairable codes.

Another approach combines locally repairable codes and regenerating codes, by allowing the codes in each set to form regenerating codes, e.g., [11], [15], [18]. By doing so, the repair bandwidth required can be reduced when the system performs local erasure repairs.

However, this method has a drawback: the repair property only works for the punctured codes in the repair sets. This means that if there are erasures beyond the local repair capability in one repair set, the repair scheme and the locality cannot simultaneously reduce the repair bandwidth. To address this issue, we propose a new combination strategy. Our approach involves repair schemes for locally repairable codes that can handle erasures beyond local recoverability.

The work of H. Cai was supported by National Key R&D Program of China under Grant 2022YFA1005000.

The work of Y. Miao was supported by JSPS Grant-in-Aid for Scientific Research (C) under Grant 24K06824

Specifically, in this paper, we propose repair schemes for the well-known Tamo-Barg codes [27], which are optimal locally repairable codes with respect to the Singleton-type bound [8], [24]. We present two proposed schemes. Firstly, in a rack-aware model where each repair set is one rack, we introduce an optimal repair scheme for the case of one failed rack, i.e., one erased repair set. Secondly, for the scenario where there are erasures within a repair set that cannot be recovered locally, we introduce a repair scheme that reduces the repair bandwidth required for recovering those failures. We prove the optimality of our schemes by modifying the well-known cut-set bound [7] to incorporate locality. Our proposed schemes generalize the rack-aware model regenerating codes [12], [13].

Due to space limitations we omit all proofs, which are available in a full version of this work [4].

II. Preliminaries

We start by introducing basic notation and definitions. For any $n \in \mathbb{N}$ we denote $[n] \triangleq \{1, 2, \ldots, n\}$. For a prime power q, let $\mathbb{F}_q$ denote the finite field of size q, $\mathbb{F}_q^* \triangleq \mathbb{F}_q \setminus \{0\}$, and let $\mathbb{F}_q[x]$ denote the set of polynomials in the indeterminate x with coefficients from $\mathbb{F}_q$. An $[n, k]_q$ linear code $\mathcal{C}$ over $\mathbb{F}_q$ is a k-dimensional subspace of $\mathbb{F}_q^n$ with a $k \times n$ generator matrix $G = (\mathbf{g}_1, \mathbf{g}_2, \ldots, \mathbf{g}_n)$, where $\mathbf{g}_i$ is a column vector of length k for all $i \in [n]$. More specifically, it is called an $[n, k, d]_q$ linear code if its minimum Hamming distance is d.

A. Generalized Reed-Solomon codes

Let $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n) \in \mathbb{F}_q^n$ contain distinct entries, where we assume $q \geqslant n$. Then the well-known *generalized Reed-Solomon (GRS) code* with parameters $[n, k, n-k+1]_q$ can be defined as

$$\begin{aligned}\mathrm{GRS}_k(\boldsymbol{\theta}, \boldsymbol{\alpha}) \triangleq &\{(\alpha_1 f(\theta_1), \alpha_2 f(\theta_2), \ldots, \alpha_n f(\theta_n)) \\ &: \ f(x) \in \mathbb{F}_q[x] \text{ with } \deg(f(x)) < k\},\end{aligned}$$

where $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_n) \in (\mathbb{F}_q^*)^n$. It is well known that the dual of an $[n, k, n-k+1]_q$ GRS code is an $[n, n-k, k+1]_q$ GRS code (e.g., see [23]).

B. Regenerating codes

An important problem in distributed storage systems is to repair an erasure by downloading as little data as possible. Dimakis et al. [7] introduced *repair bandwidth*, the amount of data downloaded during a node repair, as a metric to measure the procedure's efficiency.

Definition 1: Let $\mathcal{C}$ be an $[N, K]_q$ array code with *sub-packetization* L, that is, $c_i \in \mathbb{F}_q^L$ for any codeword $(c_1, c_2, \ldots, c_N) \in \mathcal{C}$. For an erasure pattern $\mathcal{E} \subseteq [N]$ and a D-subset $\mathcal{R} \subseteq [N] \setminus \mathcal{E}$ (whose entries are called *helper nodes*), define $B(\mathcal{C}, \mathcal{E}, \mathcal{R})$ as the *minimum repair bandwidth* for $c_i \in \mathbb{F}_q^L$ stored in node $i \in \mathcal{E}$, i.e., the smallest total number of symbols of $\mathbb{F}_q$ helper nodes need to send in order to recover c_i (where each helper node $j \in \mathcal{R}$ may send symbols that depend solely on $c_j \in \mathbb{F}_q^L$).

In [7], the well-known cut-set bound was first derived for the minimum download bandwidth.

Theorem 1 (Cut-set bound, [1], [7])**:** Let $\mathcal{C}$ be an $[N, K]_q$ MDS array code with sub-packetization L. Let D be an integer with $K \leqslant D \leqslant N - 1$. For any non-empty $\mathcal{E} \subseteq [N]$ with $|\mathcal{E}| \leqslant N - D$ and any D-subset $\mathcal{R} \subseteq [N] \setminus \mathcal{E}$, we have

$$B(\mathcal{C}, \mathcal{E}, \mathcal{R}) \geqslant \frac{DL}{D - K + |\mathcal{E}|}.$$

Definition 2: For $K < D \leqslant N - \tau$, an $[N, K]_q$ MDS array code is said to be an $[N, K]_q$ *minimum storage regenerating* (MSR) code with repair degree D, if for each $\mathcal{I} = \{i_1, i_2, \ldots, i_\tau\} \subset [N]$ there exists a D-subset $\mathcal{R}_\mathcal{I} \subseteq [N] \setminus \mathcal{I}$ such that $B(\mathcal{C}, \mathcal{I}, \mathcal{R}_\mathcal{I})$ meets the cut-set bound described above with equality. Throughout this paper, such codes are also said to have (τ, D) *optimal repair* property.

C. Locally repairable codes

Another important figure of merit is symbol locality [8], [24].

Definition 3: Let $\mathcal{C}$ be an $[n, k, d]_q$ linear code. For $j \in [n]$, the j-th code symbol, c_j, of $\mathcal{C}$, is said to have (r, δ)-locality if there exists a subset $S_j \subseteq [n]$ such that:

- $j \in S_j$ and $|S_j| \leqslant r + \delta - 1$; and
- the minimum Hamming distance of the punctured code $\mathcal{C}|_{S_j}$ is at least δ.

In that case, the set S_j is also called a *repair set* of c_j. The code $\mathcal{C}$ is said to have information (r, δ)-locality (denoted as $(r, \delta)_i$-locality) if there exists $S \subseteq [n]$ with $\mathrm{rank}(\{\mathbf{g}_i : i \in S\}) = k$ such that for each $i \in S$, the i-th code symbol has (r, δ)-locality. Similarly, $\mathcal{C}$ is said to have all symbol (r, δ)-locality (denoted as $(r, \delta)_a$-locality) if all the code symbols have (r, δ)-locality.

In [24] (and for the case $\delta = 2$, originally [8]), the following upper bound on the minimum Hamming distance of linear codes with information (r, δ)-locality is derived.

Lemma 1 ([8], [24])**:** The minimum distance, d, of an $[n, k, d]_q$ code with $(r, \delta)_i$-locality, is upper bounded by

$$d \leqslant n - k + 1 - \left(\left\lceil \frac{k}{r} \right\rceil - 1 \right) (\delta - 1).$$

Definition 4: A code is said to be an optimal locally repairable code (LRC) with $(r, \delta)_i$-locality (or $(r, \delta)_a$-locality) if its minimum distance d attains the bound of Lemma 1 with equality.

III. Rack-aware distributed storage system with locality

In this section, we introduce some basic model setting for a rack-aware distributed storage system with local parity checks.

We consider a system containing k original files, which are encoded into n files stored on n nodes. The $n = N \times L$ nodes are divided into N racks, and each rack contains L nodes. In each rack, the data of the nodes form a codeword with length L and a minimum Hamming distance of at least δ. In each rack, each node is capable of collecting information from nodes in the same rack and process these data. Denote this system as the $(n, k; L, \delta)$ rack-aware system with locality or $(n, k; L, \delta)$-RASL. If among those n nodes, k of them store original information and are named as information nodes then the system is said to be systematic. In this case, all the remaining nodes are parity checks, including the local parity checks and cross-rack parity checks.

As usual for rack-aware storage systems, we disregard the inner rack bandwidth, that is, we assume that each node can access the data on the nodes of the same rack. Since, in each rack, the data of nodes form a codeword with a minimum Hamming distance of at least δ, when the rack suffers $\tau \leqslant \delta - 1$ erasures, the rack can recover the erasures locally. Thus, we focus on racks that experience more than $\delta - 1$ node erasures. Specifically, we consider the following two models of erasures:

- Rack erasures: there are t racks erased in the systems;
- Partial erasures: there are t racks suffer more that $\delta - 1$ nodes erasures.

An intriguing challenge for these systems involves minimizing cross-rack bandwidth when facing erasures that exceed the capabilities of the parity checks within each rack. This challenge serves as the motivation for the subsequent part of our discussion. That is, we would like to determine the minimum amount of data we need to download from help racks to repair erasures and the explicit construction of codes with optimal repair schemes.

Remark 1: The previously mentioned rack-aware system with locality is a generalization of the rack-aware model in [12]. This extension is primarily driven by the practical observation that modern storage systems incorporate both parity checks and controllers with data processing capabilities within each rack.

Remark 2: The term "locality" is derived from the condition that, for (n, k) locally repairable codes with $(r, \delta)_a$-locality, if the repair sets $S_1, S_2, \cdots, S_N$ constitute a partition of $[n]$ and each set has a uniform size, i.e., $|S_i| = r + \delta - 1$ for $1 \leqslant i \leqslant N$, then each repair set can be arranged as a rack to construct the desired codes for a rack-aware system. Consequently, when the repair sets form a partition, the repair problems for the rack-aware system with locality are equivalent to the repair problems of locally repairable codes, assuming that we disregard the bandwidth within the rack or the repair set. Therefore, in the subsequent discussion, we will use these notations interchangeably.

IV. Tamo-Barg codes and redundant residue codes

In this section, we review the construction of locally repairable codes using redundant residue codes and prove that it can, in fact, explain Tamo-Barg codes.

Construction A ([27])**:** Let $h(x) \in \mathbb{F}_q[x]$ and denote $\deg(h(x)) = w$. For $y \in \mathbb{F}_q$, define $\mathrm{Roots}(y) \triangleq \{x \in \mathbb{F}_q \ : \ h(x) = y\}$ and $t_y \triangleq |\mathrm{Roots}(y)|$. Assume $m_1 \leqslant k$ and r_i, for all $i \in [m_1]$, are positive integers such that $\sum_{i=1}^{m_1} r_i = k$. We further assume that there exist two disjoint subsets of $\mathbb{F}_q$, $\{y_i\}_{i=1}^{m_1}$ and $\{y_i\}_{i=m_1+1}^{m_1+m_2}$, satisfying $t_{y_i} > r_i$ for all $i \in [m_1]$, and m_2 is a non-negative integer.

Denote $m \triangleq m_1 + m_2$, and $\mathrm{Roots}(y_i) = \{\beta_{i,1}, \beta_{i,2}, \dots, \beta_{i,t_{y_i}}\}$ for all $i \in [m]$. Let $\boldsymbol{a} = (a_{1,1}, a_{1,2}, \dots, a_{1,r_1}, a_{2,1}, \dots, \ a_{m_1,r_{m_1}}) \in \mathbb{F}_q^k$ be the information vector. Define $f_{\boldsymbol{a},i}$ as the polynomial with degree less than r_i such that $f_{\boldsymbol{a},i}(\beta_{i,j}) = a_{i,j}$ for all $i \in [m_1]$ and $j \in [r_i]$. Let $F_{\boldsymbol{a}}(x) \in \mathbb{F}_q[x]$ be a polynomial with degree less than $m_1 w$ satisfying

$$F_{\boldsymbol{a}}(x) \equiv f_{\boldsymbol{a},i}(x) \pmod{h(x) - y_i} \quad \text{for all } i \in [m_1]. \quad (1)$$

Then the code we construct is

$$\begin{aligned}\mathcal{C} = \{C_{\boldsymbol{a}} =& (F_{\boldsymbol{a}}(\beta_{1,1}), F_{\boldsymbol{a}}(\beta_{1,2}),\\ & \dots, F_{\boldsymbol{a}}(\beta_{1,t_{y_1}}) \dots, F_{\boldsymbol{a}}(\beta_{m,t_{y_m}})) \ : \ \boldsymbol{a} \in \mathbb{F}_q^k\}.\end{aligned}$$

According to (1), determining the locality of information symbols in $\mathcal{C}$ is straightforward, as discussed in [27, Theorem 5.3]. In the following, we present a lemma that is useful for determining the locality of global parity-check symbols.

Lemma 2: Consider the setting of Construction A, and let $0 \leqslant r \leqslant w$ be an integer. Suppose that there exist m_1 distinct constants $y_1, y_2, \dots, y_{m_1} \in \mathbb{F}_q$ such that

$$F_{\boldsymbol{a}}(x) \bmod (h(x) - y_i) = f_{\boldsymbol{a},i}(x) = \sum_{j=0}^{r-1} e_{i,j} x^j \text{ for all } i \in [m_1].$$

Then for any $y \in \mathbb{F}_q$,

$$F_{\boldsymbol{a}}(x) \bmod (h(x) - y) = \sum_{j=0}^{r-1} H_{\boldsymbol{a},j}(y) x^j,$$

where $H_{\boldsymbol{a},j}(x)$ is a polynomial satisfying $\deg(H_{\boldsymbol{a},j}(x)) \leqslant m_1 - 1$ and

$$H_{\boldsymbol{a},j}(y_i) = e_{i,j} \text{ for all } i \in [m_1] \text{ and } 0 \leqslant j \leqslant r-1.$$

Corollary 1: In the setting of Construction A, let $\Gamma = \bigcup_{i\in[m]} \mathrm{Roots}(y_i) \subseteq \mathbb{F}_q$. If $|\mathrm{Roots}(y_i)| > r$ for any $i \in [m]$, then the code constructed by Construction A is a locally repairable code with all symbol (r, δ)-locality, where $\delta = \min\{|\mathrm{Roots}(y_i)| + 1 - r \ : \ i \in [m]\}$.

In general, the code may not be optimal for the simple reason that there may not be enough roots in $\mathbb{F}_q$ for some $h(x) - y_i$, where $i \in [m]$, to serve as evaluation points. Therefore, to attain optimality, we consider the case where $\mathbb{F}_q$ contains the splitting field of $h(x) - y_i$, for all $i \in [m]$.

Definition 5: In the setting of Construction A, let $\Gamma = \bigcup_{i\in[m]} \mathrm{Roots}(y_i) \subseteq \mathbb{F}_q$. If $|\mathrm{Roots}(y_i)| = \deg(h(x)) = w$ for all $i \in [m]$, then the polynomial $h(x)$ is said to be a *good polynomial* over Γ.

For more details about good polynomials the readers may refer to [27] and [21].

Corollary 2 (Tamo-Barg codes,[27])**:** Consider the setting of Construction A and let $h(x)$ be a good polynomial over $\Gamma = \bigcup_{i\in[m]} \mathrm{Roots}(y_i) \subseteq \mathbb{F}_q$. If $r_i = r < w$ and $k = rm_1$, then the resulting code is an optimal $[n = mw, m_1 r, (m_2+1)w - r + 1]_q$ locally repairable code with all symbol $(r, w - r + 1)$-locality (where optimality is with respect to the bound in Lemma 1).

Corollary 2 follows directly from Lemma 2 and Corollary 1, which also can be directly derived from [27, Construction 1].

V. Repairing Tamo-Barg codes: rack erasures

In this section, we consider the repair problem for Tamo-Barg codes for the rack erasure case, where each repair sets of Tamo-Barg codes are arranged as racks. We begin with an array form of the Tamo-Barg code, where each repair set is arranged as a column in the array.

Construction B: Let $h(x) \in \mathbb{F}_q[x]$ be a good polynomial over $\Gamma = \bigcup_{i\in[m]} \mathrm{Roots}(y_i) \subseteq \mathbb{F}_q$ with $\deg(h(x)) = r + \delta - 1$, and let $\boldsymbol{a} = (a_{1,1}, a_{1,2}, \dots, a_{1,r}, a_{2,1}, \dots a_{m_1,r}) \in \mathbb{F}_q^k$ be the information vector, where $k = rm_1$. Define $f_{\boldsymbol{a},i}(x)$ as the polynomial with degree less than r such that $f_{\boldsymbol{a},i}(\beta_{i,j}) = a_{i,j}$ for all $i \in [m_1]$ and $j \in [r]$, where we assume that $\mathrm{Roots}(y_i) = \{\beta_{i,1}, \beta_{i,2}, \dots, \beta_{i,r+\delta-1}\}$ for all $i \in [m_1]$. Then for any $\boldsymbol{a} \in \mathbb{F}_q^k$ we can find a polynomial $F_{\boldsymbol{a}}(x) \in \mathbb{F}_q[x]$ with degree less than $m_1(r + \delta - 1)$ satisfying

$$F_{\boldsymbol{a}}(x) \equiv f_{\boldsymbol{a},i}(x) \pmod{h(x) - y_i} \qquad \text{for all } i \in [m_1].$$

Construct an array code as follows

$$\mathcal{A} \triangleq \left\{ \begin{aligned} A_{\boldsymbol{a}} &= (\boldsymbol{A}_1, \dots, \boldsymbol{A}_m) \\ &= \begin{pmatrix} F_{\boldsymbol{a}}(\beta_{1,1}) & \dots & F_{\boldsymbol{a}}(\beta_{m,1}) \\ F_{\boldsymbol{a}}(\beta_{1,2}) & \dots & F_{\boldsymbol{a}}(\beta_{m,2}) \\ \vdots & & \vdots \\ F_{\boldsymbol{a}}(\beta_{1,r+\delta-1}) & \dots & F_{\boldsymbol{a}}(\beta_{m,r+\delta-1}) \end{pmatrix} \ : \ \boldsymbol{a} \in \mathbb{F}_q^k \end{aligned} \right\},$$

where we define $m \triangleq m_1 + m_2$.

Theorem 2: Consider the setting of Construction B. Let q_0 be a prime power, and $\mathbb{F}_q = \mathbb{F}_{q_0}(y_1, y_2, \dots, y_m)$, $\mathbb{F}_{q_i} = \mathbb{F}_{q_0}(y_1, \dots, y_{i-1}, y_{i+1}, \dots, y_m)$. Define $w_i \triangleq [\mathbb{F}_q : \mathbb{F}_{q_i}]$ for each $i \in [m]$.

I. Define $w_i^* \triangleq \frac{r}{\gcd(w_i, r)}$. If $w_i^* \leqslant w_i$ and $m_2 \geqslant \max\{2, w_i^*\}$, then we can recover $\boldsymbol{A}_i$ by downloading $\frac{(w_i^*+m_1-1)r}{w_i^*}$ symbols from any other $w_i^* + m_1 - 1$ nodes (columns).

II. Assume w_i^* is a positive integer. If $w_i^* | w_i$ and $m_2 \geqslant \max\{2, w_i^*\}$, then we can recover $\boldsymbol{A}_i$ by downloading $\frac{(w_i^*+m_1-1)r}{w_i^*}$ symbols from any other $w_i^* + m_1 - 1$ nodes (columns).

To analyze the repair bandwidth performance for the scheme introduced in Theorem 2, we need to modify the cut-set bound to the case that each element in the array code can be locally repaired.

Theorem 3: Let $\mathcal{C}$ be an optimal $[N, K, N-K+1]_q$ array code with sub-packetization L. It provides $(r, L-r+1)_a$-locality when viewed as a scalar code with length NL, where $0 < r \leqslant L-1$. In this setting, each column corresponds to a repair set. Let D be an integer with $K \leqslant D \leqslant N-1$. For any $i \in [N]$ and any D-subset $\mathcal{R} \subseteq [N] \setminus \{i\}$, we have

$$B(\mathcal{C}, \{i\}, \mathcal{R}) \geqslant \frac{Dr}{D-K+1}.$$

Based on Theorems 2 and 3, we have the following corollary:

Corollary 3: Let q_0 be a prime power, and $\mathbb{F}_q = \mathbb{F}_{q_0}(y_1, y_2, \ldots, y_m)$, $\mathbb{F}_{q_i} = \mathbb{F}_{q_0}(y_1, \ldots, y_{i-1}, y_{i+1}, \ldots, y_m)$. Define $w_i \triangleq [\mathbb{F}_q : \mathbb{F}_{q_i}]$ for all $i \in [m]$. Furthermore, let $m_2 \geqslant \max\{2, r\}$, where $m = m_1 + m_2$ and $k = m_1 r$. If $0 < r \leqslant \min\{w_1, w_2, \ldots, w_m\}$, and $\gcd(r, w_i) = 1$ for all $i \in [m]$, then for any $i \in [m]$ we can recover $\boldsymbol{A}_i$ by downloading $r + m_1 - 1$ symbols from any other $r + m_1 - 1$ nodes, which is exactly the optimal bandwidth with respect to the cut-set bound of Theorem 3.

VI. Repairing Tamo-Barg codes: Partial erasures

In this section, we further explore the partial repair problem for the rack-aware system with locality, i.e., array codes under the assumption that each column of the array is an (r, δ)-repair set. Specifically, we consider the scenario where some repair sets have failed, meaning that certain columns contain more than $\delta - 1$ erasures. We seek to determine the minimum amount of data that needs to be downloaded from D remaining columns, and how to construct a code that achieves the minimum repair bandwidth for this model. To begin, we provide some necessary definitions.

Definition 6: Let $\mathcal{C}$ be an optimal $[N, K]_q$ locally repairable array code with sub-packetization L and $(r, L-r+1)_a$-locality in which each column corresponds to a repair set, where $0 < r \leqslant L-1$. Let $\mathcal{I} = \{i_1, i_2, \ldots, i_\tau\} \subseteq [N]$ denote the failed columns, and let $E_{i_t} \subseteq [L]$ with $|E_{i_t}| \geqslant \delta$ for $t \in [\tau]$ denote the corresponding erasures in the i_t-th column. For a D-subset $\mathcal{R} \subseteq [N] \setminus \mathcal{I}$, define $B(\mathcal{C}, \mathcal{I}, \mathcal{E}, \mathcal{R})$ as the *minimum repair bandwidth* for $\{c_{i,j} : i \in \mathcal{I}, j \in E_i\}$, i.e., the smallest number of symbols of $\mathbb{F}_q$ helper nodes need to send in order to recover the erasure pattern $\mathcal{E} = \{E_{i_t} : t \in [\tau]\}$ (where each helper node $j \in \mathcal{R}$ may send symbols that depend solely on $c_j \in \mathbb{F}_q^L$).

First, we consider the case where $\mathcal{E} = \{E_i\}$, and $E_i = E$.

Theorem 4: Consider the setting of Construction B. Let q_0 be a prime power, and $\mathbb{F}_q = \mathbb{F}_{q_0}(y_1, y_2, \ldots, y_m)$, $\mathbb{F}_{q_i} = \mathbb{F}_{q_0}(y_1, \ldots, y_{i-1}, y_{i+1}, \ldots, y_m)$. Define $w_i \triangleq [\mathbb{F}_q : \mathbb{F}_{q_i}]$ for each $i \in [m]$. For any given set $E \subseteq [L]$ with $\delta \leqslant |E| \leqslant L$, consider the case that the elements in $\boldsymbol{A}_i|_E$ are erasures.

I. Define $w_i^* \triangleq \frac{L-\delta+1}{\gcd(w_i, L-\delta+1)}$. If $w_i^* \leqslant w_i$ and $m_2 \geqslant \max\{2, w_i^*\}$, then we can recover $\boldsymbol{A}_i|_E$ by downloading $\frac{(w_i^*+m_1-1)(|E|-\delta+1)}{w_i^*}$ symbols from any other $w_i^*+m_1-1$ nodes (columns).

II. Assume w_i^* is a positive integer. If $w_i^* | w_i$ and $m_2 \geqslant \max\{2, w_i^*\}$, then we can recover $\boldsymbol{A}_i|_E$ by downloading $\frac{(w_i^*+m_1-1)(|E|-\delta+1)}{w_i^*}$ symbols from any other $w_i^*+m_1-1$ nodes (columns).

We now move on to address the challenge of repairing erasure patterns in scenarios where multiple repair sets (columns) have failed, specifically when $|E_{i_t}| \geqslant \delta$ for $t \in [\tau]$.

Theorem 5: Let $\mathcal{C}$ be the code generated by Construction B, and $\mathcal{E} = \{E_{i_1}, E_{i_2}, \ldots, E_{i_\tau}\}$ be an erasure pattern with $|E_{i_t}| \geqslant \delta$ for all $t \in [\tau]$. For a subfield $\mathbb{F}_{q_1} \subset \mathbb{F}_q$, if $m_2 \geqslant \frac{q\tau}{q_1}$ then the erasure pattern can be recovered with repair bandwidth $\frac{M}{q_1}$ by contacting all the remaining nodes, where $M = \sum_{t \in [\tau]} |E_{i_t}| - \tau(\delta - 1)$.

When the finite field is sufficiently large, it is possible to further decrease the partial repair bandwidth.

Lemma 3: Let $\boldsymbol{Y} = (y_1, y_2, \ldots, y_m) \in \mathbb{F}_{q_1}^m$ and $\mathcal{C}_1 = \mathrm{GRS}_{m_1}(\boldsymbol{1}, \boldsymbol{Y})$ be an $[m = m_1+m_2, m_1, m_2+1]_{q_1}$ GRS code. If $\mathcal{C}_1$ has (τ, D) optimal repair property with $1 \leqslant \tau \leqslant m_2$ and $m_1 < D \leqslant m - \tau$, then for any $\mathcal{E} = \{E_{i_1}, E_{i_2}, \ldots, E_{i_\tau}\}$ with $E_{i_j} \subseteq [L]$ and $|E_{i_j}| = w \geqslant \delta$ for $j \in [\tau]$, the code $\mathcal{C}$ generated by Construction B satisfies

$$B(\mathcal{C}, \mathcal{I}, \mathcal{E}, \mathcal{R}) \leqslant \frac{\tau D \ell (w - \delta + 1)}{D - k + \tau}$$

for any $|\mathcal{I}| = \tau$ and $|\mathcal{R}| = D$ with $\mathcal{R} \subseteq [m] \setminus \mathcal{I}$, where $\ell = r + \delta - 1$ is the sub-packetization of $\mathcal{C}_1$, i.e., $q_1 = q^l$.

A. A lower bound on the partial-repair bandwidth

A natural question arises regarding the partial-repair problem: what is the theoretical bound for the partial-repair bandwidth? Let L denote the number of elements stored in each node (column). Assume that there is a node, say the i-th node, suffering from $L - s_i$ erasures in the positions given in E_i. Define $\beta(L, s_i, D)$ as the number of elements the system needs to download from each of the D helper nodes to recover the $L - s_i$ erased elements.

Inspired by the idea of the information-flow graph presented in [7], we propose a solution to this problem by defining a special kind of information-flow graph called a *partial information-flow graph*. The basic idea is to allow each node that experiences erasures to have a certain amount of surviving information. When a node experiences partial erasure, the system needs to recover the erased portion of information for the goal node. Since the recovered node inherits the surviving information from the original node, it is named as an inheritor.

Definition 7: A directed acyclic graph is said to be a *partial information-flow graph* if it satisfies the following:

- A source node S, corresponding to the original data which will be stored into N initial storage nodes.
- Initial storage nodes $X^{(i)}$, each of them consists of an input node $X_{in}^{(i)}$ and an output node $X_{out}^{(i)}$. $X_{in}^{(i)}$ and

$X_{out}^{(i)}$ are connected by a directed edge $(X_{in}^{(i)}, X_{out}^{(i)})$ with capacity equal to the number of elements stored at $X^{(i)}$, i.e., $r \leqslant L$, where r is the number of original elements stored at $X_{in}^{(i)}$. S connects to each $X_{in}^{(i)}$ by a directed edge $(S, X_{in}^{(i)})$ with capacity r.

- To model the dynamic behavior of storage systems such as erasures and repair, the time factor is also considered. At any given time, nodes are either active or inactive. At the initial time step, the storage nodes $X^{(i)}$ are all active and the source node S is inactive. Later on, at any given time step, if a node suffers from a partial erasure, say the node $v = (v_{in}, v_{out})$, then the node v is set to be inactive and we create a direct inheritor (I_{in}, I_{out}), which is connected with v_{out} by an edge (v_{out}, I_{in}) with capacity $s(v)$, where $s(v)$ is the number of surviving information symbols of the node v. The node I also needs to download $\beta(L, s(v), D)$ symbols from each of D other active nodes, i.e., we add D directed edges $(v_{out}^{(j)}, I_{in}^{(i)})$ with a capacity of $\beta = \beta(L, s(v), D)$. Finally, we set the node I as active.
- A data collector node DC, corresponding to a request to construct the data. DC connects to K active nodes with subscript "out" by directed edges with infinite capacity to recover the original data.

Given positive integers $N > K$, $D \leqslant N-1$, $L \geqslant r$, $s_i \leqslant L$ for all $i \in [N]$ and a real number $\beta \geqslant 0$, let $G(N, K, r, D, \beta; \boldsymbol{s} = (s_1, s_2, \ldots, s_N))$ denote a family of partial information-flow graphs with all possible evolutions. The parameter tuple $(N, K, r, D, \beta; \boldsymbol{s})$ is said to be *feasible* if there exists a locally repairable array code with repair bandwidth β and sub-packetization L, with $L - s_i$ erasures in the i-th node.

Definition 8: A *cut* in the partial information flow graph G between S and DC is a subset of edges W such that each directed path from S to DC contains at least one edge in W. Furthermore, the minimal cut is the cut with the smallest edge capacity sum. We define the capacity of W as $C(W) = \sum_{e \in W} C(e)$, where $C(e)$ denotes the capacity of an edge e.

Theorem 6: For given positive integers $N > K$, $D \leqslant N-1$, $L \geqslant r$, $s_i \leqslant L$ for $i \in [N]$, the parameter tuple $(N, K, D, L, \beta; \boldsymbol{s})$ is feasible if and only if $Kr \leqslant c(N, K, D, \beta; \boldsymbol{s})$ under a large enough finite field, where $c(N, K, D, \beta; \boldsymbol{s})$ satisfies

$$\begin{aligned} c(N, K, D, \beta; \boldsymbol{s}) = & \sum_{i=0}^{\min\{K,D\}-1} \min\{(D-i)\beta + s_{j_i}, r\} \\ & + \sum_{i=\min\{K,D\}}^{K-1} \min\{s_{j_i}, r\}, \end{aligned}$$

where $\beta \geqslant 0$ is a real number, and $s_{j_0} \leqslant s_{j_1} \leqslant \ldots \leqslant s_{j_{N-1}}$.

As the next step, we study the relationship between r and N, K, β, D, M by solving the inequality $Kr \leqslant c(N, K, D, \beta; \boldsymbol{s})$ in Theorem 6.

Theorem 7: Let $\boldsymbol{s} = (s_1, s_2, \cdots, s_N)$ and $s_{j_0} \leqslant s_{j_1} \leqslant \ldots \leqslant s_{j_{N-1}} < s_{j_N} \triangleq +\infty$. For $1 \leqslant t \leqslant K-1$, define

$$s_{j_t}^* \triangleq \frac{\sum_{i=0}^{t-1} s_{j_i} + (K-t)s_{j_t}}{K},$$

$$g(t) \triangleq \sum_{i=0}^{t-1} s_{j_i} + t(D-K)\beta + \frac{t(t+1)\beta}{2},$$

and for $0 \leqslant t \leqslant K-1$,

$$f(t) \triangleq \frac{1}{K(D+1-K) + \frac{t(t+1)}{2} + t(K-t-1)}. \quad (2)$$

When $s_{j_{\tau-1}} < r \leqslant s_{j_\tau}$ for $\tau \in [N]$, the parameter tuple $(N, K, r, D, \beta; \boldsymbol{s})$ is feasible if and only if $r \geqslant r^*(N, K, D, \beta, \boldsymbol{s})$ and the solution can be achieved via linear codes, where

$$\begin{aligned} & r^*(N, K, D, \beta; \boldsymbol{s}) \\ & = \begin{cases} \frac{M}{K}, & \beta \in [f(0)K(r - s_{j_0}), \infty), \\ \frac{M-g(t)}{K-t}, & \beta \in (f(t)K(r - s_{j_t}^*), f(t-1)K(r - s_{j_{t-1}}^*)], \end{cases} \end{aligned}$$

for $t \in [\tau^* - 1]$ with $\tau^* \triangleq \min\{\tau, K\}$. If $r \leqslant s_{j_0}$ then the parameter tuple $(N, K, r, D, \beta; \boldsymbol{s})$ is feasible for any $\beta \geqslant 0$.

Remark 3: For the case $L > s_{j_0} \geqslant r$, i.e., the case that the number of erased symbols is less than $L - r$, we know that $\beta = 0$ is sufficient to repair those erasures since we have $(r, L - s + 1)$-locality inside each node.

Remark 4: For the case $s_{j_0} = s_{j_1} = \cdots = s_{j_{N-1}} = 0$ and $L = r$, i.e., the ordinary case without locality, the bounds in Theorems 6 and 7 are exactly the cut-set bound described in [7] and [13] for the rack-aware model.

Considering the regular case, $s_1 = s_2 = \cdots = s_N$, we have the following corollary that is implied directly from Theorem 7.

Corollary 4: Let $\mathcal{C}$ be an $[N, K]$ MDS array code with sub-packetization L, and each column is a repair set with $(r, \delta = L - r + 1)$-locality. Let D be an integer $K < D \leqslant N-1$. For any $i \in [N]$, $E_i \subseteq [L]$, and D-subset $\mathcal{R} \subseteq [N] \backslash \{i\}$, we have

$$B(\mathcal{C}, \{i\}, \{E_i\}, \mathcal{R}) \geqslant \begin{cases} \frac{D(|E_i| - \delta + 1)}{D-K+1}, & \text{if } |E_i| \geqslant \delta, \\ 0, & \text{otherwise.} \end{cases}$$

Thus, similarly to Corollary 3, we have the following conclusion for partial repairing for Tamo-Barg codes, which follows directly from Corollary 4 and Theorem 4.

Corollary 5: Consider the setting of Construction B. Let q_0 be a prime power, and $\mathbb{F}_q = \mathbb{F}_{q_0}(y_1, y_2, \ldots, y_m)$, $\mathbb{F}_{q_i} = \mathbb{F}_{q_0}(y_1, \ldots, y_{i-1}, y_{i+1}, \ldots, y_m)$. Define $w_i \triangleq [\mathbb{F}_q : \mathbb{F}_{q_i}]$ for each $i \in [m]$. For any given erasure set $E \subseteq [L]$ with $\delta \leqslant |E| \leqslant L$, we can recover $\boldsymbol{A}_i|_E$ by downloading $\frac{(r+m_1-1)(|E|-\delta+1)}{r}$ symbols from any other $r + m_1 - 1$ nodes, which is exactly the optimal bandwidth with respect to the cut-set bound according to Corollary 4.

References

[1] V. R. Cadambe, S. A. Jafar, H. Maleki, K. Ramchandran, and C. Suh, “Asymptotic interference alignment for optimal repair of MDS codes in distributed storage,” *IEEE Transactions on Information Theory*, vol. 59, no. 5, pp. 2974–2987, 2013.

[2] V. R. Cadambe and A. Mazumdar, “Bounds on the size of locally recoverable codes,” *IEEE Transactions on information theory*, vol. 61, no. 11, pp. 5787–5794, 2015.

[3] H. Cai, Y. Miao, M. Schwartz, and X. Tang, “On optimal locally repairable codes with super-linear length,” *IEEE Transactions on Information Theory*, vol. 66, no. 8, pp. 4853–4868, 2020.

[4] ——, “Repairing schemes for tamo-barg codes,” *arXiv:2312.12803*, 2023.

[5] H. Cai and M. Schwartz, “On optimal locally repairable codes and generalized sector-disk codes,” *IEEE Transactions on Information Theory*, vol. 67, no. 2, pp. 686–704, 2021.

[6] B. Chen, W. Fang, S.-T. Xia, J. Hao, and F.-W. Fu, “Improved bounds and Singleton-optimal constructions of locally repairable codes with minimum distance 5 and 6,” *IEEE Transactions on Information Theory*, vol. 67, no. 1, pp. 217–231, 2021.

[7] A. G. Dimakis, P. B. Godfrey, Y. Wu, M. J. Wainwright, and K. Ramchandran, “Network coding for distributed storage systems,” *IEEE Transactions on information theory*, vol. 56, no. 9, pp. 4539–4551, 2010.

[8] P. Gopalan, C. Huang, H. Simitci, and S. Yekhanin, “On the locality of codeword symbols,” *IEEE Transactions on Information Theory*, vol. 58, no. 11, pp. 6925–6934, 2012.

[9] V. Guruswami and M. Wootters, “Repairing Reed-Solomon codes,” *IEEE Transactions on Information Theory*, vol. 63, no. 9, pp. 5684–5698, 2017.

[10] J. Hao, S.-T. Xia, K. W. Shum, B. Chen, F.-W. Fu, and Y. Yang, “Bounds and constructions of locally repairable codes: parity-check matrix approach,” *IEEE Transactions on Information Theory*, vol. 66, no. 12, pp. 7465–7474, 2020.

[11] L. Holzbaur, S. Puchinger, E. Yaakobi, and A. Wachter-Zeh, “Partial MDS codes with local regeneration,” *arXiv preprint arXiv:2001.04711*, 2020.

[12] H. Hou, P. P. Lee, K. W. Shum, and Y. Hu, “Rack-aware regenerating codes for data centers,” *IEEE Transactions on Information Theory*, vol. 65, no. 8, pp. 4730–4745, 2019.

[13] Y. Hu, P. P. Lee, and X. Zhang, “Double regenerating codes for hierarchical data centers,” in *2016 IEEE International Symposium on Information Theory Proceedings*. IEEE, 2016, pp. 245–249.

[14] C. Huang, M. Chen, and J. Li, “Pyramid codes: Flexible schemes to trade space for access efficiency in reliable data storage systems,” *ACM Transactions on Storage (TOS)*, vol. 9, no. 1, pp. 1–28, 2013.

[15] G. M. Kamath, N. Prakash, V. Lalitha, and P. V. Kumar, “Codes with local regeneration and erasure correction,” *IEEE Transactions on Information Theory*, vol. 60, no. 8, pp. 4637–4660, 2014.

[16] G. Kim and J. Lee, “Locally repairable codes with unequal local erasure correction,” *IEEE Transactions on Information Theory*, vol. 64, no. 11, pp. 7137–7152, 2018.

[17] J. Li and X. Tang, “Optimal exact repair strategy for the parity nodes of the $(k+2, k)$ zigzag code,” *IEEE Transactions on Information Theory*, vol. 62, no. 9, pp. 4848–4856, 2016.

[18] J. Li, X. Tang, H. Hou, Y. S. Han, B. Bai, and G. Zhang, “PMDS array codes with small sub-packetization, small repair bandwidth/rebuilding access,” *IEEE Transactions on Information Theory*, vol. 69, no. 3, pp. 1551–1566, 2022.

[19] J. Li, X. Tang, and C. Tian, “A generic transformation to enable optimal repair in MDS codes for distributed storage systems,” *IEEE Transactions on Information Theory*, vol. 64, no. 9, pp. 6257–6267, 2018.

[20] X. Li, L. Ma, and C. Xing, “Construction of asymptotically good locally repairable codes via automorphism groups of function fields,” *IEEE Transactions on Information Theory*, vol. 65, no. 11, pp. 7087–7094, 2019.

[21] J. Liu, S. Mesnager, and L. Chen, “New constructions of optimal locally recoverable codes via good polynomials,” *IEEE Transactions on Information Theory*, vol. 64, no. 2, pp. 889–899, 2017.

[22] Y. Liu, J. Li, and X. Tang, “A generic transformation to enable optimal repair/access MDS array codes with multiple repair degrees,” *IEEE Transactions on Information Theory*, 2023.

[23] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes*. Elsevier, 1977, vol. 16.

[24] N. Prakash, G. M. Kamath, V. Lalitha, and P. V. Kumar, “Optimal linear codes with a local-error-correction property,” in *2012 IEEE International Symposium on Information Theory Proceedings*. IEEE, 2012, pp. 2776–2780.

[25] A. S. Rawat, O. O. Koyluoglu, N. Silberstein, and S. Vishwanath, “Optimal locally repairable and secure codes for distributed storage systems,” *IEEE Transactions on Information Theory*, vol. 60, no. 1, pp. 212–236, 2014.

[26] B. Sasidharan, G. K. Agarwal, and P. V. Kumar, “Codes with hierarchical locality,” in *2015 IEEE International Symposium on Information Theory Proceedings*. IEEE, 2015, pp. 1257–1261.

[27] I. Tamo and A. Barg, “A family of optimal locally recoverable codes,” *IEEE Transactions on Information Theory*, vol. 60, no. 8, pp. 4661–4676, 2014.

[28] I. Tamo, Z. Wang, and J. Bruck, “Zigzag codes: MDS array codes with optimal rebuilding,” *IEEE Transactions on Information Theory*, vol. 59, no. 3, pp. 1597–1616, 2012.

[29] I. Tamo, M. Ye, and A. Barg, “Optimal repair of Reed-Solomon codes: Achieving the cut-set bound,” in *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE, 2017, pp. 216–227.

[30] ——, “The repair problem for Reed-Solomon codes: Optimal repair of single and multiple erasures with almost optimal node size,” *IEEE Transactions on Information Theory*, vol. 65, no. 5, pp. 2673–2695, 2018.

[31] C. Xing and C. Yuan, “Construction of optimal locally recoverable codes and connection with hypergraph,” in *46th International Colloquium on Automata, Languages, and Programming (ICALP 2019)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2019.

[32] M. Ye and A. Barg, “Explicit constructions of high-rate MDS array codes with optimal repair bandwidth,” *IEEE Transactions on Information Theory*, vol. 63, no. 4, pp. 2001–2014, 2017.

[33] A. Zeh and E. Yaakobi, “Bounds and constructions of codes with multiple localities,” in *2016 IEEE International Symposium on Information Theory Proceedings*. IEEE, 2016, pp. 640–644.