Modified Generalized Integrated Interleaved Codes for Local Erasure Recovery

Xinmiao Zhang
The Ohio State University, Columbus, OH 43210, U.S.A.

***Abstract*—Locally recoverable codes allow erasures to be corrected from a portion of codeword symbols and are essential to address the high reliability and performance requirements of hyper-scale distributed storage. Generalized integrated interleaved (GII) codes allocate parities shared by the interleaves, and require lower redundancy to achieve target reliability. However, all interleaves need to be read in order to make use of the shared parities. In this paper, GII codes are modified to improve the locality. Without sacrificing the erasure correction capability, the proposed scheme accesses much fewer interleaves when only a small number of extra erasures need to be recovered using the shared parities, which happens in most cases. As a result, the average locality is greatly improved. Compared to existing local erasure recovery schemes, the modified GII codes achieve a good tradeoff on locality and correction capability.**

I. Introduction

Hyper-scale distributed storage is the backbone of cloud computing and big data analytics. To improve the reliability, erasure coding schemes capable of recovering from multiple drive failures are needed. Besides correction capability and redundancy, another critical issue to consider is the locality in terms of the number of codeword symbols to access for erasure recovery. Better locality enables improved data availability, reduced network traffic, and shorter recovery latency.

The traditional erasure codes used for failure recovery are maximum distance separable (MDS) codes. $t = n-k$ parities are added to k data symbols to recover up to t erasures, and k symbols are needed for recovery regardless of the actual erasure number. Most often, the number of actual erasures is smaller than t, and locally recoverable (LRC) erasure codes have been recently developed to use fewer symbols for recovery in these cases. Examples of the LRC codes include the Windows Azure codes [1], Pyramid codes [2], the codes utilizing multiplicative and additive subgroups of finite fields in [3], and the codes based on parity splitting in [4]. Existing schemes have either complicated en/decoders or correction capability that is highly dependent on the erasure patterns.

The generalized integrated interleaved (GII) codes [5] nest a set of linear block codewords, also called the interleaves, to create codewords of stronger linear block codes. When there are fewer erasures, they are corrected within individual interleaves. Therefore, the GII codes are naturally locally recoverable. Moreover, the shared parities generated from the nesting can be utilized to correct extra erasures, and hence the overall redundancy needed to achieve the required reliability is reduced. Nevertheless, all interleaves are involved in each nesting. If any interleave has extra erasures, all interleaves are accessed for recovery in the nested layer.

This abstract describes a modified nesting scheme for improving the locality of the GII codes without changing the correction capability and redundancy. The proposed nestings involve increasing numbers of interleaves according to the strengths of the codewords they generate. As a result, weaker nestings and fewer interleaves are needed for recovery when there are fewer extra erasures, which happens in most cases. To maintain the same correction capability as the GII codes and minimize the locality, an optimized procedure is also developed to choose the nestings used for decoding. The new scheme achieves significant locality improvement over the original GII codes, and has negligible hardware complexity overhead. It also compares favorably with existing LRC codes in terms of locality-correction capability tradeoff. This work has been presented in [6].

II. Modified GII codes

The GII scheme in [5] has two layers of codes, and a codeword consists of m interleaves. Each interleave belongs to the first-layer code. In the second layer, the m interleaves are nested to generate v codewords of stronger codes that are subcodes of the first-layer code. Assume narrow-sense Reed-Solomon (RS) codes are adopted. Denote the m interleaves of length n by $c_0, c_1, \cdots, c_{m-1}$. Let $\mathcal{C}_i(n, k_i)$ $0 \le i \le v < m$ be RS codes over $GF(q)$ such that $\mathcal{C}_v \subseteq \mathcal{C}_{v-1} \subseteq \cdots \subseteq \mathcal{C}_0$. A GII code is defined as in (1), in which α is a primitive element of $GF(q)$.

$$\mathcal{C} = \left\{ c = [c_0, \cdots, c_{m-1}] : c_i \in \mathcal{C}_0, \tilde{c}_j = \sum_{i=0}^{m-1} \alpha^{ij} c_i \in \mathcal{C}_{v-j}, 0 \le j < v \right\} \quad (1)$$

The nesting scheme in (1) is described as follows.

$$\begin{bmatrix} \tilde{c}_0 \\ \tilde{c}_1 \\ \vdots \\ \tilde{c}_{v-1} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & \alpha & \alpha^2 & \cdots & \alpha^{m-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \alpha^{(v-1)} & \alpha^{2(v-1)} & \cdots & \alpha^{(v-1)(m-1)} \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_{m-1} \end{bmatrix}$$

The two-dimentional matrix above is denoted by G, and is referred to as the nesting matrix. The erasure correction capability of $\mathcal{C}_l$ is $t_l = n - k_l$ ($0 \le l \le v$). Let r_i be the number of erasures in interleave i. r_i syndromes are needed to correct r_i erasures, and t_l valid syndromes can be computed from a (corrupted) codeword of $\mathcal{C}_l$. Denote the interleave and nested syndromes by $S_j^{(i)} = e_i(\alpha^j)$ and $\tilde{S}_j^{(i)} = \tilde{e}_i(\alpha^j)$, respectively. e_i and $\tilde{e}_i$ are the erasure vectors for c_i and $\tilde{c}_i$, respectively. If $r_i > t_0$, in which case the interleave is called exceptional, the $t_1 - t_0$ higher-order syndromes for the $b \le v$ exceptional interleaves, whose indices are

TABLE I
MAXIMUM ALLOWED m FOR DIFFERENT v AND FINITE FIELDS

v	$GF(2^4)$	$GF(2^5)$	$GF(2^6)$	$GF(2^7)$	$GF(2^8)$	$GF(2^9)$
2	16	32	64	128	256	512
3	5	6	12	22	28	62
4	5	6	8	12	15	20

$l_1, l_2, \cdots, l_b$, are derived from the nested syndromes by $\left[S_j^{(l_1)}, S_j^{(l_2)}, \cdots, S_j^{(l_b)}\right]^T = A^{-1}\left[\tilde{S}_j^{(0)}, \tilde{S}_j^{(1)}, \cdots, \tilde{S}_j^{(b-1)}\right]^T$. A consists of the $l_1, l_2, \cdots, l_b$ columns in the first b rows of G, and is called the syndrome conversion matrix. After the interleaves with up to t_1 erasures are corrected, the nested syndromes are updated and converted to interleave syndromes with even higher order. This process repeats.

The key to enable the correction of extra erasures utilizing the nested syndromes is to have an invertible syndrome conversion matrix A. The columns of G picked to be included in A correspond to the exceptional interleaves, and the rows depend on which nestings to use. When G is a Vandermonde matrix, any b columns in the first b rows form a square Vandermonde matrix, which is invertible. However, each nesting in (1) utilizes all interleaves, and every interleave is needed to correct any exceptional interleave.

To improve the locality, the following modified GII code was proposed in [6].

$$\mathcal{C}' = \left\{ c = [c_0, \cdots, c_{m-1}] : c_i \in \mathcal{C}_0, \tilde{c}_j = \sum_{i=j}^{m-1} \alpha^{ij} c_i \in \mathcal{C}_{v-j}, 0 \le j < v \right\}$$

It effectively replaces the lower triangular part of G by zeros. Denote this modified nesting matrix by G'. The nestings corresponding to the lower rows of G' involve fewer interleaves and are assigned relatively weaker codes. There are most likely only a few exceptional interleaves and the number of extra erasures to handle is also small. In this case, the lower nestings can be utilized to reduce the number of interleaves involved.

To maintain the same correction capability as in the original GII codes, the syndrome conversion matrix for any erasure pattern needs to be invertible. On the other hand, the nestings with the lowest yet sufficient correction capabilities should be used in order to reduce the number of interleaves involved. Analysis was carried out in [6] to choose the nestings that satisfy the above requirements and at the same time simplify the selection procedure. Assume there are $b \le v$ exceptional interleaves, and their numbers of erasures are sorted as $r_{l_1} \le r_{l_2} \le \cdots \le r_{l_b}$. If h is the smallest integer such that $t_h \ge r_{l_b}, t_{h-1} \ge r_{l_{b-1}}, \cdots t_{h+b-1} \ge r_{l_1}$, then the nestings $\tilde{c}_d, \tilde{c}_{d+1}, \cdots, \tilde{c}_{d+b-1}$ should be used for decoding. Here $d = \min\{l_1, l_2, \cdots, l_b, v-h\}$. This nesting selection brings negligible overhead to the overall decoding process. The corresponding syndrome conversion matrices are always non-singular, if the maximum number of columns in G' and hence m does not exceed the values listed in Table I. The allowed m is sufficient for most practical applications.

III. FAILURE RATE AND LOCALITY COMPARISONS

Fig. 1 shows the frame failure rates (FFRs) of various codes. The correction capabilities of the modified GII codes are identical to those of the original GII codes, and are not plotted separately. The $(N, K) = (48, 41)$ GII code has $n = 12$, $[m, v] = [4, 2]$, $t_0 = 1, t_1 = 2$, and $t_2 = 3$. The (47,40) parity-split code [4] has 4 subsets of 10 data symbols and pads one parity symbol to each (group of) subsets in a hierarchical manner. For comparison, $N = 48$ symbols are divided into 4 copies of $(12, 10)$ MDS codes. From Fig. 1, the GII codes have much better performance than the other codes with similar or lower rates. Similar results are also observable for the (48,41) GII code with $n = 6$ and its counterparts. Fig. 2 shows comparisons on average locality, which is defined as the number of symbols required for recovery averaged over all correctable erasure patterns with the same number of erasures. From Fig. 2, the modified GII codes lead to significant locality improvement over the original GII codes. Comparing the curves in Fig. 1 and 2, they also achieve a good tradeoff on the locality and correction capability.

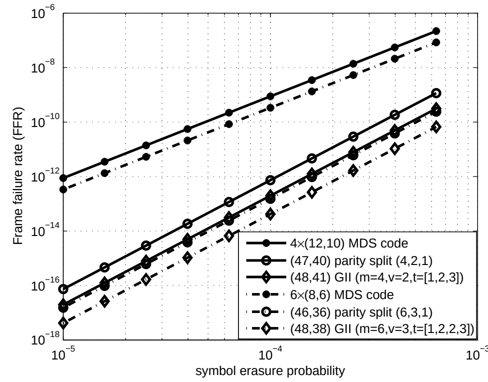

Fig. 1. Frame failure rates of erasure recovery schemes

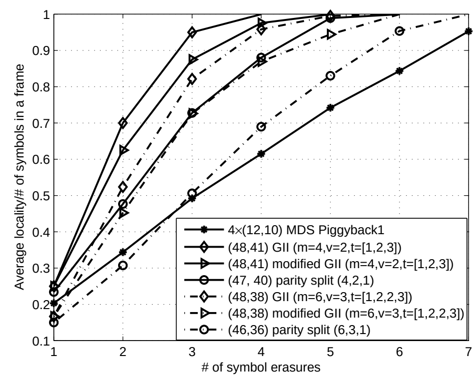

Fig. 2. Average locality for correctable erasure patterns

REFERENCES

[1] C. Huang, *et. al.*, "Erasure coding in Windows Azure Storage," *Proc. USENIX Annual Technical Conf.*, pp. 15-26. 2012.

[2] C. Huang, M. Chen, and J. Li, "Pyramid codes: flexible schemes to trade space for access efficiency in reliable data storage systems," *ACM Trans. on Storage*, vol. 9, no. 1, Mar. 2013.

[3] I. Tamo and A. Barg, "A family of optimal locally recoverable codes," *IEEE Trans. on Info. Theory*, vol. 60, no. 8, pp. 4661-4676, Aug. 2014.

[4] X. Zhang, S. Sprouse, and I. Ilani, "A flexible and low-complexity local erasure recovery scheme," *IEEE Commun. Letters*, vol. 20, no. 11, pp. 2129-2132, Nov. 2016.

[5] X. Tang and R. Koetter, "A novel method for combining algebraic decoding and iterative processing," *Proc. IEEE Int. Symp. Inf. Theory*, pp. 474-478, Seattle, WA, USA, Jul. 2006.

[6] X. Zhang, "Modified generalized integrated interleaved codes for local erasure recovery," *IEEE Commun. Letters*, vol. 21, no. 6, pp. 1241-1244, Jun. 2017.