

DOI:10.11918/202306010

可用性不等的全符号局部修复码构造

王 静, 徐忠环, 刘帅帅, 刘 哲

(长安大学 信息工程学院, 西安 710018)

摘要: 现有的可用性不等的全符号局部修复码存在参数取值受限与码率较低的问题。为了解决该问题, 基于饱和正交表构造一类可用性不等的全符号局部修复码, 实现了较高码率。具体而言, 根据饱和正交表生成关联矩阵, 对关联矩阵进行矩阵变换以及克罗内克积操作, 构造了信息符号具有高可用性的全符号局部修复码, 以及信息位可用性不等的全符号局部修复码。理论分析表明, 信息符号具有高可用性的全符号局部修复码参数取值灵活, 且在局部性 $r=2$ 时维度和码长都达到最优; 与现有的全符号局部修复码相比, 信息位可用性不等的全符号局部修复码在码率性能上更具优势。

关键词: 全符号局部修复码; 饱和正交表; 可用性不等; 码率

中图分类号: TN911.2

文献标志码: A

文章编号: 0367-6234(2026)03-0129-07

Construction of all-symbol locally repairable codes with unequal availability

WANG Jing, XU Zhonghuan, LIU Shuaishuai, LIU Zhe

(School of Information Engineering, Chang'an University, Xi'an 710018, China)

Abstract: The existing all-symbol locally repairable codes with unequal availability have limited parameter values, and low code rates. In order to solve the problems above, this paper constructs a class of all-symbol locally repairable codes with unequal availability based on saturated orthogonal arrays, which achieves higher code rates. Specifically, the association matrix is generated based on the saturated orthogonal array, and the matrix transformations as well as the Kronecker product operations are performed on the association matrix to generate all-symbol locally repairable codes with high availability of information symbols and unequal availability of information bits. Theoretical analyses show that all-symbol locally repairable codes with high availability of information symbols have flexible parameter values and are optimal in dimension and code length at locality $r=2$. Compared with the existing all-symbol locally repairable codes, the constructed all-symbol locally repairable codes with unequal availability of information bits perform better in terms of code rate.

Keywords: all-symbol locally repairable codes; saturated orthogonal array; unequal availability; code rate

随着社会数字化进程的不断推进, 全球数据量呈指数级增长, 如何高效可靠地存储和管理海量数据成为亟需解决的关键问题。分布式存储系统以其海量存储能力、高吞吐量、高可用性、高可扩展性等优势成为海量数据存储的主流方式^[1]。然而, 随着系统规模扩大, 分布式存储节点不可避免地会发生故障, 因此, 需要进一步提升系统的可靠性。为了解决这一问题, Dimakis 等^[2]将网络编码思想引入分布式存储, 提出再生码 (regenerating codes) 的概念, 有效减少了修复带宽开销, 然而再生码在节点修复时, 需要连接大量存活节点, 从而带来较高的磁盘 I/O 开销。为了降低磁盘 I/O 开销, 文献^[3]提出了局部修复码 (locally repairable codes, LRCs), 通过减少修复需要连接的节点数量, 降低修复成本, 但该类码

通常只能修复单故障节点。

为了实现多故障节点的局部修复, 文献^[4]提出了 (r, t) 局部性的概念; 故障节点可由 t 个不相交的局部修复组分别修复, 每个修复组的大小至多为 r 。若局部修复码的每个符号都具有 (r, t) 可用性, 则称为全符号局部修复码^[5-6]。进一步, 文献^[7]提出了可用性不等的全符号局部修复码, 考虑了数据访问频率的差异 (即数据的“冷热”程度), 更适用于实际的分布式存储系统。然而, 现有的可用性不等的全符号局部修复码存在参数取值受限的问题。文献^[8]和文献^[9]分别基于 Simplex 码和反码构造了可用性不等的全符号局部修复码, 该码虽然容错能力较高, 但局部性 r 仅限于 2。文献^[7]构造的可用性不等的全符号局部修复码, 使信息符号具有高

收稿日期: 2023-06-02; 录用日期: 2023-07-11; 网络首发日期: 2023-12-27

网络首发地址: <https://link.cnki.net/urlid/23.1235.T.20231226.1508.002>

基金项目: 国家自然科学基金项目 (62001059); 陕西省重点研发计划项目 (2021GY-019)

作者简介: 王 静 (1982—), 女, 教授, 硕士生导师; 徐忠环 (1999—), 女, 硕士研究生

通信作者: 徐忠环, 2891903601@qq.com

可用性,校验符号具有低可用性。与文献[8]和文献[9]相比,文献[7]构造的局部修复码在修复故障节点时选择更具多样性,但局部性 r 仍仅限取 2。

此外,现有的局部修复码的码率不高。文献[7]、[8]和[9]所构造的可用性不等的全符号局部修复码在特定参数下虽能达到维度界,但码率都较低;文献[6]基于组合设计构造了全符号局部修复码,该码构造灵活且达到了最小码长界,但码率同样不高;文献[10]和[11]基于迭代矩阵构造了具有任意局部性和可用性的全符号局部修复码;文献[12]基于超图构造了全符号局部修复码,但这些方案仍存在码率较低的问题。

针对上述问题,本文提出一类可用性不等的全符号局部修复码的构造方法。具体地,通过饱和正交表生成关联矩阵,将关联矩阵进行矩阵变换以及克罗内克积操作,构造出信息符号具有高可用性的全符号局部修复码,以及信息位可用性不等的全符号局部修复码。前者参数取值灵活,且在 $r=2$ 时达到维度和码长最优,修复故障节点时选择更多样;后者则在前者基础上进一步提升了码率。

1 饱和正交表和局部修复码

1.1 饱和正交表

定义 1 (水平正交表)^[13] 设 A 是一个 $n \times q$ 矩阵,矩阵的每一列元素由 $1, 2, 3, \dots, s$ 构成。如果矩阵 A 的任意两列构成的所有有序元素对均完整出现,且每对出现次数相同,则称矩阵 A 为 s 水平正交表,记为 $L_n(s^q)$, n 和 q 分别表示该表的行数和列数, s 表示每一列所包含的元素数目。

定义 2 (饱和正交表)^[14] 一个 $L_n(s^q)$ 型正交表,当 $q = (n-1)/(s-1)$ 时,该正交表为饱和正交表,即该正交表的列数已达到最大值。

在 $L_n(s^q)$ 型饱和正交表中,若每个水平对只出现一次,则 $n = s^2, q = \frac{s^2-1}{s-1} = s+1$, 可记为 $L_{s^2}(s^{s+1})$ 。

1.2 局部修复码

定义 3 (全符号局部修复码) C 是一个 $[n, k]_q$ 线性码, h_1, \dots, h_m 是码 C 校验矩阵的行。如果 $\text{supp}(h_j) \cap \text{supp}(h_l) = \{i\}, 1 \leq j < l \leq m, i \in [n]$, 则在第 i 个位置上 $h_1, \dots, h_m \in C^\perp$ 是正交的。在第 i 个位置,当且仅当存在 t 个校验矩阵的行 $h_1, \dots, h_t \in C^\perp$, 满足 $\text{wt}(h_l) \leq r+1, 1 \leq l \leq t$, 那么线性码 C 的第 i 个位置具有局部性 r 和可用性 t 。若一个线性码的所有位置都具有局部性 r 和可用性 t , 该线性码被称为全符号局部修复码, 记作 $(r, t)_a$ -LRCs。

定义 4 (可用性不等的全符号局部修复码)^[15]

码 C 是一个 (n, k, r, t) 全符号局部修复码, 长度为 n 的码 C 可用性分布可表示为 $t = (t_1, t_2, \dots, t_n)$, 其中 $t_i \geq 1$ 表示局部修复码第 i 个符号的可用性。如果 t 中的元素不相等, 则该局部修复码的可用性是不均衡的, 称该码字为可用性不等的全符号局部修复码。

2 可用性不等的全符号局部修复码的构造

现有的可用性不等的全符号局部修复码参数取值受限且码率较低, 为此本节主要构造可用性不等的全符号局部修复码。

2.1 信息符号具有高可用性的全符号局部修复码的构造

基于饱和正交表构造关联矩阵 N , 对关联矩阵 N 进行矩阵变换以及克罗内克积操作, 构造信息符号具有高可用性的全符号局部修复码, 码率始终为 $1/2$ 。

构造 1 信息符号具有高可用性的全符号局部修复码

首先, 构造关联矩阵 N , 具体步骤如下:

步骤 1: 令 $L_{s^2}(s^{s+1})$ 型饱和正交表每列中的 s 个元素对应关联矩阵 N 的 s 列, 则关联矩阵 N 共有 $s(s+1)$ 列。

步骤 2: 构造 $s^2 \times s(s+1)$ 阶关联矩阵 N 。若 $L_{s^2}(s^{s+1})$ 型饱和正交表中某一列的元素 $b (0 \leq b \leq s-1)$ 属于饱和正交表的 $a_i (1 \leq a_i \leq s^2, 1 \leq i \leq s)$ 行, 则元素 b 对应的关联矩阵 N 所在列的 a_i 行元素为 1, 所在列的其他行元素为 0。

其次, 基于关联矩阵 N 构造信息符号具有高可用性的全符号局部修复码, 主要分为以下两种情况:

情况 1: $s=2$ 时, 令矩阵 $B=N$, 将矩阵 B 的第 1 行删除得到矩阵 B' ; 再将矩阵 B' 的列向量按列重从大到小排列成校验矩阵 H_1 ; 由校验矩阵 H_1 生成的 BLRCs ($n=6, k=3, r=2, t_1 = \{2, 2, 2, 1, 1, 1\}$) 是信息符号具有高可用性的全符号局部修复码。

情况 2: $s \geq 3$ 且 s 是奇数时, 将关联矩阵 N 进行分块处理, 将 N 的行划分为 s 个组 (每组包含 s 行), 列划分为 $(s+1)$ 个组 (每组包含 s 列), 通过划分, 可得到 $s(s+1)$ 个 $s \times s$ 的子矩阵, 记为 $C_{i,j}$, 其中索引满足 $1 \leq i \leq s, 1 \leq j \leq s+1$ 。以这些子矩阵 $C_{i,j}$ 为分块元素, 即可构造出分块矩阵 M_1 , 具体形式为

$$M_1 = \begin{pmatrix} C_{1,1} & C_{1,2} & \cdots & C_{1,s+1} \\ C_{2,1} & C_{2,2} & \cdots & C_{2,s+1} \\ \vdots & \vdots & & \vdots \\ C_{s,1} & C_{s,2} & \cdots & C_{s,s+1} \end{pmatrix}$$

令行向量 $L_1 = (\underbrace{1, 1, 0, \dots, 0}_{s-2})$, $L_i (2 \leq i \leq s)$ 为 L_1 循环左移 $i-1$ 位得到的结果。接着构建分块矩阵:

$$M_2 = \begin{pmatrix} C_{1,1} & \cdots & C_{1,s} & C_{1,s+1} \otimes L_1 \\ C_{2,1} & \cdots & C_{2,s} & C_{2,s+1} \otimes L_2 \\ \vdots & & \vdots & \vdots \\ C_{s,1} & \cdots & C_{s,s} & C_{s,s+1} \otimes L_s \end{pmatrix}$$

其中 \otimes 为克罗内克积。令校验矩阵 $H_1 = M_2$, 此时由校验矩阵 H_1 构造得到线性码 C_1 。码 C_1 是信息符号具有高可用性、校验符号具有低可用性的全符号局部修复码,其参数满足 $n = 2s^2, k = s^2, r = s + 1, t_1 = \{\underbrace{s, \dots, s}_k, \underbrace{2, \dots, 2}_{n-k}\}$, 即信息符号的可用性为 s , 校验符号的可用性为 2。

$s \geq 3$ 且 s 为奇数时, 码长 n 可由校验矩阵 H_1 的列数直接确定。已知校验矩阵 H_1 的秩 $\text{rank}(H_1) = s^2$, 故码 C_1 的维数 $k = n - \text{rank}(H_1) = s^2$ 。校验矩阵 H_1 行重为 $s+2$, 故 C_1 局部性 $r = s+1$ 。

接下来证明 C_1 的可用性。由克罗内克积可知,

$$\begin{pmatrix} C_{1,s+1} \otimes L_1 \\ \vdots \\ C_{s,s+1} \otimes L_s \end{pmatrix}_{s^2 \times s^2}$$
 是一个行重和列重都为 2 的 $s^2 \times s^2$ 矩阵。

校验矩阵 H_1 前 s^2 列的列重为 s , s 为信息符号可用性; 校验矩阵 H_1 后 s^2 列的列重为 2, 2 为校验符号可用性。校验矩阵 H_1 中任意两列线性无关(说明 C_1 的最小距离 $d \geq 3$), 任意 s 列线性相关(说明 C_1 的最小距离 $d \leq s$), 故最小距离 d 满足 $3 \leq d \leq s$ 。

例 1 $s = 3$ 时, 由 $L_9(3^4)$ 型饱和正交表得到 9×12 阶关联矩阵 N , 表达式为

$$N = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \end{pmatrix}$$

将关联矩阵 N 按 3 行 3 列划分, 得到分块矩阵 M_1 , 其中, 第四列对应的子矩阵为:

$$C_{1,4} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, C_{2,4} = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, C_{3,4} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}.$$

$s = 3$ 时, 行向量 $L_1 = (1, 1, 0), L_2 = (1, 0, 1), L_3 = (0, 1, 1)$ 。对 M_1 的子矩阵与上述行向量做克罗内克积, 可得矩阵 M_2 :

$$M_2 = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

令校验矩阵 $H_1 = M_2$, 由该校验矩阵可得到全符号局部修复码, 参数为: $n = 18, k = 9, r = 4, t = \{\underbrace{3, \dots, 3}_{9个}, \underbrace{2, \dots, 2}_{9个}\}$ 。其中, 信息符号可用性为 3, 校验符号可用性为 2。最小距离 $d = 3$, 码率 $R = \frac{1}{2}$ 。

2.2 信息位可用性不等的全符号局部修复码的构造

与现有的全符号局部修复码相比, 构造 1 只考虑校验符号为低可用性, 而在实际分布式存储系统中信息符号也存在冷热数据差异。构造 2 利用克罗内克积构造了列重不同的子矩阵, 使热度高的信息符号具有高可用性, 热度一般的信息符号具有次高可用性, 热度低的信息符号以及校验符号的可用性为 2。

构造 2 信息位可用性不等的全符号局部修复码

步骤 1: 将构造 1 中矩阵 M_1 的子矩阵按列划分为 $s+1 (s \geq 3)$ 组, 可表示为 $M_3 = (C_1, C_2, \dots, C_s, C_{s+1})$, 其中 $C_l = (C_{1,l}, C_{2,l}, \dots, C_{s,l})^T$, 其中 $1 \leq l \leq s+1$ 。

步骤 2: 令 $L' (t' \geq 2)$ 表示调节因子, $L'_i = (\underbrace{1, 1, 0, 1, 0, 0, 1, 0, 0, 0, 1, \dots, 0}_{l'+l'(l'-1)/2}, \underbrace{0, \dots, 0}_{s-l'+l'(l'-1)/2})$, $L'_i (2 \leq i \leq s)$ 为 L'_1 循环左移 $i-1$ 位。 L' 由 L'_1 以及 $L'_i (2 \leq i \leq s)$ 共 s 个向量构成, 可表示为 $L' = \{L'_1, L'_2, \dots, L'_s\}$ 。

步骤 3: 当满足 $s \geq t'(1+t')/2$ 时, 可以从 L^2, L^3, \dots, L^s 中选择不同的调节因子调节子矩阵 C_l 的列重。若选择调节因子 L' , 矩阵列重为 s , 将矩阵 C_s 与调节因子 L' 进行克罗内克积, 具体为, $D'_s = (C_{1,s} \otimes L'_1, C_{2,s} \otimes L'_2, \dots, C_{s,s} \otimes L'_s)^T$, D'_s 的列重为 t' 。因此, 调节因子 L' 使矩阵 C_s 的列重从 s 变为 t' 。

步骤 4: 令 $1 < m \leq s$, 将矩阵 M_3 中后 m 组子矩阵 $C_{s-m+2} \sim C_{s+1}$ 与调节因子做克罗内克积, 从而调节后 m 组子矩阵 $C_{s-m+2} \sim C_{s+1}$ 的列重。为了使校验符号可用性为 2, 将含校验符号位置所在的子矩阵 C_{s+1} 与调节因子 L^2 进行克罗内克积得矩阵 D^2_{s+1} 。为了使热度较低的信息符号具有较低的可用性, 将 $s-j+1$ 组子矩阵 $C_j \sim C_s (s-m+2 \leq j \leq s)$ 与调节因子 L^2 进行克罗内克积得矩阵 $D^2_j \sim D^2_s$ 。为了使热

$4 \leq k \leq 6$ 时维度达到了最优;文献[7]构造的可用性不等的局部修复码 P-BLRCs 在 $k=4$ 时达到了维度边界;本文的构造 1 在 $k=3$ 时达到了维度最优,码长达到了 Griesmer 界。与现有的可用性不等的局部修复码相比,构造 1 最小距离较小,容错能力较弱。

但在维度 k 相同的情况下,构造 1 码率比 P-BLRCs、A-BLRCs、S-BLRCs 码率高,且构造 1 中参数取值灵活,局部性 r 不仅限于 2,且有 2 个或 2 个以上可用符号的数量比其他的多,在修复故障节点时选择更丰富。

表 1 可用性不等的局部修复码参数对比

Tab. 1 Comparison of locally repairable code parameters with unequal availability

构造方法	最优取值	r	d	R	$t \geq 2$ 符号数量
S-BLRCs	$\forall k$	2	$2^k - 1$	$\frac{k}{2^k - 1}$	$n - 2^{k-1}$
A-BLRCs	$4 \leq k \leq 6$	2	$2^{k-1} - \lfloor \frac{(k-1)^2}{4} \rfloor$	$\frac{2k}{2^{k+1} - k^2 + 3k - 2}$	$n - 2k + 1$
P-BLRCs	$k=4$	2	$2(k-1)$	$\frac{1}{k-1}$	$n - k$
构造 1	$k=3$	$\{s+1, 2\}$	$3 \leq d \leq s$	$\frac{1}{2}$	$\geq n - k$

图 1 为 4 种可用性不等的局部修复码的码率 R 随维度 k 的变化曲线,可以看出,构造 1 的码率远高于 S-BLRCs、A-BLRCs 和 P-BLRCs。

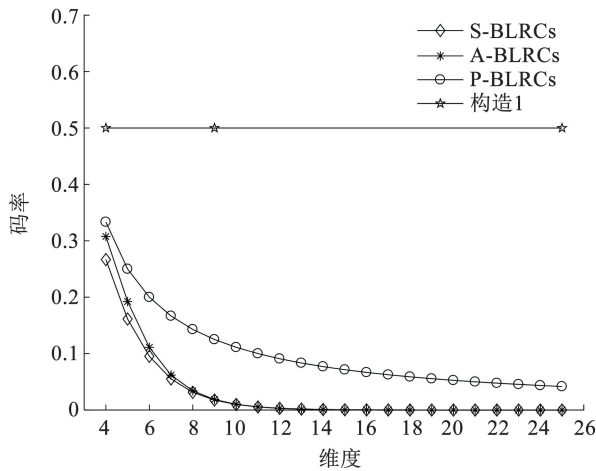


图 1 码率随维度的变化曲线

Fig. 1 Curve of code rate with dimension

图 1 为 4 种可用性不等的全符号局部修复码的码率对比。虽然文献[10-12]中的构造都为全符号局部修复码,但其构造的局部修复码的每个符号可用性相等。本文侧重构造可用性不等的全符号局部修复码,故此处没有将构造 1 的码率与文献[10-12]中所构造的局部修复码的码率进行对比。

表 2 为 s 取 3、4 和 7 时,构造 2 可用性不等的全符号局部修复码的码率。 $s = \{3, 4\}$, 调节因子只能为 L^2 , 因此构造的可用性不等的局部修复码的可

用性只能为 2 或 s , 调节因子 L^2 使冷数据的可用性为 2。将 s 和 m 取值代入码长和维度公式,可以得到 $s = \{3, 4\}$ 的码率,如表 2 所示。 $s = 7$ 时,调节因子可以从 L^2 和 L^3 选择,选择调节因子 L^2 生成热度较冷的数据对应的可用性,选择调节因子 L^3 生成热度中等信息符号对应的可用性。 $s = 7$ 时,令 $j = 7$, 可以算出表 2 中的码率。

表 2 s 取值不同时构造 2 码率的变化

Tab. 2 Variation in the code rate of the construction 2 with different values of s

s	$m=2$	$m=3$	$m=4$	$m=5$	$m=6$	$m=7$
$s=3$	15/24	21/30	—	—	—	—
$s=4$	29/44	41/56	53/6	—	—	—
$s=7$	91/140	133/182	175/224	217/266	259/308	301/350

从表 2 可以看出, $s=3, 4, 7$ 时,构造 2 的码率大于 $1/2$, 即高于现有的(文献[7-9])可用性不等的全符号局部修复码的码率;当 s 为其他取值时,构造 2 的码率也是高于文献[7-9]的码率。

现有的全符号局部修复码是具有并行读取特性且码率较好的局部修复码,因此将构造 2 的码率与文献[11]中的全符号局部修复码的码率进行对比。文献[11]中每个符号的可用性相同。图 2 是 s 取 3、4 和 7 时,构造 2 的码率以及文献[11]中基于迭代矩阵的全符号局部修复码的码率随 m 的变化。由图 2 可看出,在 $s = \{3, 4, 7\}$ 时,构造 2 的码率高于文献[11]的码率。

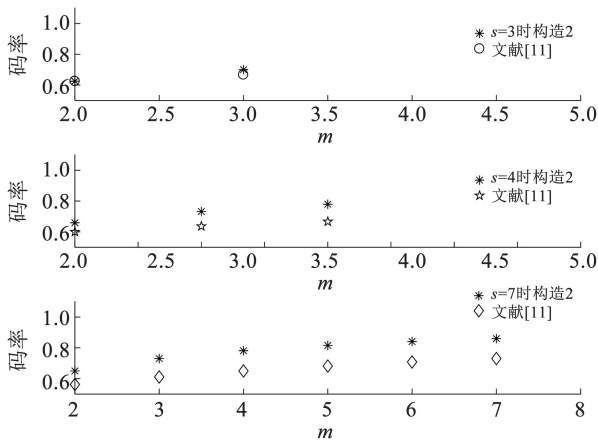


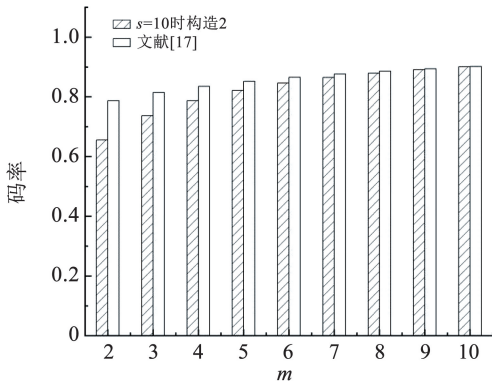
图 2 s 取不同值时的码率对比

Fig. 2 Comparison of code rates for different values of s

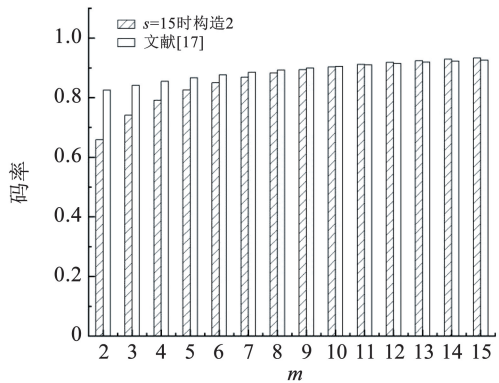
与现有的全符号局部修复码相比,构造 2 的可用性不等的全符号局部修复码的码率较高。文献[10]提出的全符号局部修复码的码率为 $R = r/(r + t)$,文献[11]提出的基于迭代矩阵的全符号局部修复码的码率也为 $R = r/(r + t)$,文献[12]基于超图的

全符号局部修复码的码率同样为 $R = r/(r + t)$ 。可见,目前构造的全符号局部修复码的码率大多为 $R = r/(r + t)$ 。而构造 2 可用性不等的全符号局部修复码,在每个符号具有局部性可用性($t_i \geq 2$)的同时,码率超过了 $R = r/(r + t)$,具有更高码率。

通过上述分析可知,构造 2 的码率高于现有的大多数全符号局部修复码的码率。文献[17]提出了全符号局部修复码的码率上界,即该码的最高码率,但现有的全符号局部修复码大多无法达到最高码率。将本文构造 2 的码率与全符号局部修复码的最高码率进行比较。令 $s = j = \{10, 15\}$,调节因子为 L^2, L^3 ,图 3 为构造 2 的码率以及文献[17]中全符号局部修复码的最高码率随 m 的变化。文献[17]是可用性相等的全符号局部修复码的码率上界,并不是可用性不等的全符号局部修复码的码率上界,且可用性不等的全符号局部修复码的码率上界目前还未被证明。



(a) s=10时, 码率随m的变化



(b) s=15时, 码率随m的变化

图 3 s 取不同值时的码率对比

Fig. 3 Comparison of code rates for different values of s

从图 3 可以看出,当 $s = j = 10$ 且调节因子为 L^2, L^3 时,随着 m 的增大,构造 2 的码率越来越接近可用性相等的全符号局部修复码的最高码率,且始终小于该最高码率。当 $s = 15, j = 15$ 且调节因子为 L^2, L^3 时,构造 2 的码率在 $m \geq 12$ 时超过了可用性相等的全符号局部修复码的最高码率。由此可以得出结论,只要选择适当的参数,就可以使构造 2 的码率高于可用性相等的全符号局部修复码的最高码率。

4 结 语

现有的可用性不等的全符号局部修复码参数

取值受限且码率较低,为此,本文基于饱和和正交表构造信息符号具有高可用性的全符号局部修复码。理论分析表明,该码的码率较高,且在 $r = 2$ 时达到维度和码长最优。进一步,考虑到实际的分布式存储系统中信息符号也存在冷热数据,构造了信息位可用性不等的全符号局部修复码,使得热度较高的信息符号具有高的可用性,热度中等的信息符号具有次高可用性,热度较低的数据的可用性为 2。与现有的全符号局部修复码相比,该码的信息符号在具有不等可用性和并行读取特性的同时,具有更高码率,更适用于实际的分布式存储系统。

参考文献

- [1] WANG Y, LI X, BAI P X, et al. An efficient big data storage service architecture [C]//2022 IEEE 25th International Conference on Computer Supported Cooperative Work in Design (CSCWD). Hangzhou: IEEE, 2022: 740
- [2] DIMAKIS A G, GODFREY P B, WU Y, et al. Network coding for distributed storage systems [J]. IEEE Transactions on Information Theory, 2010, 56(9): 4539. DOI:10.1109/TIT.2010.2054295
- [3] PAPALIOPOULOS D S, DIMAKIS A G. Locally repairable codes [C]//2012 IEEE International Symposium on Information Theory Proceedings. Cambridge, MA: IEEE, 2012: 2271
- [4] WANG A Y, ZHANG Z F. Repair locality with multiple erasure tolerance [J]. IEEE Transactions on Information Theory, 2014, 60(11): 6979. DOI:10.1109/TIT.2014.2351404
- [5] MUKHOPADHYAY D, HANSDA K, BAGCHI S. Some properties and constructions of weakly self dual LRCs [C]//2022 IEEE International Conference on Signal Processing and Communications (SPCOM). Bangalore: IEEE, 2022: 1
- [6] ZHANG Y, KAN H B. Locally repairable codes from combinatorial designs [J]. Science China Information Sciences, 2020, 63: 122304. DOI:10.1007/s11432-019-2649-5
- [7] LEE K S, PARK H, NO J S. New binary locally repairable codes with locality 2 and uneven availabilities for hot data [J]. Entropy, 2018, 20(9): 636
- [8] CADAMBE V R, MAZUMDAR A. Bounds on the size of locally recoverable codes [J]. IEEE Transactions on Information Theory, 2015, 61(11): 5787. DOI:10.1109/TIT.2015.2477406
- [9] SILBERSTEIN N, ZEH A. Optimal binary locally repairable codes via anticodes [C]//2015 IEEE International Symposium on Information Theory (ISIT). Hong Kong: IEEE, 2015: 1247
- [10] ZHANG M, LI R H. Two families of LRCs with availability based on iterative matrix [C]//2020 13th International Symposium on Computational Intelligence and Design (ISCID). Hangzhou: IEEE, 2020: 334
- [11] WANG A Y, ZHANG Z F, LIU M L. Achieving arbitrary locality and availability in binary codes [C]//2015 IEEE International Symposium on Information Theory (ISIT). Hong Kong: IEEE, 2015: 1866
- [12] KIM J H, SONG H Y. Hypergraph-based binary locally repairable codes with availability [J]. IEEE Communications Letters, 2017, 21(11): 2332. DOI:10.1109/LCOMM.2017.2730183
- [13] 杨子胥. 正交表的构造 [M]. 济南: 山东人民出版社, 1978
- YANG Zixu. The construction of orthogonal tables [M]. Jinan: Shandong People's Publishing House, 1978
- [14] 张晓. 强度 2 饱和混合正交表的汉明距离 [D]. 河南: 河南师范大学, 2019
- ZHANG Xiao. Hamming distance for intensity 2 saturated mixed orthogonal tables [D]. Henan: Henan Normal University, 2019
- [15] ZHONG L, HAN G J, HOU H X, et al. Optimizing repair-cost of locally repairable codes for hot data in cluster storage systems [C]//2022 IEEE International Conference on Big Data (Big Data). Osaka: IEEE, 2022: 3236
- [16] GRIESMER J H. A bound for error-correcting codes [J]. IBM Journal of Research and Development, 1960, 4(5): 532
- [17] TAMO I, BARG A. Bounds on locally recoverable codes with multiple recovering sets [C]//2014 IEEE International Symposium on Information Theory. Honolulu, HI: IEEE, 2014: 691

(编辑 丁晓清)