

最优局部修复码的构造

杨佳蓉¹, 王 娥¹, 李静辉¹, 余春雷²

(1. 长安大学 信息工程学院, 西安 710018; 2. 四川文理学院 智能制造学院, 四川 达州 635002)

摘要: 局部修复码 (LRCs) 作为纠删码的一种, 被广泛应用于分布式存储系统中; 针对目前局部修复码在满足最小距离最优界时码率不高且局部性的参数限制大的问题, 提出一种基于方形网络的最优局部修复码构造方法, 利用方形网络构造局部修复码的校验矩阵, 从校验矩阵入手构造局部修复码, 此码达到了最优码率界, 但是其局部性有所限制; 进一步将方形网络水平方向和垂直方向上的关联矩阵进行扩展, 用方形网络的扩展矩阵构造局部修复码的校验矩阵, 所构造的局部修复码在局部性上的性能有所提升; 和现有局部修复码进行对比分析, 构造的局部修复码不仅满足最小距离最优界, 同时达到了局部修复码的码率最优界, 可适用于任意局部性的情况, 对二元最优局部修复码的构造具有借鉴意义。

关键词: 分布式存储系统; 局部修复码; 方形网络; 最小距离; 码率

Construction of Optimal Locally Repairable Codes

YANG Jiarong¹, WANG E¹, LI Jinghui¹, YU Chunlei²

(1. School of Information Engineering, Chang'an University, Xi'an 710018, China;

2. School of Intelligent Manufacturing, Sichuan University of Arts and Sciences, Dazhou 635002, China)

Abstract: As one of erasure codes, locally repairable codes (LRCs) are widely used in distributed storage systems. Aiming at the problem that current LRCs have low code rate and large locality parameter restriction when meeting the minimum distance and optimal bound, an optimal locally repairable code construction method based on square network is proposed. The square network is used to construct the check matrix of the LRCs, and the locally repairable codes are constructed by starting with the check matrix, which achieves the optimal rate bound, but its locality is limited. By further extending the incidence matrix in horizontal and vertical directions of the square network, the check matrix of the LRCs is constructed by using the extended matrix of square network, the performance of the constructed LRCs is improved in locality. And compared with existing LRCs, the constructed LRCs not only meet the minimum distance optimal bound, but also reach the rate optimal bound of the LRCs, which can be applied to any locality, it has reference significance for the construction of bivariate optimal LRCs.

Keywords: distributed storage system; LRCs; square network; minimum distance; code rate

0 引言

在大数据时代, 由于数据的急剧增长, 分布式存储系统变得越来越重要^[1]。分布式存储技术是将互联网上产生的各种数据和信息同步、分散地存储在多个独立设备中的技术。它可以通过使用网络将数千个存储节点连接在一起, 以支持数据的持续增长, 同时分布式系统还具有高访问、高性能、低成本等优点。但是分布式存储系统中存在不同种类的设备, 包括存储节点、路由器和交换机, 这些设备可能会不时出现故障, 导致系统无法正常工作。为了保证存储节点故障的可靠性, 在分布式存储系统中采用了各种编码技术。

最简单的方法是直接复制, 但是复制方案会导致较大的存储开销, 与复制相比, 纠删码在保证数据可靠性的同时, 可以大大降低存储负载, 极大地提高了存储设备的利用率。然而在故障节点的修复过程中, 经典的纠删码在访

问的辅助节点数量、修复带宽等方面效率低下。为了保证数据的可靠性和节点的有效修复, 后续的研究中陆续提出了再生码和局部修复码 (LRCs, locally repairable codes)。再生码可以通过提高修复程度和减少从每个节点下载的数据来减少修复带宽, 而局部修复码则可以有效修复故障节点, 通过在修复过程中使用少量的节点来降低修复成本^[2]。

许多学者已经对局部修复码进行了研究, Gopalan 等在文献 [3] 中首次提出局部修复码, 并证明了适用于任意码字的 Singleton-like 限。文献 [4] 和文献 [5] 分别介绍了修复局部性的概念, 如果码字的第 i 个码元能够被其它至多 r 个码元修复, 则称该码元具有修复局部性 r 。文献 [6] 证明了在分布式存储系统中使用局部修复码时可以更有效地修复故障点, 还提出了一个最优的和显式的局部修复码, 实现了任意高的数据率。在文献 [7-8] 中, Shahabineja 等研究了局部修复码的信息位的最优平均局部性, 实现了

收稿日期: 2022-11-04; 修回日期: 2022-11-10。

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

作者简介: 杨佳蓉 (1998-), 女, 山西长治人, 硕士研究生, 主要从事分布式存储中的网络编码方向的研究。

引用格式: 杨佳蓉, 王 娥, 李静辉, 等. 最优局部修复码的构造[J]. 计算机测量与控制, 2023, 31(2): 249-255.

奇偶校验位的最优最大局部性,但在构造算法中复杂度较高。Wang 等人^[9]推导出了每个修复组中包含多个奇偶校验符号时的最小距离界。Tamo 等人在文献 [10] 中提出了所有符号具有 (r,t) 局部性的局部修复码的最小距离界。在文献 [11] 中, Pamies-Juarez 等人利用部分几何图构造了具有多个不相交修复组的局部修复码^[11]。

在构造最优局部修复码方面, Tamo 在文献 [12] 中提出了一组使用多项式构造最优局部修复码的方法,但码率不高。在文献 [13] 和文献 [14] 中, Tamo 和 Luo 等人提出了环状局部修复码的构造方法,其最小距离达到了最优界,但是码率较小。在文献 [15] 中 Hao 等人给出了最优三元局部修复码的分类。Kruglik 等人在文献 [16] 中从二分图的角度出发构造二元 LRCs,并对 Tamo 等人提出的所有符号具有 (r,t) 局部性的局部修复码的最小距离的界进行了改进,达到了最小距离最优,但是其构造算法略复杂。Jung-Hyun Kim 等人提出一种基于超图的二进制局部修复码的构造^[17],给出了具有 (r,t) 局部性的超图存在的必要条件,增加了最小距离,但是码率受超图参数的影响较大。在文献 [18] 中, Wang 等人利用区组设计中的 DBBD 设计构造二元局部修复码,其最小距离和码率都达到了最优界,但是局部性有所限制。Silberstein 等人在文献 [19] 中提出了基于各种反编码的二元最优局部修复码,但是其局部性只能取 2 或 3,不能适用于任意局部性的情况。张茂等利用不相交局部修复组、sunflower 和射影几何等理论构造了一般域上的两类局部修复码^[20],所构造的两类码在相同的码的最小距离和局部度下提高了码率,但是码率没有达到最优。

针对上述问题,本文提出一种基于方形网络的局部修复码的构造方法。具体地,利用方形网络中顶点与边的对应关系构造关联矩阵,然后在关联矩阵右侧添加单位矩阵得到局部修复码的校验矩阵,进而基于该校验矩阵构造局部修复码,所构造的局部修复码最小距离和码率都达到了最优界;进一步地,将方形网络中水平方向和垂直方向的关联矩阵分别进行扩展,新构造的局部修复码参数选择灵活,算法复杂度较低,在修复故障节点时可以实现故障节点的快速精确修复。与现有局部修复码相比,本文构造局部修复码在满足最小距离最优界的同时码率也达到了最优,且适用于任意局部性。

1 基础知识

局部修复码属于线性码,设码 C 是有限域 F_q 上的 $[n,k]$ 线性码,码 C 的第 i 位具有局部性 r 是指第 i 位是其它 r 位在有限域 F_q 中的线性组合,如果码 C 的任意位都具有局部性 r ,则称码 C 是具有局部性 r 的局部修复码。局部修复码是一种特殊的纠删码,当分布式存储系统的一个节点发生故障时,该节点可以通过其它不超过 r 个存活节点进行修复, r 称为码的修复局部性。除了局部性之外,局部修复码的另一个重要特性是可用性,如果码字符符号可以从 t 个不相

交修复集中恢复,则这个符号具有 (r,t) 可用性。

定义 1^[21-22]: 如果一个码字符符号 c_i 满足以下三个属性则说明它具有 (r,t) 局部性:

- 1) 存在 t 个子集 $R_1^{(i)}, \dots, R_t^{(i)} \subset [n] \setminus \{i\}$, 使得 c_i 可以被 $R_j^{(i)}$ 中的符号修复;
- 2) $|R_j^{(i)}| \leq r, j \in [t]$;
- 3) $R_j^{(i)} \cap R_l^{(i)} = \emptyset, j \neq l \in [t]$ 。

如果只有 k 个信息符号具有 (r,t) 局部性,则称该码为具有信息局部性的 (n,k,r,t) -LRCs。

简单来说, LRCs 是纠删码的一种,可以用 n 表示码长, k 表示维度, d 表示码的最小距离。如果局部修复码的所有信息位码元都具有 (r,t) 可用性,则可以称该码具有 (r,t) 可用性的局部修复码,记作 (n,k,r,t) -LRCs。如果其信息位码元的每个修复集都只有一个校验位码元,那么称该码为单校验 (n,k,r,t) -LRCs,如果局部修复码的所有码元都具有 (r,t) 可用性,那么这个码称为全符号 (n,k,r,t) -LRCs,本文主要研究单校验 (n,k,r,t) -LRCs。

令 u 和 v 是线性码 C 中任意两个非零码字,两个码字之间不相等,码 C 的最小距离可以表示为 $d = \min\{d(u,v)\}$,也就是码字之间的最小汉明距离,对于码 (n,k,d) ,如果 d 个节点同时出现故障,分布式存储系统就不能被修复^[23]。

定理 1^[24]: 若局部修复码信息位码元的每个修复集中只含有一个校验位,那么该单校验 (n,k,r,t) -LRCs 的最小距离满足:

$$d \leq n - k - \left\lfloor \frac{kt}{r} \right\rfloor + t + 1 \quad (1)$$

称达到式 (1) 中边界的局部修复码是最小距离最优的单校验 (n,k,r,t) -LRCs。

定理 2^[25]: 若线性分组码 C 是信息位具有局部性 r 和可用性 t 的局部修复码,最小距离满足:

$$d \geq t + 1 \quad (2)$$

如果上式中局部修复码的最小距离 d 取等号,则此局部修复码为最小距离最优的局部修复码。

定理 3^[26]: (n,k,d) LRC 码的局部参数 r 满足:

$$r \geq \left\lfloor \frac{k}{n - k - d + 2} \right\rfloor \quad (3)$$

定理 4^[27]: 特别地, Prakash 等人提出了可用性 $t = 2$ 的最优码率的 $(n,k,r,t = 2)$ LRCs,该码满足:

$$R \leq \frac{r}{r+2} \quad (4)$$

2 最优局部修复码的构造

2.1 方形网络

定义 2^[28]: 由若干个单位边长的正方形拼接而成的网络称为方形网络,方形网络中水平向上和垂直向上的直线相等。

给定一个方形网络,其中顶点 p^2 ($p \geq 2$) 个, p 为正整数。 p 表示方形网络中水平方向和垂直方向的每条直线上的

顶点数量, 由此可知, 方形网络中水平和垂直方向上的直线也分别为 p 条。令方形网络的顶点表示分布式存储系统中的数据包, 水平方向和垂直方向的直线代表分布式存储系统中的存储节点, 我们将这样的 $p \times p$ 方形网络称为分布式存储系统中的方形网络。

给出分布式存储系统中的一个 3×3 的方形网络, 这里 $p = 3$, 如图 1 所示。该方形网络在水平方向和垂直方向上共有 6 条直线 v_1, v_2, \dots, v_6 , 代表分布式存储系统中的 6 个存储节点, 方形网络中的 9 个顶点 d_1, d_2, \dots, d_9 , 代表分布式存储系统中节点存储的数据包。

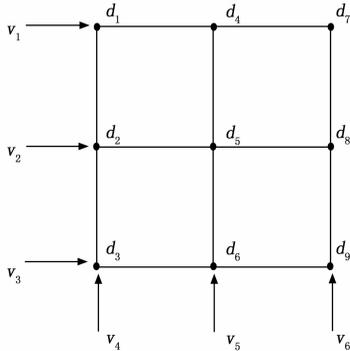


图 1 3×3 方形网络

2.2 基于方形网络构造局部修复码

本小节基于方形网络构造局部修复码, 将方形网络的顶点与分布式存储中需要存储的数据块相对应, 方形网络中水平方向和垂直方向的直线对应于分布式存储系统中的存储节点。

构造 1: 利用方形网络构造局部修复码的具体步骤如下:

步骤 1: 给定一个 $p \times p$ 方形网络, p 为正整数且 $p \geq 2$ 。将方形网络中的顶点 $d_j (1 \leq j \leq p^2)$ 按照先从上到下再从左到右的规则进行编号, 直线 $v_i (1 \leq i \leq 2p)$ 也按照同样的规则进行编号。

步骤 2: 根据 2.1, 给定的方形网络中一共有 $2p$ 条直线和 p^2 个顶点, 令第 i 条直线对应于局部修复码中的第 i 个存储节点, 第 j 个顶点对应局部修复码中第 j 个的数据块。

步骤 3: 构造基于方形网络的局部修复码的校验矩阵 $H = [M | I_{2p}]$, 其中关联矩阵 $M = (n_{ij})_{2p \times p^2}$, 当方形网络的顶点 $d_j (1 \leq j \leq p^2)$ 正好在其直线 $v_i (1 \leq i \leq 2p)$ 上时, 第 i 个节点存储顶点上的数据块 d_j , 基于方形网络的关联矩阵中 $n_{ij} = 1$, 否则为 0。

步骤 4: 由校验矩阵构造的码 C 是信息位具有局部性 r 和可用性 t 的局部修复码, 码 C 的参数满足条件 $n = 2p + p^2, k = p^2, r = p, t = 2$ 。由于 M 是基于方形网络的关联矩阵, 每一列的汉明权重都等于 2, 这意味着每个信息符号都有 $t = 2$ 个修复集, M 的每一行都有汉明权重 p , 且 M 的任意两个不同的行最多有一个位置是相同的, 因此该码具有局部性 $r = p$, 每个信息符号的修复集是不相交的, 每个修

复集只包含一个校验位。

例 1: 给定一个方形网络, 其中 $p = 3$, 如图 1 所示。根据方形网络中顶点与直线的对应关系可以构造关联矩阵:

$$M = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix} \quad (5)$$

在关联矩阵 M 的右侧添加单位矩阵可以得到校验矩阵 $H = [M | I_b]$, 从而得到参数为 $n = 15, k = 9, r = 3, t = 2$ 的局部修复码。其中, 节点发生故障时的修复方法为: 如果信息位 c_1 发生故障, 由校验矩阵 H 可知, c_1 对应的修复关系为 $c_1 = c_{10} - c_4 - c_7 = c_{13} - c_2 - c_3$, 所以信息位 c_1 的修复集可以对应表示为 $R_1^{(1)} = \{4, 7, 10\}, R_1^{(2)} = \{2, 3, 13\}$, 类似地, 可以得到其他信息符号的修复集。

若 c_1 和 c_{10} 同时故障, 由于信息位和校验位同时发生故障且在同一个修复集中, 这种情况应该选择顺序修复, 即先在修复集 $R_1^{(2)} = \{2, 3, 13\}$ 进行信息位 c_1 的修复, 然后利用 c_1, c_4, c_7 进行异或来修复校验位 c_{10} 。如果信息位 c_1 和 c_2 同时发生故障, 则这两个信息位的修复过程互不干扰, 此时进行并行修复, 可以同时从 c_1 的修复集 $R_1^{(1)} = \{4, 7, 10\}$ 或 $R_1^{(2)} = \{2, 3, 13\}$ 中任选一个修复信息位 c_1 , 从 c_2 的修复集 $R_2^{(1)} = \{5, 8, 11\}$ 或 $R_2^{(2)} = \{1, 3, 13\}$ 中任选一个修复信息位 c_2 。当其余的信息位发生故障时, 也可以用同样的方式进行修复。

定理 5: 基于方形网络构造的 $(n = 2p + p^2, k = p^2, r = p, t = 2)$ 局部修复码为最小距离最优的局部修复码, 并且 $d = t + 1$ 。

证明: 基于方形网络构造的 $(n = 2p + p^2, k = p^2, r = p, t = 2)$ 局部修复码, 其信息位码元的每一个修复集中只包含一个校验位, 根据定理 1, 将其参数代入边界条件:

$$d \leq n - k - \left\lfloor \frac{kt}{r} \right\rfloor + t + 1 = 2p + p^2 - p^2 - \left\lfloor \frac{2p^2}{p} \right\rfloor + t + 1 = t + 1 \quad (6)$$

即 $d \leq t + 1$, 又根据定理 2 可得 $d \geq t + 1$, 因此所构造的局部修复码的最小距离 $d = t + 1$, 达到了最小距离最优界, 是最小距离最优的局部修复码。

定理 6: 基于方形网络构造的 $(n = 2p + p^2, k = p^2, r = p, t = 2)$ 局部修复码的码率是最优的。

证明: $(n = 2p + p^2, k = p^2, r = p, t = 2)$ 局部修复码的码率:

$$R = \frac{k}{n} = \frac{p^2}{2p + p^2} = \frac{p}{2 + p} = \frac{r}{r + 2} \quad (7)$$

其码率满足定理 4 中 Prakash 等人提出的可用性 $t = 2$ 时局部修复码的边界, 所以此码是码率最优的局部修复码。

2.3 基于方形网络构造扩展局部修复码

上节中基于方形网络构造的局部修复码是最小距离最优的码，并且达到了码率最优界，但是该局部修复码的局部性有所限制。为进一步提高局部度，本节运用方形网络构造扩展关联矩阵，以此构造局部修复码的校验矩阵，可以得到适用于任意局部性的局部修复码。

构造 2：基于方形网络构造扩展局部修复码的具体步骤如下：

步骤 1：首先将方形网络中的顶点 $d_j (1 \leq j \leq p^2)$ 按照构造 1 的规则进行编号，直线 $v_i (1 \leq i \leq 2p)$ 也按照同样的规则进行编号。其中第 i 条直线对应于局部修复码中的第 i 个存储节点，第 j 个顶点对应局部修复码中第 j 个数据块。

步骤 2：如构造 1 步骤 3 中直线与顶点的存储关系，将基于方形网络中水平方向构造的关联矩阵定义为 M_1 ，垂直方向构造的关联矩阵定义为 M_2 ，另外用 I_p 表示 p 阶单位矩阵， S_p 表示副对角线为 1 的单位矩阵。基于方形网络构造扩展局部修复码的关联矩阵为：

$$M' = \begin{bmatrix} M_1 & I_p \\ M_2 & S_p \end{bmatrix} \quad (8)$$

步骤 3：将矩阵 M' 与单位矩阵级联作为方形网络的扩展校验矩阵 $H' = [M' | I_{2p}]$ ，由此扩展校验矩阵可以构造参数为 $(n = p^2 + 3p, k = p^2 + p, r = p + 1, t = 2)$ 的局部修复码。

由于 M' 是基于方形网络扩展的关联矩阵，由方形网络的性质可以知道每一列的汉明权重都等于 2，这意味着每个信息符号都有 $t = 2$ 个修复集， M' 的每一行都有汉明权重 $p + 1$ ， M' 的任意两个不同的行最多有一个位置是相同的，因此基于方形网络扩展矩阵构造的局部修复码具有局部性 $r = p + 1$ ，每个修复集只包含一个校验位。

例 2：给定一个方形网络，其中 $p = 3$ ，如图 1 所示，基于方形网络得到的关联矩阵为 M ，根据水平方向构造关联矩阵 M_1 和垂直方向构造关联矩阵 M_2 ：

$$M_1 = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \end{bmatrix} \quad (9)$$

$$M_2 = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix} \quad (10)$$

将 M_1 和 M_2 上下级联，并在新矩阵的右方级联一个 3 阶单位矩阵和一个副对角线为 1 的 3 阶矩阵，进而可得矩阵：

$$M' = \begin{bmatrix} M_1 & I_p \\ M_2 & S_p \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \end{bmatrix} \quad (11)$$

最后将 M' 作为方形网络的扩展关联矩阵，和单位矩阵 I_{2p} 级联生成校验矩阵 $H' = [M' | I_{2p}]$ ，由此可以构造出 $(n = 18, k = 12, r = 4, t = 2)$ 的局部修复码，从两种构造的参数对比可知，与例 1 相比，基于方形网络扩展矩阵的局部修复码适用于任意局部性的情况。

其中，例 2 中局部修复码的节点发生故障时的修复方法为：如果信息位 c_1 发生故障，由校验矩阵 H 可知， c_1 对应的修复关系为 $c_1 = c_{13} - c_{10} - c_7 - c_4 = c_{16} - c_{12} - c_3 - c_2$ ，所以信息位 c_1 的修复集可以对应表示为 $R_1^{(1)} = \{4, 7, 10, 13\}$ ， $R_1^{(2)} = \{2, 3, 12, 16\}$ ，类似地，可以得到其他信息符号的修复集。

c_1 和 c_{13} 同时发生故障时，信息位和校验位同时发生故障且在同一个修复集中，这种情况应该选择顺序修复，即先在修复集 $R_1^{(2)} = \{2, 3, 12, 16\}$ 进行信息位 c_1 的修复，然后利用 c_1, c_4, c_7, c_{10} 进行异或来修复校验位 c_{13} 。如果信息位 c_1 和 c_5 同时发生故障，则这两个信息位的修复过程互不干扰，此时进行并行修复，可以同时从 c_1 的修复集 $R_1^{(1)} = \{4, 7, 10, 13\}$ 或 $R_1^{(2)} = \{2, 3, 12, 16\}$ 中任选一个修复信息位 c_1 ，从 c_5 的修复集 $R_1^{(5)} = \{2, 8, 11, 14\}$ 或 $R_1^{(5')} = \{4, 6, 11, 17\}$ 中任选一个修复信息位 c_5 。当其余的信息位发生故障时，也可以用同样的方式进行修复。

定理 7：基于方形网络的扩展矩阵构造的 $(n = p^2 + 3p, k = p^2 + p, r = p + 1, t = 2)$ 局部修复码为最小距离最优的局部修复码，并且 $d = t + 1$ 。

证明：基于方形网络的扩展矩阵构造的 $(n = p^2 + 3p, k = p^2 + p, r = p + 1, t = 2)$ 局部修复码，根据定理 1，将其参数代入边界条件：

$$d \leq n - k - \left\lfloor \frac{kt}{r} \right\rfloor + t + 1 =$$

$$3p + p^2 - p^2 - p - \left\lfloor \frac{2(p^2 + p)}{p + 1} \right\rfloor + t + 1 = t + 1 \quad (12)$$

即 $d \leq t + 1$ ，又根据定理 2 可得 $d \geq t + 1$ ，因此所构造的局部修复码的最小距离 $d = t + 1$ ，正好达到最小距离最优界，是最小距离最优的局部修复码。

定理 8：基于方形网络的扩展矩阵构造的 $(n = p^2 + 3p, k = p^2 + p, r = p + 1, t = 2)$ 局部修复码的码率是最优的。

证明： $(n = p^2 + 3p, k = p^2 + p, r = p + 1, t = 2)$ 局部修复码的码率：

$$R = \frac{k}{n} = \frac{p^2 + p}{3p + p^2} = \frac{p + 1}{p + 3} = \frac{r}{r + 2} \quad (13)$$

其码率满足定理 4 中 Prakash 等人提出的可用性 $t = 2$ 时局部修复码的边界，所以此码是码率最优的局部修复码。

3 性能分析

3.1 码率分析

码率是局部修复码中一个重要的参数，码率表示码字中信息码元所占的比例。码的码率越大，表示其传输效率越高。本小节主要对基于方形网络构造的局部修复码的码

率进行最优化分析。

3.1.1 与校验矩阵构造的 LRCs 比较

本小节将构造 1、构造 2 与基于射影平面和基于仿射平面的局部修复码^[24]进行比较。这四种局部修复码都是由校验矩阵构造的,但是构造方式有所不同,基于射影平面的局部修复码是利用低密度奇偶校验码的校验矩阵构造,基于仿射平面的局部修复码是将仿射平面的一个固定的点和经过这个点的线删除之后得到关联矩阵,由关联矩阵来构造局部修复码的校验矩阵。

表 1 是由校验矩阵构造的不同局部修复码的参数对比。从表 1 中可以看出,基于射影平面的局部修复码的码率为 $\frac{r^2}{r^2 + rt + 1}$,与本文构造的局部修复码的码率 $\frac{r}{r+2}$ 相比,构造 1 和构造 2 的码率略高于基于射影平面的局部修复码的码率。基于仿射平面的局部修复码的码率为 $\frac{r^2 - 1}{2r^2 - 1}$,与本文构造的局部修复码的码率 $\frac{r}{r+2}$ 相比, $t \geq 2$ 时,构造 1 和构造 2 在码率上有所提升。

表 1 由校验矩阵构造的 LRCs 参数表

构造方法	n	k	d	R
构造 1	$2p + p^2$	p^2	$t + 1$	$\frac{r}{r+2}$
构造 2	$3p + p^2$	$p^2 + p$	$t + 1$	$\frac{r}{r+2}$
基于射影平面的 LRCs ^[24]	$r^2 + rt + 1$	r^2	$t + 2$	$\frac{r^2}{r^2 + rt + 1}$
基于仿射平面的 LRCs ^[24]	$2r^2 - 1$	$r^2 - 1$	$t + 2$	$\frac{r^2 - 1}{2r^2 - 1}$

由于本文构造的局部修复码的可用性 $t = 2$,图 2 给出了 $t = 2$ 时,构造 1、构造 2、基于射影平面的局部修复码和基于仿射平面的局部修复码的码率,以及 Prakash 等人提出的最优码率界。构造 1 和构造 2 的码率 $R = \frac{r}{r+2}$,达到了 Prakash 等人提出的局部修复码在可用性 $t = 2$ 时的最优码率界,从图 2 也可以看出,本文构造的局部修复码的码率曲线与 Prakash 等人提出的最优码率上界重合,且始终大于基于射影平面的局部修复码的码率和基于仿射平面的局部修复码的码率,所以本文构造的码是码率最优的局部修复码。

3.1.2 与任意 (r, t) 局部性的 LRCs 的比较

将本文构造的局部修复码与可达到任意局部性 (r, t) 的局部修复码进行比较,码的参数比较如表 2 所示。

基于直积码的局部修复码^[10]是由 t 个二元单校验 $(r + 1, t)$ 码生成的,其码率为 $(\frac{r}{r+1})^t$,当本文构造的局部修复码与基于直积码的局部修复码局部性 r 相同且 $t > 1$ 时,有

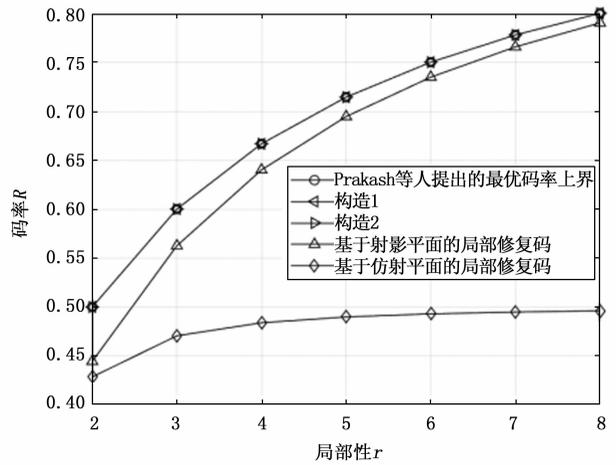


图 2 $t = 2$ 时基于校验矩阵的 LRCs 的码率 R 和局部性 r 的关系

表 2 与可达任意局部性的 LRCs 参数比较表

构造方法	n	k	d	R
构造 1	$2p + p^2$	p^2	$t + 1$	$\frac{r}{r+2}$
构造 2	$3p + p^2$	$p^2 + p$	$t + 1$	$\frac{r}{r+2}$
基于直积码的 LRCs ^[10]	$(r + 1)^2$	r^t	$t + 1$	$(\frac{r}{r+1})^t$

$$\left(1 + \frac{1}{r}\right)^t > \left(1 + \frac{t}{r}\right), \text{因此可得 } \left(\frac{r}{r+1}\right)^t = \frac{1}{\left(1 + \frac{1}{r}\right)^t} < \frac{1}{\left(1 + \frac{t}{r}\right)} = \frac{r}{r+t}, \text{所以经过分析可得构造 1 和构造 2 的码率要高于基于直积码的局部修复码的码率。}$$

如图 3 所示,为最优码率界,将本文构造的局部修复码与可达任意局部性的基于直积码的局部修复码进行比较。可以看出, $t = 2$ 时构造 1 和构造 2 的码率与 Prakash 等人提出的最优码率界重合,且始终大于基于直积码的局部修复码的码率。

3.2 局部性分析

局部修复码中的局部性 r 意味着它最多可以从 r 个其它码字符号中修复故障节点,局部性将直接影响磁盘 I/O 和修复过程中需要连接的节点数量。将构造 1 和构造 2 进行对比分析可知,构造 1 的局部性 $r = p$,而构造 2 的局部性 $r = p + 1$,构造 2 所构造的局部修复码在不牺牲码率和最小距离等关键参数的情况下,适用于任意局部性的情况。

表 3 是不同局部修复码局部性的参数对比分析,将本文构造的局部修复码与基于区组设计的局部修复码^[18]、基于反编码构造的局部修复码^[19]以及基于循环移位的局部修复码^[29]相比,基于区组设计的局部修复码的局部性为 2,限制了局部性的可选范围,从表 3 中可以看出基于区组设

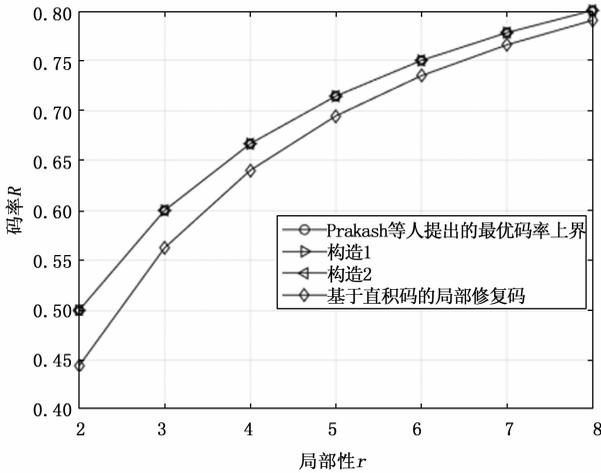


图 3 $t = 2$ 时可达任意局部度的 LRCs 的码率 R 和局部性 r 的关系

计的局部修复码的码率较低且不能适用于任意局部性的情况。

表 3 不同 LRCs 关于局部性 r 的参数对比表

构造方法	k	d	r	R
构造 1	p^2	$t + 1$	p	$\frac{r}{r + 2}$
构造 2	$p^2 + p$	$t + 1$	$p + 1$	$\frac{r}{r + 2}$
基于区组设计的 LRCs ^[18]	$p + 1$	$t + 1$	2	$\frac{2}{t + 2}$
基于反编码的 LRCs ^[19]	m	$2^{m-1} - x$	2 或 3	$\frac{m}{2^{m-1}}$
基于循环移位的 LRCs ^[29]	m	$2m - 1$	2	$\frac{1}{m - 1}$

基于反编码构造的局部修复码的局部性为 2 或者 3，局部性 r 的取值较小，只适用于具有小局部性的局部修复码，除此之外，此码的局部性 r 的可选择范围也被限制，无法灵活地构造不同局部性下的局部修复码。

基于循环移位的局部修复码的局部性也为 2，其中大部分的局部修复组都只有一个信息符号和两个校验符号，每个信息符号的局部修复不能通过访问其它信息符号来实现，只能通过访问校验符号来实现，局部性有所限制。相比之下，本文构造的局部修复码的局部性可以达到任意值，参数选择更加灵活。

4 结束语

本文针对已有的局部修复码在最小距离最优的条件下码率不高，局部性无法适用于任意情况的问题，提出了基于方形网络构造局部修复码的方法，首先利用方形网络中

顶点与边的对应关系来构造局部修复码的校验矩阵，所构造的局部修复码达到了最优码率界。为了进一步提升局部性，将方形网络水平方向和垂直方向上关联矩阵进行扩展，用扩展的关联矩阵构造局部修复码的校验矩阵，所构造的新的局部修复码不仅在最小距离最优界时达到了码率最优，并且提高了局部性。经过对码率和局部性的性能分析可知，本文构造的局部修复码不仅满足最小距离最优，并且达到了码率最优界，参数选择灵活，适用于任意局部性的情况。

参考文献:

- [1] HUANG C, SIMITCI H, XU Y, et al. Erasure coding in windows azure storage [C] // Proceedings of the 2012 USENIX Annual Technical Conference, Berkeley: USENIX Association, 2012: 15 - 26.
- [2] KOPPARTY S, SARAF S, YEKHANIN S. High-rate codes with sublinear-time decoding [C] // Proceedings of 43rd ACM Symposium on Theory of Computing. New York: ACM, 2011: 167 - 176.
- [3] GOPALAN P, HUANG C, SIMITCI H, et al. On the locality of codeword symbols [J]. IEEE Transactions on Information theory, 2012, 58 (11): 6925 - 6934.
- [4] OGGIER F, DATTA A. Self-repairing homomorphic codes for distributed storage systems [C] // International Conference on Computer Communications, 2010.
- [5] PAPAILIOPOULOS D S, LUO J, DIMAKIS, et al. Simple regenerating codes: network coding for cloud storage [C] // International Conference on Computer Communications, IEEE, 2011.
- [6] PAPAILIOPOULOS D S, DIMAKIS, et al. Locally repairable codes [J]. IEEE Transactions on Information Theory, 2014, 60 (10): 5843 - 5855.
- [7] WANG A, ZHANG Z, LIU M, et al. Achieving arbitrary locality and availability in binary codes [C] // 2015 IEEE International Symposium on Information Theory (ISIT), IEEE, 2015: 1866 - 1870.
- [8] SHAHABINEJAD M, KHABBAZIAN M, ARDAKANI M. Locally repairable codes with the optimum average information locality [C] // Proceedings of the 2017 IEEE International Symposium on Information Theory, Piscataway: IEEE, 2017: 181 - 185.
- [9] WANG A, ZHANG Z. Repair locality with multiple erasure tolerance [J]. IEEE Transactions on Information Theory, 2014, 60 (11): 6979 - 6987.
- [10] TAMO I, BARG A. Bounds on locally recoverable codes with multiple recovering sets [C] // 2014 IEEE International Symposium on Information Theory, IEEE, 2014: 691 - 695.
- [11] PAMIES-JUAREZ L, HOLLMANN H D, OGGIER F. Locally repairable codes with multiple repair alternatives [C] // 2013 IEEE International Symposium on Information Theory,

- IEEE, 2013; 892–896.
- [12] TAMO I, BARG A. A family of optimal locally recoverable codes [J]. *IEEE Transactions on Information Theory*, 2014, 60 (8): 4661–4676.
- [13] TAMO I, BARG A, GOPARAJU S, et al. Cyclic LRC codes and their subfield subcodes [C] //2015 IEEE International Symposium on Information Theory (ISIT), IEEE, 2015; 1262–1266.
- [14] LUO Y, XING C, YUAN C. Optimal locally repairable codes of distance 3 and 4 via cyclic codes [J]. *IEEE Transactions on Information Theory*, 2018, 65 (2): 1048–1053.
- [15] HAO J, XIA S T, CHEN B. On optimal ternary locally repairable codes [C] //2017 IEEE International Symposium on Information Theory (ISIT), IEEE, 2017; 171–175.
- [16] KRUGLIK S, NAZIRKHANOVA K, FROLOV A. New bounds and generalizations of locally recoverable codes with availability [J]. *IEEE Transactions on Information Theory*, 2019, 65 (7): 4156–4166.
- [17] KIM J H, SONG H Y. Hypergraph-based binary locally repairable codes with availability [J]. *IEEE Communications Letters*, 2017, 21 (11): 2332–2335.
- [18] WANG J, SHEN K Q, LIU X Y, et al. Construction of binary locally repairable codes with optimal distance and code rate [J]. *IEEE Communications Letters*, 2021, 25 (7): 2109–2113.
- [19] SILBERSTEIN N, ZEH A, et al. Optimal binary locally repairable codes via anticodes [C] //2015 IEEE International Symposium on Information Theory (ISIT), IEEE, 2015; 1247–1251.
- [20] 张 茂, 李瑞虎, 郑尤良, 等. 基于 sunflower 的局部修复码 (上接第 248 页)
- [7] 陈 晨, 胡明成, 孙灿顶, 等. 一种载波监听门限与竞争窗口联合控制的方法 [P]. 中国: CN104936303B; 2018–09–21.
- [8] 李明菊, 朱亚军, 张云飞. 基于竞争窗口的信道检测方法及其装置 [P]. 中国: CN105338652A; 2016–02–17.
- [9] 高祎宁. 基于改进的退避机制的 MAC 协议性能分析 [D]. 兰州: 兰州理工大学, 2019.
- [10] 孙 晟, 徐正勤. 一种 CSMA-CA 半窗方案的系统和方法 [P]. 中国: CN105284079B; 2016–01–27.
- [11] SU Y S, FU X M, HAN G Y, et al. Implementation of a cross-layer sensing medium-access control scheme [J]. *Sensors*, 2017, 17 (4): 816–915.
- [12] LEE J H, CHO S H. Tree TDMA MAC algorithm using time and frequency slot allocations in tree-based WSNs [J]. *Wireless Personal Communications*, 2017, 95 (3): 2575–2597.
- [13] WU W, LI Y, ZHANG Y H, et al. Distributed queueing based random access protocol for LoRa networks [J]. *IEEE Internet of Things Journal*, 2020, 7 (1): 763–772.
- [14] CHENG R G, BECVAR Z, YANG P H. Modeling of distributed queueing-based random access for machine type communications [J]. *计算机应用*, 2021, 41 (03): 763–767.
- [21] RAWAT A S, PAPALIOPOULOS D S, DIMAKIS A G, et al. Locality and availability in distributed storage [J]. *IEEE Transactions on Information Theory*, 2016, 62 (8): 4481–4493.
- [22] TAN P, ZHOU Z, SIDORENKO V, et al. Two classes of optimal LRCs with information (r, t) -locality [J]. *Designs, Codes and Cryptography*, 2020, 88 (9): 1741–1757.
- [23] 王新梅, 肖国镇. 纠错码: 原理与方法 [M]. 西安: 西安电子科技大学出版社, 1991.
- [24] HAO J, XIA S T, CHEN B. On the single-parity locally repairable codes with availability [C] //2016 IEEE/CIC International Conference on Communications in China (ICCC), IEEE, 2016; 1–4.
- [25] TAN P, ZHOU Z, SIDORENKO V, et al. Constructions of optimal locally repairable codes with information (R, T) -Locality [C] //The Eleventh International Workshop on Coding and Cryptography (WCC), Saint-Jacut-de-la-Mer, France, 2019.
- [26] SHAHABINEJAD M, KHABBAZIAN M, ARDAKANI M. On the average locality of locally repairable codes [J]. *IEEE Transactions on Communications*, 2018, 66 (7): 2773–2783.
- [27] PRAKASH N, LALITHA V, KUMAR P V. Codes with locality for two erasures [C] //2014 IEEE International Symposium on Information Theory, IEEE, 2014; 1962–1966.
- [28] 沈克勤. 分布式存储系统中优化编码研究 [D]. 西安: 长安大学, 2021.
- [29] LEE K S, PARK H, NO J S, et al. New binary locally repairable codes with locality 2 and uneven availabilities for hot data [J]. *Entropy*, 2018, 20 (9): 636.
- communications in mobile networks [J]. *IEEE Communications Letters*, 2018, 22 (1): 129–132.
- [15] GARCIA-LUNA-ACEVES J J. Carrier-tone multiple access with collision avoidance and detection [C] //2018 IEEE Wireless Communications and Networking Conference (WCNC), Barcelona, Spain, 2018; 1–6.
- [16] PERVEZ Z, SHAH S F, HUSSAIN Z, et al. An application of negative binomial distribution of order k in sensitive surveys [J]. *Communications in Statistics*, 2017, 46 (8): 6654–6660.
- [17] KILGOUR D M. Approval elections with a variable number of winners [J]. *Theory & Decision*, 2016, 81 (2): 1–13.
- [18] 赵国峰. 5G 异构网络非授权频段共存机制研究 [D]. 北京: 北京交通大学, 2019.
- [19] LIU W, HE X X, HUANG Z T, et al. Transmission capacity characterization in VANETs with enhanced distributed channel access [J]. *Electronics*, 2019, 8 (3): 340–357.
- [20] CHARFI E, GUEGUEN C, CHAARI L, et al. Dynamic frame aggregation scheduler for multimedia applications in IEEE 802.11n networks [J]. *Transactions on Emerging Telecommunications Technologies*, 2017, 28 (2): 2942–2958.