为 Reed-Solomon 码设计紧凑修复组

Thi Xinh Dinh*†、Serdar Boztas*‡、Son Hoang Dau* 和 Emanuele Viterbo§
*澳大利亚墨尔本皇家墨尔本理工大学，维多利亚州 3000，澳大利亚， {s3880660@student.rmit, serdar.boztas@rmit, sonhoang.dau@rmit}.edu.au
†越南西原大学，dinhthixinh@ttn.edu.vn
‡大学研究基金会，美国马里兰州格林贝尔特市，邮编 20770，boztas@urf.com§莫纳什大学，澳大利亚维多利亚州克莱顿市，邮编 3800，emanuele.viterbo@monash.edu

arXiv:2305.06600v1 [cs.IT] 11 May 2023

摘要——**受 Reed-Solomon 码在 Storj 和 Filebase/Sia 等新兴去中心化存储系统中的应用所启发，我们研究如何设计紧凑修复组，以便以去中心化的方式恢复多个故障。这里，紧凑性意味着这些辅助节点组所对应的迹修复方案可以由一个或少数几个种子修复方案生成，从而节省查找和存储这些方案所需的时间与空间。目标是设计能够容忍尽可能多故障的紧凑修复组。结果表明，一组修复组所能容忍的最大故障数，等于有限域 $\mathbb{F}_q$minusone 中一组子集的最小击中集的大小。当每个符号的修复组由一个单个子空间生成时，我们为这种最小击中集的大小建立了一对渐近紧确的下界和上界。利用伯恩赛德引理和 M¨obius 反演公式，我们确定了多个子空间的数量；当修复组由多个子空间生成时，这些子空间共同达到最小击中集大小的上界。**

I. 引言

与 Google File System II [1] 或 Facebook 的 f4 [2]，等传统分布式存储系统不同，在这些系统中，存储文件或数据对象的服务器发生的故障大多是单点故障 [3], ；而在近期兴起的去中心化存储系统（如基于区块链的 Storj [4] 和 Filebase/Sia [5],）中，多点故障才是常态。分布式系统与去中心化系统之间的这一差异，源于去中心化系统中的存储节点不受任何中心化组织控制，因此可以更自由地加入或离开系统，也可以随时上线或下线。相比之下，传统分布式存储系统中的所有存储节点均由中心化服务提供商管理，极少发生故障或离开系统。

因此，去中心化系统中采用的 Reed-Solomon 码，其参数范围也显著不同。例如，虽然大多数主流分布式存储系统使用具有三到四个校验位的短码，如 RS(9,6) 或 RS(14,10)，但 Storj 和 Filebase/Sia 等去中心化存储系统依赖冗余度更高的长码，如 RS(30,10)、RS(40,20) 或 RS(80,40) [6, 表 I]。较高的冗余度使这些码能够容忍去中心化环境中常见的更多节点故障。

当存储与特定数据对象相对应的码字符号的节点发生故障（或退出系统）时，其他节点可以加入，作为该对象的替换节点。这些节点可以是刚刚加入系统的新节点，也可以是已准备好提供额外存储空间的现有节点。必须有一种机制来处理这些节点以恢复丢失的符号。当前的去中心化系统（如 Storj [4] ）正在采用“惰性修复”机制 [7]–[9],，该机制会等待发生一定数量的故障后，再应用朴素集中式修复方法，一次性修复所有故障。该方法会随机指定一个节点作为修复节点，由其从 k 个任意可用节点（k 为码维数）收集码字符号，一次性恢复所有丢失的码字符号，并将其发送至相应的替换节点（见图 1）。我们注意到，朴素集中式修复方法的修复带宽，即修复过程中存储节点之间通信的总比特数，有可能通过采用文献 [10]–[16] 中提出的针对多重故障的迹修复方法来降低。此外，由于针对多重故障寻找低带宽修复方案的（启发式）搜索问题计算量极大，因此代数构造提供了一种更为可行的方法，但其适用的参数范围有限。

本工作研究第二种方法，即去中心化修复 方法。在这种方法中，不同的替换节点独立地执行修复过程，不一定同时进行。在其朴素版本中，每个替换节点都执行朴素修复方案，从 k 辅助节点下载数据并恢复丢失的符号。我们关注一种带宽效率更高的版本，其中应用了迹修复方法 [10], [11]: 每个替换节点从超过 k 个节点下载数据，但与朴素方案相比，这样可以降低修复带宽。与 Storj 的朴素集中式 修复方法相比，这种方法可能需要更高的总修复带宽[1], , 但恢复任务会分散到多个节点上，从而消除单个修复节点上的通信与计算瓶颈。这有望缩短系统的总恢复时间，并降低因指定修复节点自身发生故障而导致恢复过程受阻的可能性。此外，主动而非延迟地修复节点，可以降低因相关故障（软件漏洞或病毒攻击）造成数据丢失的风险。

然而，使用迹修复来降低修复带宽会带来额外开销：寻找迹修复方案时的计算开销（见 [6,第 IV. B 节]），以及存储迹修复方案参数的存储开销（见 [6,第 V. B 节]）。一种极端的解决方案是

[1]对于 RS(n, k) 码，修复 e 个失效节点时，Storj 的修复方案需要的总带宽为 $k\ell+(e-1)\ell$ 个符号，而我们的方案需要 $ek\ell(1-s)$ 个符号。其中， s 表示与朴素方案相比，轨迹修复方法节省或减少的带宽，例如 $s = 0.3$。

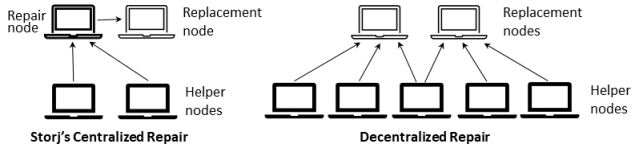


图 1：在维数为 $k=2$ 的码中修复 $e=2$ 个失效节点时，Storj 的集中式方法与去中心化方法的示意图。

对每个码字符号使用由固定的 d 个辅助节点组成的单个修复组，从而将计算／存储开销降至最低。遗憾的是，这种方案甚至无法容忍一次故障：一旦该修复组中的一个节点发生故障，相应的修复方案便不再有效。另一种极端方案是，通过选择所有$\binom{n-1}{d}$ 个不同的修复组，并寻找／存储相应的修复方案，最多容忍 $n-1-d$ 次故障。显然，这种方案在计算和存储方面都需要指数级开销。我们的目标是为每个码字符号设计紧凑的修复组，在实现低计算和存储开销的同时，尽可能多地容忍故障。在本研究范围内，我们仅考虑长度为 $n=q^\ell$ 的 $\mathbb{F}_{q^\ell}$ 上的全长度码。

构造紧凑修复组的关键思路如下：我们不使用大量互不相关的修复组（这会带来较高的计算和存储开销），而是精心选择彼此相关的修复组，使其对应的修复方案能够通过一个或少数几个“种子”修复方案插值得到，从而降低构造和存储这些修复组的开销。具体而言，为修复码字符号 c_j，我们选取一个 d-子集 $S \ni 0$，称为种子，并将其所有不同的乘法陪集 $\mathcal{C}(S^*) \triangleq \{bS^*\colon\ b \in \mathbb{F}_{q^\ell}^*\}$（其中 $S^* = S \setminus \{0\}$ 和 $\mathbb{F}_{q^\ell}^* = \mathbb{F}_{q^\ell} \setminus \{0\}$）经过适当平移后用作不同的修复组。事实证明，所有这些修复组对应的修复方案都可以通过对应于 S^* 的单个修复方案轻松生成。此外，这些紧凑修复组所能容忍的最大失效数，等于 $\mathcal{C}(S^*)$ 的最小击中集大小减一。最后，为进一步提高这一数值，我们可以从多个种子开始。为获得易于处理的结果，我们研究 $\mathbb{F}_{q^\ell}$ 的一个 $\mathbb{F}_q$-子空间 S，并考虑全长码。

我们的贡献总结如下。

- 首先，我们介绍了一个新问题：在去中心化场景下为 Reed-Solomon 码设计紧凑修复组，并将其与最小击中集问题联系起来。
- 其次，当修复组由单个种子 S 生成时，我们建立了 $\mathcal{C}(S^*)$ 的最小击中集大小的渐近匹配的下界和上界（当 $q \to \infty$ 时）。值得注意的是，对于任意集合族的最小击中集大小，其最佳近似比通常为 $\log(n) = \log(q^\ell) = \ell\log(q)$ [17], [18] 的量级。
- 最后，我们证明，可以通过使用 s 种子达到可容忍故障数量的上界，其中s 是 $\mathcal{S}_\delta^*$的轨道数； $\mathcal{S}_\delta^*$是由所有 S^*构成的集合，其中S是 $\mathbb{F}_{q^\ell}$中的 δ维 $\mathbb{F}_q$子空间， $\delta = \log_q d$，并且这些轨道是在乘法群 $\mathbb{F}_{q^\ell}^*$的作用下得到的。我们利用伯恩赛德引理和莫比乌斯反演公式推导出 s 的显式公式。

II. 预备知识

A. 定义与符号约定

设 q 是一个素数幂， $\mathbb{F}_q$ 是含有 q 个元素的有限域， $\mathbb{F}_{q^\ell}$ 是 $\mathbb{F}_q$的 ℓ 次扩域。我们用 $[n]$ 表示集合 $\{1,2,\dots,n\}$，并用 $a \mid b$ 表示对于两个整数 a 和 b， a 整除b。对于集合 S，令 $S^* \triangleq S \setminus \{0\}$。

设 C 是 $\mathbb{F}_{q^\ell}$上的一个 线性 $[n,k]$ 码。那么 C 是$\mathbb{F}_{q^\ell}^n$ 的一个 k 维 $\mathbb{F}_{q^\ell}$ 子空间。 C 的一个码字是一个元素 $\vec{\boldsymbol{c}} = (\boldsymbol{c}_1, \boldsymbol{c}_2, \dots, \boldsymbol{c}_n) \in C$，其码字符号是各个分量 $\boldsymbol{c}_j$、 $j \in [n]$。码 C 的对偶码是 C 的正交补 $C^\perp$ ，即 $C^\perp = \{\vec{\boldsymbol{g}} \in \mathbb{F}_{q^\ell}\colon \langle \vec{\boldsymbol{c}}, \vec{\boldsymbol{g}} \rangle = 0, \forall \vec{\boldsymbol{c}} \in C\}$，其中 $\langle \vec{\boldsymbol{c}}, \vec{\boldsymbol{g}} \rangle$ 是 $\vec{\boldsymbol{c}}$ 与 $\vec{\boldsymbol{g}}$ 的数量积。码 $C^\perp$ 是$\mathbb{F}_{q^\ell}^n$ 的一个 $\mathbb{F}_{q^\ell}$ 子空间，其维数为 $n-k$。 $C^\perp$ 中的元素称为对偶码字。数值 $n-k$ 称为码 C 的冗余度。

定义 1. 设 $A = \{\boldsymbol{\alpha}_j\}_{j=1}^n$ 是 $\mathbb{F}_{q^\ell}$中大小为 n 的一个子集。维数为 k 、求值点为 A 的 ***Reed-Solomon*** 码 $\mathrm{RS}(A,k) \subseteq \mathbb{F}_{q^\ell}^n$ 定义为

$$\mathrm{RS}(A,k) = \big\{\big(f(\boldsymbol{\alpha}_1)\dots, f(\boldsymbol{\alpha}_n)\big)\colon\ f \in \mathbb{F}_{q^\ell}[x], \deg(f) < k\big\},$$

其中 $\mathbb{F}_{q^\ell}[x]$ 是 $\mathbb{F}_{q^\ell}$上的多项式环。我们还使用记号 $\mathrm{RS}(n,k)$，省略求值点。

广义 Reed-Solomon 码 $\mathrm{GRS}(A,k,\vec{\boldsymbol{\lambda}})$，其中 $\vec{\boldsymbol{\lambda}} = (\boldsymbol{\lambda}_1, \dots, \boldsymbol{\lambda}_n) \in \mathbb{F}_{q^\ell}^n$，是码字 $\big(\boldsymbol{\lambda}_1 g(\boldsymbol{\alpha}_1) \dots, \boldsymbol{\lambda}_n g(\boldsymbol{\alpha}_n)\big)$的集合，其中 $\boldsymbol{\lambda}_j \neq 0$对所有 $j \in [n]$，$g \in \mathbb{F}_{q^\ell}[x]$均成立，且 $\deg(g) < n-k$。Reed-Solomon 码 $\mathrm{RS}(A,k)$ 的对偶码是一个广义 Reed-Solomon 码 $\mathrm{GRS}(A, n-k, \vec{\boldsymbol{\lambda}})$，其中乘子向量 $\vec{\boldsymbol{\lambda}}$（[19, 第 10] 章）。有时我们使用记号 $\mathrm{GRS}(n,k)$，省略 A和 $\vec{\boldsymbol{\lambda}}$。

设 $f(x)$ 为与 Reed-Solomon 码 $C = \mathrm{RS}(A,k)$ 的一个码字相对应的多项式， $g(x)$ 为次数至多为 $n-k-1$、与对偶码 $C^\perp$ 的一个码字相对应的多项式。则 $\sum_{j=1}^n g(\boldsymbol{\alpha}_j)\big(\boldsymbol{\lambda}_j f(\boldsymbol{\alpha}_j)\big) = 0$, 我们称多项式 $g(x)$ 为 C 的校验多项式。

对于子集 $S \subseteq \mathbb{F}_{q^\ell}$ 和 $a \in \mathbb{F}_{q^\ell}$，令 $aS \triangleq \{as\colon\ s \in S\}$，$a + S \triangleq \{a+s\colon\ s \in S\}$， $\mathcal{C}(a,\ S) \triangleq \{a + bS\colon\ b \in \mathbb{F}_{q^\ell}^*\}$，且$\mathcal{C}(S) \triangleq \{bS\colon\ b \in \mathbb{F}_{q^\ell}^*\}$。令 $\mathcal{S}$ 为集合族 $\{S_i\}_{i \in I}$。 $\mathcal{S}$ 的击中集是一个与所有 S_i 都有非空交集的集合，其中 $i \in I$。若 $\mathcal{S}$ 的击中集在所有击中集中大小最小，则称其为 最小击中集 （MHS）。为简便起见，我们用 $\mathrm{MHS}(\mathcal{S})$ 表示 $\mathcal{S}$ 的一个 MHS。

B. 迹修复方法

设 $\mathrm{RS}(n,k)$ 为定义在 $\mathbb{F}_{q^\ell}$ 上、评估点为 A, $\vec{\boldsymbol{c}}$ 的 Reed-Solomon 码，其中一个对应于多项式 $f(x)$ 的码字为 $f \in \mathbb{F}_{q^\ell}[x]$， $\deg(f) < k$，且 $\boldsymbol{c}^* = f(\boldsymbol{\alpha}^*)$，其中 $\boldsymbol{\alpha}^* \in A$ 为 $\vec{\boldsymbol{c}}$ 的一个码字符号/节点。针对 $\boldsymbol{c}^*$ 的一个（线性）迹修复方案对应于一组 ℓ 个校验多项式 $\{g_i(x)\}_{i \in [\ell]}$, $g_i \in \mathbb{F}_{q^\ell}[x]$ $g_i < n-k$ ， $\deg(\,)$，其满足满秩条件：$\mathrm{rank}_{\mathbb{F}_q}\{g_i(\boldsymbol{\alpha}^*)\}_{i \in [\ell]} = \ell$。如果我们要求辅助节点集合对应于 $S \setminus \{\boldsymbol{\alpha}^*\}$，其中 $\boldsymbol{\alpha}^* \in S \subseteq A$，那么对于每个 $i \in [\ell]$ 和 $\boldsymbol{\alpha} \notin S$，必须满足 $g_i(\boldsymbol{\alpha}) = 0$。此类修复方案的修复带宽（以 $\mathbb{F}_q$ 符号为单位）为 $\mathrm{bw} = \sum_{\boldsymbol{\alpha} \in S \setminus \{\boldsymbol{\alpha}^*\}} \mathrm{rank}^{\mathbb{F}_q}(\{g_i(\boldsymbol{\alpha})\}_{i \in [\ell]})$。

要修复 $\vec{\boldsymbol{c}}$的所有 n 个组件，我们需要 n 个这样的修复方案（可能会有重复）。例如，参见 [20], ，其中详细说明了上述方案为何有效，并给出了一个示例。

A. 问题描述Th

我们假设在发生多次故障的去中心化存储系统中，采用了一个全长度 Reed-Solomon 码 $RS(n,k)$。一些存储节点，即码字符号，丢失并需要恢复。为了修复每个故障节点，相应的替换节点会联系存活节点中的一组 $d \le n-1$ 辅助节点（独立于其他替换节点）。在用于修复过程之前，必须确定并存储与每组此类 d辅助节点及丢失节点相对应的低带宽修复方案。

问题。迹修复方法 [10], [11] 可用于构建低带宽的修复方案。由于通过启发式算法为任意修复组寻找低带宽修复方案可能需要耗费大量时间（参见 [6]），因此系统应离线寻找所有修复方案，并预先存储其参数。使用的修复组越多，这些修复组能够容忍的故障数量就越多（即至少有一个修复组保持完整）。然而，使用更多修复组也意味着需要更高的计算/存储成本。本文旨在解决这一两难问题。

B. 总体策略

我们的核心思路是使用多个相关的修复组，即尽管会使用多个修复方案，但它们都仅由一个或少数几个种子修复方案生成。我们将此类修复组称为紧凑型修复组。对于这类修复组，我们只需存储一组或少数几组求值点（称为种子）及其对应的修复方案，随后即可利用它们生成所有其他修复方案。

更具体地说，对于单个种子的情况，假设我们已经有一个针对 $f(0)$ 的修复方案，其辅助节点对应于 $S^* \triangleq S \setminus \{0\}$，其中 $0 \in S \subseteq \mathbb{F}_{q^\ell}^*$。那么，对于每个 $\boldsymbol{\alpha}^* \in \mathbb{F}_{q^\ell}$，我们都可以构造一个针对 $f(\boldsymbol{\alpha}^*)$ 的修复方案，其带宽相同，且其辅助节点对于每个 $b \in \mathbb{F}_{q^\ell}^*$ 都对应于$\boldsymbol{\alpha}^* + bS^*$。如引理 1 所示（另见 [6, 推论 4]），通过对针对 $f(0)$ 的（种子）修复方案进行适当的伸缩和平移，即可得到针对 $f(\boldsymbol{\alpha}^*)$ 的修复方案。对于多个种子，我们只需从多个集合 $\{S_t\}_{t\in[s]}$ 开始。请注意，所有这些结论均在所考虑的 RS 码为满长码的假设下成立，因此，对于每个 $\boldsymbol{\alpha}^*$、b 和 S，$\boldsymbol{\alpha}^* + bS^*$ 都对应于求值点的一个子集，即一个有效的辅助节点集合。

引理 1. 假设 $0 \in S \subseteq \mathbb{F}_{q^\ell}$。设 $\{g_i(x)\}_{i=1}^{\ell} \subset \mathbb{F}_{q^\ell}[x]$ 构成一个带宽为bw 的修复方案，且辅助节点对应于 *RS* 码的 $f(0)$ 时的 S^*，其求值点为 S。那么，对于每个 $b \in \mathbb{F}_{q^\ell}^*$，$\{h_i(x) \triangleq g_i\big((x-\boldsymbol{\alpha}^*)/b\big)\}_{i=1}^{\ell}$构成一个带宽为 bw 的修复方案，且辅助节点对应于维数相同、求值点为 $\boldsymbol{\alpha}^* + bS$ 的 *RS* 码在 $f(\boldsymbol{\alpha}^*)$ 时的 $\boldsymbol{\alpha}^* + bS^*$。

证明。由于对于每个 $i \in [\ell]$，$h_i(\boldsymbol{\alpha}^*) = g_i(0)$，我们有 $\mathsf{rank}_{\mathbb{F}_q}\big(\{h_i(\boldsymbol{\alpha}^*)\}_{i=1}^{\ell}\big) = \ell$，这意味着满足满秩条件。此外，对于 $\boldsymbol{a} \in \boldsymbol{\alpha}^* + bS^* = (\boldsymbol{\alpha}^* + bS) \setminus \{\boldsymbol{\alpha}^*\}$，我们有 $\boldsymbol{a} = \boldsymbol{\alpha}^* + b\boldsymbol{\alpha}$，其中某个 $\boldsymbol{\alpha} \in S^*$。因此，$h(\boldsymbol{a}) = h(\boldsymbol{\alpha}^* + b\boldsymbol{\alpha}) = g(\boldsymbol{\alpha})$，从而，

$$\sum_{\boldsymbol{a}\in\boldsymbol{\alpha}^*+bS^*} \mathsf{rank}_{\mathbb{F}_q}\big(\{h_i(\boldsymbol{a})\}_{i=1}^{\ell}\big) = \sum_{\boldsymbol{\alpha}\in S^*} \mathsf{rank}_{\mathbb{F}_q}\big(\{g_i(\boldsymbol{\alpha})\}_{i=1}^{\ell}\big) = \mathsf{bw}.$$

因此，$\{h_i(x)\}_{i=1}^{\ell}$ 构成了针对 $f(\boldsymbol{\alpha}^*)$ 的一种修复方案，其带宽与针对 $f(0)$ 的修复方案 $\{g_i(\boldsymbol{\alpha})\}_{i=1}^{\ell}$ 相同。■

令 $\mathcal{C}(S^*) \triangleq \{bS^* : b \in \mathbb{F}_{q^\ell}^*\}$ 为 S^* 的互不相同的乘法陪集集合，并令 $\mathcal{C}(\boldsymbol{\alpha}^*, S^*) \triangleq \{\boldsymbol{\alpha}^* + bS^* : b \in \mathbb{F}_{q^\ell}^*\}$。根据上述讨论，由单个种子 S 生成的 $f(\boldsymbol{\alpha}^*)$ 的修复方案数量为 $|\mathcal{C}(\boldsymbol{\alpha}^*, S^*)| = |\mathcal{C}(S^*)|$。发生节点故障时，所有涉及至少一个故障节点作为辅助节点的修复方案都将无法使用。我们称这些修复方案被这些节点故障击中。因此，这些修复组所能容忍的最大故障数，比 $\mathcal{C}(\boldsymbol{\alpha}^*, S^*)$ 以及 $\mathcal{C}(S^*)$ 的最小击中集大小少一。我们在定义 2 和引理 2 中总结这一讨论，并将其推广至多种子情形。

定义 2. 令 $\mathcal{C}\big(\boldsymbol{\alpha}^*, \{S_t^*\}_{t\in[s]}\big)$ 表示集合$\cup_{t\in[s]}\{\boldsymbol{\alpha}^* + bS_t^* : b \in \mathbb{F}_{q^\ell}^*\}$，其中 s 为种子 $\{S_t\}_{t\in[s]}$。那么，该集合所能容忍的最大故障数，记为 $\mathcal{F}\big(\mathcal{C}\big(\boldsymbol{\alpha}^*, \{S_t^*\}_{t\in[s]}\big)\big)$，定义为满足以下条件的最大数值 e：对于每一个 $E \subset \mathbb{F}_{q^\ell} \setminus \{\boldsymbol{\alpha}^*\}$，$|E| = e$，集合中至少存在一个集合与 E不相交。

引理 2。设 $\{S_t\}_{t\in[s]}$ 是 $\mathbb{F}_{q^\ell}$ 的 s 个子集 $\big|(\mathcal{C}(\boldsymbol{\alpha}^* \{S_t^*\}_{t\in[s]}))\big| - 1 = \big|(\mathcal{C}(\{S_t^*\}_{t\in[s]}))\big| - 1$ 的集合，包含 。那么，MHS，MHS，其中 $\mathcal{C}\big(\{S_t^*\}_{t\in[s]}\big) \triangleq \cup_{t\in[s]}\{bS_t^* : b \in \mathbb{F}_{q^\ell}^*\}$。

证明。显然，$\big|\mathsf{MHS}\big(\mathcal{C}\big(\boldsymbol{\alpha}^*, \{S_t^*\}_{t\in[s]}\big)\big)\big|$ 是覆盖$\mathcal{C}\big(\boldsymbol{\alpha}^*, \{S_t^*\}_{t\in[s]}\big)$中所有修复组所需的最少失效数量。显然还有$\big|\mathsf{MHS}\big(\mathcal{C}\big(\boldsymbol{\alpha}^*, \{S_t^*\}_{t\in[s]}\big)\big)\big| = \big|\mathsf{MHS}\big(\mathcal{C}\big(\{S_t^*\}_{t\in[s]}\big)\big)\big|$。因此，与 $\mathcal{C}(\boldsymbol{\alpha}^*, S^*)$相对应的修复组最多能够容忍 $\big|\mathsf{MHS}\big(\mathcal{C}\big(\boldsymbol{\alpha}^*, \{S_t^*\}_{t\in[s]}\big)\big)\big| - 1$次失效。■

得益于引理 2，我们已将确定由种子 $\{S_t\}_{t\in[s]}$ 生成的紧凑修复组集合所能容忍的故障数量这一问题，转化为寻找 $\mathcal{C}\big(\{S_t^*\}_{t\in[s]}\big)$ 的最小击中集这一问题。

C. 从单个种子构造紧凑修复组

寻找任意集合族的 MHS 通常是一个 NP-hard 问题 [21]。我们重点研究一种特殊情况，即种子是 $\mathbb{F}_{q^\ell}$ 的 $\mathbb{F}_q$-子空间。值得注意的是，文献中已经提出了求值点构成一个子空间的 Reed-Solomon 码的低带宽修复方案 [15],[22]。本节研究 $\big|\mathsf{MHS}\big(\mathcal{C}(S^*)\big)\big|$，其中 S 是一个单独的种子，同时也是 $\mathbb{F}_{q^\ell}$ 的一个 δ-维 $\mathbb{F}_q$-子空间。我们为 $\big|\mathsf{MHS}\big(\mathcal{C}(S^*)\big)\big|$ 建立了一对渐近紧致的下界和上界（定理 1），当 $q \to \infty$ 时成立。值得注意的是，在一般情况下，这是不可能实现的。因此，对于某些特殊情况，可以确定 MHS 的精确值（推论 2、推论 3）。

对于上界，给定一个 δ 维 $\mathbb{F}_q$ 子空间 S，它属于 $\mathbb{F}_{q^\ell}$，我们首先将 $\big|\mathsf{MHS}\big(\mathcal{C}(S^*)\big)\big|$ 上界为 $|\mathsf{MHS}(\mathcal{S}_\delta^*)|$，其中 $\mathcal{S}_\delta^*$ 是 $\mathbb{F}_{q^\ell}$ 中所有非零 δ 维 $\mathbb{F}_q$ 子空间的集合；随后将 $\mathsf{MHS}(\mathcal{S}_\delta^*)$ 与 Tur´an 设计的一种特殊情形联系起来 [23]。

引理 3. 若 S 是 δ维的 $\mathbb{F}_q$子空间 S，且$\mathcal{S}_\delta^* \triangleq \{T^* : T$是 $\mathbb{F}_{q^\ell}\}$ 的 δ维 $\mathbb{F}_q$子空间，则 $\big|\mathsf{MHS}\big(\mathcal{C}(S^*)\big)\big| \le \big|\mathsf{MHS}\big(\mathcal{S}_\delta^*\big)\big|$。

证明。由于 $\mathbb{F}_{q^\ell}$ 的 $\mathbb{F}_q$子空间的一个乘法陪集仍是维数相同的 $\mathbb{F}_q$子空间，若 $\dim_{\mathbb{F}_q}(S) = \delta$，则 $\mathcal{C}(S^*) \subseteq \mathcal{S}_\delta^*$。引理得证。■

事实证明，$\mathcal{S}_\delta^*$ 的 MHS 对应于一种特殊的 Tur´an 设计 [23]。下面我们回顾 Tur´an 设计的定义，并讨论其与 $\mathcal{S}_\delta^*$ 的 MHS 之间的关系。

定义 3. 令 $\mathcal{S}_\delta$ 和 $\mathcal{S}_r$ 分别为 $\mathbb{F}_{q^\ell}$中维数为 δ和 r的所有 $\mathbb{F}_q$-子空间的集合。一个 (q,ℓ,δ,r)-图兰´设计 是 $\mathcal{S}_r$ 的一个子集 $\mathcal{D}$，使得 $\mathcal{S}_\delta$ 中的每个子空间都至少包含 $\mathcal{D}$中的一个子空间。图兰设计的最小规模称为 (q,ℓ,δ,r)-图兰´数，记为 $\mathcal{T}_q(\ell,\delta,r)$。

在引理 4 中，我们证明 $|\mathrm{MHS}(\mathcal{S}_\delta^*)|$ 与上文定义的图兰数（带有 $r=1$）相同。

引理 4. 对于每个素数幂 q 和每个 $1 \le \delta \le \ell$，我们都有 $|\mathrm{MHS}(\mathcal{S}_\delta^*)| = \mathcal{T}_q(\ell,\delta,1)$，其中 $\mathcal{S}_\delta^*$是 $\mathbb{F}_{q^\ell}$ 中所有维数为 δ 的 $\mathbb{F}_q$-子空间的集合（不包括零元素），而 $\mathcal{T}_q(\ell,\delta,1)$ 表示定义 *3* 中的 *Tur´an* 数。因此，$|\mathrm{MHS}(\mathcal{S}_\delta^*)| = \left(q^{\ell-\delta+1}-1\right)/(q-1)$。

证明。首先，我们证明，对于 $\mathcal{S}_\delta^*$的每个击中集 H，都存在一个大小为 $|\mathcal{D}_H| \le |H|$的 $(q,\ell,\delta,1)$-Tur´an 设计 $\mathcal{D}_H$。事实上，令 $\mathcal{D}_H$ 为由 H 中元素生成的所有（互不相同的）一维 $\mathbb{F}_q$-子空间的集合。由于每个 $S^* \in \mathcal{S}_\delta^*$ 至少包含一个元素 $h \in H$，因此 S 包含相应的一维子空间 $\mathrm{span}_{\mathbb{F}_q}(\{h\})$。因此，$\mathcal{D}_H$ 是一个大小为 $|\mathcal{D}_H| \le |H|$的 $(q,\ell,\delta,1)$-Tur´an 设计。

其次，我们证明，如果 $\mathcal{D}$ 是一个 $(q,\ell,\delta,1)$-Tur´an 设计，那么我们可以构造一个大小为 $|H| = |\mathcal{D}|$的 $\mathcal{S}_\delta^*$ 的击中集 H。事实上，令 H 为 $\mathcal{D}$ 中所有一维子空间的生成元集合。那么 $|H| = |\mathcal{D}|$ 和 H 击中了整个集合 $S^* \in \mathcal{S}_\delta^*$。

因此，$|\mathrm{MHS}(\mathcal{S}_\delta^*)| = \mathcal{T}_q(\ell,\delta,1) = \left(q^{\ell-\delta+1}-1\right)/(q-1)$（见 [23, 定理 4.10]）。■

$\left|\mathrm{MHS}\left(\mathcal{C}(S^*)\right)\right|$ 的一个上界可以直接由引理 3 和引理 4 推得。

推论 1. 设 S 是 $\mathbb{F}_{q^\ell}$的一个 δ 维 $\mathbb{F}_q$ 子空间。那么 $\left|\mathrm{MHS}\left(\mathcal{C}(S^*)\right)\right| \le \left(q^{\ell-\delta+1}-1\right)/(q-1)$。

现在我们讨论 $\left|\mathrm{MHS}\left(\mathcal{C}(S^*)\right)\right|$的一个下界。

引理 5。 设 S 是 $\mathbb{F}_{q^\ell}$的一个 δ 维 $\mathbb{F}_q$ 子空间。那么 $\left|\mathrm{MHS}\left(\mathcal{C}(S^*)\right)\right| \ge \left\lceil \frac{q^\ell-1}{q^\delta-1} \right\rceil$。

证明。我们首先注意到，$\mathbb{F}_{q^\ell}^*$ 中的任意两个元素 x 和 y 在 $\mathcal{C}(S^*)$ 中属于相同数量的陪集，因为如果$x = bS^*$，那么 $y = \frac{y}{x}x \in \left(\frac{y}{x}b\right)S^*$，反之亦然。通过对集合 $\left\{(x,\ bS^*)\right.$进行双重计数：$\left.x \in bS^*,\ b \in \mathbb{F}_{q^\ell}^*\right\}$，注意到对于每个 $b \in \mathbb{F}_{q^\ell}^*$，都有 $|bS^*| = |S^*|$，因此 $\mathbb{F}_{q^\ell}^*$中的每个元素恰好出现在 $r \triangleq |\mathcal{C}_{(S^*\,||S^*|)}$ 个陪集中。现在，假设 H是 $\mathcal{C}(S^*)$的一个击中集。如上所证，H 中的元素合计至多能击中 $r|H|$ 个陪集。这意味着 $r|H| \ge |\mathcal{C}(S^*)|$，进而表明 $|H| \ge \left\lceil \frac{q^\ell-1}{q^\delta-1} \right\rceil$，从而证明了该引理。■

所推导出的$\left|\mathrm{MHS}\left(\mathcal{C}(S^*)\right)\right|$的下界和上界，以及由此得出的在 $\mathcal{C}(\boldsymbol{\alpha}^*,\ S^*)$ 中修复组所能容忍的最大故障数量（对于每个 $\boldsymbol{\alpha}^* \in \mathbb{F}_{q^\ell}$），均在定理 1 中给出。值得注意的是，当 $q \to \infty$ 时，这两个界在渐近意义下相匹配。一般而言，已知的是集合族的 MHS 大小只能近似到其底层集合大小的对数比例，即 $\log(q^\ell) = \ell\log\ q$（参见例如 [17], [18]）。对于任意子空间 S，确定 $\mathcal{C}(S^*)$ 的 MHS 的精确大小仍是一个开放问题。

定理 1. 设 S 是 $\mathbb{F}_{q^\ell}$的一个 δ维 $\mathbb{F}_q$子空间。那么

$$\left\lceil \frac{q^\ell-1}{q^\delta-1} \right\rceil \le \left|\mathrm{MHS}\left(\mathcal{C}(S^*)\right)\right| \le \frac{q^{\ell-\delta+1}-1}{q-1}.$$

因此，对于每个 $\boldsymbol{\alpha}^* \in \mathbb{F}_{q^\ell}$,

$$\left\lceil \frac{q^\ell-1}{q^\delta-1} \right\rceil - 1 \le \left|\mathcal{F}\left(\mathcal{C}(\boldsymbol{\alpha}^*,S^*)\right)\right| \le \frac{q^{\ell-\delta+1}-1}{q-1} - 1.$$

证明。该定理直接由推论 1、引理 2 和引理 5 得出。■

当两个界限重合时，我们可以针对一些特殊情形显式确定 $\left|\mathrm{MHS}\left(\mathcal{C}(S^*)\right)\right|$。

推论 2。 设 S 是 $\mathbb{F}_{q^\ell}$ 中子域 $\mathbb{F}_{q^\delta}$ 的一个乘法陪集。那么，$\left|\mathrm{MHS}\left(\mathcal{C}(S^*)\right)\right| = \frac{q^\ell-1}{q^\delta-1}$。

证明。我们在乘法群 $\mathbb{F}_{q^\ell}^*$上定义一个等价关系 $\sim$，如下所示：对于 $a,b \in \mathbb{F}_{q^\ell}^*$，$a \sim b$，当且仅当 $b^{-1}a \in \mathbb{F}_{q^\delta}^*$时成立。该关系将 $\mathbb{F}_{q^\ell}^*$划分为 $\frac{q^\ell-1}{q^\delta-1}$个互不相交的等价类，这些等价类是 $\mathbb{F}_{q^\ell}$中 $\mathbb{F}_{q^\delta}$的乘法陪集。由于对于所有 $b \in \mathbb{F}_{q^\ell}^*$，均有 $\mathcal{C}(S^*) = \mathcal{C}(b\mathbb{F}_{q^\delta}^*) = \mathcal{C}(\mathbb{F}_{q^\delta}^*)$，且各等价类互不相交，因此 $\left|\mathrm{MHS}\left(\mathcal{C}(S^*)\right)\right| = |\mathcal{C}(\mathbb{F}_{q^\delta}^*)| = \frac{q^\ell-1}{q^\delta-1}$。■

推论 3。 假设$(\ell-\delta)$ | 最大公约数(ℓ,δ)，且 S 是 $\mathbb{F}_{q^\ell}$ 的一个维数为 $\delta/(\ell-\delta)$ 的 $\mathbb{F}_{q^{\ell-\delta}}$-子空间。那么 $\left|\mathrm{MHS}\left(\mathcal{C}(S^*)\right)\right| = q^{\ell-\delta}+1$。

证明。令 $q \triangleq q^{\ell-\delta}, \ell \triangleq \ell/(\ell-\delta)$ 和 $\delta \triangleq \delta/(\ell-\delta)$。在定理 1 中用 q、ℓ 和 δ 替换 q、ℓ、δ，左右两边均等于 $q+1$ 或 $q^{\ell-\delta}+1$，注意到 $\ell-\delta=1$。■

示例 1. 考虑定义在 $\mathbb{F}_{2^4}$ 上的 RS(16,2) 码。集合 $S = \{0, z^2, z^7, z^{12}\}$（其中 z 是 $\mathbb{F}_{2^4}$ 的本原元）是子域 $\mathbb{F}_{2^2}$的一个陪集，并具有 $\left|\mathrm{MHS}\left(\mathcal{C}(S^*)\right)\right| = |\mathcal{C}(\mathbb{F}_{2^2})| = 5$。我们可以从 S^* 为 $f(\boldsymbol{\alpha}^*)$ 生成五种修复方案，这些方案能够容忍四个 节点故障：当四个节点发生故障时，五组辅助节点中至少有一组仍然存活。例如，当 $\boldsymbol{\alpha}^* = z^5$时，这五种修复方案分别对应于

$\mathcal{C}\ z^5\ S^* = \{\{z, z^{13}, z^{14}\}, \{z^{11}, z^4, z^7\}, \{z^8, z^6, z^{12}\}, (,)$ 中 $\{z^{10}, 0, 1\},\ \{z^2, z^9, z^3\}\}$　　$S = \{0, z^4, z^5, z^8\}$

的五个求值点集。然而，如果选择，则 $\left|\mathrm{MHS}\left(\mathcal{C}(S^*)\right)\right| = 6$，并且可以生成十五种修复方案，分别对应于 $\mathcal{C}(z^5, S^*)$ 中的十五个求值点集。综合起来，它们能够容忍五个 节点故障。

D. 基于多个种子的紧凑修复组

尽管上一节讨论的单种子紧凑修复组能够渐近地达到最大可容忍故障数的上界，但当 q 较小时，差距仍可能很大。本节讨论多种子方法，并确定一组能够达到定理 1 所述上界的种子。我们仍假设种子是 $\mathbb{F}_{q^\ell}$的子空间。一般（非子空间）种子的情形留待未来研究。

假设我们利用s子空间种子 $\{S_t\}_{t\in[s]}$构造紧凑修复组 $\mathcal{C}(\{S_t^*\}_{t\in[s]})$。如果对所有 $t\in[s]$ 都有 $\dim_{\mathbb{F}_q}(S_t)=\delta$，那么 $\mathcal{C}(\{S_t^*\}_{t\in[s]})\subseteq\mathcal{S}_\delta^*$，定理 1 中的上界仍然成立。也就是说，$\left|\mathrm{MHS}(\mathcal{C}(\{S_t^*\}_{t\in[s]}))\right| \le \left|\mathrm{MHS}(\mathcal{S}_\delta^*)\right|$。因此，令 s 满足 $\cup_{t\in[s]}\mathcal{C}(\{S_t^*\}_{t\in[s]}))=\mathcal{S}_\delta^*$，那么基于这些 s种子的修复组将达到上界$\left|\mathrm{MHS}(\mathcal{S}_\delta^*)\right|$。请注意，这是 s的充分条件，但非必要条件。寻找最小的 s 仍然是一个未解决的问题。**引理 6**。令 $s\triangleq|\mathcal{S}_\delta^*/\mathbb{F}_{q^\ell}^*|$ 为考虑群 $\mathbb{F}_{q^\ell}^*$ 在 $\mathcal{S}_\delta^*$ 上（采用标准域乘法）作用时轨道的数量。令 $\{S_t^*\}_{t\in[s]}\subseteq\mathcal{S}_\delta^*$对应于 s 个互不相交的轨道的集合。那么 $\left|\mathrm{MHS}(\mathcal{C}(\{S_t^*\}_{t\in[s]}))\right|=\left|\mathrm{MHS}(\mathcal{S}_\delta^*)\right|$。

证明。显然，每个 $S^*\in\mathcal{S}_\delta^*$ 的轨道都是 $\{bS^*:\ b\in\mathbb{F}_{q^\ell}^*\}\equiv\mathcal{C}(S^*)$。因此，$\mathcal{S}_\delta^*$ 可以表示为 s 个陪集 $\mathcal{C}(\{S_t^*\}_{t\in[s]})$的不相交并集，这就解释了该引理。■

借助引理 6，我们接下来的任务是明确计算轨道的数量 $|\mathcal{S}_\delta^*/\mathbb{F}_{q^\ell}^*|$。我们通过伯恩赛德引理和莫比乌斯反演公式来完成这一任务。回顾一下，对于 $S^*\in\mathcal{S}_\delta^*$，$\mathbb{F}_{q^\ell}^*$ 中 S^* 的稳定子是集合 $\mathrm{Stab}^{\mathbb{F}_{q^\ell}^*}(S^*)\triangleq\{b\in\mathbb{F}_{q^\ell}^*:\ bS^*=S^*\}$。

引理 7 *[24]* $|\mathcal{S}_\delta^*/\mathbb{F}_{q^\ell}^*|$(伯恩赛德引理)。$=\frac{1}{|\mathbb{F}_{q^\ell}^*|}\sum_{S^*\in\mathcal{S}_\delta^*}|\mathbb{F}_{q^\ell}^*:\mathrm{Stab}(S^*)|$。

我们首先需要确定 $S^*\in\mathcal{S}_\delta^*$ 的 $\mathrm{Stab}^{\mathbb{F}_{q^\ell}^*}(S^*)$。

定义 4。设 S 是 $\mathbb{F}_{q^\ell}$ 的一个 $\mathbb{F}_q$ 子空间，m 是使得 S 为 $\mathbb{F}_{q^\ell}$ 的一个 $\mathbb{F}_{q^m}$ 子空间的最大正整数。我们称 S 具有基 $\mathbb{F}_{q^m}$，并记作 $\mathrm{base}(S)=\mathbb{F}_{q^m}$。

引理 8。设 S 是 $\mathbb{F}_{q^\ell}^*$ 的一个子空间，其基为 $\mathbb{F}_{q^m}$。则 $\mathrm{Stab}^{\mathbb{F}_{q^\ell}^*}(S^*)=\mathbb{F}_{q^m}^*$。因此，$|\mathrm{Stab}^{\mathbb{F}_{q^\ell}^*}(S^*)|=q^m-1$。

证明。假设 S 是 $\mathbb{F}_{q^\ell}^*$ 的一个子空间，其基为 $\mathbb{F}_{q^m}$。由于 S 是一个 $\mathbb{F}_{q^m}$-子空间，因此对所有 $\alpha\in\mathbb{F}_{q^m}^*$，均有 $\alpha S=S$。因此，$\mathbb{F}_{q^m}^*\subseteq\mathrm{Stab}^{\mathbb{F}_{q^\ell}^*}(S^*)$。只需证明反向结论也成立，即若对于 $\alpha\in\mathbb{F}_{q^\ell}$ 有 $\alpha S=S$，则 $\alpha\in\mathbb{F}_{q^m}^*$。事实上，反设对于 $\alpha\notin\mathbb{F}_{q^m}^*$ 成立。令 $\mathbb{F}=\mathbb{F}_{q^m}(\alpha)\supsetneq\mathbb{F}_{q^m}$，这是关于 α 的 $\mathbb{F}_{q^m}$ 的扩域。于是 $\mathbb{F}=\left\{\sum_{i=0}^{\Delta-1}c_i\alpha^i:\ c_i\in\mathbb{F}_{q^m}\right\}$，其中 Δ 是 α 在 $\mathbb{F}_{q^m}$ 上的最小多项式的次数。由于对于每个 $i\ge0$ $u\in S$ $(\sum_{i=0}^{\Delta-1}c_i\alpha^i)u=$ 和，均有 $\sum_{i=0}^{\Delta-1}c_i\,\alpha^iu\in S$ S $\mathbb{F}$ $\mathbb{F}_{q^\ell}$ $\alpha^iu\in S$，我们推得 ( )。因此， 也是 的一个 - 子空间，其中 $\mathbb{F}$ 大于 $\mathbb{F}_{q^m}$，这与 $\mathbb{F}_{q^m}$ 是 S 的基这一假设矛盾。因此，$\alpha\in\mathbb{F}_{q^m}^*$。■

根据引理 8，我们知道，相对于 $\mathbb{F}_{q^\ell}^*$，$S^*\in\mathcal{S}_\delta^*$ 的稳定子群为 $\mathrm{base}(S)$。为了应用伯恩赛德引理，我们还需要针对每个 $m\mid\gcd(\ell,\delta)$，计算具有基 $\mathbb{F}_{q^m}$ 的 $S^*\in\mathcal{S}_\delta^*$ 的数量。注意，q 元高斯系数

$$\begin{bmatrix}\ell\\ \delta\end{bmatrix}_q\triangleq\frac{(q^\ell-1)(q^\ell-q)(q^\ell-q^2)\dots(q^\ell-q^{\delta-1})}{(q^\delta-1)(q^\delta-q)(q^\delta-q^2)\dots(q^\delta-q^{\delta-1})}$$

仅计算了 $\mathbb{F}_{q^\ell}$ 中维数为 δ 的 $\mathbb{F}_q$ 子空间的数量，其中包含基底不同的子空间。例如，对于 $\ell=48$、$\delta=24$，$\gcd(\ell,\delta)=24$，其因数为 2、3、4、6、8、12、24。一个 24 维的 $\mathbb{F}_q$ 子空间可以具有基底 $\mathbb{F}_{q^2}$、$\mathbb{F}_{q^3}$、$\mathbb{F}_{q^4}$、$\mathbb{F}_{q^6}$、$\mathbb{F}_{q^8}$、$\mathbb{F}_{q^{12}}$ 或 $\mathbb{F}_{q^{24}}$。尽管高斯系数不能直接给出 $\mathbb{F}_{q^\ell}$ 中具有基底 $\mathbb{F}_{q^m}$ 的 δ 维 $\mathbb{F}_q$ 子空间的数量，但仍然可以利用莫比乌斯反演公式推导出这一数量。

M¨obius 函数 $\mu(v)$（参见 [25, 第十六章]）定义在正整数上，如下所示：$\mu(v)\triangleq0$ 当 v 不是无平方因子的数时成立，而当 v 是 r 个互异素数的乘积时，$(-1)^r$ 成立。

引理 9 (M¨ Möbius 反演公式)。 *[25,*第 *XVI* 章*]*设f和 g 是定义在 $\mathbb{Z}^+$ 和 $f(n)=\sum_{v|n}g(v)$上的函数。则 $g(n)=\sum_{v|n}\mu(v)f(n/v)$，其中 $\mu(v)$ 是 *Möbius* 函数。

定理 2。设 $N_{q,\ \ell,\ \delta}(m)$ 为以 $\mathbb{F}_{q^m}$ 为基的 $\mathbb{F}_{q^\ell}$ 中 δ 维 $\mathbb{F}_q$ 子空间的数量。设 $(a,b)\triangleq$最大公约数(a,b)。那么

$$N_{q,\ell,\delta}(m)=\sum_{v|\frac{(\ell,\delta)}{m}}\mu(v)\begin{bmatrix}\ell/mv\\ \delta/mv\end{bmatrix}_{q^{mv}}. \tag{1}$$

Using Burnside's lemma, we have

$$|\mathcal{S}_\delta^*/\mathbb{F}_{q^\ell}^*|=\frac{1}{q^\ell-1}\sum_{m|(\ell,\delta)}(q^m-1)\left(\sum_{v|\frac{(\ell,\delta)}{m}}\mu(v)\begin{bmatrix}\ell/mv\\ \delta/mv\end{bmatrix}_{q^{mv}}\right). \tag{2}$$

证明。为了证明 (1)，我们注意到，高斯系数 $\begin{bmatrix}\ell/m\\ \delta/m\end{bmatrix}_{q^m}$ 统计了 $\mathbb{F}_{q^\ell}$ 中所有 $\mathbb{F}_{q^m}$ 维 δ/m 子空间的数量，其中包括所有具有如下基的子空间：$\mathbb{F}_{q^p}$，其中 $p\ge m$、$m\mid p$ 和 $p\mid(\ell,\ \delta)$。因此，

$$\begin{bmatrix}\ell/m\\ \delta/m\end{bmatrix}_{q^m}=\sum_{p\ge m,\ m|p,\ p|(\ell,\delta)}N_{q,\ell,\delta}(p). \tag{3}$$

固定 q,ℓ一个 或每个 m et $n=$, (ℓ,δ)和 $(\ell,\delta)/v=(\ell,\delta)/p$。由于 $m\mid p$，我们还有 $v\mid n$。显然，$m=(\ell,\delta)/n$ 且 $=\ \ell\ \delta\ v$。定义 $p\quad(\ ,\ )/$

$$f(n)\triangleq\begin{bmatrix}\ell/m\\ \delta/m\end{bmatrix}_{q^m}=\begin{bmatrix}\ell n/(\ell,\delta)\\ \delta n/(\ell,\delta)\end{bmatrix}_{q^{\frac{(\ell,\delta)}{n}}},$$

$$g(v)\triangleq N_{q,\ell,\delta}(p)=N_{q,\ell,\delta}\Big(\frac{(\ell,\delta)}{v}\Big).$$

由（3），我们有 $f(n)=\sum_{v|n}g(v)$，应用莫比乌斯反演公式，可得 $g(n)=\sum_{v\ n}\mu(v)f(n/v)$，或

$$\begin{aligned}N_{q,\ell,\delta}(m)=N_{q,\ell,\delta}\left(\frac{(\ell,\delta)}{n}\right)&=\sum_{v|n}\mu(v)\begin{bmatrix}\ell\frac{n}{v}/(\ell,\delta)\\ \delta\frac{n}{v}/(\ell,\delta)\end{bmatrix}_{q^{\frac{(\ell,\delta)}{\frac{n}{v}}}}\\&=\sum_{v|\frac{(\ell,\delta)}{m}}\mu(v)\begin{bmatrix}\ell/mv\\ \delta/mv\end{bmatrix}_{q^{mv}},\end{aligned}$$

这就证明了（1）。

由伯恩赛德引理和引理 8 可得

$$\begin{aligned}|\mathcal{S}_\delta^*/\mathbb{F}_{q^\ell}^*|&=\frac{1}{|\mathbb{F}_{q^\ell}^*|}\sum_{S^*\in\mathcal{S}_\delta^*}|\mathrm{Stab}_{\mathbb{F}_{q^\ell}^*}(S^*)|\\&=\frac{1}{q^\ell-1}\sum_{m|(\ell,\delta)}\ \sum_{S^*\in\mathcal{S}_\delta^*,\ \mathrm{base}(S^*)=\mathbb{F}_{q^m}^*}(q^m-1)\\&=\frac{1}{q^\ell-1}\sum_{m|(\ell,\delta)}(q^m-1)N_{q,\ell,\delta}(m),\end{aligned}$$

这意味着式（2）成立。■

利用定理 2，我们得到了生成一组修复组所需的最少种子数量的一个上界，该组修复组能够达到定理 1 中的上界。确定其精确值仍然是一个开放问题。

推论 4。设定理 *2* 中给出的 $s=|\mathcal{S}_\delta^*/\mathbb{F}_{q^\ell}^*|$ 为 $\mathbb{F}_{q^\ell}^*$ 在 $\mathcal{S}_\delta^*$ 上作用所对应的轨道数，且 $\{S_t^*\}_{t\in[s]}\subseteq\mathcal{S}_\delta^*$ 对应于 s 个互不相交的轨道集合。则 $\left|\mathrm{MHS}(\mathcal{C}(\{S_t^*\}_{t\in[s]}))\right|=\frac{q^{\ell-\delta+1}-1}{q-1}$。使用这些$s$ 种子，所得紧凑修复组可容忍 $\mathcal{F}(\mathcal{C}(\boldsymbol{\alpha}^*,\{S_t^*\}_{t\in[s]}))=\frac{q^{\ell-\delta+1}-1}{q-1}-1$ 次故障。

致谢。本研究得到 ARC DECRA Grant DE180100768 和 ARC DP200100731 的资助。

参考文献

[1] A. Fikes，《Colossus：Google 文件系统的后继者》，https://cloud.google.com/files/storage architecture and challenges.pdf，2010 年。[2] S. Muralidhar、W. Lloyd、S. Roy、C. Hill、E. Lin、W. Liu、S. Pan、S. Shankar、V. Sivakumar、L. Tang 和 S. Kumar，“f4：Facebook 的温数据块存储系统”，载于第 11 届 ACM/USENIX 操作系统设计与实现研讨会（OSDI）论文集，2014 年，第 383–398 页。[3] A. Dimakis、P. Godfrey、Y. Wu、M. Wainwright 和 K. Ramchandran，“分布式存储系统的网络编码”，《IEEE 信息论汇刊》，第 56 卷，第 9 期，第 4539–4551 页，2010 年。[4] Storj 白皮书，https://www.storj.io/whitepaper。[5] Filebase/Sia，https://docs.filebase.com/storage-networks/sia。[6] T. X. Dinh、L. Y. Nhi Nguyen、L. J. Mohan、S. Boztas、T. T. Luong 和 S. H. Dau，“Reed-Solomon 码修复中的实际考量”，载于 2022 年 IEEE 国际信息论研讨会（ISIT），2022 年，第 2607–2612 页。[7] R. Bhagwan、K. Tati、Y.-C. Cheng、S. Savage 和 G. M. Voelker，“全面恢复：支持自动化可用性管理的系统”，载于第 1 届网络系统设计与实现研讨会会议论文集，第 1 卷，NSDI’04 系列，2004 年，第 25 页。[8] M. Silberstein、L. Ganesh、Y. Wang、L. Alvisi 和 M. Dahlin，“延迟即智能：降低纠删码分布式存储中的修复带宽成本”，载于国际系统存储会议（SYSTOR）论文集，2014 年，第 15:1–15:7 页。[9] M. Luby、R. Padovani、T. J. Richardson、L. Minder 和 P. Aggarwal，“ 液态云存储 ”，《 ACM 存储汇刊》，第 15 卷，第 1 期，2019 年 2 月。[10] V. Guruswami 和 M. Wootters，“修复 Reed-Solomon 码”，载于计算理论年度研讨会（STOC）论文集，2016 年。[11] ——，“修复 Reed-Solomon 码”，《IEEE 信息论汇刊》，第 63 卷，第 9 期，第 5684–5698 页，2017 年。[12] H. Dau、I. Duursma、H. M. Kiah 和 O. Milenkovic，“修复具有多个擦除的 Reed-Solomon 码”，《IEEE 信息论汇刊》，第 54 卷，第 10 期，第 6567–6582 页，2018 年。[13] B. Bartan 和 M. Wootters，“标量 MDS 码的多重故障修复”，载于第 55 届 Allerton 通信、控制与计算年度会议（Allerton）论文集，2017 年。[14] J. Mardia、B. Bartan 和 M. Wootters，“标量 MDS 码的多重故障修复”，《IEEE 信息论汇刊》，第 65 卷，第 5 期，第 2661–2672 页，2018 年。[15] W. Li、Z. Wang 和 H. Jafarkhani，“关于 Reed-Solomon 码的子包化大小和修复带宽”，《IEEE 信息论汇刊》，第 65 卷，第 9 期，第 5484–5502 页，2019 年。[16] I. Tamo、M. Ye 和 A. Barg，“Reed-Solomon 码的修复问题：以近乎最优的节点大小实现单重和多重擦除的最优修复”，《IEEE 信息论汇刊》，第 65 卷，第 5 期，第 2673–2695 页，2019 年。[17] C. Lund 和 M. Yannakakis，“关于最小化问题近似求解的困难性”，《ACM 杂志》，第 41 卷，981，1994 年 5 月，第 960 页。[18] U. Feige，“集合覆盖近似求解的 ln n 阈值”，《ACM 杂志》，第 45 卷，第 4 期，第 634–652 页，1998 年。[19] F. J. MacWilliams 和 N. J. A. Sloane，《纠错码理论》。阿姆斯特丹：North-Holland，1977 年。[20] S. H. Dau、T. X. Dinh、H. M. Kiah、T. T. Luong 和 O. Milenkovic，“通过子空间多项式修复 Reed-Solomon 码”，《IEEE 信息论汇刊》，第 67 卷，第 10 期，第 6395–6407 页，2021 年。[21] B. Korte 和 J. Vygen，《组合优化：理论与算法》。Springer，2012 年。[22] A. Berman 、 S. Buzaglo 、 A. Dor 、 Y. Shany 和 I. Tamo，“ 修复在子空间上求值的 Reed-Solomon 码”，载于 IEEE 国际信息论研讨会（ISIT）论文集，2021 年，第 867–871 页。[23] T. Etzion 和 A. Vardy，“Steiner 系统和覆盖设计的 q 类比”，《通信数学进展》，第 5 卷，第 2 期，第 161–176 页，2011 年。[24] J. D. H. Smith，《抽象代数导论》。CRC Press，2009 年。[25] G. H. Hardy 和 E. M. Wright，《数论导论》。牛津大学出版社，1975 年。