私人信息检索系统中存储-检索权衡的香农理论方法

Chao Tian、Hua Sun 和 Jun Chen

arXiv:2301.02155v1 [cs.IT] 5 Jan 2023

摘要

我们采用香农理论方法，研究私人信息检索（PIR）系统中的存储-检索率权衡。我们的研究主要聚焦于经典的双消息双数据库情形，并提出一种基于随机码本生成和分箱技术的编码方案。该编码方案揭示了私人信息检索与经典多描述源编码问题之间隐藏的联系。我们首先证明，在保持检索率最优的情况下，所提出的非线性方案能够优于任何线性方案。此外，在这一极端点与另一个由检索全部策略实现的最优极端点之间进行空间共享之外，还能实现非平凡的存储-检索率权衡。我们进一步证明，借助一种类似于删余技术的方法，可以从随机码中提取出零错误 PIR 码。我们还研究并比较了外界，以确立非线性码优于线性码。

1 引言

私人信息检索（PIR）研究的是这样一种情形：在 N 个数据库中存储 K 条消息，每条消息为 L 位，同时要求向任一数据库（或任一小型数据库子集）请求的消息身份保持私密。早期研究大多属于计算机科学理论领域 [1]，其中 $L = 1$，而核心问题是检索率关于 (K, N) 的标度律。

编码与信息论领域的研究者从多个角度研究了私人信息检索系统中的存储开销，主要采用两种问题表述。Shah 等人 [2] 研究了允许 N 随 L 和 K 变化的情形，并得出了一些确定性结果。类似地，对于 $L = 1$，Fazeli 等人 [3] 提出了一种技术，通过增加 N，将任意线性私人信息检索码转换为存储开销较低的新码。沿此思路的其他重要结果见 [4–9]。

文献 [10], 研究了 PIR 问题的信息论表述，其中允许 L 增加，同时固定 (N, K)。文献 [10], 指出了存储率与检索率之间权衡的重要性质，并提出了一种线性码构造。在这一表述中，即使没有存储开销约束，要刻画私人信息检索系统中的最小检索速率也并非易事，而这一容量问题已在文献 [11]中得到解决。Tajeddine等人 [12] 研究了消息通过最大距离可分（MDS）码在各数据库间编码时的容量问题，该问题后来由 Banawan 和 Ulukus [13]解决。文献 [14, 15]给出了消息大小最优的容量达到型码设计。文献 [16]研究了服务器可以合谋的系统。此外还有各种扩展和推广，近期的综述文章 [17] 全面概述了沿着这一信息论表述开展的研究工作。

在许多已有工作中，存储组件和私人信息检索组件大多是分开设计的，通常通过对其中一个施加某些结构约束来实现；例如，对存储组件提出 MDS 编码要求 [13], ，或采用无编码存储 [18]；此外，几乎所有码构造都是线性的。据我们所知，少数例外包括 [19–21]。本文考虑私人信息检索问题的信息论表述，不对这两个组件施加任何额外的结构约束，并明确研究存储-检索权衡区域。我们主要关注

此处的情形 $N = K = 2$ ，因为它能提供最重要的直觉；我们将其称为 (2，2) 私人信息检索系统。我们的方法自然支持通过线性或非线性 方案对这两个组件进行联合设计。

[19] 中的研究与我们的工作密切相关。该研究考察了在检索率必须达到最优时，单轮和多轮私人信息检索系统中的存储开销。尽管多轮 PIR 与单轮 PIR 的容量相同，但研究表明，在最小检索率下，多轮、ϵ错误的非线性码确实能够突破最优单轮零错误线性码的存储性能壁垒。所有这三项差异是否都是突破这一壁垒所必需的，仍是一个开放问题。

在这项工作中，我们证明，在一系列存储率范围内，非线性码在单轮零错误 (2，2) PIR 系统中的性能优于最优线性码。我们通过提供一种基于随机码本生成和分箱技术的香农理论编码方案来实现这一点。所提出的方案在最小检索率下概念上更为简单，我们将以一个具体示例加以说明。随后，我们给出一般内界，并证明，通过最小检索率码与另一个最优极端点之间的空间共享，可以实现更优的权衡。借助一种类似删余技术的方法，我们进一步证明，可以从随机 ϵ错误 PIR 码中提取出零错误确定性 PIR 码。我们还研究了一般码和线性码的外界，从而确凿地证明了非线性码优于线性码。我们的工作实质上回答了 [19], 中的开放问题，并表明，事实上，只有非线性这一点是突破上述壁垒所必需的。

本工作的初版部分内容最初发表于 [22]。在这篇更新后的文章中，我们提出了一种更通用的随机编码方案，揭示了它与多描述源编码问题 [23] 之间隐含的联系。直观而言，我们可以将检索到的消息视为对完整消息集合的某种部分重构，而非对单条消息的完整重构。因此，来自服务器的回答可以视为对完整消息集合的描述；这些描述要么直接存储在服务器上，要么在收到请求时生成。由此，多描述编码中的技术自然适用于私人信息检索场景。自初版 [22], 发表以来，已有多项后续研究探讨私人信息检索场景中的存储-检索权衡，并给出了更强、更通用的信息论外界以及若干新的线性码构造 [20, 21, 24]。然而，对于 (2，2) 情形， [22] 中给出的香农信息论随机编码方案仍然具有最佳性能，这促使我们在本工作中提出通用编码方案，并更明确地阐述其与多描述源编码之间的联系。我们希望，这种联系能够推动将多描述问题中现有的编码技术应用于私人信息检索问题的研究。

2 预备知识

我们考虑的问题本质上与 [11], 中的问题相同，此外还考虑了数据库的存储开销约束。我们采用更传统的香农理论语言给出形式化的问题定义，以便于后续处理。本节还简要回顾了该问题的一些相关结果。

2.1 问题定义

该系统中有两条独立消息，分别记为 W_1 和 W_2，每条消息均在有限域 $\mathbf{F}_2^L$ 上均匀随机生成，即每条消息都是一个 L 位序列。有两个数据库用于存储这些消息，它们由两个作用于（W_1，W_2）的编码函数生成

$$\phi_n : \mathbb{F}_2^L \times \mathbb{F}_2^L \to \mathbb{F}_2^{\alpha_n}, \quad n = 1, 2,$$

其中， α_n 是数据库-n中的存储符号数量， $n = 1$，2，它是L的确定性函数，即我们对存储采用定长码。我们记为 $S_1 = \phi_1(W_1, W_2)$和 $S_2 = \phi_2(W_1, W_2)$。

当用户请求消息-k时，会在字母表 $\mathcal{Q}\times\mathcal{Q}$中随机生成两个查询（$Q_1^{[k]}$ ， $Q_2^{[k]}$ ），并将其发送给两个数据库。注意，联合分布满足以下条件

$$P_{W_1,W_2,Q_1^{[k]},Q_2^{[k]}}=P_{W_1,W_2}P_{Q_1^{[k]},Q_2^{[k]}},\quad k=1,2, \tag{1}$$

即，消息与查询相互独立。边缘分布 $P_{W_1,\ W_2}$ 和 $P_{Q_1^{[k]}\ ,\ Q_2^{[k]}}$,$k=1$，2 因此完全确定了系统中的随机性。

数据库收到查询后，通过一组确定性函数生成对查询的回答

$$\varphi_n^{(q)}:\mathbb{F}_2^{\alpha_n}\to\mathbb{F}_2^{\beta_n^{(q)}},\quad q\in\mathcal{Q},\ n=1,2. \tag{2}$$

我们也将回答记作 $A_n^{[k]}=\varphi_n^{(Q_n^{[k]})}(S_n)$，$n=1$，2。用户利用检索到的信息，希望通过一组解码函数复现所需的消息

$$\psi^{(k,q_1,q_2)}:\mathbb{F}_2^{\beta_1^{(q_1)}}\times\mathbb{F}_2^{\beta_2^{(q_2)}}\to\mathbb{F}_2^{L}. \tag{3}$$

函数 $\hat{W}_k=\psi^{(k,Q_1^{[k]},Q_2^{[k]})}(A_1^{[k]}\ ,\ A_2^{[k]})$ 的输出本质上就是检索到的消息。我们要求系统正确检索消息（零错误），即， $W_k=W_k$ 对于 $k=1$，2。

或者，我们可以要求系统具有较小的错误概率。将 PIR 码的平均编码错误概率记为

$$P_e=0.5\sum_{k=1,2}P_{W_1,W_2,Q_1^{[k]},Q_2^{[k]}}(W_k\neq\hat{W}_k). \tag{4}$$

一个（L、α_1、α_2、β_1、β_2） ϵ-错误私人信息检索码的定义与（零错误）私人信息检索码类似，不同之处在于，正确性条件被替换为错误概率 $P_e\leq\epsilon$ 的条件。

最后，隐私 约束规定必须满足同分布条件:

$$P_{Q_n^{[1]},A_n^{[1]},S_n}=P_{Q_n^{[2]},A_n^{[2]},S_n},\quad n=1,2. \tag{5}$$

注意，一个显而易见的推论是，对于 $n=1$、2， $P_{Q_n^{[1]}}=P_{Q_n^{[2]}}\triangleq P_{Q_n}$。

我们将由两个概率分布 $P_{Q_1^{[k]}}$、$_{Q_2^{[k]}}$，$k=1$、2 和一组有效编码函数 $\{\phi_n,\varphi_n^{(q)},\psi^{k,q_1,q_2}\}$ 指定、且满足正确性和隐私约束的码称为（L、α_1、α_2、β_1、β_2）私人信息检索码，其中 $\beta_n=\mathrm{E}_{Q_n}[\beta^{(Q_n)_n}\ \ ]$，对于 $n=1$、2。

定义 1. 如果对于任意 $\epsilon>0$ 以及充分大的 L，都存在一个 $(L,\alpha_1,\alpha_2,\beta_1,\beta_2)$ *PIR*码，使得归一化存储-检索速率对 $(\bar{\alpha},\beta)$ 是可达的

$$L(\bar{\alpha}+\epsilon)\geq\frac{1}{2}(\alpha_1+\alpha_2),\ L(\bar{\beta}+\epsilon)\geq\frac{1}{2}(\beta_1+\beta_2). \tag{6}$$

所有可达的归一化存储-检索速率对 $(\bar{\alpha},\beta)$ 构成可达存储-检索速率区域，记为 $\mathcal{R}$。

除非另有明确说明，速率区域 $\mathcal{R}$ 均用于零错误 PIR 设定。在上面的定义中，我们使用了跨数据库的平均速率（$\bar{\alpha},\beta$），而不是各数据库的单独速率向量$\frac{1}{n}$（$\alpha_1,\alpha_2,\mathrm{E}_{Q_1}[\beta^{(Q_1)}{}_1\ \ ],\mathrm{E}_{Q_2}[\beta^{(Q_2)}{}_2\ \ ]$）。这一点可由以下引理说明。

引理 1. 如果存在一个 $(L,\alpha_1,\alpha_2,\beta_1,\beta_2)$ *PIR*码，那么也存在一个 $(2L,\alpha,\alpha,\beta,\beta)$ *PIR*码，其中

$$\alpha=\alpha_1+\alpha_2,\quad \beta=\beta_1+\beta_2. \tag{7}$$

这个引理基本上可以通过空间共享论证来证明，具体细节见 [19]。使用传统的空间共享论证也可以直接得到以下引理。

引理 2。 该区域 $\mathcal{R}$是凸的。

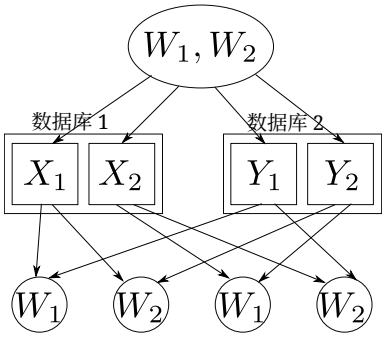

图1：一种可能的编码结构。

2.2 一些相关已知结果

具有 K 条消息和 N 个数据库的一般私人信息检索系统的容量在 [11] 中被确定为

$$C = \frac{1 - 1/N}{1 - 1/N^K}, \tag{8}$$

¯这在我们的定义中对应于 β 最小化的情形，而所提出的线性码可达 ($\bar{\alpha}$， $\bar{\beta}$) =(K，($1 - 1/N^K$)/($N - 1$))。MDS码私人信息检索系统的容量已在 [13] 中得到确立。在存储-检索权衡的背景下，这一结果可视为给出了可达的权衡对

$$(\bar{\alpha}, \bar{\beta}) = \left(t, \frac{1 - t^K/N^K}{N - t} \right), t = 1, 2, \ldots, N. \tag{9}$$

然而，当应用于 (2，2) 私人信息检索问题时，与检索所有内容的平凡码和 [11] 中的码之间的空间共享策略相比，这并未带来任何改进。通过对 [11], 中的码进行专门化， [19] 表明，对于 (2，2) 私人信息检索问题，在检索值 $_\beta = 0.75$ 最小的情况下，使用单轮零错误线性码可达到存储率 $\bar{\alpha}_l = 1.5$ ； 事实上，这是任何单轮零错误线性码所能达到的最优存储率。

[19] 中的一个关键观察是，在（2，2）私人信息检索场景中，一种特殊的编码结构似乎是主要难点，如图 1 所示。消息 W_1 可以通过（X_1， Y_1）或（X_2， Y_2）恢复，消息 W_2 可以通过（X_1， Y_2）或（X_2， Y_1）恢复；（X_1， X_2）本质上是 S_1 ，存储在数据库 1 中；（Y_1， Y_2）本质上是 S_2 ，存储在数据库 2 中。显然，我们可以采用以下策略来满足隐私约束：请求消息 W_1时，用户以概率 1/2查询（X_1，Y_1）或（X_2， Y_2）；对于消息 2，用户以概率 1/2查询（X_1， Y_2）或（X_2， Y_1）。更准确地说，可以采用以下概率分布 $P_{Q_1^{[1]}}$、$_{Q_2^{[1]}}$ 和 $P_{Q_1^{[2]}}$、$_{Q_2^{[2]}}$：

$$P_{Q_1^{[1]}, Q_2^{[1]}} = \begin{cases} 0.5 & (Q_1^{[1]}, Q_2^{[1]}) = (11) \\ 0.5 & (Q_1^{[1]}, Q_2^{[1]}) = (22) \end{cases}, \tag{10}$$

and

$$P_{Q_1^{[2]}, Q_2^{[2]}} = \begin{cases} 0.5 & (Q_1^{[2]}, Q_2^{[2]}) = (12) \\ 0.5 & (Q_1^{[2]}, Q_2^{[2]}) = (21) \end{cases}. \tag{11}$$

2.3 多描述信源编码

多描述源编码问题 [23]考虑将无记忆信源 S 压缩为总共 M 个描述，即 M 个压缩比特序列，使得这些描述中任意子集的组合都可用于重建信源 S，以满足特定的质量要求。这个问题的动机主要是解决通信网络中数据包可能随机丢失的情况。

将每个描述的编码率记为 R_i, $i=1$、2、.....、M。文献 [25], 提出了一种编码方案，可得到以下速率区域。设 U_1、U_2、.....、U_M为与 S联合分布的 M随机变量，则以下速率 (R_1、R_2、.....、R_M) 和失真 ($D_{\mathcal{A}}$、$\mathcal{A}\subseteq\{1,2,\ldots,M\}$) 是可达的:

$$\sum_{i\in\mathcal{A}} R_i \geq \sum_{i\in\mathcal{A}} H(U_i) - H(\{U_i, i\in\mathcal{A}\}|S), \quad \mathcal{A}\subseteq\{1,2,\ldots,M\}, \tag{12}$$

$$D_{\mathcal{A}} \geq \mathbb{E}[d(S, f_{\mathcal{A}}(U_i, i\in\mathcal{A}))], \quad \mathcal{A}\subseteq\{1,2,\ldots,M\}. \tag{13}$$

这里，$f_{\mathcal{A}}$ 是从随机变量 $\{U_i, i\in\mathcal{A}\}$ 到重构域的重构映射，$d(\cdot,\cdot)$ 是用于度量失真的失真度量，而 $D_{\mathcal{A}}$ 是使用集合 $\mathcal{A}$中的描述可达到的失真。粗略地说，该编码方案要求对每个 $i=1$，2，.....，M，使用边缘分布 U_i 以独立同分布的方式生成约 2^{nR_i} 个长度为 n 的码字；速率约束确保当 n 足够大时，以极高的概率存在一组 M个码字 ($u_1^n, u_2^n, ...,u_M^n$)，它们分别来自先前构造的码本之一，并且与源向量 S^n联合典型。在该编码方案中，描述就是这些码字在各自码本中的码字索引。对于给定的联合分布 ($S,U_1,U_2,.....,U_M$)，我们将式 (12) 中的速率区域称为多描述速率区域 $\mathcal{R}_{MD}(S,U_1,U_2,.....,U_M)$，并将相应的随机码构造称为与 ($S,U_1,U_2,.....,U_M$) 相关联的多描述码本。

分箱技术 [26] 可应用于多描述问题，以进一步提升性能；尤其是在并非所有描述组合都需要满足某些性能约束、而只需其中一部分满足时。此前已有研究在 [27] 和 [28]中将此技术用于这一目的。假设只有描述子集 $\mathcal{A}_1$、$\mathcal{A}_2$、.....、$\mathcal{A}_T\subseteq\{1,2,\ldots,M\}$ 对应的重构具有失真要求，使用这些描述进行重构时的失真要求分别记为 $D_{\mathcal{A}_i}$、$i=1$、2、.....、T。考虑与 ($S,U_1,U_2,\ldots,U_M$) 相关联、编码率为 ($R_1', R_2', \ldots, R_M'$) $\in\mathcal{R}_{MD}(S,U_1,U_2,\ldots,U_M)$ 的多描述码本，然后将第 i 个码本中的码字均匀随机地分配到 2^{nR_i} 个分箱中，其中 $0\leq R_i\leq R_i'$。同时满足以下约束的编码率和失真是可达的：对于所有 $\mathcal{A}_i, i=1$、2、.....、T,

$$\sum_{j\in\mathcal{J}}(R_j' - R_j) \leq \sum_{j\in\mathcal{J}} H(U_j) - H\left(\{U_j, j\in\mathcal{J}\}\,\middle|\,\{U_{j'}, j'\in\mathcal{A}_i\setminus\mathcal{J}\}\right), \quad \forall\mathcal{J}\subseteq\mathcal{A}_i, \tag{14}$$

$$D_{\mathcal{A}_i} \geq \mathbb{E}[d(S, f_{\mathcal{A}_i}(U_j, j\in\mathcal{A}_i))]. \tag{15}$$

我们将这类速率向量（R_1、R_2、.....、R_M、R_1'、R_2'、.....、R_M'）的集合记为 $\mathcal{R}_{MD}^*$（（S、U_1、U_2、.....、U_M），（$\{U_j, j\in\mathcal{A}_i\}$、$i=1$、2、.....、$T$））, 并将相应的码本称为与随机变量（$S$、$U_1$、$U_2$、.....、$U_M$）及重构集合（$\mathcal{A}_1$、$\mathcal{A}_2$、.....、$\mathcal{A}_T$）相关联的 MD* 码本。

3 一个特殊情形：用于最小检索速率的 Slepian-Wolf 编码

本节中，我们考虑最小检索速率情形，并说明非线性码和香农信息论编码的优势。这里我们会比较宽松地处理，略去一些细节，以便更好地传达直觉。具体而言，我们暂且忽略随机编码论证通常涉及的渐近趋零错误概率，但会在第 4 节中对此进行更仔细的讨论。

将 L 比特消息改写为

$$W_k = (V_k[1], \ldots, V_k[L]) \triangleq V_k^L, \quad k = 1, 2. \tag{16}$$

这些消息可视为由离散无记忆信源 $P_{V_1,\ V_2} = P_{V_1} \cdot P_{V_2}$ 生成，其中 V_1 和 V_2 是相互独立且服从均匀分布的伯努利随机变量。

考虑以下辅助随机变量

$$\begin{aligned} X_1 &\triangleq V_1 \wedge V_2, \quad X_2 \triangleq (\neg V_1) \wedge (\neg V_2), \\ Y_1 &\triangleq V_1 \wedge (\neg V_2), \quad Y_2 \triangleq (\neg V_1) \wedge V_2, \end{aligned} \tag{17}$$

其中，$\neg$ 表示二元取反，$\wedge$ 表示二元“与”运算。该特定分布满足图 1 所示的编码结构，其中 $(V_1,\ V_2)$ 起到 $(W_1,\ W_2)$ 的作用，且这种关系是非线性的。[19] 中使用了相同的分布来构造多轮私有信息检索码。这种非线性映射似乎能使所得代码比线性码更高效。

我们希望以无损方式将 $(X_1^L,\ X_2^L)$ 存储在第一个数据库中，同时仅存储关于 Y_1^L 和 Y_2^L 的某些必要信息，以便于恢复 W_1 或 W_2。为此，我们将按如下方式对消息进行编码：

- 在数据库 1 处，无损压缩并存储 $(X_1^L,\ X_2^L)$；
- 在数据库 2 处，使用 Slepian-Wolf 码对 Y_1^L进行编码（更准确地说，使用具有不确定性边信息 [29]的 Sgarro 码），解码器处采用 X_1^L或 X_2^L，所得码索引记为C_{Y_1}；以相同方式对 Y_2^L进行编码，且独立于 Y_1^L，其码索引记为C_{Y_2}。

显然，对于数据库 1，我们需要大约 $\bar{\alpha}_1 = H(X_1,\ X_2)$。在数据库 2 处，为了保证 Slepian-Wolf 码能够成功解码，我们可以选择大约

$$\begin{aligned} \bar{\alpha_2} &= \max(H(Y_1|X_1), H(Y_1|X_2)) + \max(H(Y_2|X_1), H(Y_2|X_2)) \\ &= 2H(Y_1|X_1), \end{aligned} \tag{18}$$

其中第二个等号源于概率分布的对称性。因此我们发现，该码达到

$$\begin{aligned} \bar{\alpha}_{nl} &= 0.5[H(X_1, X_2) + 2H(Y_1|X_1)] \\ &= 0.75 + 0.75H(1/3, 2/3) \\ &= 0.25 + 0.75\log_2 3 \approx 1.4387. \end{aligned} \tag{19}$$

根据图 1 中的编码结构，检索策略显而易见，其中 $(V_1^L,\ V_2^L,\ X_1^L,\ X_2^L,\ C_{Y_1},\ C_{Y_2})$ 分别充当 $(W_1,\ W_2,\ X_1,\ X_2,\ Y_1,\ Y_2)$ 的角色，因此确实满足隐私约束。检索速率大致如下

$$\bar{\beta}_1^{(1)} = \bar{\beta}_1^{(2)} = H(X_1) = H(X_2), \tag{20}$$

$$\bar{\beta}_2^{(1)} = \bar{\beta}_2^{(2)} = H(Y_1|X_1), \tag{21}$$

这意味着

$$\bar{\beta} = 0.5[H(X_1) + H(Y_1|X_1)] = 0.5H(Y_1, X_1) = 0.75.$$

因此，在最优检索率 $\beta = 0.75$ 下，我们有

$$\bar{\alpha}_l = 1.5 \text{ vs. } \bar{\alpha}_{nl} \approx 1.4387, \tag{22}$$

显然，所提出的非线性香农理论码优于最优线性码。我们注意到，[19] 中已证明，通过采用多轮方法，可以进一步降低存储率 $\bar{\alpha}$；不过，这一问题超出了本文的研究范围。在本文余下部分，我们将以这一特殊情形中的直觉为基础，对编码方案进行推广和强化。

4 主要结果

4.1 一般内界

我们首先给出存储-检索权衡区域的一般内界。令 $(V_1,\ V_2)$ 为相互独立且在 $\mathrm{F}_2^t \times \mathrm{F}_2^t$ 上均匀分布的随机变量。定义区域 $\mathcal{R}_{in}^{(t)}$ 为所有满足以下条件的 $(\bar{\alpha},\ \bar{\beta})$ 对的集合：存在与 $(V_1,\ V_2)$ 联合分布的随机变量 $(X_0,\ X_1,\ X_2,\ Y_1,\ Y_2)$，使得：

1. 存在确定性函数 $f_{1,\ 1}$、$f_{1,\ 2}$、$f_{2,\ 1}$ 和 $f_{2,\ 2}$，使得

$$V_1 = f_{1,1}(X_0, X_1, Y_1) = f_{2,2}(X_0, X_2, Y_2), \quad V_2 = f_{1,2}(X_0, X_1, Y_2) = f_{2,1}(X_0, X_2, Y_1); \tag{23}$$

2. 存在非负编码率

$$\begin{aligned}&(\beta_1^{(0)}, \beta_1^{(1)}, \beta_1^{(2)}, \beta_2^{(1)}, \beta_2^{(2)}, \gamma_1^{(0)}, \gamma_1^{(1)}, \gamma_1^{(2)}, \gamma_2^{(1)}, \gamma_2^{(2)}) \\ &\quad \in \mathcal{R}_{MD}^{*}\left(((V_1, V_2), X_0, X_1, X_2, Y_1, Y_2), (\{X_0, X_1, Y_1\}, \{X_0, X_1, Y_2\}, \{X_0, X_2, Y_1\}, \{X_0, X_2, Y_2\})\right);\end{aligned} \tag{24}$$

3. 存在非负存储率（$\alpha_1^{(0)}$，$\alpha_1^{(1)}$，$\alpha_1^{(2)}$，$\alpha_2^{(1)}$，$\alpha_2^{(2)}$），使得

$$\alpha_1^{(0)} \le \beta_1^{(0)}, \alpha_1^{(1)} \le \beta_1^{(1)}, \alpha_1^{(2)} \le \beta_1^{(2)}, \alpha_2^{(1)} \le \beta_2^{(1)}, \alpha_2^{(2)} \le \beta_2^{(2)}, \tag{25}$$

且如果

$$\gamma_1^{(0)} - \beta_1^{(0)} + \gamma_1^{(1)} - \beta_1^{(1)} + \gamma_1^{(2)} - \beta_1^{(2)} < H(X_1) + H(X_2) + H(X_3) - H(X_0, X_1, X_2), \tag{26}$$

选择

$$(\alpha_1^{(0)}, \alpha_1^{(1)}, \alpha_1^{(2)}, \gamma_1^{(0)}, \gamma_1^{(1)}, \gamma_1^{(2)}) \in \mathcal{R}_{MD}^{*}\left(((V_1, V_2), X_0, X_1, X_2), (\{X_0, X_1, X_2\})\right); \tag{27}$$

否则，选择 $(\alpha_1^{(0)},\ \alpha_1^{(1)},\ \alpha_1^{(2)}) = (\beta_1^{(0)},\ \beta_1^{(1)},\ \beta_1^{(2)})$。类似地，如果

$$\gamma_2^{(1)} - \beta_2^{(1)} + \gamma_2^{(2)} - \beta_2^{(2)} < I(Y_1; Y_2), \tag{28}$$

选择

$$(\alpha_2^{(1)}, \alpha_2^{(2)}, \gamma_2^{(1)}, \gamma_2^{(2)}) \in \mathcal{R}_{MD}^{*}\left(((V_1, V_2), Y_1, Y_2), (\{Y_1, Y_2\})\right), \tag{29}$$

否则 $(\alpha_2^{(1)},\ \alpha_2^{(2)}) = (\beta_1^{(1)},\ \beta_1^{(2)})$；

4. 归一化平均检索率和存储率

$$2t\bar{\alpha} \ge \alpha_1^{(0)} + \alpha_1^{(1)} + \alpha_1^{(2)} + \alpha_2^{(1)} + \alpha_2^{(2)}, \tag{30}$$

$$4t\bar{\beta} \ge 2\beta_1^{(0)} + \beta_1^{(1)} + \beta_1^{(2)} + \beta_2^{(1)} + \beta_2^{(2)}. \tag{31}$$

于是我们得到以下定理。

定理1。 $\mathcal{R}_{in}^{(t)} \subseteq \mathcal{R}$。

事实上，通过取 $\cup_{t=1}^{\infty}\mathcal{R}_{in}^{(t)}$，我们有可能扩大可达区域。然而，除非 $\mathcal{R}_{in}^{(t+1)} \subseteq \mathcal{R}_{in}^{(t)}$ 对所有 $t \ge 1$均成立，否则区域 $\cup_{t=1}^{\infty}\mathcal{R}_{in}^{(t)}$ 的刻画将更加困难。尽管如此，对于每个固定的t，我们可以通过指定一组可行的随机变量 $X_0,\ X_1,\ X_2,\ Y_1,\ Y_2$来确定内界。

我们不直接证明该定理，而是证明以下定理。该定理证明了存在错误概率渐减的 PIR 码，随后我们将使用删余技术提取一个零错误 PIR 码。

定理 2。 考虑任意 $(\bar{\alpha}, \bar{\beta}) \in \mathcal{R}_{in}^{(t)}$。对于任意 $\epsilon > 0$，且当 L充分大时，存在一个$^{-}$ $^{-}(L, L($ $\bar{\alpha}+\epsilon)$ $L(\bar{\alpha}+\epsilon)$ $L(\beta+\epsilon)$ $L(\beta+\epsilon))$ ϵ-错误 *PIR* 码，其查询分布由式 *(10)* 和 *(11)* 给出。

建立该定理的关键观察是，在此设定中有五个描述；然而，检索和存储对不同的描述组合施加了不同的约束，而且有些描述实际上可以先存储，再重新压缩，最后检索。这种压缩和重新压缩可能带来存储节省。基于 X_0 的描述可视为 X_1 和 X_2的某种共同信息，从而使我们能够权衡存储速率与检索率。

定理 *2* 的证明。码本生成：使用基于分布 $((V_1、\ V_2)\ X_0、\ X_1、\ X_2、\ Y_1、\ Y_2)$ 的多描述码本构建码本。

存储码：码本的分箱索引存储在两台服务器上： $X_0、\ X_1$和 X_2的分箱索引分别以 $\alpha_1^{(0)}$ 、 $\alpha_1^{(1)}$ 和 $\alpha_1^{(2)}$ 的速率存储在服务器-1 上； Y_1 和 Y_2 的分箱索引则分别以 $\alpha_2^{(1)}$ 和 $\alpha_2^{(2)}$ 的速率存储在服务器-2 上。注意，在这些速率下， $X_0、\ X_1$和 X_2 对应的码字能够以极高的概率联合恢复，而 Y_1 和 Y_2 对应的码字也能够以极高的概率联合恢复。

检索码： 在检索过程中，会再次根据 MD* 码本检索出一组不同的码本分箱索引：服务器 1 以速率 $\beta_1^{(0)}$ 、 $\beta_1^{(1)}$ 和 $\beta_1^{(2)}$ 分别检索 $X_0、\ X_1$ 和 X_2 对应的索引；服务器 2 则以速率 $\beta_2^{(1)}$ 和 $\beta_2^{(2)}$ 分别检索 Y_1 和 Y_2 对应的索引。请注意，在这些速率下，可以联合恢复 X_0 、 X_1 和 Y_1 的码字，从而利用这三个相应的码字，以极高概率恢复所需的 V_1 源向量。类似地，$(X_0,\ X_1,\ Y_2) \to V_2$、$(X_0,\ X_2,\ Y_1) \to V_2$ 和 $(X_0,\ X_2,\ Y_2) \to V_2$ 这三种检索模式也都将以极高概率成功。

存储与检索速率： 将参数 t 归一化后，即可直接计算这些速率。

□

接下来我们用它来证明定理1。

定理*1*的证明。 给定一个 $\epsilon > 0$，根据命题 2，对于某个足够大的 L，我们可以找到一个 $(L, L(\bar{\alpha}+\epsilon)$ $L(\bar{\alpha}+_{-}$ $_{-}\epsilon)$ $L(\beta+\epsilon)$ $L(\beta+\epsilon))$ ϵ-错误的私人信息检索码。该码的错误概率可改写为

$$P_e = 0.5 \sum_{k=1,2} \sum_{(w_1,w_2)} 2^{-2L} P_{Q_1^{[k]}, Q_2^{[k]}|(w_1,w_2)}(w_k \neq \hat{W}_k).$$

对于一个固定的 (w_1, w_2) 对，记存在某个 $(q_1, q_2) \in \{(11), (22)\}$ 的事件为，即当 $(Q_1^{[1]}, Q_2^{[1]}) = (q_1, q_2)$ 时，使得 $\hat{}w_1 \neq w_1$ as $E_{w_1,w_2}^{(1)}$，并且存在某个 $(q_1, q_2) \in \{(12),(21)\}$，使得 $\hat{}w_2 \neq w_2$ as$E_{w_1,w_{2\circ}}^{(2)}$ 由于 $(Q_1^{[k]}$，$Q_2^{[k]})$ 与 (W_1, W_2) 相互独立，如果 $P(E_{w_1,w_2}^{(k)}) \neq 0$，我们必有 $P(E_{w_1,w_2}^{(k)}) \geq 0.5$。因此

$$P_e \geq 0.25 \sum_{(w_1,w_2)} 2^{-2L} \mathbf{1}(E_{w_1,w_2}^{[1]} \cup E_{w_1,w_2}^{[2]}), \tag{32}$$

其中 $(\cdot)$ 是示性函数。这意味着对于任意 $\epsilon \leq 0.125$，至多有 2^{2L-1} 对 $(w_1,\ w_2)$ 会导致编码错误。我们可以通过任意但固定的一一映射，使用剩余 2^{2L-1} 对 L 位序列对中的任意 2^{2L-2} 对，来存储一对 $(L-1)$ 位消息。这种新编码的归一化编码速率会增加一个 $1 + 1/(L-1)$ 的因子；当 L 很大时，这种增加可以忽略不计。因此，我们渐近地找到了一个归一化速率与 ϵ-错误编码相同的零错误 PIR 码，证明至此完成。

□

4.2 外界

接下来，我们将关注 $\mathcal{R}$ 的外界，其内容总结于以下定理。

定理 3。 任何 $(\bar{\alpha},\beta) \in \mathcal{R}$ 都必须满足

$$\bar{\beta} \geq 0.75, \quad \bar{\alpha}+\bar{\beta} \geq 2, \quad 3\bar{\alpha}+8\bar{\beta} \geq 10. \tag{33}$$

此外，如果$(\bar{\alpha},\beta) \in \mathcal{R}$ 可以由线性码实现，则它必须满足

$$\bar{\alpha}+6\bar{\beta} \geq 6. \tag{34}$$

不等式 $\beta \geq 0.75$ 由 [11], 得出，而 (33) 中另外两个界已在 [24] 中证明。因此，我们只需证明 (34)。

定理 *3* 的证明。 按照 [19], ，我们作出以下不失一般性的简化假设。定义 $\mathsf{Q} = \{Q_1^{[1]}, Q_1^{[2]}, Q_2^{[1]}, Q_2^{[2]}\}$。

$$1.\ \ Q_1^{[1]} = Q_1^{[2]} \ \Rightarrow\ A_1^{[1]} = A_1^{[2]}, \tag{35}$$

$$2.\ \ H(A_1^{[1]}|\mathbb{Q}) = H(A_2^{[1]}|\mathbb{Q}) = H(A_2^{[2]}|\mathbb{Q}), \quad H(S_1) = H(S_2) \tag{36}$$

$$\Rightarrow\ H(A_1^{[1]}|\mathbb{Q}) \leq \beta \leq (\bar{\beta}+\epsilon)L, \quad H(S_2) \leq \alpha \leq (\bar{\alpha}+\epsilon)L. \tag{37}$$

假设 1 指出，无论目标消息索引为何，向第一个数据库发送的查询都相同。这一假设可由隐私条件得到证明：向一个数据库发送的查询与目标消息索引相互独立。假设 2 指出，经过引理 1 中的对称化操作后，方案是对称的（证明参见 [19] 中的定理 3）。(37) 源于这样一个事实：描述 S_2，$A_1^{[1]}$ 所需的比特数不能少于其熵值，以及定义 1。

下文中，我们用 (c) 表示正确性条件，用 (i) 表示查询与消息相互独立这一约束，用 (a) 表示回答是存储变量与相应查询的确定性函数这一约束，并用 (p) 表示隐私条件。

根据 $A_1^{[1]}$，$A_2^{[1]}$，Q，我们可以解码 W_1。

$$\begin{aligned}
H(A_1^{[1]}, A_2^{[1]}|W_1, \mathbb{Q}) &= H(A_1^{[1]}, A_2^{[1]}, W_1|\mathbb{Q}) - H(W_1|\mathbb{Q}) && (38)\\
&\overset{(c)(i)}{=} H(A_1^{[1]}, A_2^{[1]}|\mathbb{Q}) - L && (39)\\
&\overset{(36)}{\leq} 2H(A_1^{[1]}|\mathbb{Q}) - L. && (40)
\end{aligned}$$

接下来，考虑英格尔顿不等式。

$$\begin{aligned}
I(A_2^{[1]}; A_2^{[2]}|\mathbb{Q}) &\leq I(A_2^{[1]}; A_2^{[2]}|W_1, \mathbb{Q}) + I(A_2^{[1]}; A_2^{[2]}|W_2, \mathbb{Q}) && (41)\\
&= 2I(A_2^{[1]}; A_2^{[2]}|W_1, \mathbb{Q}) && (42)\\
&= 2\big(H(A_2^{[1]}|W_1, \mathbb{Q}) + H(A_2^{[2]}|W_1, \mathbb{Q}) - H(A_2^{[1]}, A_2^{[2]}|W_1, \mathbb{Q})\big) && (43)\\
&\overset{(p)}{=} 2\big(2H(A_2^{[1]}|W_1, \mathbb{Q}) - H(A_2^{[1]}, A_2^{[2]}|W_1, \mathbb{Q})\big) && (44)\\
&\leq 2\big(2H(A_2^{[1]}|W_1, \mathbb{Q}) + H(A_1^{[1]}, A_2^{[1]}|W_1, \mathbb{Q})\\
&\quad - H(A_1^{[1]}, A_2^{[1]}, A_2^{[2]}|W_1, \mathbb{Q}) - H(A_2^{[1]}|W_1, \mathbb{Q})\big) && (45)\\
&\overset{(c)(35)}{=} 2\big(H(A_2^{[1]}|W_1, \mathbb{Q}) + H(A_1^{[1]}, A_2^{[1]}|W_1, \mathbb{Q})\\
&\quad - H(A_1^{[1]}, A_2^{[1]}, A_2^{[2]}, W_2|W_1, \mathbb{Q})\big) && (46)
\end{aligned}$$

$$\overset{(i)}{\leq} \quad 2\big(2H(A_1^{[1]}, A_2^{[1]}|W_1, \mathbb{Q}) - H(W_2)\big) \tag{47}$$

$$\overset{(40)}{\leq} \quad 2\big(2(2H(A_1^{[1]}|\mathbb{Q}) - L) - L\big) \tag{48}$$

其中，(42) 是基于如下观察得出的：将消息索引互换后，可用界定第一项的相同方法来界定第二项。关于 (44) 的更详细推导见 [19] 中的 (79)。(45) 则源于熵的次模性。

注意到

$$I(A_2^{[1]}; A_2^{[2]}|\mathbb{Q}) \quad = \quad H(A_2^{[1]}|\mathbb{Q}) + H(A_2^{[2]}|\mathbb{Q}) - H(A_2^{[1]}, A_2^{[2]}|\mathbb{Q}) \tag{49}$$

$$\overset{(36)}{\geq} \quad 2H(A_1^{[1]}|\mathbb{Q}) - (\bar{\alpha} + \epsilon)L \tag{50}$$

其中在（50）中，而第二项有如下界：

$$H(A_2^{[1]}, A_2^{[2]}|\mathbb{Q}) \leq H(A_2^{[1]}, A_2^{[2]}, S_2|\mathbb{Q}) \overset{(a)}{=} H(S_2|\mathbb{Q}) \overset{(37)}{\leq} (\bar{\alpha} + \epsilon)L. \tag{51}$$

结合（48）和（50），可得

$$2H(A_1^{[1]}|\mathbb{Q})/L - (\bar{\alpha} + \epsilon) \geq 2(4H(A_1^{[1]}|\mathbb{Q})/L - 3)$$

$$\Rightarrow \quad \bar{\alpha} + \epsilon + 6H(A_1^{[1]}|\mathbb{Q})/L \geq 6 \tag{52}$$

$$\overset{(37)}{\Rightarrow} \quad \bar{\alpha} + 6\bar{\beta} \geq 6. \tag{53}$$

证明完毕。 □

4.3 内界的特化

定理1给出的内界具有一般性，但较为复杂，因此我们可以通过多种方式对其进行特化，以简化表达。以下是一种特别有趣的方法。定义区域 $\tilde{\mathcal{R}}_{in}^{(t)}{}^{-}$ 为满足以下条件的 $(\bar{\alpha}, \beta)$ 对的集合：存在随机变量 $(X_0, X_1, X_2, Y_1, Y_2)$，其与 (V_1, V_2) 联合分布，并且满足

1. 分布可因式分解如下

$$P_{V_1,V_2,X_0,X_1,X_2,Y_1,Y_2} = P_{V_1,V_2} P_{X_0|V_1,V_2} P_{X_1|V_1,V_2} P_{X_2|V_1,V_2} P_{Y_1|V_1,V_2} P_{Y_2|V_1,V_2};$$

2. 存在确定性函数 $f_{1,\ 1}$、 $f_{1,\ 2}$、 $f_{2,\ 1}$和 $f_{2,\ 2}$ 使得

$$V_1 = f_{1,1}(X_0, X_1, Y_1) = f_{2,2}(X_0, X_2, Y_2), \tag{54}$$

$$V_2 = f_{1,2}(X_0, X_1, Y_2) = f_{2,1}(X_0, X_2, Y_1); \tag{55}$$

3. 一组速率

$$\gamma_1^{(0)} = I(V_1, V_2; X_0),\ \gamma_1^{(1)} = I(V_1, V_2; X_1),\ \gamma_1^{(2)} = I(V_1, V_2; X_2), \tag{56}$$

$$\gamma_2^{(1)} = I(V_1, V_2; Y_1),\ \gamma_2^{(2)} = I(V_1, V_2; Y_2), \tag{57}$$

$$\beta_1^{(0)} = \gamma_1^{(0)},\ \beta_1^{(1)} = I(V_1, V_2; X_1|X_0),\ \beta_1^{(2)} = I(V_1, V_2; X_2|X_0), \tag{58}$$

$$\beta_2^{(1)} = \max(I(V_1, V_2; Y_1|X_0, X_1), I(V_1, V_2; Y_1|X_0, X_2)), \tag{59}$$

$$\beta_2^{(2)} = \max(I(V_1, V_2; Y_2|X_0, X_1), I(V_1, V_2; Y_2|X_0, X_2)), \tag{60}$$

以及（$\alpha_1^{(0)} = \gamma_1^{(0)}$， $\alpha_1^{(1)}$， $\alpha_1^{(2)}$， $\alpha_2^{(1)}$， $\alpha_2^{(2)}$），如一般区域 $\mathcal{R}^{(t)}$ 的第 3 项中所定义；

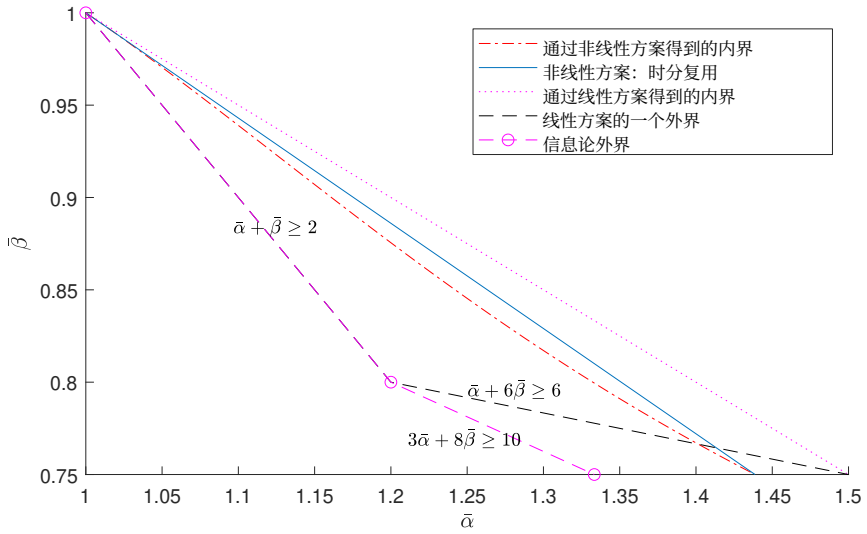

图 2：内界和外界示意图。

4. 归一化平均检索率和存储率

$$2t\bar{\alpha} \geq \alpha_1^{(0)} + \alpha_1^{(1)} + \alpha_1^{(2)} + \alpha_2^{(1)} + \alpha_2^{(2)}, \tag{61}$$

$$4t\bar{\beta} \geq 2\beta_1^{(0)} + \beta_1^{(1)} + \beta_1^{(2)} + \beta_2^{(1)} + \beta_2^{(2)}. \tag{62}$$

于是得到以下推论。

推论 1. $\tilde{\mathcal{R}}_{in}^{(t)} \subseteq \mathcal{R}$.

图 2 将此内界与外界一并展示。

证明。 与定理1的主要区别在于，$(X_0, X_1, X_2, Y_1, Y_2)$ 与 (V_1, V_2) 联合分布时具有特殊的依赖结构，即马尔可夫结构。我们验证这些速率分配满足定理1中的所有约束。由于 $(X_0, X_1, X_2, Y_1, Y_2)$ 与 (V_1, V_2) 联合分布时具有特殊的依赖结构，因此不难验证

$$(\gamma_1^{(0)}, \gamma_1^{(1)}, \gamma_1^{(2)}, \gamma_2^{(1)}, \gamma_2^{(2)}) \in \mathcal{R}_{MD}((V_1, V_2), X_0, X_1, X_2, Y_1, Y_2).$$

接下来，我们验证按上述选择时 (24) 成立。由于结构具有对称性，我们只需确认随机变量的一个子集，即 $\{X_0, X_1, Y_1\}$，另外三个子集 $\{X_0, X_1, Y_2\}$、$\{X_0, X_2, Y_1\}$ 和 $\{X_0, X_2, Y_2\}$ 也可同理验证。与此子集 $\{X_0, X_1, Y_1\}$ 相关、形式为 (14) 的条件总共有 7 个。注意，

$$\gamma_1^{(0)} - \beta_1^{(0)} = 0,\ \gamma_1^{(1)} - \beta_1^{(1)} = I(X_1; X_0),\ \gamma_2^{(2)} - \beta_2^{(2)} \leq I(Y_1; X_0, X_1),$$

事实上，当 $\mathcal{J}$ 为单元素集时，这验证了七个条件中的三个。接下来，当 $\mathcal{J}$ 有两个元素时，我们验证

$$\begin{aligned}
\gamma_1^{(0)} - \beta_1^{(0)} + \gamma_1^{(1)} - \beta_1^{(1)} &= I(X_1; X_0) = H(X_0) + H(X_1) - H(X_0, X_1) \\
&\leq H(X_0) + H(X_1) - H(X_0, X_1|Y_1), \\
\gamma_1^{(0)} - \beta_1^{(0)} + \gamma_2^{(1)} - \beta_2^{(1)} &\leq I(Y_1; X_0, X_1) = H(Y_1) + H(X_0, X_1) - H(X_0, X_1, Y_1) \\
&\leq H(Y_1) + H(X_0) + H(X_1) - H(X_0, X_1, Y_1)
\end{aligned} \tag{63}$$

表 1：推论 2 中使用的条件分布 $P_{X_0|W_1,W_2}$

(w_1,w_2)	$x_0=(00)$	$x_0=(01)$	$x_0=(10)$	$x_0=(11)$
(00)	1/2			1/2
(10)	$(1-p)/2$	p		$(1-p)/2$
(01)	$(1-p)/2$		p	$(1-p)/2$
(11)	1/2			1/2

$$= H(X_0) + H(Y_1) - H(X_0, Y_1|X_1), \tag{64}$$

$$\gamma_1^{(1)} - \beta_1^{(1)} + \gamma_2^{(1)} - \beta_2^{(1)} \leq I(X_1; X_0) + I(Y_1; X_0, X_1) = H(X_1) + H(Y_1) - H(X_1, Y_1|X_0). \tag{65}$$

最后，当 $\mathcal{J}$ 包含全部三个元素时，我们有

$$\begin{aligned}
&\gamma_1^{(0)} - \beta_1^{(0)} + \gamma_1^{(1)} - \beta_1^{(1)} + \gamma_2^{(1)} - \beta_2^{(1)} \\
&= I(X_0; X_1) + I(V_1, V_2; X_1) - \max(I(V_1, V_2; Y_1|X_0, X_1), I(V_1, V_2; Y_1|X_0, X_2)) && (66)\\
&\leq I(X_0; X_1) + I(V_1, V_2; X_1) - I(V_1, V_2; Y_1|X_0, X_1) && (67)\\
&= H(X_0) + H(X_1) + H(Y_1) - H(X_0, X_1, Y_1). && (68)
\end{aligned}$$

因此，取值如 (56)-(60) 时，(24) 确实成立。这实际上完成了证明。 □

我们可以使用任意显式分布 $(X_0,\ X_1,\ X_2,\ Y_1,\ Y_2)$，得到 $\tilde{\mathcal{R}}_{in}^{(t)}$ 的显式内界，而下一个推论给出了一个这样的非平凡内界。为方便起见，我们将概率质量 $(p_1,\ \ldots\ldots,\ p_t)$ 的熵函数记作 $H(p_1,\ \ldots\ldots,\ p_t)$。

推论 2. 对于任意 $p \in [0,1]$，以下 $(\bar{\alpha}, \beta) \in \mathcal{R}$：

$$\begin{aligned}
\bar{\alpha} =& \frac{9}{4} - H(\frac{1}{4}, \frac{3}{4}) + \frac{1}{4}H(\frac{1-p}{2}, \frac{1-p}{2}, \frac{p}{2}, \frac{p}{2}) \\
&+ \frac{1}{2}H(\frac{2-p}{4}, \frac{2-p}{4}, \frac{p}{2}) - \frac{3}{4}H(\frac{3-2p}{6}, \frac{3-2p}{6}, \frac{p}{3}, \frac{p}{3}), \\
\bar{\beta} =& \frac{5}{8} + \frac{1}{4}H(\frac{2-p}{4}, \frac{2-p}{4}, \frac{p}{2}) - \frac{1}{8}H(\frac{1-p}{2}, \frac{1-p}{2}, p).
\end{aligned}$$

证明。这些权衡对通过应用推论 1 并选择 $t = 1$，再按 (17) 设定 $(X_1,\ X_2,\ Y_1,\ Y_2)$ 得到，并令 X_0 如表 1 所定义。注意，该联合分布确实满足所需的马尔可夫结构，在此情形下 $\alpha_2^{(1)} = \beta_2^{(1)}$ 和 $\alpha_2^{(2)} = \beta_2^{(2)}$。

□

5 结论

我们采用香农理论方法研究私人信息检索问题。本文提出一种基于随机编码和分箱的新编码方案，该方案揭示了其与多描述问题之间隐藏的联系。研究表明，对于 (2,2) 私人信息检索设置，这种非线性编码方案能够实现已知的检索率与存储率之间的最佳权衡，且严格优于采用线性码所能达到的结果。我们进一步研究了该设置下零错误 PIR 码与 ϵ-错误 PIR 码之间的关系，并表明二者不会对该问题设置造成实质性差异。我们希望，与多描述编码之间的这种隐藏联系能够为设计更高效的 PIR 码带来新的思路。

参考文献

[1] B. Chor、O. Goldreich、E. Kushilevitz 和 M. Sudan，“私人信息检索”，载于 计算机科学基础，*1995*年。会议论文集，第*36*届年度研讨会，1995年10月，第41–50页。[2] N. Shah、K. Rashmi 和 K. Ramchandran，“多下载一位即可确保完全私密的信息检索”，载于 *2014*年 *IEEE* 国际信息论研讨会（*ISIT*）论文集，2014年6月至7月，第856–860页。[3] A. Fazeli、A. Vardy 和 E. Yaakobi，“低存储开销的分布式私人信息检索码”，载于*2015*年 *IEEE* 国际信息论研讨会（*ISIT*）论文集，2015年6月，第2852–2856页。[4]S. Rao 和 A. Vardy，“私人信息检索码冗余度的下界”，*arXiv*预印本 *arXiv:1605.01869*，2016年。[5] S. R. Blackburn 和 T. Etzion，“具有最优虚拟服务器速率的私人信息检索阵列码”，*IEEE* 信息论汇刊，第65卷，第10期，第6136–6145页，2019年。[6] S. R. Blackburn、T. Etzion 和 M. B. Paterson，“下载复杂度小且存储需求低的私人信息检索方案”，*IEEE* 信息论汇刊，第66卷，第1期，第557–571页，2019年。[7] Y. Zhang、X. Wang、H. Wei 和 G. Ge，“关于私人信息检索阵列码”，*IEEE* 信息论汇刊，第65卷，第9期，第5565–5573页，2019年。[8] M. Vajha、V. Ramkumar 和 P. V. Kumar，“用于编码私人信息检索的二元缩短射影 Reed–Muller 码”，载于 *2017*年 *IEEE* 国际信息论研讨会（*ISIT*），2017 年，第 2648–2652 页。[9] H. Asi 和 E. Yaakobi，“ 近乎最优的私人信息检索码和批处理码构造”，*IEEE* 信息论汇刊，第65卷，第2期，第947–964页，2018年。[10] T. H. Chan、S.-W. Ho 和 H. Yamamoto，“编码存储的私人信息检索”，载于*2015*年 *IEEE* 国际信息论研讨会（*ISIT*）论文集，2015年6月，第2842–2846页。[11] H. Sun 和 S. A. Jafar，“私人信息检索的容量”，*IEEE* 信息论汇刊，第63卷，第7期，第4075–4088页，2017年7月。[12] R. Tajeddine、O. W. Gnilke 和 S. El Rouayheb，“从分布式存储系统中的 MDS 编码数据进行私人信息检索”，*IEEE* 信息论汇刊，第64卷，第11期，第7081–7093页，2018年。[13] K. Banawan 和 S. Ulukus，“从编码数据库进行私人信息检索的容量”，*IEEE* 信息论汇刊，第64卷，第3期，第1945–1956页，2018年3月。[14] C. Tian、H. Sun 和 J. Chen，“具有最优消息大小和上传成本的容量可达私人信息检索码”，*IEEE* 信息论汇刊，第65卷，第11期，第7613–7627页，2019年11月。[15] R. Zhou、C. Tian、H. Sun 和 T. Liu，“基于消息大小最小的 MDS 编码数据库构造容量可达私人信息检索码”，*IEEE* 信息论汇刊，第66卷，第8期，第4904–4916页，2020年。

[16]H. Sun 和 S. A. Jafar，“串谋数据库下鲁棒私人信息检索的容量”，*IEEE* 信息论汇刊，第 64 卷，第 4 期，第 2361–2370 页，2018年。[17] S. Ulukus、S. Avestimehr、M. Gastpar、S. Jafar、R. Tandon 和 C. Tian，“私有检索、计算与学习：近期进展与未来挑战”，*IEEE* 通信领域精选期刊，2022年。[18] M. A. Attia、D. Kumar 和 R. Tandon，“受约束数据库中从无编码存储进行私人信息检索的容量”，*IEEE* 信息论汇刊，第 66 卷，第 11 期，第 6617–6634 页，2020年。[19] H. Sun 和 S. A. Jafar，“多轮私人信息检索：容量与存储开销”，*IEEE* 信息论汇刊，第 64 卷，第 8 期，第 5743–5754 页，2018年。[20] H. Sun 和 C. Tian，“通过联合存储编码突破 MDS-PIR 容量壁垒”，*Information*，第 10 卷，第 9 期，第 265 页，2019年。[21] T. Guo、R. Zhou 和 C. Tian，“私人信息检索系统中存储-检索权衡的新结果”，*IEEE Journal on Selected Areas in Information Theory*，第 2 卷，第 1 期，第 403–414 页，2021年。[22] C. Tian、H. Sun 和 J. Chen，“PIR 系统中存储-检索权衡的香农理论方法”，载于 *2018*年 *IEEE* 国际信息论研讨会 *(ISIT)*，2018年，第 1904–1908 页。[23] A. Gamal 和 T. Cover，“多描述的可达速率”，*IEEE Transactions on Infor-mation Theory*，第 28 卷，第 6 期，第 851–857 页，1982年。[24] C. Tian，“私人信息检索的存储成本”，*IEEE* 信息论汇刊，第 66 卷，第 12 期，第 7539–7549 页，2020年。[25] R. Venkataramani、G. Kramer 和 V. K. Goyal，“多信道下的多描述编码”，*IEEE* 信息论汇刊，第 49 卷，第 9 期，第 2106–2114 页，2003年。[26] A. Wyner 和 J. Ziv，“解码器端具有边信息的信源编码率失真函数”，*IEEE Transactions on information Theory*，第 22 卷，第 1 期，第 1–10 页，1976年。[27] S. S. Pradhan、R. Puri 和 K. Ramchandran，“n 信道对称多描述——第一部分：(n, k) 信源-信道擦除码”，*IEEE* 信息论汇刊，第 50 卷，第 1 期，第 47–61 页，2004年。[28] C. Tian 和 J. Chen，“对称 k描述问题的新编码方案”，*IEEE Trans-actions on Information Theory*，第 56 卷，第 10 期，第 5344–5365 页，2010年。[29] A. Sgarro，“多个解码器端具有边信息的信源编码”，*IEEE Transactions on Infor-mation Theory*，第 23 卷，第 2 期，第 179–182 页，1977年。