arXiv:2302.02230v2 [cs.IT] 5 May 2023

k 服务器拜占庭容错 PIR 方案：最优下载率和最优文件大小

Stanislav Kruglik*、Son Hoang Dau†、Han Mao Kiah*、Huaxiong Wang*
* 新加坡南洋理工大学物理与数学科学学院 † 澳大利亚墨尔本皇家墨尔本理工大学 STE
MCollege 计算技术学院 stanislav.kruglik@ntu.edu.sg, sonhoang.dau@rmit.edu.au,
hmkiah@ntu.edu.sg, hxwang@ntu.edu.sg

摘要——我们研究这样一个问题：如何为存储在 m 个数据库服务器上的 k 个文件设计私有信息检索（PIR）方案，这些服务器可能串通，甚至可能返回错误答案。我们的目标是在确保数据库服务器无法获知所检索消息身份的同时，正确检索特定消息。我们研究该问题的渐近信息论容量，该容量定义为：当存储文件数量足够大时，正确检索的符号数与下载符号数之比的最大值。我们提出了一种文件大小较小的可实现方案，并证明对于固定的检索符号数，这样的文件大小已是最小值，从而解决了 Banawan 和 Ulukus 指出的问题。

I. 引言

私有信息检索（PIR）方案是一种工具，用于从数据库 $\mathbf{x} = (f^{(1)}, \ldots, f^{(m)})$ 中检索给定文件 f_ι，同时对数据库服务器 [1], [2] 隐藏其身份 $\iota \in [m]$。这种设置有许多相关的实际应用，包括保护投资基金所查看的股票市场记录的身份，因为表明对特定记录感兴趣可能会对股价产生负面影响。首个 PIR 方案由 Chor 等人在开创性论文中提出 [3]。对于单服务器情形，作者还指出，为了保证所检索文件索引的信息论隐私，用户必须下载整个数据库。因此，要在信息论设置中降低通信成本，我们必须转向多服务器设置。在该模型中，客户端向每个 k 服务器各发送一次查询，同时确保所检索文件的身份不会被多达 t 个诚实但好奇的服务器获知。在 PIR 文献中，这种方案称为 t-private k-server PIR 方案，这种性质称为 t-privacy。

计算机科学中对 PIR 问题的表述假设文件大小为 1，并通过查询长度之和（上传成本）和响应长度之和（下载成本）来衡量性能 [4]–[6]。出于实际应用的考虑（在这些应用中，消息的大小可以任意大），信息论领域的研究者重新审视了 PIR 问题。下载成本成为主要的性能指标，而最大可实现下载率——定义为检索到的文件大小与用户下载的信息量之比——则成为一系列研究论文的关注焦点 [7]–[10]。

目前大多数PIR方案都假设服务器是诚实但好奇的，并会提供正确的答案。然而，在云环境中无法保证这一假设。这一事实引出了一个有趣的问题：应如何应对服务器给出的错误答案？本文对这一问题给出了三种不同的解释及其形式化定义。

- s-可验证性 *[在 [11]]*中被称为 s-安全性。客户端可以检测出最多有 s 个服务器试图诱使客户端输出错误结果（[11]–[14]）。
- a-可问责性。客户端可以识别至多 a 个诱使客户端输出错误结果的服务器（[15], [16]）。
- b-拜占庭容错/b-拜占庭鲁棒性。即使存在至多 b 个诱使客户端输出错误结果的服务器，客户端仍能检索到正确结果（[17]–[20]）。

显然，a-问责性意味着 a-可验证性，而b-拜占庭容错性则同时意味着 b-问责性和 b-可验证性 [11]。然而，在某些实际应用中，用户需要能够不受服务器恶意行为的影响，正确重建所需消息。这促使我们考虑最强形式的 b-拜占庭容错 PIR [17]–[20]。

b-拜占庭容错 PIR 方案在 $2b + t < k$ 情况下的容量，在 [17] 中被证明等于

$$C_m(t, b, k) = \frac{k - 2b}{k} \cdot \frac{1 - \frac{t}{k-2b}}{1 - \left(\frac{t}{k-2b}\right)^m}. \tag{1}$$

[17] 的作者还提出了一种基于 MDS码的一般可达方案。该方案将每个文件划分为多个子分包，其数量称为子分包化。在 [17], 中，该方案的子分包化值为 $(k - 2b)^m$，而如何求得其达到容量所需的最小数量仍是一个开放问题。值得注意的是，每个子分包通常对应一个有限域元素，而后者的大小会显著影响实现成本 [21]。因此，本文关注文件总大小。由于文件数量很多，我们关注渐近容量值；在 [17] 的方案中，文件大小极其庞大。因此，我们令 $m \to \infty$，对于 $2b + t < k$ 的情况，有

$$C(t, b, k) \triangleq \lim_{m\to\infty} C_m(t, b, k) = \frac{k - 2b}{k} \cdot \left(1 - \frac{t}{k - 2b}\right). \tag{2}$$

针对不同的 PIR 设置，已有大量研究致力于降低子分包层级和文件大小，包括 1-合谋复制PIR [22], 1、 [23]–[25] 合谋的 MDS 编码 PIR，以及 t合谋复制PIR [26]。然而，据我们所知，尚无论文研究拜占庭容错 PIR 中的类似问题。为填补这一空白，本文提出一种非通用的 b-拜占庭容错 k 服务器 PIR 方案，在文件数量渐近增大时具有最优下载率和较小的文件大小

。我们还从形式上证明，后者在所有达到容量的方案中具有最小的文件大小。我们方法的关键要素是近期提出的基于迹恢复框架 [27] 的通信高效秘密共享方案，以及在存在错误迹时修复 Reed-Solomon 码的技术 [28]。

II. 预备知识

A. 符号约定

对于任意整数 $n > 0$，记作 $[n] = \{1, \ldots, n\}$。对于任意素数幂 q，将含有 q^s 个元素的扩展有限域记作 $\mathbb{F}_{q^s}$。含有 q 个元素的基域记作 $\mathbb{F}_q$。对于任意 $\xi \in \mathbb{F}_{q^s}$，定义从 $\mathbb{F}_{q^s}$ 到 $\mathbb{F}_q$ 的迹函数为 $\mathrm{Tr}(\xi) = \sum_{i=0}^{s-1} \xi^{q^i}$。注意，它是 $\mathbb{F}_q$-线性函数。用 $\mathbb{F}_{q^s}[\xi]$ 表示 $\mathbb{F}_{q^s}$ 上的多项式环。上标 T 表示向量的转置。用 $\mathbf{M}_{ij}$ 表示矩阵 $\mathbf{M}$ 的第 (i, j) 个元素。用 $\langle \mathbf{M}, \mathbf{N} \rangle$ 表示 $\mathbf{M}$ 和 $\mathbf{N}$ 的 Frobenius 内积，即 $\langle \mathbf{M}, \mathbf{N} \rangle = \sum_{i,\ j} M_{ij} N_{ij}$。用 $H(X)$ - 表示离散随机变量 X 的熵。

B. k-服务器 PIR 方案

下面正式定义 k-服务器 PIR 方案。设数据库 $\mathbf{x}$ 由 m 个文件 $f^{(1)}, \ldots, f^{(m)}$ 组成，并在每台服务器上复制。用户希望通过向每台服务器发送查询 $\mathbf{q}_1, \ldots, \mathbf{q}_k$ 来获取文件 ι。每台服务器根据收到的查询 $\mathbf{q}_j$，计算答案 $\mathbf{a}_j$ 并将其发回用户。在拜占庭PIR场景中，存在一个用户未知的服务器集合，其中最多有 b 台服务器会对查询提供错误答案。介绍完毕后，我们可以定义 k-服务器、 t-隐私、 b-拜占庭鲁棒 PIR。

定义 1（k-服务器、 t-隐私、 b-拜占庭鲁棒 PIR）。k-服务器、 t-隐私、 b-拜占庭鲁棒 PIR 是一种满足以下性质的方案:

1)（隐私性）该方案具有 t-隐私性，即任意 t个或更少的查询都不会泄露有关文件身份的任何信息。2)（正确性）该方案具有正确性且能够抵御 b个拜占庭故障，即使 b 个答案错误，用户也始终能够根据任意 k 个查询及其对应答案成功解码文件。我们注意到，用户事先并不知道哪些响应是 b 个错误响应。

备注 2. 通过设置 $b = 0$，该定义可简化为 k 服务器 t 私有 PIR 方案。

定义 3（检索阈值）。定义 1 中的 k 服务器 t 私有 b 拜占庭容错 PIR 方案具有检索阈值 r，如果对于任意包含 r 个及以上答案的集合，即使其中有 b 个答案错误，用户也始终能够根据这些答案及相应的查询成功解码文件。与之前一样，我们注意到，用户事先并不知道包含 b 个错误响应的集合。

C. 通信高效的 PIR 方案

我们采用 [27] 中通信高效的秘密共享方案，构造具有最优下载率的 k-服务器 t-隐私 PIR 方案。为简单起见，我们考虑非通用情形，即仅请求恰好 $k \geq r$ 个服务器的响应，其中 r 是恢复阈值，且 $(r - t)$ 可整除 $(k - t)$。与 Reed-Solomon修复问题类似，我们可以通过增加参与的服务器数量 [29]来降低总下载成本。

方案 Π_1： k 服务器 t 私有 PIR

令 t、 r、 k 为满足 $t < r < k \leq q$、$\Delta = r - t$ 和 $\Delta | (k - t)$ 的正整数。记 s、 $\frac{k-t}{\Delta}$。令 $\Omega_\alpha = \{\alpha_1, \ldots, \alpha_\Delta\} \subset \mathbb{F}_{q^s}$、$\Omega_\chi = \{\chi_1, \ldots, \chi_t\} \subset \mathbb{F}_{q^s}$ 和 $\Omega_\beta = \{\beta_1, \ldots, \beta_k\} \subseteq \mathbb{F}_q$ 为公开已知且互不相交的集合，使得 Ω_α 中的所有元素都是 $\mathbb{F}_q$ 上次数为 s 的不同首一不可约多项式的根。

我们将含有 m 个文件的数据库 $\mathbf{x}$ 表示为一个 $m \times \Delta$ 数组。令 $\mathbf{x}$ 的第 (i, j) 个条目为 $x_j^{(i)}$。然后我们设置文件 $f^{(i)}$，$[x_1^{(i)}, \ldots, x_\Delta^{(i)}]$。因此，

$$\mathbf{x} = \begin{bmatrix} x_1^{(1)} & x_2^{(1)} & \cdots & x_\Delta^{(1)} \\ x_1^{(2)} & x_2^{(2)} & \cdots & x_\Delta^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{(m)} & x_2^{(m)} & \cdots & x_\Delta^{(m)} \end{bmatrix} = \begin{bmatrix} f^{(1)} \\ f^{(2)} \\ \vdots \\ f^{(m)} \end{bmatrix}.$$

接着，我们定义 $\mathbf{e}^{(i,j)}$ 为数据库组件的 $m \times \Delta$ 指示数组。换言之，$\mathbf{e}^{(i,j)}$ 的第 (i, j) 个条目为 1，而 $\mathbf{e}^{(i,j)}$ 的所有其他条目均为 0。我们将数据库 $\mathbf{x}$ 复制到 k 个服务器上。

- 查询生成算法：为检索文件 $\iota \in [m]$，用户随机生成 t 个 $(m \times \Delta)$ 数组 $\mathbf{r}^{(1)}$.....、$\mathbf{r}^{(t)}$，并抽取一条随机的 $t + \Delta - 1$ 次曲线

$$\mathbf{g}(\xi) \triangleq \sum_{j=1}^{\Delta} \prod_{\ell \in [\Delta] \setminus \{j\}} \left(\frac{\xi - \alpha_\ell}{\alpha_j - \alpha_\ell} \right) \prod_{\ell=1}^{t} \left(\frac{\xi - \chi_\ell}{\alpha_j - \chi_\ell} \right) \mathbf{e}_{(\iota,j)} + \sum_{h=1}^{t} \prod_{\ell=1}^{\Delta} \left(\frac{\xi - \alpha_\ell}{\chi_h - \alpha_\ell} \right) \prod_{\ell \in [t] \setminus \{h\}} \left(\frac{\xi - \chi_\ell}{\chi_h - \chi_\ell} \right) \mathbf{r}^{(h)} \tag{3}$$

该曲线位于 $\mathbb{F}_{q^s}^{m \times \Delta}$ 中，并经过点 $(\alpha_1, \mathbf{e}^{(\iota,1)})$ .....、$(\alpha_\Delta, \mathbf{e}^{(\iota,\Delta)})$。发送给服务器 $j \in [k]$ 的查询为 $\mathbf{g}(\beta_j)$。我们注意到，$\mathbf{g}$ 和 $\mathbf{r}^{(h)}$ 都依赖于所检索的索引 ι，但为便于阅读，我们省略下标 ι。

- 答案生成算法：收到查询 $\mathbf{g}(\beta_j)$ 后，服务器 $j \in [k]$ 计算 Frobenius 内积 $\langle \mathbf{g}(\beta_j), \mathbf{x} \rangle$。我们可以观察到

$$\langle \mathbf{g}(\xi), \mathbf{x} \rangle = \sum_{j=1}^{\Delta} \prod_{\ell \in [\Delta] \neq \{j\}} \left(\frac{\xi - \alpha_\ell}{\alpha_j - \alpha_\ell} \right) \prod_{\ell=1}^{t} \left(\frac{\xi - \chi_\ell}{\alpha_j - \chi_\ell} \right) x_j^{(\iota)} + \sum_{h=1}^{t} \prod_{\ell=1}^{\Delta} \left(\frac{\xi - \alpha_\ell}{\chi_h - \alpha_\ell} \right) \prod_{\ell \in [t] \setminus \{h\}} \left(\frac{\xi - \chi_\ell}{\chi_h - \chi_\ell} \right) \langle \mathbf{r}^{(h)}, \mathbf{x} \rangle,$$

这是一个关于 ξ 的 $\Delta + t - 1 = r - 1$次多项式。我们将这个多项式称为 $\phi(\xi)$，并进一步观察到 $\phi(\quad \alpha_i = x_i^{(\iota)} \quad i \in [\Delta]\quad)$，其中 .

– 从 r 个服务器的答案中检索时，服务器 j返回 $\mathbf{a}_j = \phi(\beta_j) \in \mathbb{F}_{q^s}$ 的值。– 从 k 个服务器的答案中检索时，服务器 re-

sponds with

$$\mathbf{a}_j = \mathrm{Tr}(v_j \phi(\beta_j)) \in \mathbb{F}_q, \tag{4}$$

其中

$$v_j = \prod_{\ell=1}^{\Delta} (\beta_j - \alpha_\ell)^{-1} \times \prod_{\ell \in [k] \setminus \{j\}} (\beta_j - \beta_\ell)^{-1}. \tag{5}$$

- 文件检索算法
  – 若要从 r 台服务器的答案中检索，用户将应用拉格朗日插值公式。
  – 若要从 k 台服务器的答案中检索，用户将为 $\mathbb{F}_{q^s}$ 在 $\mathbb{F}_q$ 上准备一个基 $\{\theta_1, \ldots, \theta_s\}$ 及其迹正交基 $\{\eta_1, \ldots, \eta_s\}$。之后，用户为所有 $i \in [\Delta]$ 和 $\delta \in [s]$ 选择次数小于 s的多项式 $h_{i\delta} \in \mathbb{F}_q[\xi]$ ，使得

$$h_{i\delta}(\alpha_i) = u_i^{-1} \eta_\delta \prod_{\ell \in [\Delta] \setminus \{i\}} \tilde{f}_\ell^{-1}(\alpha_i), \tag{6}$$

其中， $f_\ell(\xi)$是 α_ℓ 在 $\mathbb{F}_q$上的最小多项式，并且

$$u_i = \prod_{\ell \in [\Delta] \setminus \{i\}} (\alpha_i - \alpha_\ell)^{-1} \times \prod_{j=1}^{k} (\alpha_i - \beta_j)^{-1}. \tag{7}$$

用户通过以下方式检索感兴趣的文件

$$x_i^{(\iota)} = \phi(\alpha_i) = -\sum_{\delta=1}^{s} \theta_\delta \left(\sum_{j=1}^{k} h_{i\delta}(\beta_j) \mathrm{Tr}(v_j \phi(\beta_j)) \cdot \prod_{\ell \in [\Delta] \setminus \{i\}} \tilde{f}_\ell(\beta_j) \right), \tag{8}$$

for all $i \in [\Delta]$.

定理 1。方案 Π_1 是基于 $\mathbb{F}_{q^s}$ 的 k服务器 t隐私 PIR，文件大小为 $(k-t)\log(q)$，恢复阈值为 r，并且对于 $b=0$ 以及任意给定的 Δ 和 r，可达到渐近容量 (2)，使得

$$t < r < k \le q,\ \Delta = r - t \ \textit{and}\ \Delta | (k-t),$$

and $s = \frac{k-t}{\Delta}$.

证明。根据 k服务器 t隐私 PIR 的定义，我们将证明隐私性和正确性，并说明来自 r 台服务器的响应足以完成文件检索。证明与 [27]中的证明非常相似。为使本文内容完整，我们在此详细给出证明。

为证明安全性，我们需要说明

$$I(\mathbf{g}(\beta_{l_1}), \ldots, \mathbf{g}(\beta_{l_t}); \mathbf{e}_{\iota,1}, \ldots, \mathbf{e}_{\iota,\Delta}) = 0, \tag{9}$$

对于任意服务器子集 $\{l_1, \ldots, l_t\} \subset [k]$ 以及任意文件索引 $\iota \in [m]$。

由于矩阵 $\mathbf{g}(\xi)$ 的每个元素都与其他元素分别编码，且对应的随机符号彼此独立，因此，$\mathbf{g}(\xi)$ 的第 (i,j) 个条目仅取决于 $\mathbf{e}^{(i,j)}$，并且在给定条件下与其他所有内容独立。因此，我们的方案等价于通过 $m\Delta$ 个独立信道 [30] 进行传输，因此我们有

$$I(\mathbf{g}(\beta_{l_1}), \ldots, \mathbf{g}(\beta_{l_t}); \mathbf{e}_{\iota,1}, \ldots, \mathbf{e}_{\iota,\Delta}) \le \sum_{i=1}^{m} \sum_{j=1}^{\Delta} I(\mathbf{g}(\beta_{l_1})_{(ij)}, \ldots, \mathbf{g}(\beta_{l_t})_{(ij)}; (\mathbf{e}_{\iota,1})_{(ij)}, \ldots, (\mathbf{e}_{\iota,\Delta})_{(ij)}). \tag{10}$$

容易看出，对于每个i、 j， $\mathbf{g}(\beta_{l_1})_{(ij)}$， ....., $\mathbf{g}(\beta_{l_t})_{(ij)}$是$t$个随机多项式 $\psi_{(ij)}$ 在 $\mathbb{F}_{q^s}$ 上、次数为 $t+\Delta-1$ 的多项式在 l_t 个不同点 β_{l_1}， ....., β_{l_t}处的取值。因此，对于给定的任意值 $(\mathbf{e}_{\iota,1})_{ij}$， ....., $(\mathbf{e}_{\iota,\Delta})_{ij}$，根据拉格朗日插值公式，我们可以得到 $\mathbb{F}_{q^s}$ 上的唯一多项式 $\psi_{(ij)}$，使得 $\psi_{(ij)}(\alpha_1) = (\mathbf{e}_{\iota,1})_{(ij)}$， ....., $\psi_{(ij)}(\alpha_\Delta) = (\mathbf{e}_{\iota,\Delta})_{(ij)}$，并且 $\psi_{(ij)}(\beta_{l_1}) = \mathbf{g}(\beta_{l_1})_{(ij)}$， ....., $\psi_{(ij)}(\beta_{l_t}) = \mathbf{g}(\beta_{l_t})_{(ij)}$。这意味着$I(\mathbf{g}(\beta_{l_1})_{(ij)}, \ldots, \mathbf{g}(\beta_{l_t})_{(ij)}; (\mathbf{e}_{\iota,1})_{(ij)}, \ldots, (\mathbf{e}_{\iota,\Delta})_{(ij)}) = 0$，由 (10) 可知，隐私性质成立。

来自 r 个服务器的响应足以完成文件检索，这一性质可直接由以下事实得出：服务器的响应是多项式 ϕ 在 $\mathbb{F}_{q^s}$ 上的取值，该多项式的次数为 $r-1$，因此 $\phi(\alpha_j) = x_j^{(\iota)}$ （对于 $j \in [\Delta]$）；我们可以使用拉格朗日插值公式来恢复它们。

我们来证明方案 Π_1 的正确性。显然，值 $(\phi(\alpha_1) \ldots, \phi(\alpha_\Delta)\ \phi(\beta_1) \ldots, \phi(\beta_k))$ 可以看作 Reed-Solomon 码的一个码字

$$\mathrm{RS}_r(\Omega_\alpha \cup \Omega_\beta) = \{(\phi(\alpha_1), \ldots, \phi(\alpha_\Delta), \phi(\beta_1), \ldots, \phi(\beta_k)) | \phi \in \mathbb{F}_{q^s}[\xi], \deg(\phi) < r\}. \tag{11}$$

RS_r 的对偶是一个广义里德-所罗门码 [31]，其定义如下

$$\mathrm{GRS}_{k+\Delta-r}(\Omega_\alpha \cup \Omega_\beta) = \{u_1 h(\alpha_1), \ldots, u_\Delta h(\alpha_\Delta), v_1 h(\beta_1), \ldots, v_k h(\beta_k)) | h \in \mathbb{F}_{q^s}[\xi], \deg(h) < k + \Delta - r = k - t\}, \tag{12}$$

其中 $u_i = \prod_{\ell \in [\Delta] \setminus \{i\}} (\alpha_i - \alpha_\ell)^{-1} \times \prod_{j=1}^{k} (\alpha_i - \beta_j)^{-1}$，且 $v_j = \prod_{\ell=1}^{\Delta} (\beta_j - \alpha_\ell)^{-1} \times \prod_{\ell \in [k] \setminus \{j\}} (\beta_j - \beta_\ell)^{-1}$，对于 $i \in [\Delta]$ 和 $j \in [k]$。

由于每个 $\alpha_j, j \in [\Delta]$ 都是不同的首一不可约多项式 f_j 在 $\mathbb{F}_q$ 上的 s 次根，因此我们有

$$\tilde{f}_j(\alpha_j) = 0 \quad \tilde{f}_j(\alpha_i) \neq 0 \quad \text{for } i \in [\Delta], j \neq i \tag{13}$$

$$\prod_{j \in [\Delta], j \neq i} \tilde{f}_j(\alpha_n) = 0 \quad \text{for } n \in [\Delta], n \neq i. \tag{14}$$

令 $\{\theta_1, \ldots, \theta_s\}$ 为 $\mathbb{F}_q$ 上的 $\mathbb{F}_{q^s}$ 基，$\{\eta_1, \ldots, \eta_s\}$ 为其迹正交基。对于每个 $\delta \in [s]$ 和 $i \in [\Delta]$，我们可以将元素 $u_i^{-1} \eta_\delta \prod_{\ell \in [\Delta], \ell \neq i} \tilde{f}_\ell^{-1}(\alpha_i)$ 表示为次数小于 s 的函数 $h_{i\delta}(\xi) \in \mathbb{F}_q[\xi]$ 在点 α_i 处的值。显然，$\deg(h_{i\delta} \prod_{\ell \in [\Delta] \setminus \{i\}} f_\ell) < \Delta s \le k - t$，因此这类函数属于对偶广义 Reed-Solomon 码 (12)。此外，我们有

$$h_{i\delta}(\alpha_i) \prod_{\ell \in [\Delta] \setminus \{i\}} \tilde{f}_\ell(\alpha_i) = u_i^{-1} \eta_\delta \tag{15}$$

and

$$h_{i\delta}(\alpha_n) \prod_{\ell \in [\Delta] \setminus \{i\}} \tilde{f}_\ell(\alpha_n) = 0 \quad \text{for all } n \in [\Delta], n \neq i. \tag{16}$$

因此，

$$\big(u_1 h_{i\delta}(\alpha_1) \prod_{\ell\in[\Delta]\setminus\{i\}} \tilde{f}_\ell(\alpha_1), \ldots, u_\Delta h_{i\delta}(\alpha_\Delta) \prod_{\ell\in[\Delta]\setminus\{i\}} \tilde{f}_\ell(\alpha_\Delta),$$
$$v_1 h_{i\delta}(\beta_1) \prod_{\ell\in[\Delta]\setminus\{i\}} \tilde{f}_\ell(\beta_1), \ldots, v_k h_{i\delta}(\beta_k) \prod_{\ell\in[\Delta]\setminus\{i\}} \tilde{f}_\ell(\beta_k)\big)$$
$$\cdot (\phi(\alpha_1), \ldots, \phi(\alpha_\Delta), \phi(\beta_1), \ldots, \phi(\beta_k))^T = 0. \quad (17)$$

利用函数 $h_{i\delta}(\xi) \in \mathbb{F}_q[\xi]$ 的性质，我们有

$$\eta_\delta \phi(\alpha_i) + v_1 h_{i\delta}\phi(\beta_1) \prod_{\ell\in[\Delta]\setminus\{i\}} \tilde{f}_\ell(\beta_1) + \ldots +$$
$$v_k h_{i\delta}\phi(\beta_k) \prod_{\ell\in[\Delta]\setminus\{i\}} \tilde{f}_\ell(\beta_k) = 0 \quad (18)$$

and

$$\eta_\delta \phi(\alpha_i) = -\sum_{j=1}^{k} \left(v_j h_{i\delta}(\beta_j)\phi(\beta_j) \prod_{\ell\in[\Delta]\setminus\{i\}} \tilde{f}_\ell(\beta_j) \right). \quad (19)$$

对方程-˜（19）两边应用迹映射函数，并利用以下事实： $h_{i\delta}(\xi) \in \mathbb{F}_q[\xi]$, $f_\ell(\xi) \in \mathbb{F}_q[\xi]$ 以及对所有 $i, \ell \in [\Delta]$, $\delta \in [s]$, $j \in [k]$ 均有 $\beta_j \in \mathbb{F}_q$，再结合迹映射函数的线性性，我们得到

$$\mathrm{Tr}(\eta_\delta \phi(\alpha_i)) = -\sum_{j=1}^{k} \mathrm{Tr}\left(v_j \phi(\beta_j) h_{i\delta}(\beta_j) \prod_{\ell\in[\Delta]\setminus\{i\}} \tilde{f}_\ell(\beta_j) \right) =$$
$$-\sum_{j=1}^{k} h_{i\delta}(\beta_j) \left(\mathrm{Tr}\left(v_j \phi(\beta_j)\right) \prod_{\ell\in[\Delta]\setminus\{i\}} \tilde{f}_\ell(\beta_j) \right) \quad (20)$$

由 $\{\theta_1, \ldots, \theta_s\}$ 和 $\{\eta_1, \ldots, \eta_s\}$ 是 $\mathbb{F}_q$ 上 $\mathbb{F}_{q^s}$ 的迹正交基这一事实可知（例如，参见 [32][第 2]章）：

$$x_i^{(\iota)} = \phi(\alpha_i) = \sum_{\delta=1}^{s} \theta_\delta \mathrm{Tr}(\eta_\delta \phi(\alpha_i)) \quad (21)$$

因此，通过从所有相关服务器 $j = 1, \ldots, k$ 获取 $\mathrm{Tr}(v_j\phi(\beta_j))$，即可恢复所有 $\phi(\alpha_1) \ldots, \phi(\alpha_\Delta)$。

注意到每个文件由 Δ 个 $\mathbb{F}_{q^s}$ 元素组成（$s = \frac{k-t}{\Delta}$），且下载率等于 $\frac{k-t}{k}$，证明得证。

□

III. 拜占庭容错 PIR 方案

通过修改第 II-C 节中的构造，我们来构造一个具有 t 个合谋服务器和 b 个拜占庭服务器的 k 服务器 PIR 方案。为简单起见，我们考虑非通用情形：恰好向 $k \geq r$ 个服务器请求响应，其中 r 是恢复阈值，且 $(r - 2b - t)$ 整除 $(k - 2b - t)$。这里，我们采用 [28] 的思想，在 PIR 方案中加入纠错能力。

方案 Π_2: k 服务器、t 私有、可容忍 b 个拜占庭服务器的 PIR 方案

令 t、r、k 为满足 $t < r-2b < k-2b \leq k \leq q$、$\Delta = r-2b-t$ 和 $\Delta|$（$k-2b-t$）的正整数。记 s、$\frac{k-2b-t}{\Delta}$。令 $\Omega_\alpha = \{\alpha_1, \ldots, \alpha_\Delta\} \subset \mathbb{F}_{q^s}$、$\Omega_\chi = \{\chi_1, \ldots, \chi_t\} \subset \mathbb{F}_{q^s}$ 和 $\Omega_\beta = \{\beta_1, \ldots, \beta_k\} \subseteq \mathbb{F}_q$ 为公开已知且互不相交的集合，使得 Ω_α 的所有元素都是 $\mathbb{F}_q$ 上次数为 s 的不同首一不可约多项式的根。

将包含 m 个文件的数据库 $\mathbf{x}$ 表示为一个 $m \times \Delta$ 数组。令 $\mathbf{x}$ 的第 (i, j) 个条目为 $x_j^{(i)}$。然后我们设文件 $f^{(i)}$，$[x_1^{(i)}, \ldots, x_\Delta^{(i)}]$。因此，

$$\mathbf{x} = \begin{bmatrix} x_1^{(1)} & x_2^{(1)} & \cdots & x_\Delta^{(1)} \\ x_1^{(2)} & x_2^{(2)} & \cdots & x_\Delta^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{(m)} & x_2^{(m)} & \cdots & x_\Delta^{(m)} \end{bmatrix} = \begin{bmatrix} f^{(1)} \\ f^{(2)} \\ \vdots \\ f^{(m)} \end{bmatrix}.$$

然后我们定义 $\mathbf{e}^{(i,j)}$ 为数据库组件的指示数组。换言之，$\mathbf{e}^{(i,j)}$的第（i, j）个元素为 1，而 $\mathbf{e}^{(i,j)}$ 的所有其他元素均为 0。我们将数据库 $\mathbf{x}$ 复制到 k 台服务器上。

- 查询生成算法：为了检索文件 $\iota \in [m]$，用户随机生成 t 个（$m \times \Delta$）数组 $\mathbf{r}^{(1)}$.....$\mathbf{r}^{(t)}$，并绘制一条随机的 $t + \Delta - 1$ 次曲线

$$\mathbf{g}(\xi) \triangleq \sum_{j=1}^{\Delta} \prod_{\ell\in[\Delta]\setminus\{j\}} \left(\frac{\xi - \alpha_\ell}{\alpha_j - \alpha_\ell}\right) \prod_{\ell=1}^{t} \left(\frac{\xi - \chi_\ell}{\alpha_j - \chi_\ell}\right) \mathbf{e}_{(\iota,j)}$$
$$+ \sum_{h=1}^{t} \prod_{\ell=1}^{\Delta} \left(\frac{\xi - \alpha_\ell}{\chi_h - \alpha_\ell}\right) \prod_{\ell\in[t]\setminus\{h\}} \left(\frac{\xi - \chi_\ell}{\chi_h - \chi_\ell}\right) \mathbf{r}^{(h)} \quad (22)$$

该曲线位于 $\mathbb{F}_{q^s}^{m\times\Delta}$中，并经过点（$\alpha_1$, $\mathbf{e}^{(\iota,1)}$）.....（α_Δ, $\mathbf{e}^{(\iota,\Delta)}$）。发送给服务器 $j \in [k]$ 的查询为 $\mathbf{g}(\beta_j)$。我们注意到，$\mathbf{g}$ 和 $\mathbf{r}^{(h)}$ 都取决于待检索的索引 ι，但为便于阅读，我们省略下标 ι。

- 答案生成算法：收到查询 $\mathbf{g}(\beta_j)$后，服务器 $j \in [k]$ 计算 Frobenius内积$\langle \mathbf{g}(\beta_j)$ 与 $\mathbf{x}\rangle$。我们可以观察到

$$\langle \mathbf{g}(\xi), \mathbf{x}\rangle = \sum_{j=1}^{\Delta} \prod_{\ell\in[\Delta]\neq\{j\}} \left(\frac{\xi - \alpha_\ell}{\alpha_j - \alpha_\ell}\right) \prod_{\ell=1}^{t} \left(\frac{\xi - \chi_\ell}{\alpha_j - \chi_\ell}\right) x_j^{(\iota)}$$
$$+ \sum_{h=1}^{t} \prod_{\ell=1}^{\Delta} \left(\frac{\xi - \alpha_\ell}{\chi_h - \alpha_\ell}\right) \prod_{\ell\in[t]\setminus\{h\}} \left(\frac{\xi - \chi_\ell}{\chi_h - \chi_\ell}\right) \langle \mathbf{r}^{(h)}, \mathbf{x}\rangle,$$

这是一个关于 ξ的次数为 $\Delta + t - 1 = r - 2b - 1$ 的多项式。我们将该多项式记为 $\phi(\xi)$，并进一步观察到，对于 $i \in [\Delta]$，有 $\phi(\alpha_i) = x_i^{(\iota)}$。

– 从 r 台服务器的答案中检索时，服务器 j以 $\mathbf{a}_j = \phi(\beta_j) \in \mathbb{F}_{q^s}$ 的值作为响应。

– 从 k 台服务器的答案中检索时，服务器的响应为

$$\mathbf{a}_j = \mathrm{Tr}(v_j \phi(\beta_j)) \in \mathbb{F}_q, \quad (23)$$

其中

$$v_j = \prod_{\ell=1}^{\Delta} (\beta_j - \alpha_\ell)^{-1} \times \prod_{\ell\in[k]\setminus\{j\}} (\beta_j - \beta_\ell)^{-1}. \quad (24)$$

- 文件检索算法

–为了从 r 个服务器的答案中检索信息，用户会应用任意 Reed-Solomon 码译码算法

（例如，参见 [31]）。

– 为从 k 台服务器返回的答案中检索信息，用户将向量解码为

$$\left(\prod_{\ell=1}^{\Delta}\tilde{f}_\ell(\beta_1)\mathrm{Tr}(v_1\phi(\beta_1)),\ldots,\prod_{\ell=1}^{\Delta}\tilde{f}_\ell(\beta_k)\mathrm{Tr}(v_k\phi(\beta_k))\right), \tag{25}$$

其中，$\tilde{f}_\ell(\xi)$ 是 α_ℓ 在 $\mathbb{F}_q$ 上的最小多项式，作为 $\mathbb{F}_q$ 上的广义 Reed-Solomon 码字，使用任意译码算法（例如参见 [31]）并提取以下值

$$\mathrm{Tr}(v_j\phi(\beta_j)),\ \text{for}\ j\in[k]. \tag{26}$$

之后，用户为 $\mathbb{F}_{q^s}$ 在 $\mathbb{F}_q$ 上选取一个基 $\{\theta_1,\ldots,\theta_s\}$ 及其迹正交基 $\{\eta_1,\ldots,\eta_s\}$。随后，用户为所有 $i\in[\Delta]$ 和 $\delta\in[s]$ 选择次数小于 s 的多项式 $h_{i\delta}\in\mathbb{F}_q[\xi]$，使得

$$h_{i\delta}(\alpha_i)=u_i^{-1}\eta_\delta\prod_{\ell\in[\Delta]\setminus\{i\}}\tilde{f}_\ell^{-1}(\alpha_i), \tag{27}$$

and

$$u_i=\prod_{\ell\in[\Delta]\setminus\{i\}}(\alpha_i-\alpha_\ell)^{-1}\times\prod_{j=1}^{k}(\alpha_i-\beta_j)^{-1}. \tag{28}$$

用户通过以下方式检索感兴趣的文件

$$x_i^{(\iota)}=\phi(\alpha_i)=-\sum_{\delta=1}^{s}\theta_\delta\left(\sum_{j=1}^{k}h_{i\delta}(\beta_j)\mathrm{Tr}(v_j\phi(\beta_j))\cdot\prod_{\ell\in[\Delta]\setminus\{i\}}\tilde{f}_\ell(\beta_j)\right), \tag{29}$$

for all $i\in[\Delta]$.

定理 2。方案 Π_2 是在 $\mathbb{F}_{q^s}$ 上具有 k 台服务器、t 私有、b 拜占庭容错的 PIR 方案，文件大小为 $(k-2b-t)\log(q)$，恢复阈值为 r，并且对于任意给定的 Δ 和 r，实现渐近容量 (2)，使得

$$t<r-2b<k-2b\le q,\ \Delta=r-2b-t\ \textit{and}\ \Delta|(k-2b-t),\ \textit{and}\ s=\tfrac{k-2b-t}{\Delta}.$$

证明。根据 k-服务器 t-隐私 b-拜占庭容错 PIR 的定义，我们将证明其隐私性和正确性，并说明即使存在至多 b 个错误响应，来自 r 个服务器的响应也足以完成文件检索。隐私性证明与定理 1 的隐私性证明相同，故此处省略。

来自 r 个服务器的响应足以完成文件检索，这一性质源于以下事实：值 $(\phi(\alpha_1)\ldots,\phi(\alpha_\Delta)\ \phi(\beta_1)..,\phi(\beta_k))$ 可视为 Reed-Solomon 码的一个码字

$$\mathrm{RS}_{r-2b}(\Omega_\alpha\cup\Omega_\beta)=\{(\phi(\alpha_1),\ldots,\phi(\alpha_\Delta),\phi(\beta_1),\ldots,\phi(\beta_k))|\phi\in\mathbb{F}_{q^s}[\xi],\deg(\phi)<r-2b\}. \tag{30}$$

因此，通过多项式 ϕ 在任意 r 个点上的值，即使存在 b 个错误值，我们也能正确地对其进行插值，并使用任意 Reed-Solomon 译码算法（例如参见 [31]）。

下面证明方案 Π_2 的正确性。显然，RS_{r-2b} 的对偶是一个广义里德-所罗门码 [31]，其定义如下

$$\mathrm{GRS}_{k+\Delta-(r-2b)}(\Omega_\alpha\cup\Omega_\beta)=\{u_1h(\alpha_1),\ldots,u_\Delta h(\alpha_\Delta),v_1h(\beta_1),\ldots,v_kh(\beta_k))|h\in\mathbb{F}_{q^s}[\xi],\deg(h)<k+\Delta-r+2b=k-t\}, \tag{31}$$

其中 $u_i=\prod_{\ell\in[\Delta]\setminus\{i\}}(\alpha_i-\alpha_\ell)^{-1}\times\prod_{j=1}^{k}(\alpha_i-\beta_j)^{-1}$，且 $v_j=\prod_{\ell=1}^{\Delta}(\beta_j-\alpha_\ell)^{-1}\times\prod_{\ell\in[k]\setminus\{j\}}(\beta_j-\beta_\ell)^{-1}$，对于 $i\in[\Delta]$ 和 $j\in[k]$。

由于每个 $\alpha_j,j\in[\Delta]$ 都是 $\mathbb{F}_q$ 上不同的首一不可约次 f_j 多项式 s 的根，因此我们有

$$\tilde{f}_j(\alpha_j)=0\quad\tilde{f}_j(\alpha_i)\neq0\quad\text{for}\ i\in[\Delta],j\neq i \tag{32}$$

$$\prod_{j\in[\Delta]\setminus\{i\}}\tilde{f}_j(\alpha_n)=0\quad\text{for}\ n\in[\Delta],n\neq i. \tag{33}$$

设 $\{\theta_1,\ldots,\theta_s\}$ 是 $\mathbb{F}_q$ 上的 $\mathbb{F}_{q^s}$ 基，$\{\eta_1,\ldots,\eta_s\}$ 是其迹正交基。对于每个 $\delta\in[s]$ 和 $i\in[\Delta]$，我们可以将元素 $u_i^{-1}\eta_\delta\prod_{\ell\in[\Delta]\setminus\{i\}}\tilde{f}_\ell^{-1}(\alpha_i)$ 表示为次数小于 s 的函数 $h_{i\delta}(\xi)\in\mathbb{F}_q[\xi]$ 在点 α_i 处的取值。显然，$\deg(h_{i\delta}\prod_{\ell\in[\Delta]\setminus\{i\}}f_\ell)<\Delta s<k-t$，因此这类函数属于对偶广义 Reed-Solomon 码（31）。此外，我们有

$$h_{i\delta}(\alpha_i)\prod_{\ell\in[\Delta]\setminus\{i\}}\tilde{f}_\ell(\alpha_i)=u_i^{-1}\eta_\delta \tag{34}$$

and

$$h_{i\delta}(\alpha_n)\prod_{\ell\in[\Delta]\setminus\{i\}}\tilde{f}_\ell(\alpha_n)=0\quad\text{for all}\ n\in[\Delta],n\neq i. \tag{35}$$

因此，

$$\begin{aligned}&\Big(u_1h_{i\delta}(\alpha_1)\prod_{\ell\in[\Delta]\setminus\{i\}}\tilde{f}_\ell(\alpha_1),\ldots,u_\Delta h_{i\delta}(\alpha_\Delta)\prod_{\ell\in[\Delta]\setminus\{i\}}\tilde{f}_\ell(\alpha_\Delta),\\&v_1h_{i\delta}(\beta_1)\prod_{\ell\in[\Delta]\setminus\{i\}}\tilde{f}_\ell(\beta_1),\ldots,v_kh_{i\delta}(\beta_k)\prod_{\ell\in[\Delta]\setminus\{i\}}\tilde{f}_\ell(\beta_k)\Big)\\&\cdot(\phi(\alpha_1),\ldots,\phi(\alpha_\Delta),\phi(\beta_1),\ldots,\phi(\beta_k))^T=0.\end{aligned} \tag{36}$$

利用函数 $h_{i\delta}(\xi)\in\mathbb{F}_q[\xi]$ 的性质，我们可以写出

$$\begin{aligned}&\eta_\delta\phi(\alpha_i)+v_1h_{i\delta}\phi(\beta_1)\prod_{\ell\in[\Delta]\setminus\{i\}}\tilde{f}_\ell(\beta_1)+\ldots+\\&v_kh_{i\delta}\phi(\beta_k)\prod_{\ell\in[\Delta]\setminus\{i\}}\tilde{f}_\ell(\beta_k)=0\end{aligned} \tag{37}$$

and

$$\eta_\delta\phi(\alpha_i)=-\sum_{j=1}^{k}\left(v_jh_{i\delta}(\beta_j)\phi(\beta_j)\prod_{\ell\in[\Delta]\setminus\{i\}}\tilde{f}_\ell(\beta_j)\right). \tag{38}$$

对等式（38）两边应用迹映射函数，并利用以下事实：$h_{i\delta}(\xi)\in\mathbb{F}_q[\xi]$，$f_\ell(\xi)\in\mathbb{F}_q[\xi]$ 以及 $\beta_j\in\mathbb{F}_q$ 对所有 $i,\ell\in[\Delta],\delta\in[s],j\in[k]$ 均成立，

结合迹映射函数的线性性，我们得到

$$\mathrm{Tr}(\eta_\delta\phi(\alpha_i)) = -\sum_{j=1}^{k}\mathrm{Tr}\left(v_j\phi(\beta_j)h_{i\delta}(\beta_j)\prod_{\ell\in[\Delta]\setminus\{i\}}\tilde{f}_\ell(\beta_j)\right) = -\sum_{j=1}^{k}h_{i\delta}(\beta_j)\left(\mathrm{Tr}\left(v_j\phi(\beta_j)\right)\prod_{\ell\in[\Delta]\setminus\{i\}}\tilde{f}_\ell(\beta_j)\right) \quad (39)$$

由 $\{\theta_1,\ldots,\theta_s\}$和 $\{\eta_1,\ldots,\eta_s\}$是 $\mathbb{F}_q$ 上的 $\mathbb{F}_{q^s}$ 迹正交基这一事实，显然可知（例如，参见[32][第 2]章）：

$$x_i^{(\iota)} = \phi(\alpha_i) = \sum_{\delta=1}^{s}\theta_\delta\mathrm{Tr}(\eta_\delta\phi(\alpha_i)) \quad (40)$$

因此，通过从所有相关服务器 $j=1,\ldots,k$获取 $\mathrm{Tr}(v_j\phi(\beta_j))$，即可恢复所有的 $\phi(\alpha_1)\ldots,\phi(\alpha_\Delta)$。

下面证明，即使 $\mathrm{Tr}(v_j\phi(\beta_j))$ 中至多有 b 个错误值，我们仍能正确恢复 $\phi(\alpha_1)\ldots,\phi(\alpha_\Delta)$。依照 [28], 中的思路，我们将上述推导中的函数 $h_{i\delta}$替换为函数

$$\tilde{h}_e(\xi) = \xi^e\prod_{\ell\in[\Delta]}\tilde{f}_\ell(\xi)\in\mathbb{F}_q[\xi]. \quad (41)$$

˜显然，对所有 $i\in[\Delta]$， $h_e(\alpha_i)=0$，且 $\deg(h_e(\xi)) < \Delta s+e$。因此，对所有 $e<2b$，都有 $\deg(h_e(\xi)) < k-t$；由此，这些函数属于对偶广义 Reed-Solomon 码 (31)。因此，

$$\begin{aligned}&(u_1\tilde{h}_e(\alpha_1),\ldots,u_\Delta\tilde{h}_e(\alpha_\Delta),v_1\tilde{h}_e(\beta_1),\ldots,v_k\tilde{h}_e(\beta_k))\cdot\\&(\phi(\alpha_1),\ldots,\phi(\alpha_\Delta),\phi(\beta_1),\ldots,\phi(\beta_k))^T=\\&v_1\tilde{h}_e(\beta_1)\phi(\beta_1)+\ldots+v_k\tilde{h}_e(\beta_k)\phi(\beta_k)=0.\end{aligned} \quad (42)$$

˜由于 $h_e(\xi)\in\mathbb{F}_q[\xi]$和 $\{\beta_1,\ldots,\beta_k\}\subseteq\mathbb{F}_q$，对等式 (42) 两边应用迹映射函数，并利用其线性性，我们得到

$$\begin{aligned}&\beta_1^e\prod_{\ell\in[\Delta]}\tilde{f}_\ell(\beta_1)\mathrm{Tr}(v_1\phi(\beta_1))+\ldots\\&+\beta_k^e\prod_{\ell\in[\Delta]}\tilde{f}_\ell(\beta_k)\mathrm{Tr}(v_k\phi(\beta_k))=0,\end{aligned} \quad (43)$$

˜其中 $e=0$、1、.....、$2b-1$，以及 β_j、$\prod_{\ell\in[\Delta]}f_\ell(\beta_k)$、$\mathrm{Tr}(v_j\phi(\beta_j))$ 对所有 $j\in[k]$均属于 $\mathbb{F}_q$。因此，元素 ˜$\prod_{\ell\in[\Delta]}f_\ell(\beta_k)\mathrm{Tr}(v_k\phi(\beta_k))$ 构成了长度为 k、维数为 $k-2b$ 的 $\mathbb{F}_q$ 上的广义 Reed-Solomon 码的一个码字。该码最多可以纠正 b 个错误，用户可以从服务器响应中获取所有 $j\in[k]$ [31] 对应的 $\mathrm{Tr}(v_jf(\beta_j))$ 的正确值。

注意到每个文件由 Δ 个 $\mathbb{F}_{q^s}$ 中的元素组成（$s=\frac{k-t-2b}{\Delta}$），且下载率等于 $\frac{k-t-2b}{k}$，即可完成证明。

□

IV. 文件大小的下界

对于任意给定大小为 $q\geq k$ 的基域 $\mathbb{F}_q$，文件 Δs 的大小（以比特为单位）为 $\log_2 q$，取决于参数 Δs。本节将沿用 [27，第 V 节]，中的推导，介绍适用于文件数渐近趋于无穷大的拜占庭容错私有信息检索方案类别，然后证明我们方案中文件的大小是最优的。

定义 4（平衡型拜占庭容错 PIR）。若客户端从每个参与的服务器下载一个相同子域 $\mathbb{F}_{q^R}$ 中的元素，则该拜占庭容错 PIR 方案是平衡的。

定义 5（速率最优拜占庭容错 PIR）。对于任意 i 个文件 $f^{(i)}H(f^{(i)})=(r-2b-t)s=s\Delta$，若拜占庭容错 PIR 方案的速率最优，则称该方案为速率最优。

在介绍了必要的定义之后，我们可以给出本节的主要定理。

定理 3. 对于一个平衡的速率最优拜占庭容错 PIR 方案，若其在文件数量足够大时达到最小下载率，则以下结论成立：

$$(r-2b-t)|(k-2b-t) \quad (44)$$

$$s\Delta\geq k-2b-t \quad (45)$$

证明。由于所考虑的方案是速率最优的，$\Delta=r-2b-t$。考虑客户端从 k 个服务器中的每个服务器下载 $\mathbb{F}_{q^R}$ 的一个元素，且该方案达到渐近容量的情形。由此可得

$$k\log_q q^R = \frac{k}{k-2b-t}\Delta s. \quad (46)$$

因此 $\Delta s = R(k-2b-t)$。由于 s 是正整数，且 $\mathbb{F}_{q^R}$ 是 $\mathbb{F}_{q^s}$的子域，因此 R 整除 s，从而得出定理结论。□

推论 4. 方案 Π_2 是一种均衡的、速率最优且抗拜占庭的 PIR 方案，具有最优文件大小。

V. 比较

本节将 PIR 方案 Π_1和 Π_2 与文献 [33]中的 Staircase-PIR 进行比较。我们将后者记为方案 $\mathbb{A}_1$，并通过以下定理给出其参数。

定理 5. 方案 $\mathbb{A}_1$ 是基于 $\mathbb{F}_q$ 的 k服务器 t私有 PIR 方案，文件大小为 $(r-t)(k-t)\log(q)$，恢复阈值为 r；对于 $b=0$ 以及任意给定的 r，该方案可实现渐近容量 (2)，满足

$$t<r<k\leq q.$$

方案 $\mathbb{A}_1$ 假设服务器是诚实但好奇的，并会提供正确答案。为了实现 b-拜占庭抗性，我们可以像文献 [34]中那样，利用带下划线的阶梯码的纠错能力。所得方案 $\mathbb{A}_2$ 的参数将在以下定理中给出。

定理 6. 方案 $\mathbb{A}_2$ 是在 $\mathbb{F}_q$上实现的 k 服务器、 t 隐私、 b 拜占庭容错 PIR，文件大小为 $(r-2b-t)(k-2b-t)\log(q)$，恢复阈值为 r，可针对任意给定的 r 达到渐近容量 (2)，使得

$$t<r-2b<k-2b<k\leq q.$$

为了说明忽略上传成本是合理的，我们将每个方案重复l次，并在表 1 中总结参数。我们注意到

表 1： 具有最优下载率的 PIR 方案的参数。所有参数均以比特为单位 .

	Π_1	Π_2	$\mathbb{A}_1$	$\mathbb{A}_2$
文件大小	$l(k-t)\log(q)$	$l(k-2b-t)\log(q)$	$l(r-t)(k-t)\log(q)$	$l(r-2b-t)(k-2b-t)\log(q)$
域	$\mathbb{F}_{q^s}$, 其中 $s=\frac{k-t}{r-t}$	$\mathbb{F}_{q^s}$, 其中 $s=\frac{k-2b-t}{r-2b-t}$	$\mathbb{F}_q$	$\mathbb{F}_q$
下载成本	$lk\log(q)$	$lk\log(q)$	$lk(r-t)\log(q)$	$lk(r-2b-t)\log(q)$
下载率	$1-\frac{t}{k}$	$1-\frac{2b+t}{k}$	$1-\frac{t}{k}$	$1-\frac{2b+t}{k}$
容量	$1-\frac{t}{k}$	$1-\frac{2b+t}{k}$	$1-\frac{t}{k}$	$1-\frac{2b+t}{k}$
拜占庭容错	0	$b\neq 0$	0	$b\neq 0$

我们的方案基于扩域运行，而方案 $\mathcal{A}_1$和 $\mathbb{A}_2$ 基于基域运行。尽管如此，在 $\mathbb{A}_1$ 和 $\mathbb{A}_2$中，文件的每个分量都由多个域符号组成，导致文件大小和检索延迟较大。

第六节：结论

我们研究了如何设计一种能够抵御服务器对抗行为的私有信息检索方案。我们聚焦于下载成本最小化，并针对文件数量趋于无穷大且文件大小固定的情形，提出了一种文件大小较小的非通用容量达成方案。我们还正式证明了此类文件大小是最优的，从而解决了 Banawan 和 Ulukus 在[17]中提出的问题。将所提出的框架扩展到通用情形和有限文件数量，是值得研究的开放问题。

致谢。

本研究/项目获新加坡国家研究基金会战略能力研究中心资助计划、新加坡教育部学术研究基金第二级资助（MOE2019-T2-2-083 和 MOE-T2EP20121-0007），以及 ARC 资助项目（DE180100768 和 DP200100731）的支持。本文所表达的任何观点、研究结果、结论或建议均为作者所有，不代表新加坡国家研究基金会的观点。

参考文献

[1] S. Ulukus、S. Avestimehr、M. Gastpar、S. A. 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，第40卷，第3期，第729–748页，2022年。[2] R. Ostrovsky 和 W. E. Skeith，“单数据库私有信息检索综述：技术与应用”，2007年，第393–411页。[3] B. Chor、O. Goldreich、E. Kushilevitz 和 M. Sudan，“私有信息检索”，载于IEEE第36届计算机科学基础年会，1995年，第41–50页。[4] D. Woodruff 和 S. Yekhanin，“信息论私有信息检索的几何方法”，《计算理论杂志》，第37卷，第4期，第1046–1056页，2007年。[5] S. Yekhanin，“迈向次指数长度的3查询局部可解码码”，ACM期刊，第55卷，第1期，2008年。[6] Z. Dvir 和 S. Gopi，“采用次多项式通信的双服务器私有信息检索”，ACM期刊，第63卷，第4期，2016年。

[7] N. B. Shah、K. V. 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》，2014年，第856–860页。[8] T. H. Chan、S.-W. Ho 和 H. Yamamoto，《面向编码存储的私有信息检索》，载于《2015年IEEE国际信息论研讨会（ISIT）》，2015年，第2842–2846页。[9] R. Tajeddine、O. W. Gnilke 和 S. El Rouayheb，《从分布式存储系统中的MDS编码数据进行私有信息检索》，《IEEE信息论汇刊》，第64卷，第11期，第7081–7093页，2018年。[10] H. Sun 和 S. A. Jafar，《串谋数据库下鲁棒私有信息检索的容量》，《IEEE信息论汇刊》，第64卷，第4期，第2361–2370页，2018年。[11] L. F. Zhang 和 H. Wang，《低次多项式的多服务器可验证计算》，载于《2022年IEEE安全与隐私研讨会（SP）》，2022年，第596–613页。[12] P. Ke 和 L. F. Zhang，《具有结果验证的双服务器私有信息检索》，载于《2022年IEEE国际信息论研讨会（ISIT）》，2022年，第408–413页。[13] S. Kruglik、S. H. Dau、H. M. Kiah 和 H. Wang，《具有结果验证的双服务器私有信息检索》，载于《2023年IEEE国际信息论研讨会（ISIT）》，2023年。[14] Q. Cao、H. Y. Tran、S. H. Dau、X. Yi、E. Viterbo、C. Feng、Y.-C. Huang、J. Zhu、S. Kruglik 和 H. M. Kiah，《承诺式私有信息检索》，2023年。 [在线]。获取地址：https://arxiv.org/abs/2302.01733。[15] L. F. Zhang 和 R. Safavi-Naini，《可验证多服务器私有信息检索》，载于《应用密码学与网络安全》，2014年。[16] L. Zhao、X. Wang 和 X. Huang，《基于带二元误差的带误差学习的可验证单服务器私有信息检索》，《信息科学》，第546卷，第897–923页，2021年。[17] K. Banawan 和 S. Ulukus，《从拜占庭式串谋数据库进行私有信息检索的容量》，《IEEE信息论汇刊》，第65卷，第2期，第1206–1219页，2019年。[18] K. Kurosawa，《如何纠正多服务器PIR中的错误》，载于《密码学进展——ASIACRYPT 2019》，2019年，第564–574页。[19] K. Devet、I. Goldberg 和 N. Heninger，《最优鲁棒私有信息检索》，载于《第21届USENIX安全研讨会》，2012年，第269–283页。

[20] A. Beimel，《鲁棒私有信息检索》，载于《密码学、安全与隐私百科全书》，S. Jajodia、P. Samarati 和 M. Yung 编，施普林格·柏林·海德堡，2019年，第1–3页。[21] J. Plank、K. Greenan 和 E. L. Miller，《利用英特尔SIMD扩展实现极速伽罗瓦域算术》，载于《第11届USENIX文件与存储技术会议（FAST '13）》，2013年。[22] C. Tian、H. Sun 和 J. Chen，《具有最优消息大小和上传成本的容量达成私有信息检索码》，《IEEE信息论汇刊》，第65卷，第11期，第7613–7627页，2019年。[23] J. Xu 和 Z. Zhang，《MDS编码非串通服务器容量达成PIR方案的子分包化与访问次数》，《中国科学：信息科学》，第61卷，第1–16页，2018年。[24] J. Zhu、Q. Yan、C. Qi 和 X. Tang，《一种具有（几乎）最优文件长度的新型编码服务器容量达成私有信息检索方案》，《IEEE信息取证与安全汇刊》，第15卷，第1248–1260页，2020年。[25] R. Zhou、C. Tian、H. Sun 和 T. Liu，《具有最小消息大小的MDS编码数据库容量达成私有信息检索码》，《IEEE信息论汇刊》，第66卷，第8期，第4904–4916页，2020年。[26] Z. Zhang 和 J. Xu，《串谋服务器，”《IEEE信息论汇刊》，第65卷，第5期，第2723–2735页，2019年。[27] J. Ding、C. Lin、H. Wang 和 C. Xing，《小份额大小的通信高效秘密共享》，《IEEE信息论汇刊》，第68卷，第1期，第659–669页，2022年。[28] Z. Chen、M. Ye 和 A. Barg，《为里德-所罗门码的修复实现最优访问和纠错》，《IEEE信息论汇刊》，第66卷，第12期，第7439–7456页，2020年。[29] V. Guruswami 和 M. Wootters，《修复里德-所罗门码》，《IEEE信息论汇刊》，第63卷，第9期，第5684–5698页，2017年。[30] T. M. Cover 和 J. A. Thomas，《信息论基础》，Wiley-Interscience，2006年。[31] R. M. Roth，《编码理论导论》，剑桥大学出版社，2006年。[32] R. Lidl 和 H. Niederreiter，《有限域》，剑桥大学出版社，1996年。[33] R. Bitar 和 S. E. Rouayheb，《Staircase-PIR：通用鲁棒私有信息检索》，载于2018年IEEE信息论研讨会（ITW），2018年，第1–5页。[34] R. Bitar 和 S. Jaggi，《存在恶意对手时的通信高效秘密共享》，载于2020年IEEE国际信息论研讨会（ISIT），2020年，第548–553页。