

基于PIR增强的可搜索加密方案研究

唐亮¹, 马利民²

(1. 北京教育科学研究院, 北京 100036; 2. 北京信息科技大学计算机学院, 北京 102206)

摘要: 采用可搜索加密技术对数据加密后再存储云服务器, 虽能在一定程度上兼顾数据的安全性与可用性, 但基于该技术的系统在访问模式或搜索模式层面存在泄露用户隐私的风险。针对这一问题, 提出了一种隐私信息检索增强的可搜索加密方案。在方案预处理阶段, 首先对数据库进行分区及预处理, 同步完成陷门与关键字密文的计算, 并存储分区相关信息。这一设计不仅能降低系统的计算开销与通信开销, 还能提升后续查询效率。在实际查询流程中, 服务器接收客户端的查询请求后, 会先对陷门进行验证以实现精准的访问控制, 再对经过同态加密的数据执行运算。从整体执行逻辑来看, 攻击者无法通过系统交互过程推断出用户查询的具体内容, 从而有效解决了访问模式泄露问题; 同时, 服务器无法建立陷门、关键字密文与查询内容之间的关联关系, 有效避免了搜索模式泄露。实验结果表明, 当数据量持续增长时, 该方案的计算开销优势逐渐凸显。与传统隐私信息检索技术相比, 所提方案的查询时间减少了约80%。

关键词: 可搜索加密; 隐私信息检索; 同态加密; 隐私保护

中图分类号: TP393

文献标志码: A

DOI: 10.11959/j.issn.2096-109x.2025054

Research on searchable encryption scheme based on enhanced PIR

TANG Liang¹, MA Limin²

1. Beijing Academy of Educational Science, Beijing 100036, China

2. College of Computer Science, Beijing Information Science and Technology University, Beijing 102206, China

Abstract: Data was encrypted with searchable encryption technology before being uploaded to the cloud, which could, to a certain extent, balance data security and usability. However, systems based on this technology were found to carry the risk of exposing user privacy at the access or search pattern level. To address this issue, an enhanced searchable encryption scheme for privacy information retrieval was proposed. In the preprocessing stage of the scheme, the database was first partitioned and preprocessed. The trapdoor and keyword ciphertext were calculated simultaneously, and partition-related information was stored. This design was shown to not only reduce the computational and communication costs of the system but also improve subsequent query efficiency. In the actual

收稿日期: 2025-02-05; 修回日期: 2025-05-29

基金项目: 国家重点研发计划(2022YFC3320903); 未来区块链与隐私计算高精尖创新中心(GJJ-24-016)

Foundation Items: The National Key R&D Program of China (2022YFC3320903), Beijing Advanced Innovation Center for Future Blockchain and Privacy Computing (GJJ-24-016)

引用格式: 唐亮, 马利民. 基于PIR增强的可搜索加密方案研究[J]. 网络与信息安全学报, 2025, 11(5): 115-125.

Citation Format: TANG L, MA L M. Research on searchable encryption scheme based on enhanced PIR[J]. Chinese Journal of Network and Information Security, 2025, 11(5): 115-125.

query process, after the query request from the client was received by the server, the trapdoor was verified to achieve precise access control. Then, operations were performed on the data that had undergone homomorphic encryption. From the overall execution logic, the specific content of the user's query could not be inferred by attackers through the system interaction process, thus effectively solving the problem of access pattern leakage. Simultaneously, the association relationship between the trapdoor, keyword ciphertext, and the query content could not be established by the server, effectively avoiding search pattern leakage. Experimental results show that as the data volume continues to increase, the computational cost advantage of this scheme becomes increasingly prominent. Compared with traditional privacy information retrieval protocols, the query time is reduced by approximately 80% in the proposed scheme.

Keywords: searchable encryption, privacy information retrieval, homomorphic encryption, privacy protection

0 引言

随着云服务的不断发展,越来越多的数据被存储在云服务器上,频发的数据泄露事件已引起公众对数据安全的高度关注^[1]。数据加密后存储到云服务器虽能保障安全性,但给检索带来了困难。可搜索加密(searchable encryption, SE)技术能够实现对密文数据的检索,有效保护用户隐私,但存在搜索模式与访问模式泄露问题^[2]。隐私信息检索(private information retrieval, PIR)技术则可在不暴露查询内容的前提下检索数据库信息,不过其计算与通信开销较大^[3]。鉴于此,本文通过融合可搜索加密与基于同态加密的PIR技术,提出了一种PIR增强的可搜索加密方案(PIR-enhanced searchable encryption scheme, PESES)。该方案既解决了模式泄露问题,又提升了系统的检索效率。

本文主要研究内容及贡献如下。

1) 结合PIR技术以增强模式隐私保护。在可搜索加密的基础上引入PIR技术,具体实现流程为:客户端生成加密查询向量并将其发送给服务器,服务器接收查询请求后执行相关计算并返回结果。在查询过程中,所有数据均参与计算,使攻击者无法推断出具体被访问的数据项,从而有效防止访问模式泄露。同时,服务器在验证客户端发送的查询请求中的陷门后,对同态加密的数据组进行运算,从而确保攻击者无法分析出陷门、关键字密文与查询内容之间的关联性,避免搜索模式泄露。

2) 基于分区的预处理来优化查询效率。在预处理阶段,客户端先对数据库进行分区,记录目标数据的分区位置信息,用于构建高效的PIR查

询;服务器则依据客户端发送的PIR查询参数,在数据库中检索相应数据并返回结果。该策略通过显著减少单次查询中参与同态计算的数据量及查询请求本身的大小,有效缩减了计算开销与通信开销,大幅提升了系统的查询效率。

1 相关工作

为解决云存储中的数据泄露问题,一种方案是将数据加密后再存储于云服务器^[4]。该方案虽能保障数据安全,却导致加密数据无法直接高效处理,必须完全解密后才能使用。针对这一问题,Wang等^[5]提出一种基于Paillier加密的新型可搜索加密方案,该方案在降低信息泄露风险的同时,还能提供更强的隐私保障。Amorim等^[6]提出在可搜索加密方案中使用同态加密,以解决传统加密方案难以支持加密数据的有效搜索与检索的问题。此外,Cheng等^[7]提出了一种可抵抗内部关键字猜测攻击、支持关键字搜索的公钥认证加密方案,旨在增强此类方案的安全性。

PIR技术能够在不泄露用户具体查询内容的前提下完成信息检索。Beimel等^[8]提出了一个通用的PIR框架,并在此框架下设计了TinySet技术。然而,大多数PIR技术因计算复杂度高而难以实际部署。针对效率瓶颈,多项研究提出了改进方案。Xu等^[9]提出一种高效的PIR技术,该技术融合了单密文全同态加密与拉格朗日插值多项式技术,提升了通信效率,并降低了通信成本。Ren等^[10]通过优化查询提示的构造与使用,提出了一种简单且轻量级的单服务器亚线性PIR方案。该方案具有较小的平均响应时间、接近最优的在线响应性能及相对较小的通信开销。Zhou等^[11]提出了基于比特切

片技术的动态高效可搜索加密方案。该技术通过将S盒简化为逻辑电路,大幅提升了S盒的计算效率。袁承昊等^[12]提出利用陷门搜索记录构建密文索引。该方法通过比较新旧陷门的搜索范围并结合密文索引,有效减少了搜索过程中需要比对的密文数量,从而实现快速搜索。

2 相关技术

2.1 可搜索加密技术

可搜索加密技术主要用于解决明文搜索过程中潜在的隐私泄露问题。其实现方式主要分为对称可搜索加密和非对称可搜索加密两类。该技术的典型工作流程如图1所示。

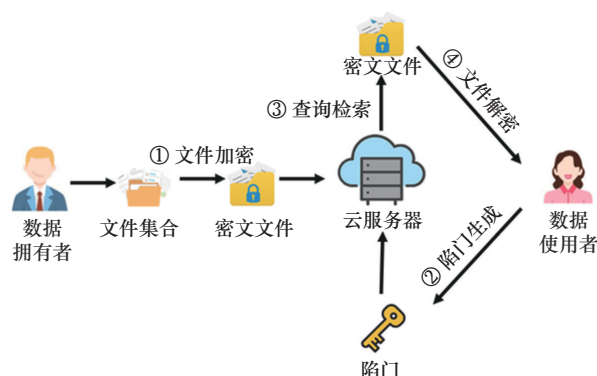


图1 可搜索加密技术的典型工作流程

Figure 1 Typical workflow of searchable encryption technology

可搜索加密技术的核心操作步骤如下。

1) 文件加密: 数据拥有者使用本地加密密钥对待上传至云服务器的文件进行加密,以保障文件在传输及云服务器存储过程中的机密性。

2) 陷门生成: 经数据拥有者授权的数据使用者,利用特定密钥对本地待查询的关键字进行计算,生成陷门。这一过程可确保关键字信息不被泄露,生成的陷门随后将被发送至云服务器。

3) 查询检索: 云服务器在接收数据使用者提交的陷门后,基于该陷门对每个上传文件的索引表进行检索,并返回包含对应查询关键字的密文文件。在此过程中,云服务器无法知晓陷门的具体含义,从而确保用户查询的隐私性与安全性。

4) 文件解密: 数据使用者收到云服务器返回的密文文件后,使用解密密钥对其进行解密。解密过程仅在客户端本地进行,确保文件内容只有授权用户可访问。

模式泄露是可搜索加密技术面临的主要安全问题,主要包括以下两类。

1) 访问模式泄露: 指攻击者通过分析用户的查询历史记录及数据访问模式(如查询频率、时间戳、访问的数据项数量等),推断出用户的访问行为特征,进而对用户隐私构成威胁。

2) 搜索模式泄露: 指攻击者通过分析用户的查询内容或搜索关键字的关联性,识别用户的搜索意图或关键字关联模式,这会严重威胁用户隐私安全。

2.2 PIR 技术

PIR 技术是保护用户查询隐私的关键技术之一,其技术路线有3种。信息论私有信息检索(information theoretical private information retrieval, IT-PIR)可以提供绝对的信息隐私保护^[13],该技术要求多台非共谋服务器存储相同的数据库副本,客户端将查询分发至各服务器,使任一服务器均无法独立推断出客户端的查询目标。但由于其计算开销与通信开销较大,实际应用受到限制。计算性隐私信息检索(computational PIR, C-PIR)依赖计算复杂性假设(如离散对数难题、椭圆曲线密码学或大整数分解难题^[14]),确保即便攻击者具备多项式时间计算能力,也无法破坏隐私。与IT-PIR相比,C-PIR通常具有更高的效率,但其隐私保护强度取决于计算难题的困难性假设。基于硬件的隐私信息检索(hardware-based PIR, HW-PIR)利用专用硬件来加速隐私信息检索过程,同时能在一定程度上保障用户的隐私。

本文重点关注基于同态加密的单服务器PIR技术,其检索流程如图2所示。

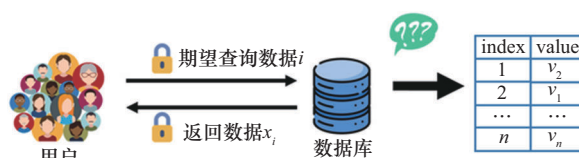


图2 PIR 检索流程

Figure 2 PIR retrieval flowchart

1) 客户端生成加密的查询向量,并将其发送给服务端。

2) 服务端接收到查询向量后,在本地执行同态运算,并将结果返回给客户端。

3) 客户端收到响应后, 通过解密操作得到查询结果。

在整个过程中, 云服务器仅依据技术执行运算, 无法知晓具体查询的是哪条数据。

2.3 同态加密技术

同态加密 (homomorphic encryption, HE) 是一种允许在密文状态下执行特定计算操作的密码学技术^[15], 其计算结果解密后等同于对相应明文执行相同操作的结果。根据支持计算能力的差异, HE 可分为部分同态加密和完全同态加密两类。部分同态加密仅支持有限类型的同态运算 (如加法或乘法), 而完全同态加密可支持任意算术运算。部分同态加密算法 (如 Paillier、基于 RSA 的乘法同态加密等算法) 的安全性, 依赖于判断性合数剩余假设困难问题、大整数分解问题。完全同态加密在选择明文攻击下的不可区分性 (indistinguishability under chosen-plaintext attack, IND-CPA) 安全性, 主要基于格密码学中的容错学习 (learning with errors, LWE) 及其变种环上容错学习 (ring-LWE, RLWE) 等困难问题假设。

同态加密的核心机制是通过概率加密算法实现密文的随机化生成 (即便对同一明文进行多次加密, 每次生成的密文均具有统计独立性), 从而有效防止攻击者通过分析密文模式关联明文信息。在本文提出的方案中, 同态加密技术被用于构造混淆陷门集合、实现 PIR 技术及对加密数据进行直接计算, 确保数据在整个处理过程中始终处于加密状态。

3 PIR 增强可搜索加密方案

3.1 PESES 单服务器模型

本文提出的 PESES 模型采用单服务器模式, 主要由客户端、云服务器和数据库构成, 如图 3 所示。该模型假设云服务器具有“诚实但好奇”的特性。

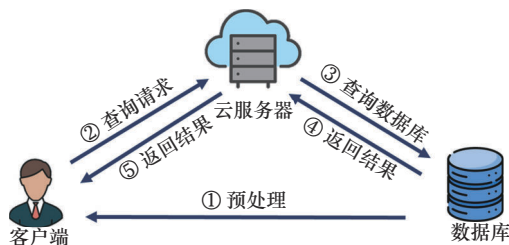


图 3 PESES 单服务器模型的架构

Figure 3 PESES single server model architecture

1) 客户端: 由客户端发起查询操作。初始化完成后, 客户端须定期与数据库同步状态信息, 在本地执行数据预处理任务, 并存储分区后的数据索引、公共参数、密钥信息及陷门信息。

2) 云服务器: 负责存储客户端查询的关键词密文。在接收到客户端的查询请求后, 云服务器严格遵循 PESES 技术规范执行操作, 查询数据库并将检索结果返回给客户端。

3) 数据库: 充当核心数据存储介质, 负责将数据按预设策略进行分区。记录每个分区的区号、对应的数据块及其索引。与客户端协同完成数据预处理, 并同步相关信息。

3.2 数据预处理

将数据库划分为 N 个独立的区, 每个分区被视为一个数据库子集。值得注意的是, 单个分区内的数据量对于云服务器而言仍然是相当多的。分区号及其对应的数据索引被同步至客户端本地存储。当客户端发起查询时, 仅需向云服务器发送目标分区的区号及相关索引信息作为查询参数。云服务器据此仅需访问并处理指定分区的数据, 从而显著减少了计算开销与通信开销。此外, 还限定了云服务器响应查询的时间。假设未分区场景下云服务器的平均响应时间为 t , 分区后云服务器的平均响应时间不得超过 $2t/N$ (N 为分区数)。

客户端在首次生成并存储数据时会定义与数据相关的关键词, 将其作为可搜索加密的参数, 并将关键词的密文上传至云服务器存储。同时, 客户端在本地预先利用公钥和私钥计算查询所需的陷门, 供后续查询时使用, 从而减少查询过程的计算开销与通信开销。

3.3 PESES 检索过程

PESES 检索过程主要包括系统初始化、生成查询请求、处理查询请求和接收并解密数据 4 个步骤, 如图 4 所示。

(1) 系统初始化

给定一个安全参数 σ , 输出双线性对 $e: G \times G \rightarrow G_T$ 。定义 g_1 为群 G 的生成元, G 为 z 的阶循环群。计算 $h = e(g, g)$ 。定义哈希函数组 $H_1: \{0, 1\}^* \rightarrow G$, $H_2: G_T \rightarrow \{0, 1\}^{\log \sigma}$ 。随机选择两个大素数 p 和 q , 且满足 p 和 $2q + 1$ 均为素数。利用式(1)计算出模数 n , 利用式(2)计算出 λ , 利用

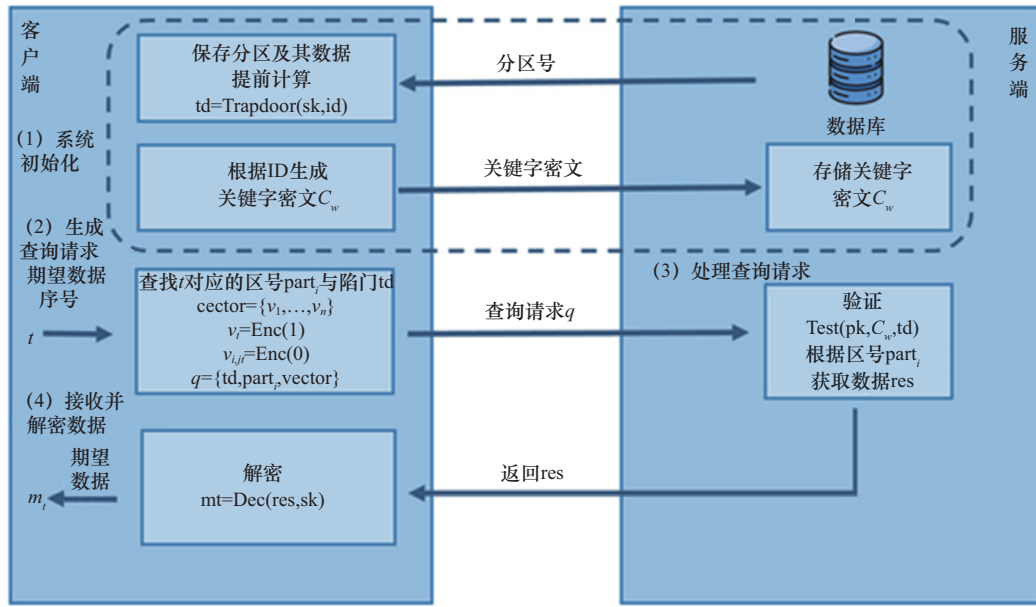


图4 PESES检索流程

Figure 4 PESES retrieval process

式(3)计算出 μ 。

$$n = p \times q \quad (1)$$

$$\lambda = \text{lcm}(p-1, q-1) \quad (2)$$

$$\mu = (\lambda^{-1} \bmod n) \quad (3)$$

选择一个生成元 g_2 ，使得 g_2 是模 n 的一个原根，并公布系统参数 $\text{params} = \{G, G_T, z, g_1, g_2, e, h, p, q, n, \lambda, \mu, H_1, H_2\}$ 。选择随机数 $y \in \mathbb{Z}_q^*$ 。可搜索加密私钥 $\text{SK}_1 = y$ ，公钥 $\text{PK}_1 = g_1^y$ ；同态加密私钥 $\text{SK}_2 = (\lambda, \mu)$ ，公钥 $\text{PK}_2 = (n, g_2)$ ；数据库中存储的所有数据 f 均运用Paillier算法使用公钥 PK_2 加密，生成密文 C_m 。

$$C_m = g^m \cdot r^n \bmod n^2 \quad (4)$$

以 l 为长度对数据库中的数据进行均匀分区，将分区号 part 及其对应的数据索引 idx 同步到客户端本地进行存储。使用公钥 PK_1 和关键字 ε ，随机选择 $x \in \mathbb{Z}_q^*$ ，根据式(5)计算 t 。

$$t = e(H_1(\varepsilon), h^x) \quad (5)$$

根据式(6)计算关键字密文 C_w 。

$$C_w = (g_1^x, H_2(t)) \quad (6)$$

将关键字密文 C_w 上传至服务器进行存储。在本地客户端预处理阶段，需要计算陷门。输入私钥 SK_1 和关键字 ε ，根据式(7)计算陷门 td 。

$$\text{td} = H_1(\varepsilon)^y \quad (7)$$

保存所有陷门 td 与对应的关键字 ε 。

(2) 生成查询请求

当客户端需要查询某关键字 ε 时，系统会先在客户端本地查找该关键字对应的陷门 td_i 、分区号 part_i 和索引 idx_i 。若未找到这些信息，则表明数据库中不存在包含该关键字的记录。此时，系统将终止查询流程，以避免后续不必要的计算开销。如果成功找到对应的陷门 td_i 、分区号 part_i 和索引 idx_i ，那么利用这些参数构建查询请求。具体构建步骤如下：①生成一个二进制数组 v ，在此数组中，仅将需要查询的第 i 位设置为1，其他比特位均设置为0；②对于每个元素 m_i （ m_i 为一条明文数据），执行式(8)的计算操作。

$$C_v = g_2^{v_i} \cdot r_i^n \bmod n^2 \quad (8)$$

将查询请求 $Q = \{\text{td}_i, \text{part}_i, C_v\}$ 发送至服务器。

(3) 处理查询请求

服务器在接收到查询请求 Q 后，首先将查询请求中的陷门 td_i 与存储的关键字密文 C_w 进行匹配，验证式(9)是否成立。

$$H_2(e(T_\varepsilon, g_1^x)) = H_2(t) \quad (9)$$

若成立，则证明用户的查询权限是合法的。据此实现对用户的访问控制，防止非授权用户滥用查询导致系统效率降低。验证通过后，服务器根据分区号 part_i 从数据库获取该分区下的所有

数据。对分区内的每条数据 C_v 和密文 C_m , 执行式(10)的计算操作。

$$C_{\text{res}} = \prod_{i=1}^N (C_{m_i})^{v_i} \bmod n^2 \quad (10)$$

将查询结果 C_{res} 返回给客户端。

(4) 接收并解密数据

客户端接收到服务器返回的加密查询结果 C_{res} 后, 使用客户端私钥 SK_2 解密数据, 还原为明文 m 。

$$m = L(C_{\text{res}} \bmod n^2) \cdot \mu \bmod n \quad (11)$$

其中, $L(x) = \frac{x-1}{n}$, 用于将大数调整为在模数域 n 内的一个值。

4 实验与分析

4.1 安全性分析

4.1.1 正确性分析

本节对 PESES 方案的正确性进行分析和验证。在方案中, 密文 $E_k(d_i)$ 被分为 l/k 块, 分别为 $D_{i,1}, D_{i,2}, \dots, D_{i,l/k}$ 。假设客户端请求检索第 r 个密文 $E_k(d_r)$, 客户端将生成查询请求 $v = (\varepsilon(0), \varepsilon(0), \dots, \varepsilon(1), \dots, \varepsilon(0))$ 。云服务器接收到该查询请求后, 执行特定的同态计算操作, 得到:

$$R = (\varepsilon(0)^{E_k(d_1)}, \varepsilon(0)^{E_k(d_2)}, \dots, \varepsilon(0)^{E_k(d_r)}, \dots, \varepsilon(0)^{E_k(d_n)}) \quad (12)$$

由于 $\varepsilon(x)$ 采用同态加密, 根据其同态性质 $\varepsilon(m_1) \cdot \varepsilon(m_2) = \varepsilon(m_1 + m_2)$ 、 $\varepsilon(m_1)^k = \varepsilon(k \cdot m_1)$, 可将式(12)化简为:

$$\begin{aligned} R &= \varepsilon(0)^{E_k(d_1)} \cdot \varepsilon(0)^{E_k(d_2)} \cdot \dots \cdot \varepsilon(0)^{E_k(d_r)} \cdot \dots \cdot \varepsilon(0)^{E_k(d_n)} = \\ &= \varepsilon(0) \varepsilon(0) \dots \varepsilon(E_k(d_r)) \dots \varepsilon(0) = \\ &= \varepsilon(0 + 0 + \dots + E_k(d_r) + \dots + 0) = \\ &= \varepsilon(E_k(d_r)) \end{aligned} \quad (13)$$

对 R 进行解密即可得到 $E_k(d_r)$ 。因此, 服务器能够通过执行 PESES 技术正确获取所需的密文。该过程验证了方案满足正确性要求。

4.1.2 搜索模式安全性分析

对于搜索模式的安全性分析, 主要聚焦于计算隐私保护。当客户端向云服务器检索数据时, 云服务器在多项式时间内无法区分客户端提交的

任意两个查询请求, 则称该系统满足查询不可区分性。该安全性要求可用如下的形式化定义表示。

$$|\Pr[\text{Query}(i_1)] - \Pr[\text{Query}(i_2)]| \leq \frac{1}{n^c} \quad (14)$$

其中 c 为大于 1 的整数。这意味着, 对于任意两个不同的查询请求, 云服务器在多项式时间内无法区分 i_1 和 i_2 的位置和内容。

此类安全性证明通常基于数学难题和若干安全假设, 具体分析如下。

1) 陷门与关键字密文的匹配安全。系统采用基于公钥的可搜索加密机制: 关键字密文使用公钥加密生成, 陷门则由客户端使用其本地存储的私钥生成。由于私钥始终驻留于客户端, 攻击者无法获取, 从而确保了陷门的不可伪造性; 同时, 根据离散对数难题的困难性假设, 攻击者无法从陷门中提取有效信息。

2) 基于同态加密的 PIR 技术安全。查询请求包含基于 Paillier 同态加密的二进制数组, 其中仅目标位置为 1, 其余位置为 0。根据 Paillier 同态加密的语义安全性 (其安全性依赖于大整数分解问题的困难性), 即使攻击者已知公钥, 也无法推导出构成公钥的大素数 p 和 q , 以及加密数组中 0 和 1 对应的明文值。此外, 同态加密算法在执行加密时引入了随机数 r , 使得相同明文的不同加密结果在语义上不可区分。

结合上述两点, 云服务器在处理查询时仅能看到同态加密后的密文, 无法通过多项式时间算法区分不同查询请求的差异。这有效保护了用户的查询隐私, 并能够抵御搜索模式泄露攻击。

4.1.3 访问模式安全性分析

访问模式是指云服务器响应特定查询请求所返回的数据集。在传统的可搜索加密方案中, 客户端将陷门发送至云服务器, 云服务器将陷门与存储的关键字密文进行匹配, 若匹配成功, 则返回相应的加密数据。尽管数据本身是加密的, 但由于相同的关键字查询总是触发云服务器返回相同的密文结果集, 因此攻击者 (或好奇的服务器) 可以通过观察和分析这些返回模式的规律来推测用户的查询内容, 从而导致访问模式泄露。

在本方案中, 云服务器在接收到客户端发送的查询请求后, 并非直接匹配并返回结果, 而是对分区内的每条数据执行模指数运算。由于查询

请求本身（陷门、加密数组）不直接暴露客户端要检索的目标条目，且分区内的所有数据都参与同态计算，因此即使攻击者知晓分区信息，在计算能力受限（多项式时间）的情况下，也难以从计算过程或中间结果中推断出客户端具体查询了哪条数据，从而有效避免了访问模式泄露。

云服务器将执行同态计算后的密文结果返回给客户端，客户端在本地对返回的密文结果进行解密，最终获得所需的明文数据。在整个检索过程中，客户端在本地解密查询结果，云服务器仅执行预定的同态运算，无法获取客户端查询的具体数据，从而有效解决了访问模式泄露问题。

4.1.4 安全存储分析

为证明 PESES 的整体安全性，需逐一验证各核心模块的安全性，包括公钥可搜索加密（public key encryption with keyword search, PEKS）、非对称加密（RSA）及 Paillier 同态加密。其中 PEKS 中关键字密文与陷门的安全性，基于判定性双线性 Diffie-Hellman（decisional bilinear Diffie-Hellman, DBDH）困难性假设，在适应性选择关键字攻击（indistinguishability under adaptive chosen-keyword attack, IND-CKA）下，该 PEKS 方案被证明是安全的；RSA 的安全性主要依赖整数分解问题（integer factorization problem, IFP）的困难性假设；而 Paillier 同态加密算法的安全性主要依赖合数剩余类问题的困难性假设。各模块的具体安全性证明过程如下。

定理 2 给定整数 n 和 $x \in \mathcal{Z}_n^*$ ，求解 m 使得 $x = g^m \cdot r^n \bmod n^2$ （其中 g 为生成元， r 为随机数）在计算上是困难的。对任意多项式时间算法 A ，其成功概率 ε 是可忽略的，即：

$$\Pr[A(n, g, x) = m] \leq \varepsilon \quad (15)$$

证明 本定理的安全性基于 Paillier 加密所依赖的判定性合数剩余类问题困难性假设，可归约为证明 Paillier 加密方案在 IND-CPA 下，即多项式时间敌手无法通过密文区分两个明文。

1) 系统初始化。挑战者选择两个大素数 p 和 q ，计算 $n = pq$ 和 $\lambda = \text{lcm}(p-1, q-1)$ （ lcm 是最小公倍数）。选取生成元 $g \in \mathcal{Z}_n^*$ （通常选择 $g = n+1$ ），计算私钥参数 $\mu = \left(L(g^\lambda \bmod n^2)\right)^{-1} \bmod n$ ， $L(x) = \frac{x-1}{n}$ ，

计算系统公钥 $\text{PK} = (n, g)$ 、私钥 $\text{SK} = (\lambda, \mu)$ 。

2) 询问阶段（阶段 1）。敌手 B 作为敌手 A 的挑战者，对敌手 A 发出的询问进行回答。敌手 A 提交明文 m' 进行询问，敌手 B 计算密文 $C = g^{m'} \cdot r^n \bmod n^2$ ，并将 C 返回给敌手 A 。

3) 挑战阶段。敌手 A 随机选择两个明文 m'_1 和 m'_2 提交给敌手 B 。敌手 B 接收一个 Paillier 问题实例 (n, g, x) ，随机选取比特 $b \in \{0, 1\}$ ，计算挑战密文 $c'_1 = g^{m'_1} \cdot r_1^n \bmod n^2$ 或 $c'_2 = g^{m'_2} \cdot r_2^n \bmod n^2$ ，将挑战密文 c^* 返回给敌手 A 。

4) 二次询问阶段（阶段 2）。重复阶段 1 的询问过程，敌手 A 可继续提交明文（除挑战明文外）进行加密询问，敌手 B 仍按阶段 1 的规则进行响应。

5) 猜测阶段。敌手 A 输出对 b 的猜测 $b' \in \{0, 1\}$ 。若 $b = b'$ ，则敌手 B 返回 $m' = m_b$ 。敌手 A 成功概率：

$$\Pr[A(c^*, m_1, m_2) = b] = \frac{1}{2} + \varepsilon \quad (16)$$

敌手 B 成功概率也为 $\frac{1}{2} + \varepsilon$ 。因为如果敌手 A 能够成功区分密文，那么敌手 B 就可以成功判断 x 是否为某个整数的 n 次幂。这与 Paillier 问题的困难性假设矛盾，说明本方案 Paillier 算法在 IND-CPA 下是安全的。

4.2 实验数据分析

4.2.1 功能性分析

文献[16]提出的可搜索加密方案，通过将搜索关键字与预先加密上传的关键字进行匹配，成功匹配后返回密文数据，实现了密文数据在不泄露明文信息前提下的查询。该方案虽支持在不解密密文的前提下完成查询，但存在搜索模式与访问模式泄露的安全隐患。文献[17]采用多分少补的方法消除文档编号统计特征，以抵御关键字猜测攻击，在一定程度上能够缓解访问模式泄露问题。然而，当面对海量参差不齐的数据时，该方法会显著增加系统的存储开销。在特定场景下，甚至出现用于补齐的字节量超过原始数据量的情况。为有效解决可搜索加密访问模式和搜索模式泄露问题，本文提出 PESES 方案。上述 3 种方案的功能性对比如表 1 所示。

表 1 3 种方案的功能性对比
Table 1 Functional comparison of three schemes

方案	抵御访问模式泄露	抵御搜索模式泄露
文献[16]	无法抵御访问模式泄露	无法抵御搜索模式泄露
文献[17]	采用多分少补的方式实现数据长度对齐, 在一定程度上能够缓解访问模式泄露问题	无法抵御搜索模式泄露
PESES	引入 PIR 技术, 可抵御访问模式泄露	采用同态加密, 可抵御搜索模式泄露

4.2.2 理论性分析

本节将从存储开销、通信开销和计算开销 3 个维度, 对 PESES、文献[16]和文献[18]方案的理论模型进行详细的对比分析。

(1) 存储开销

PESES 由于需要对数据进行预处理, 客户端需要存储数据库分区后的元数据和陷门信息。假设数据库包含 m 条 Key-Value 数据, 按长度 d 划分为 (m/d) 个区域; 存储 m 条数据对应陷门的开销为 $O(m_{\text{id}})$ 。因此, 客户端的总存储开销为 $O(m + m/d + m_{\text{id}})$ 。服务器端除存储原始的 m 条数据库数据外, 还需要存储关键字的密文索引, 设其存储开销为 $O(m_w)$, 因此服务器端的总存储开销为 $O(m + m_w)$ 。

文献[16]方案和文献[18]方案均无须预处理, 仅在服务器端存储原始数据库数据, 总存储开销为 $O(m)$ 。

(2) 通信开销

PESES 在初始化阶段需要完成两部分通信交互: 数据库将分区号及其对应数据发送给云服务器; 客户端将关键字密文发送给云服务器。因此, PESES 初始化阶段的总通信开销为 $O(mc_i + mc_{\text{id}})$ 。其中, c_i 表示每个加密文件的大小, c_{id} 表示每个关键字密文的大小。文献[16]和文献[18]方案初始化阶段无须通信。

PESES 在搜索阶段也需要完成两部分通信交互: 客户端向云服务器发送查询请求, 该请求包含一个陷门、一个区号和一个长度为 d 的加密数组, 通信开销为 $O(\text{len}_{\text{id}} + c_d + 1)$; 云服务器返回大小为 v 的加密数据库, 通信开销为 $O(v)$ 。文献[16]方案采用双链 (联盟链与私有链) 存储架构。搜索时须先向联盟链发送查询请

求以获取安全索引, 再根据安全索引向私有链发送查询请求以获取密文索引, 最后在本地解密查询结果。其通信开销为 $O(\text{len}_{\text{id}} + \text{len}_{\text{tx}})$ 。文献[18]方案的通信开销为 $O(c_m)$ 。

(3) 计算开销

文献[18]方案基于 Paillier 同态加密实现 PIR 技术, 云服务器将 Buffer 组成一个 $\sqrt{M} \times \sqrt{M}$ 的矩阵。客户端生成一个 $\sqrt{M}$ 维的查询向量, 并利用 Paillier 同态加密对其中的每个元素进行加密, 此过程需要执行 $\sqrt{M}$ 次模幂运算。云服务器根据客户端发送的加密查询向量计算响应值, 如式(17)所示。

$$Z_j = \prod_{i=1}^s v_i^{M_{ij}} (j \in [1, t]) \quad (17)$$

该计算包含 M 次乘法运算。随后, 云服务器将 $\sqrt{M}$ 个计算结果返回给客户端, 客户端需要执行一次模幂运算对其进行解密。由于同态加密针对每比特操作, 需要乘以密文长度 l , 且布隆过滤器引入 k 次额外运算, 一次模幂运算的时间开销大约是 100 次乘法运算的时间开销, 因此客户端的加密开销为:

$$\left\{ \left(\sqrt{M} + \frac{M}{100} \right) \alpha + \left(\frac{2l\sqrt{M}}{\beta} \right) \right\} kl \quad (18)$$

其中, α 表示一次模幂运算的时间开销, β 表示客户端和数据库之间的带宽。式(18)可简化为:

$$\left(\sqrt{M} + \frac{M}{100} \right) \alpha kl \quad (19)$$

客户端获得文件存储地址后, 向数据库检索文件的开销为:

$$nl\alpha + \frac{(n-1)l\alpha}{100} \quad (20)$$

PESES 进行预处理后, 不需要将整个 $\sqrt{M}$ 维向量都参与运算。查询向量中包含提前生成的陷

门与区号，仅需生成长度为 d 的同态加密数组（云服务器仅需执行 d 次乘法运算），因此，PESES 的总计算开销为：

$$\alpha + \frac{d+l}{\beta} + t_b \quad (21)$$

其中， t_b 表示可搜索加密中陷门与云服务器存储的关键词密文进行匹配的计算开销。初始化阶段生成陷门和关键词密文的计算开销为 $m(t_{id} + t_w)$ 。

文献[16]方案结合可搜索加密和代理重加密实现安全数据共享。其查询过程涉及：先在联盟链上执行可搜索加密的陷门与关键词密文匹配，获取目标数据的索引；再将在私有链上获取的索引（或伪身份）匹配到具体的加密数据。因此，其总计算开销为：

$$t_b + 2t_e + t_{id} + t_w \quad (20)$$

其中， t_b 表示联盟链上的计算开销， $2t_e$ 表示私有链上的计算开销， $t_{id} + t_w$ 表示可搜索加密生成陷门和关键词密文的计算开销。3种方案的系统开销对比如表2所示。

文献[16]与文献[18]的方案因不涉及预处理环节，在存储开销和通信开销上的表现优于 PESES。不过，PESES 的预处理可在系统空闲时段执行，从而避免在业务运行时占用计算资源。从计算开销的分布来看，系统的主要开销集中在查询阶段。文献[16]方案未采取有效措施保护访问模式与搜索模式，省去了隐私增强相关的计算步骤，因此计算开销最小，但安全性存在明显短板。文献[18]方案在查询时需要调用全量数据参与运算，计算开销最大。PESES 通过分区机制，仅需目标分区内的所有数据参与运算，大幅减少了查询阶段的计算量。尽管该方案引入了额外的存储开销，但这种“以空间换时间”的策略能够显著提升查询效率，在实际应用中具有可接受性。尤其是在存储成本持续下降、用户对查询响应速度要求日益提

高的背景下，该权衡更具实用价值。

4.2.3 效率分析

本文实验在 Ubuntu 虚拟机中运行。本地机为 Windows 10 操作系统，配备 32 GB 内存，处理器为 Intel Core i5-10500 CPU @ 3.10 GHz。采用 Python 语言在 Pycharm 平台编译，使用的开源数据库包括 PBC（pairing-based cryptography）和 NumPy。数据库分区采用固定分区策略，将数据集划分为 8 个（即 2^3 个）区。为探究不同数据区块大小对时间开销的影响，将分区数据量设置为 4 个级别，分别为 2^9 字节、 2^{10} 字节、 2^{11} 字节和 2^{12} 字节。每组实验重复进行 5 次，取平均值作为最终结果，实验结果如表 3 所示。

表3 不同数据区块大小的时间开销
Table 3 Computational time cost of different data block sizes

区块大小/字节	初始化时间开销/s	查询时间开销/s
$N=2^{12}$	164.09	0.811
$N=2^{13}$	328.19	1.613
$N=2^{14}$	657.66	3.201
$N=2^{15}$	1 312.76	6.368

在初始化阶段，需要对数据库进行分区，并将分区元数据同步至客户端。由于这是仿真实验，暂不考虑网络节点间的通信开销。客户端生成关键词对应的陷门和关键词密文，其中关键词密文需要上传至云服务器存储。初始化阶段的开销集中在陷门与关键词密文的生成过程，其中涉及的“哈希到群”运算开销较大，具体耗时与算法的实现方式密切相关。

在查询阶段，客户端须先在本地存储中检索目标序号或关键词，匹配对应的分区号与陷门；随后根据目标分区的数据量，生成长度匹配的加

表2 3种方案的系统开销对比

Table 2 Comparison of system costs for the three schemes

方案	存储开销	通信开销	计算开销
文献[16]	$O(m)$	$O(\text{len}_{id} + \text{len}_{ix})$	$t_b + 2t_e + t_{id} + t_w$
文献[18]	$O(m)$	$O(c_m)$	$\left(\sqrt{M} + \frac{M}{100}\right)akl + n\alpha + \frac{(n-1)l\alpha}{100}$
PESES (本文)	$O(m + m/d + m_{id}) + O(m + m_w)$	$O(mc_i + mc_{id}) + O(\text{len}_{id} + c_d + 1)$	$\alpha + \frac{d+l}{\beta} + t_b + m(t_{id} + t_w)$

密数组；最后整合分区号、陷门及加密数组，构造完整的查询请求并发送给云服务器。云服务器在接收请求后，须执行可搜索加密的匹配过程以验证陷门与关键字密文的一致性，并对目标分区内的所有数据执行同态加密运算（包括若干次模乘法和 1 次模加法），最终将加密计算结果返回给客户端。在此过程中，查询时间与分区数据量呈严格线性正比关系。这种线性依赖性是 PIR 技术在海量数据检索中效率受限的核心原因。

在数据量分别为 2^{12} 、 2^{13} 、 2^{14} 、 2^{15} 字节的情况下，对本文方案与文献[18]方案的查询过程进行模拟实验。本文方案的查询时间包含 PEKS 的陷门与关键字密文匹配过程耗时，以及 Paillier 同态加密运算过程耗时两部分。文献[18]的方案因无须预处理，其查询时间仅包含 Paillier 同态加密运算过程耗时，但每次查询请求需要处理全量数据（而非分区数据）。实验结果为每组算法运行 5 次的平均值，不同数据量下两种方案的查询时间如表 4 所示，查询时间对比如图 5 所示。

表 4 查询时间对比

Table 4 System query time cost comparison

方案	查询时间			
	$N=2^{12}$	$N=2^{13}$	$N=2^{14}$	$N=2^{15}$
文献[18]	5 032	9 790	20 296	38 723
本文	995	1 960	4 108	7 699

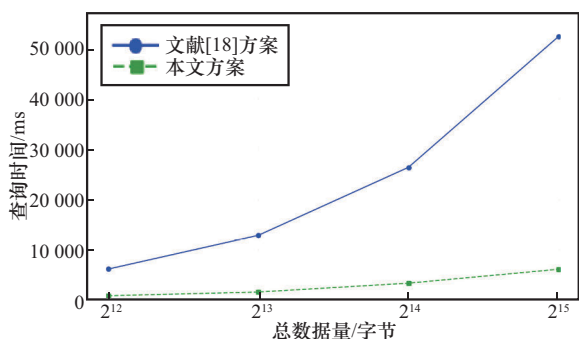


图 5 不同数据量下两种方案的查询时间对比

Figure 5 Comparison of query times for the two schemes under different data volumes

从图 5 和表 4 的实验结果可以看出，随着数据量的增加，文献[18]方案与本文方案的查询时间均呈现线性增长趋势。具体而言，文献[18]方案的查询时间随数据量增加呈快速增长态势，且

整体耗时显著高于本文方案。相比之下，本文方案的效率更高，查询时间提升约 80%，并且数据量越大，本文方案的优势越明显。

5 结束语

针对传统可搜索加密方案中普遍存在的搜索模式与访问模式泄露问题，本文提出一种基于 PIR 增强的可搜索加密方案。该方案通过引入 PIR 技术，有效解决了搜索模式泄露和访问模式泄露问题；同时，通过数据库分区与客户端预处理机制，显著减少了系统开销，有效缓解了可搜索加密与 PIR 技术在实际应用中的高计算开销问题。此外，该方案结合可搜索加密（保护数据存储与传输隐私、提供访问控制）与 PIR 技术（隐藏查询意图与结果关联），实现全链路隐私保护。

未来工作思路如下：首先，研究 PESES 在分布式环境及多云环境中的扩展性，探索跨平台部署时不同云服务商之间的技术兼容性与数据一致性保障机制；其次，针对大规模数据场景下的计算开销与通信开销优化问题，计划结合机器学习方法预测用户查询模式，以优化预处理策略；最后，研究硬件加速技术在密文检索过程中的应用，探索基于 GPU/FPGA 的并行计算架构对方案性能的提升潜力。上述研究方向将有助于推动隐私保护检索技术在实际复杂场景中的落地应用。

参考文献：

- [1] FENG H, JIAO H L, ZHANG C Y, et al. The model and method of information security data leakage detection and prevention[C]//Proceedings of the 2024 International Conference on Data Science and Network Security (ICDSNS). Piscataway: IEEE Press, 2024: 1-5.
- [2] 高汉成, 吉浩宇, 高晨昊, 等. 基于区块链的个人健康护照存储、共享与验证方案[J]. 网络与信息安全学报, 2024, 10(2): 95-105.
- [3] GAO H C, JI H Y, GAO C H, et al. Individual health passport storage, sharing and verification scheme based on blockchain[J]. Chinese Journal of Network and Information Security, 2024, 10(2): 95-105.
- [4] LU Y X, JAFAR S A. On single server private information retrieval with private coded side information[J]. IEEE Transactions on Information Theory, 2023, 69(5): 3263-3284.
- [5] ANDOLA N, GAHLOT R, YADAV V K, et al. Searchable encryption on the cloud: a survey[J]. The Journal of Supercomputing, 2022, 78(7): 9952-9984.
- [6] WANG Y L, SUN S F, WANG J F, et al. Achieving searchable encryption scheme with search pattern hidden[J]. IEEE Transactions on

- Services Computing, 2022, 15(2): 1012-1025.
- [6] AMORIM I, COSTA I. Leveraging searchable encryption through homomorphic encryption: a comprehensive analysis[J]. Mathematics, 2023, 11(13): 2948.
- [7] CHENG L X, MENG F. Public key authenticated encryption with keyword search from LWE[C]//Proceedings of the 27th European Symposium on Research in Computer Security. Berlin: Springer, 2022: 303-324.
- [8] BEIMEL A, STAHL Y. Robust information-theoretic private information retrieval[J]. Journal of Cryptology, 2007, 20(3): 295-321.
- [9] XU W J, WANG B C, LU R X, et al. Efficient private information retrieval protocol with homomorphically computing univariate polynomials[J]. Security and Communication Networks, 2021(1): 5553256.
- [10] REN L, MUGHEES M H, SUN I. Simple and practical amortized sublinear private information retrieval using dummy subsets[C]//Proceedings of the 2024 on Conference on Computer and Communications Security. New York: ACM Press, 2024:1420-1433.
- [11] ZHOU J C, ZHU X, CHEN C. Efficient dynamic searchable encryption scheme based on bitslicing[J]. Netinfo Security, 2024, 24(9): 1309-1316.
- [12] 袁承昊, 李勇, 任爽. 多关键词动态可搜索加密方案[J]. 网络与信息安全学报, 2023, 9(2): 143-153.
- YUAN C H, LI Y, REN S. Dynamic multi-keyword searchable encryption scheme[J]. Chinese Journal of Network and Information Security, 2023, 9(2): 143-153.
- [13] MAKKONEN O, KARPUK D A, HOLLANTI C. Algebraic geometry codes for cross-subspace alignment in private information retrieval[C]//Proceedings of the 2024 IEEE International Symposium on Information Theory (ISIT). Piscataway: IEEE Press, 2024: 2874-2879.
- [14] 邵航, 高思琪, 钟离, 等. 同态加密在隐私计算中的应用综述[J]. 信息通信技术与政策, 2022, (8): 75-88.
- SHAO H, GAO S Q, ZHONG L, et al. Homomorphic encryption protocols and applications in privacy preserving computation[J]. Information and Communications Technology and Policy, 2022(8): 75-88.
- [15] LIU J F, LIN S. An efficient and more secure searchable encryption algorithm based on fully homomorphic encryption[C]//Proceedings of the 2023 International Conference on Automation, Control and Electronics Engineering (CACEE). Piscataway: IEEE Press, 2023: 83-92.
- [16] 牛淑芬, 刘文科, 陈俐霞, 等. 基于联盟链的可搜索加密电子病历数据共享方案[J]. 通信学报, 2020, 41(8): 204-214.
- NIU S F, LIU W K, CHEN L X, et al. Electronic medical record data sharing scheme based on searchable encryption via consortium blockchain[J]. Journal on Communications, 2020, 41(8): 204-214.
- [17] 刘炜, 白晓丹, 余维, 等. 基于倒排索引的可搜索加密数据共享方案[J]. 计算机工程与应用, 2023, 59(10): 270-279.
- LIU W, BAI X D, SHE W, et al. Searchable encrypted data sharing scheme based on inverted index[J]. Computer Engineering and Applications, 2023, 59(10): 270-279.
- [18] BONEH D, KUSHILEVITZ E, OSTROVSKY R, et al. Public key encryption that allows PIR queries[C]//Proceedings of the 27th Annual International Cryptology Conference(CRYPTO 2007). Berlin: Springer, 2007: 50-67.

[作者简介]



唐亮 (1982-), 男, 江苏宿迁人, 北京教育科学研究院高级工程师, 主要研究方向为教育数字化与网络安全、区块链、隐私计算。



马利民 (1983-), 男, 山东新泰人, 北京信息科技大学副教授, 主要研究方向为密码技术、区块链、隐私计算和数据安全。