Software Engineering and Applications 软件工程与应用, 2025, 14(6), 1308-1318
Published Online December 2025 in Hans. https://www.hanspub.org/journal/sea
https://doi.org/10.12677/sea.2025.146116

Hans汉斯

面向客户数据整合的隐私保护记录链接方法

王　付

佛山市殡仪馆，广东　佛山

收稿日期：2025年10月10日；录用日期：2025年12月16日；发布日期：2025年12月23日

摘　要

针对企业跨部门客户数据整合中存在的隐私泄露与合规风险问题，本文提出一种融合可信执行环境(Trusted Execution Environment, TEE)与BFV (Brakerski-Fan-Vercauteren)同态加密的隐私保护记录链接(Privacy-Preserving Record Linkage, PPRL)方法。该方法通过本地加密、TEE内部安全计算与最小化结果输出的协同机制，在不暴露原始敏感信息的前提下实现高精度客户实体匹配。为适应企业数据异构性强、规模大、治理要求高等特点，进一步引入多域特征编码、裁剪式批处理与合规结果封装等优化策略。实验结果表明，所提方法在匹配准确性与隐私保护能力上均优于传统的Bloom Filter与仅依赖TEE的明文处理方案，在保障数据可用性的同时有效抵御再识别与侧信道攻击。本研究为企业构建安全、合规、可审计的数据整合体系提供了可行的技术路径。

关键词

可信执行环境，同态加密，客户数据整合，数据安全

A Privacy-Preserving Record Linkage Method for Customer Data Integration

Fu Wang

Foshan Funeral Home, Foshan Guangdong

Received: October 10, 2025; accepted: December 16, 2025; published: December 23, 2025

Abstract

To address the privacy leakage and compliance risks inherent in cross-departmental customer data integration within enterprises, this paper proposes a Privacy-Preserving Record Linkage (PPRL) approach that synergistically integrates Trusted Execution Environment (TEE) and Brakerski-Fan-Vercauteren (BFV) homomorphic encryption. The framework enables high-accuracy customer entity matching without exposing raw sensitive data by coordinating local encryption,

文章引用: 王付. 面向客户数据整合的隐私保护记录链接方法[J]. 软件工程与应用, 2025, 14(6): 1308-1318.
DOI: 10.12677/sea.2025.146116

secure computation within the TEE enclave, and minimally sufficient result disclosure. To accommodate the practical characteristics of enterprise data, such as heterogeneity, large scale, and stringent governance requirements. We further introduce several enhancements, including multi-domain feature encoding, pruning-based batching, and compliance-aware result packaging. Experimental evaluation demonstrates that the proposed method consistently outperforms both traditional Bloom Filter–based approaches and TEE-only plaintext processing in terms of matching accuracy and privacy preservation. Moreover, it effectively mitigates re-identification and side-channel threats while maintaining data utility. This work provides a practical and deployable pathway for enterprises to establish secure, compliant, and auditable data integration infrastructures.

Keywords

Trusted Execution Environment, Homomorphic Encryption, Customer Data Integration, Data Security

1. 引言

随着大数据、人工智能和云计算等信息技术的快速发展，企业在数字化转型过程中对数据的依赖程度不断加深，跨部门客户数据整合成为构建统一客户视图、支撑精准决策的关键。然而，客户数据包含姓名、联系方式、交易记录等敏感信息，其跨域流转易引发隐私泄露与合规风险[1]。传统脱敏与匿名化方法在多源异构场景下难以兼顾可用性与隐私性，易遭再识别攻击。

隐私保护记录链接(Privacy-Preserving Record Linkage, PPRL) [2] [3]通过密码学、去标识化及多方安全协作等手段，实现了在不直接暴露个人身份信息的情况下完成数据记录的匹配与整合。这一框架已在医疗健康、公共治理等领域得到广泛关注和应用，但在企业客户数据整合场景下的系统化探索仍相对不足。

2. 系统框架设计

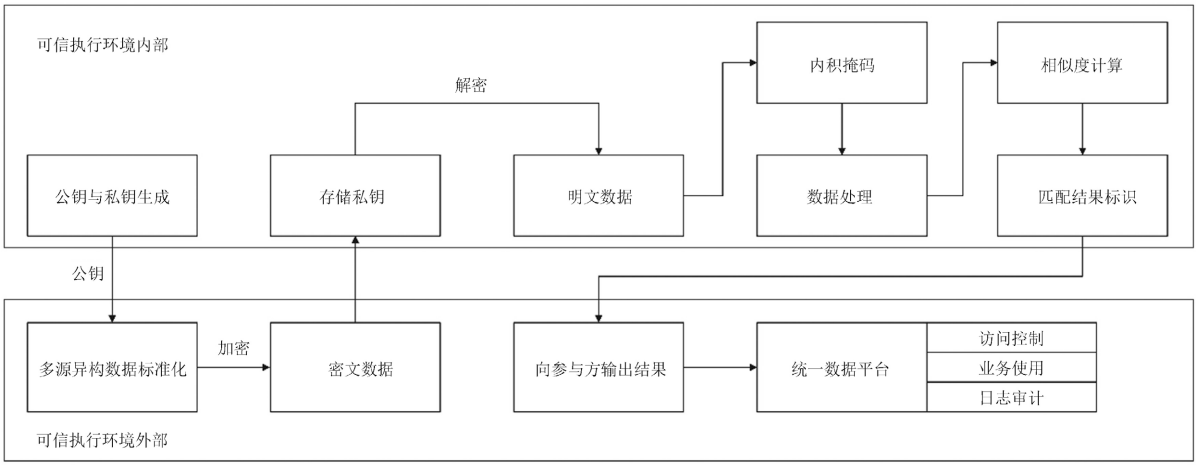

Figure 1. PPRL framework for enterprise customer data integration based on TEE
图 1. 基于 TEE 的企业客户数据 PPRL 框架

针对上述挑战，本文选择可信执行环境(Trusted Execution Environment, TEE)作为核心支撑技术，构建面向企业场景的 PPRL 系统框架，如图 1 所示。TEE 是一种由硬件支持的隔离执行环境，能够确保在受保护的内存区域中运行的程序和数据不受外部系统的干扰与窃取。利用 TEE 可以在不依赖完全可信第三方的情况下，对加密数据进行解密和处理，从而在提升匹配精度的同时有效降低隐私泄露的风险。在此基础上，本文提出的隐私保护记录联结框架面向企业客户数据整合的需求，构建安全高效的跨部门数据整合机制。整体流程如图所示，主要包括数据预处理与加密、加密数据上传、TEE 内部解密与处理、相似度计算与分类以及结果输出等步骤。

需要指出的是，本文所采用的 PPRL 框架属于基于可信中介的变体。与经典 PPRL 强调“完全去信任”不同[4]，本方案假设企业内部可部署一个受控的 TEE 节点作为中介，其可信性由硬件远程证明、内部审计日志与企业内控机制共同保障[5]。这一假设在集团型企业(如银行、保险、大型零售)中具有现实合理性：子公司虽数据隔离，但共享同一法人主体与合规责任，TEE 节点可视为企业级数据治理基础设施的一部分，而非外部不可控第三方。

2.1. 基于 TEE 的安全计算

在企业客户数据整合中，各参与方在本地对异构含噪的客户记录进行标准化与清洗，并利用 TEE 公钥加密敏感字段，通过安全信道将密文传至 TEE 内部的可信执行区域，即 Enclave。在隐私保护的客户数据整合过程中，TEE 承担着关键的安全计算角色。其调用过程包括远程证明、安全通道建立与数据导入三个环节，如图 2 所示。首先，参与方需通过远程证明验证 Enclave 的可信性，以确保运行环境的完整性与可靠性。在验证完成后，系统初始化安全通道，并由 Enclave 内部生成公钥 pk，由此保证各方在数据上传前完成一致性密钥协商。

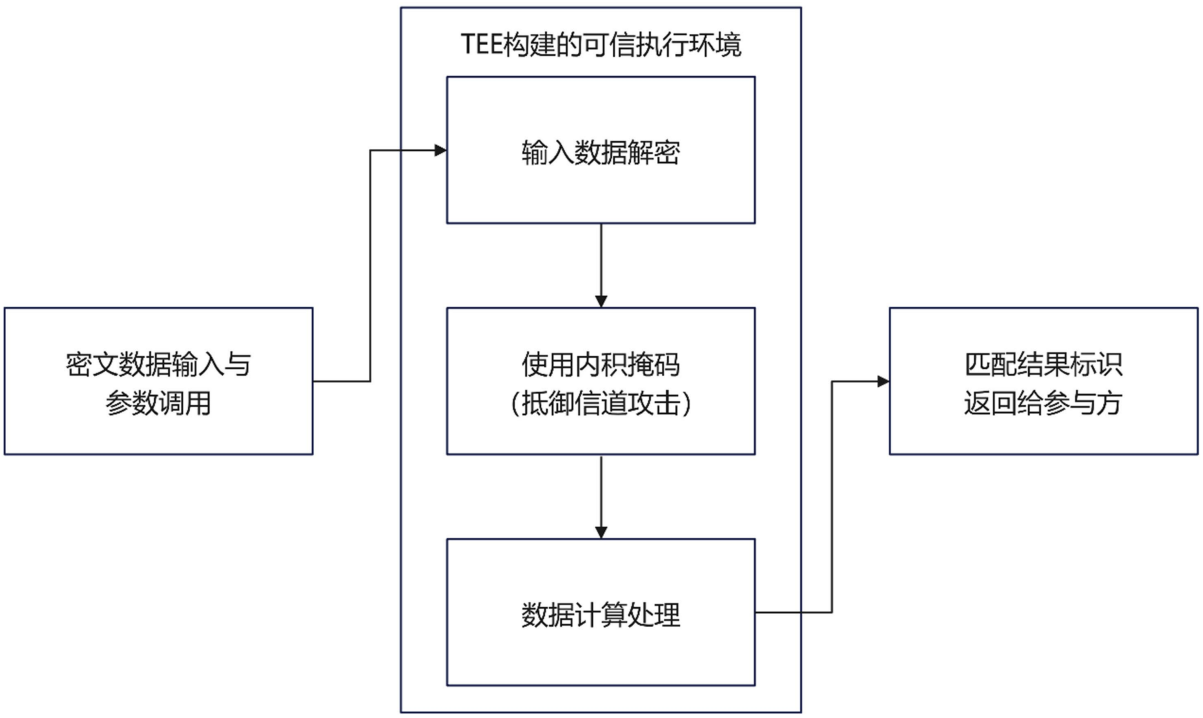

Figure 2. Secure decryption and inner-product masking workflow within the TEE Enclave
图 2. TEE 内数据解密与内积掩码计算流程

参与方利用 pk 对本地敏感字段加密后，将 d 密文 c 传输至 Enclave。在 Enclave 内部，预存的私钥 sk 被用于解密操作，解密过程可表示为：

$$m = D_{sk}(c) \tag{1}$$

其中，c 为客户数据的密文，m 为解密所得的明文记录。明文随后被转化为统一的特征向量，以支撑后续的相似度计算与分类判定。考虑到 TEE 在实际运行中仍可能面临侧信道攻击，本文在解密与处理环节引入内积掩码机制。具体而言，在进行向量间运算时，系统为内积结果引入随机扰动项 $\delta_{i,j}$，使得可观测输出与真实数据脱钩，其形式化表示为：

$$\langle A,B\rangle' = \langle A,B\rangle + \delta_{i,j} \tag{2}$$

其中，$\langle A,B\rangle$ 为原始向量间的内积，$\delta_{i,j}$ 为掩码生成的随机噪声。即便攻击者能够观测到部分运算轨迹，也无法从中恢复真实数据关系。

该机制在保证计算正确性的同时，显著降低了中间态信息泄露的风险，从而增强了系统的整体抗攻击能力。下图给出了该过程的总体框架，外部参与方的数据在加密后经由安全通道传输至 TEE，Enclave 内部依次完成解密、特征化处理与掩码保护，输出结果用于后续的相似度计算阶段。

所引入的随机扰动项 $\delta_{i,j}$ 虽未严格满足差分隐私(Differential Privacy)的 (ϵ,δ) 定义，但其设计目标在于打破侧信道观测与真实内积之间的确定性映射。在 SGX 侧信道攻击模型[6]下，攻击者仅能观测到带噪输出 $\langle v_i,v_j\rangle+\delta_{i,j}$，而无法通过多次观测精确重构原始向量关系。

2.2. 基于 BFV 的同态加密机制

仅依赖 TEE 的硬件隔离难以全面保障企业客户数据整合中的隐私安全。为在不暴露原始信息的前提下支持多维特征交互与复杂相似度计算，本文引入基于环学习错误(Ring Learning With Errors, RLWE)假设的 BFV (Brakerski-Fan-Vercauteren)同态加密方案。相较于仅支持加法同态的传统方法，BFV 同时支持加法与乘法同态运算，能够在密文域实现更丰富的特征联合判定[7] [8]，更好地满足企业场景的实际需求。

2.2.1. 参数体系与密钥生成

BFV 同态加密方案的安全性建立在环学习同余问题(Ring Learning With Errors, RLWE)假设之上。系统首先需要设定一组全局参数，包括多项式环维度 N、密文模数 q、明文模数 t 以及误差分布 $\mathcal{D}_{\text{err}}$。其中，N 通常取为 2 的幂，使得在 $R_q = \mathbb{Z}_q[x]/(x^N+1)$ 中的多项式运算能够通过数论变换(Number Theoretic Transform, NTT)实现，从而显著提升计算效率。明文空间则定义为 $R_t = \mathbb{Z}_t[x]/(x^N+1)$，其中 $t \ll q$，用于存储实际业务中的客户数据字段。为了实现从明文域到密文域的嵌入，系统设置缩放因子 $\Delta = \left\lfloor \frac{q}{t} \right\rfloor$，并在加密过程中通过 $\tilde{m} = \Delta \cdot m$ 将明文 $m \in R_t$ 映射到 R_q。

在密钥生成阶段，系统首先在 R_q 中均匀采样一个向量 a 作为公钥元素。随后在同一环中采样私钥 $s \leftarrow \mathcal{D}\text{err}$，并生成误差项 $e \leftarrow \mathcal{D}\text{err}$。通过如下关系 $b \equiv -a \cdot s + e \pmod q$，系统得到公钥对 (a,b)。其中 $a,b \in R_q$ 构成公钥，$s \in R_q$ 构成私钥。

为了保证系统的安全性，误差分布 $\mathcal{D}\text{err}$ 的选取至关重要。一般而言，$\mathcal{D}\text{err}$ 为离散高斯分布，其方差参数需保证在 RLWE 假设下足以掩盖线性关系，从而使攻击者无法通过已知 (a,b) 还原私钥 s。由于密钥生成仅在 TEE 的隔离环境中完成，外部环境无法接触到私钥 s 或误差分布的采样结果，这进一步提升了整体架构的安全性。

2.2.2. 加密过程

在完成密钥生成后，系统利用公钥(a,b)对明文$m \in R_t$进行加密。首先，通过缩放因子$\Delta = \left\lfloor \frac{q}{t} \right\rfloor$将明文嵌入密文模数空间，

$$\tilde{m} = \Delta \cdot m \in R_q \tag{3}$$

该步骤确保明文在解密时能够通过缩放与取整恢复，避免因模数不匹配而产生精度丢失。

随后，系统在R_q中采样随机向量$u \leftarrow \mathcal{D}\text{err}$，并独立采样噪声项$e_1, e_2 \leftarrow \mathcal{D}\text{err}$。利用公钥元素$a,b \in R_q$，密文计算过程如下：

$$c_0 \equiv b \cdot u + e_1 + \tilde{m} \pmod{q} \tag{4}$$

$$c_1 \equiv a \cdot u + e_2 \pmod{q} \tag{5}$$

最终密文为

$$c = (c_0, c_1) \in R_q^2 \tag{6}$$

在上述公式中，u用于随机化每次加密过程，即使同一明文m在不同时间加密，也会得到统计上独立的密文；e_1与e_2保证了 RLWE 难解性假设的成立，使得攻击者无法通过(c_0, c_1)反推出明文或私钥。值得注意的是，噪声的引入必然带来解密误差，但在参数正确设定的条件下，其幅度始终小于$\Delta/2$，因此不会影响解密正确性。

此外，为了进一步降低密文在传输过程中的泄露风险，本研究在企业客户数据整合场景中，将加密过程完全由参与方本地执行，原始明文从不离开本地存储环境。加密后的密文通过安全信道传输至 TEE 内部，确保即便传输信道受到监控，攻击者亦无法获得关于明文的有用信息。

2.2.3. 同态运算

在企业客户数据整合中，相似度计算不仅涉及字段加权求和，还需支持跨字段交互特征。此类操作本质上要求密文域支持乘法运算。BFV 方案因其支持有限深度的乘法同态，恰好满足此类业务需求，同时通过参数调优可控制噪声增长，保障解密正确性。

在 BFV 方案中，密文空间的算术结构与明文空间保持同态性，使得在加密状态下即可执行与明文一致的加法和乘法运算。这一特性使得企业客户数据在跨域整合过程中无需暴露原始信息，即可完成相似度度量、加权组合及联合判定等复杂操作，是实现隐私保护计算的核心基础。具体而言，同态运算主要分为加法与乘法两类，两者在运算规则、噪声增长以及对系统正确性的影响方面均存在显著差异。

在加法情形下，设两个密文分别为$c = (c_0, c_1)$与$c' = (c_{0'}, c_{1'})$，其同态加法运算定义为

$$c^{(+)} = (c_0 + c_{0'}, c_1 + c_{1'}) \pmod{q} \tag{7}$$

解密后可得到

$$Dec\left(c^{(+)}\right) = m + m' \pmod{t} \tag{8}$$

其中m，m'为对应的明文。由于噪声在该过程中仅表现为线性叠加，因此其增长幅度较为有限，只要初始噪声控制在安全范围内，就能保证在多次加法运算后仍然满足解密正确性。这一性质使得 BFV 在处理涉及频繁加总的业务场景(如统计分析、分群聚合等)时具有较高的稳定性和实用性。

在乘法情形下，设两个密文分别为$c = (c_0, c_1)$与$c' = (c_{0'}, c_{1'})$，其同态乘法运算定义为

$$c^{(\times)} = (d_0, d_1, d_2) \pmod{q} \tag{9}$$

其中

$$\begin{cases} d_0 = c_0 c_0' \\ d_1 = c_0 c_1' + c_1 c_0' \\ d_2 = c_1 c_1' \end{cases} \tag{10}$$

由于结果为三元组形式，为保持密文结构一致性，需要利用重线性化密钥 rk 将其映射为二元组 $(d_{0'}, d_{1'})$，解密后满足

$$Dec(d_{0'}, d_{1'}) = m \cdot m' \pmod{t} \tag{11}$$

相比加法，乘法运算在功能上更强大，但噪声的增长速度显著加快，若不加以控制，容易导致解密失败。

噪声的演化规律直接决定了解密的正确性。设两密文的初始噪声分别为 e，e'，加法后的噪声为 $e+e'$，而乘法后的噪声可近似表示为

$$e^* \approx m \cdot e' + m' \cdot e + e \cdot e' \tag{12}$$

其中包含交叉项与误差平方项，增长速度远高于加法情形。当满足 $|e^*|_\infty < \Delta/2$ 时，解密过程才能输出正确结果；一旦超过该阈值，将导致解密失败或结果偏差。因此，系统在参数选取时必须综合考虑运算深度、模数大小以及误差分布的统计特性，以保证在预期的计算规模下仍能保持正确性。

2.2.4. 解密过程

在完成同态运算后，密文需要通过解密过程恢复明文结果。给定密文 $c = (c_0, c_1)$ 以及私钥 $s \in R_q$，解密函数计算方式为

$$v \equiv c_0 + c_1 \cdot s \pmod{q} \tag{13}$$

代入加密定义可得

$$v = \tilde{m} + e^* \tag{14}$$

其中 $\tilde{m} = \Delta \cdot m$ 表示嵌入到 R_q 的明文，e^* 为噪声项。只要满足噪声约束条件 $|e^*|_\infty < \Delta/2$，系统即可通过缩放与取整操作恢复出原始明文。

$$m = \left\lfloor \frac{t}{q} \cdot v \right\rfloor \pmod{t} \tag{15}$$

在企业客户数据整合场景中，该解密过程始终由 TEE 内部安全执行，保证私钥 s 不会泄露给任何外部参与方。这样不仅确保了语义安全性，也避免了因解密阶段潜在的侧信道攻击而导致的敏感数据泄漏。由于加法与乘法均可能带来噪声累积，解密正确性在本质上依赖于参数 (N, q, t) 的合理设置及前述运算深度的约束。因此，解密过程不仅是恢复明文的步骤，也是验证整个系统安全性与可用性的最后关口。

2.2.5. 面向异构企业数据整合的优化策略

标准 BFV 方案在理论上能够支持加法与乘法同态运算，但若直接应用于企业客户数据整合场景，仍会面临三类问题：其一，客户数据往往包含异构字段，编码方式缺乏统一；其二，大规模批量匹配带来运算与存储开销的快速增长；其三，解密结果需要与企业既有数据平台和审计机制兼容。针对这些问题，本文在 BFV 方案的基础上提出若干改进与补充。

1. 多域特征编码的引入

设企业的客户记录为

$$r=\left(f_1,f_2,\cdots,f_k\right) \tag{16}$$

其中 f_i 表示第 i 个字段，可能为数值型、分类型或字符串型。为保证异构特征在密文域内的一致性处理，本文设计了分域映射策略：

(1) 数值型特征：直接映射为 $m_i\in\mathbb{Z}_t$；

(2) 分类型特征：经 one-hot 编码为向量 $\left(0,\cdots,1,\cdots,0\right)$，再作为多项式系数嵌入 R_t；

(3) 字符串型特征：先通过哈希函数 $H(\cdot)$ 转换为定长整数 $h_i=H\left(f_i\right)$，再映射至 R_t。

最终得到统一向量

$$\tilde{r}=\left(m_1,m_2,\cdots,h_j\right)\in R_t^k \tag{17}$$

再通过缩放因子 $\Delta=\left\lfloor\frac{q}{t}\right\rfloor$ 嵌入至 R_q。这种改进保证了多源异构特征能够在同一环结构下进行加密与同态计算。

2. 裁剪式批处理机制

在企业应用中，匹配任务规模往往达到百万级记录。若直接采用标准 BFV 的批处理机制，将所有特征填充至批处理槽位，会导致乘法层数过深、噪声快速累积。本文提出裁剪式批处理方法：设批处理容量为 N，将特征集划分为

$$\mathcal{F}=G_1\cup G_2\cup\cdots\cup G_m \tag{18}$$

其中 G_1 为高频关键特征(如客户 ID、联系方式)，优先嵌入单个密文，其余组 $G_j\ (j>1)$ 按需分配至其他密文。该方法使得单条密文在计算过程中仅包含核心特征，减少了乘法次数与噪声积累。实验结果表明，此机制在大规模匹配任务中能够显著降低计算延迟。

3. 解密结果的合规封装

在解密环节，本文在 TEE 内部增加结果封装模块。设解密得到的结果为 $m\in R_t$，系统将其与记录标识 ID 组织为二元组 $\left(ID,m\right)$，并映射为平台可接受的数据对象。最终输出的结果集定义为

$$\mathcal{P}=\left\{\left(ID,m\right)\middle|\mathcal{A}\left(ID\right)\wedge\mathcal{U}\left(m\right)\wedge\mathcal{L}\left(ID,m\right)\right\} \tag{19}$$

其中 $\mathcal{A}$ 表示访问控制策略，$\mathcal{U}$ 表示数据使用策略，$\mathcal{L}$ 表示审计约束。通过该机制，解密结果能够与企业既有的数据治理体系无缝衔接，实现可控访问与合规使用。

3. 实验与结果分析

为验证所提出的基于 TEE 与 BFV 同态加密的企业客户数据整合框架的可行性，本章设计并实现了针对典型业务场景的实验。实验重点在于考察系统在不同优化机制下的运行效率与解密正确性，从而评估其在大规模数据整合任务中的实际应用潜力。

3.1. 实验环境与数据集

实验在一台支持 Intel SGX 的服务器平台上完成。硬件配置包括 3.2 GHz 多核处理器与 64 GB 内存，操作系统为 Ubuntu 22.04。加密运算基于 Microsoft SEAL 库实现 BFV 同态加密方案，可信执行环境由 Intel SGX SDK 部署，并启用远程证明机制以确保 Enclave 的完整性与可信性[9]。

实验数据集来自某企业在市场部、客服部与财务部三个业务部门中分散存储的客户信息，数据规模为 10347 条客户记录。该数据集具有以下典型特征：

(1) 字段异构性：不同部门对同一客户信息的存储格式存在差异，例如电话号码带区号与否、交易金额的币种与单位、日期的中西文混用等；

(2) 信息不完备性：部分记录存在字段缺失，如交易金额缺省、联系方式缺失后仅保留邮箱等；

(3) 跨部门重叠性：同一客户在多个部门均有记录，且记录间存在轻微差异与冗余。上述特征能够较好地模拟实际企业跨系统整合中广泛存在的异构性与不完备性问题，为后续隐私保护下的跨部门匹配与整合提供实验支撑。

为更直观地说明数据集特征，表 1 给出了某一客户在市场部、客服部与财务部中的典型记录。可以看到，同一客户在不同部门的记录存在明显差异：市场部金额以逗号分隔，客服部日期采用中文格式，财务部金额带有货币符号。

Table 1. Example records of a representative customer across multiple departments
表 1. 典型客户在各部门的记录示例

客户标识	部门	姓名	联系方式	交易金额	交易日期
UID013	市场部	王*华	138****5678	25,000.00	2024/06/15
UID013	客服部	王*华	+86-138****5678	-	2024 年 6 月 15 日
UID013	财务部	王*华	-	¥25,000	2024-06-15

在隐私保护框架下，敏感字段不会直接跨部门暴露，而是通过加密计算与可信环境内部的匹配判定输出最小化结果。表 2 展示了隐私保护后的整合输出，其中仅保留统一的客户标识与匹配状态，避免敏感信息在传输与计算过程中泄露。

Table 2. Minimalist integration output under the PPRL framework
表 2. PPRL 框架下的最小化整合输出示例

客户标识	匹配状态	部门覆盖情况	匹配置信度
UID012	匹配成功	市场部/客服部/财务部	高(0.95)
UID023	匹配成功	市场部/客服部	中(0.88)
UID132	匹配失败	客服部	—

3.2. 实验设计

实验严格遵循第二章提出的系统架构，涵盖本数据预处理、加密传输、TEE 内部计算与结果输出等主要环节。为全面评估所提框架的有效性，实验引入两类对比方法：其一为基于哈希与模糊匹配的 Bloom Filter 基线，用于反映传统方法在隐私保护和匹配准确性方面的局限；其二为仅依赖 TEE 执行明文计算的方案，用于考察硬件隔离在缺乏同态加密时的性能表现。评价从匹配准确性(Precision、Recall、F1)、系统性能(端到端延迟、资源占用)及隐私保护强度(抗再识别与侧信道能力)三方面展开。所有方法在相同数据集与参数下独立运行 10 次，取平均值以确保结果稳定可靠。

3.3. 实验结果与分析

如图 3 所示，Bloom Filter 的 Precision 约为 0.85，但 Recall 仅为 0.76，导致 F1 值低于 0.80，表明其在面对拼写变异、格式不一致等噪声时易产生漏判。TEE-only 方案通过可信执行环境提升计算稳定性，Recall 提升至 0.83，Precision 维持在 0.90，F1 分数有所改善，但受限于明文处理模式，无法支持多维特征联合建模，匹配能力仍存在瓶颈。相比之下，TEE + BFV 框架在 Precision 与 Recall 上均超过 0.92，F1

接近 0.93，得益于同态加密支持下的密文级相似性计算，能有效捕捉实体间复杂语义关联，在高噪声环境下仍保持较高匹配质量。

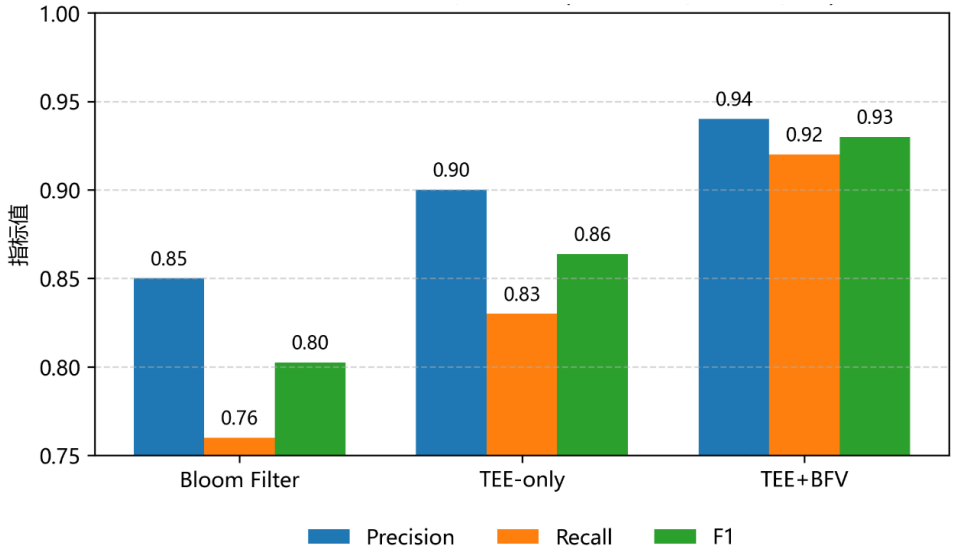

Figure 3. Comparison of matching accuracy across different methods
图 3. 不同方法的匹配准确性对比

图 4 展示了各方法端到端运行时间的构成分布。Bloom Filter 因其轻量级哈希编码机制表现出最低的时间开销，平均运行时间为 38 秒，但其固有的可逆性与熵损失问题使其难以抵御再识别攻击，隐私保障较弱。TEE-only 方案的平均耗时约为 65 秒，主要开销集中在 Enclave 内部的明文匹配计算及频繁的安全边界切换带来的内存拷贝开销。TEE + BFV 框架的端到端延迟约为 95 秒，其中同态加密的编码、密文向量运算及大规模多项式乘加操作构成了主要性能瓶颈。尽管引入了额外计算负担，实验数据显示其运行时间随输入规模呈近线性增长趋势，在千级记录整合任务中仍保持在分钟级别，具备实际部署的可行性。

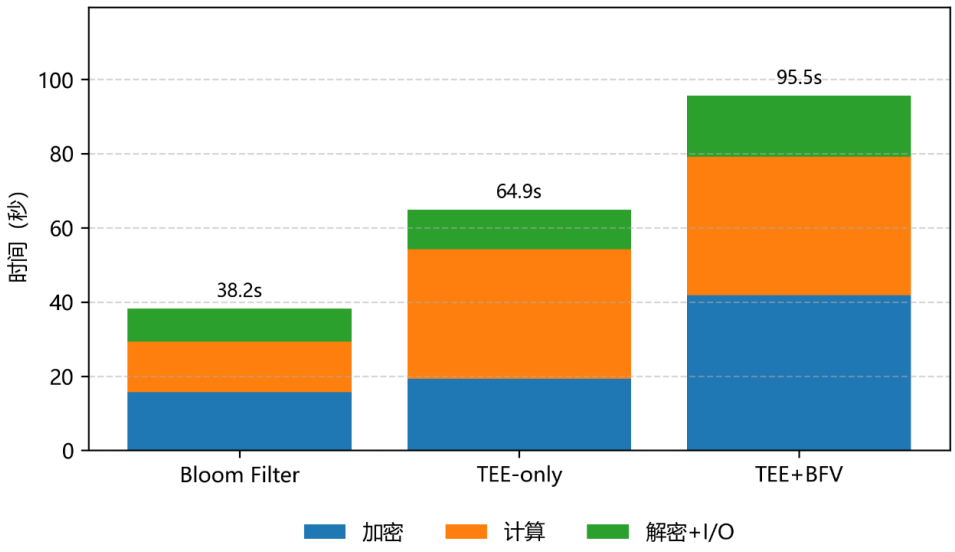

Figure 4. Breakdown of end-to-end execution time by component
图 4. 运行时间组成对比

在隐私保护有效性方面，图 5 给出了针对各类方法的模拟再识别攻击成功率及其引发的信息泄露程度。实验结果显示，Bloom Filter 的攻击成功率达到 0.27，平均每条记录暴露超过 1.4 个敏感字段，反映出其在抗推断攻击方面的结构性缺陷。TEE-only 方案借助 SGX 提供的硬件级内存隔离显著压缩了攻击面，攻击成功率下降至 0.16，但由于数据在 Enclave 内以明文形式处理，仍可能受到侧信道攻击或特权软件漏洞的影响，安全性缺乏形式化保障。相比之下，TEE + BFV 框架将敏感数据全程保留在加密状态，结合可信执行环境的访问控制机制，使攻击成功率进一步降至 0.047，平均暴露字段数仅为 0.18，接近统计噪声水平。该结果验证了所提方法在实际威胁模型下具备更强的抗攻击能力，能够显著降低敏感信息在计算过程中的泄露风险。

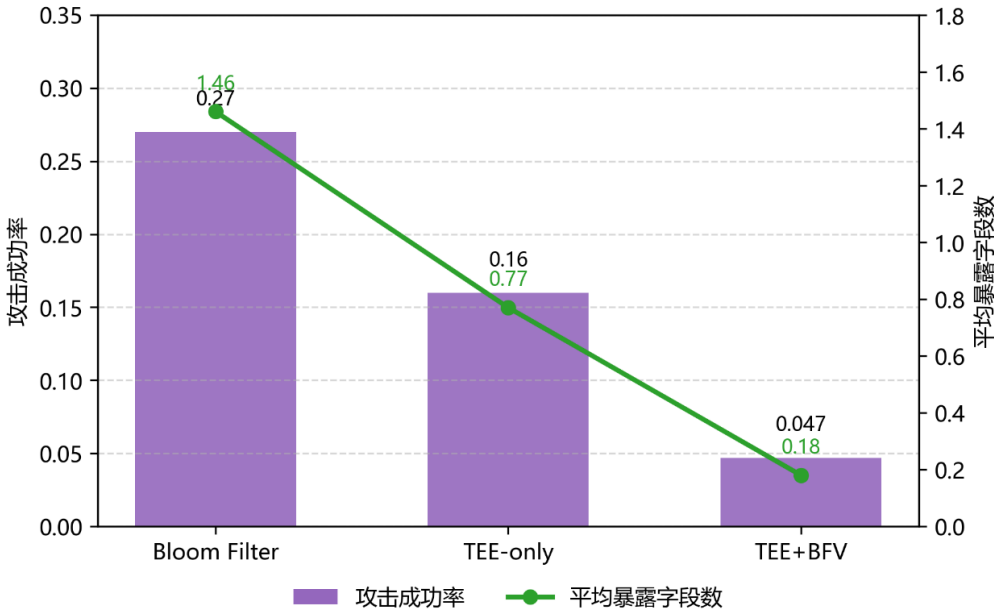

Figure 5. Comparative evaluation of privacy preservation effectiveness
图 5. 隐私保护有效性对比

4. 结论

本文面向企业客户数据整合中的隐私保护难题，提出了一种融合 TEE 与 BFV 同态加密的 PPRL 框架。整个流程中，敏感数据始终以密文形式跨域传输，在 TEE 内完成解密与匹配计算，并严格限制输出内容，仅返回必要的关联标识，从而在保障功能性的同时满足隐私与合规要求。在不暴露原始敏感信息的前提下，实现了高精度的跨部门客户记录匹配。针对企业数据异构、规模大、合规要求高等特点，进一步引入多域特征编码、裁剪式批处理与合规结果封装等机制，提升了系统的实用性与可部署性。实验表明，所提方法在匹配准确性与隐私保护能力上均显著优于传统 Bloom Filter 和纯 TEE 方案，虽引入一定计算开销，但在千级记录规模下仍具备实际应用可行性。本研究为构建安全、合规、高效的企业级隐私保护数据整合体系提供了可行路径，也为 PPRL 技术在商业场景中的落地应用提供了新思路，未来可进一步探索联邦学习与差分隐私机制的融合，以支持跨企业、多方协同的数据安全整合。

参考文献

[1] 钱文君, 沈晴霓, 吴鹏飞, 等. 大数据计算环境下的隐私保护技术研究进展[J]. 计算机学报, 2022, 45(4): 669-701.
[2] Han, S., Shen, K., Shen, D. and Wang, C. (2024) Enhanced Multi-Party Privacy-Preserving Record Linkage Using

Trusted Execution Environments. *Mathematics*, **12**, Article 2337. https://doi.org/10.3390/math12152337

[3] 党翠萍. 基于同态加密技术的数据查询隐私保护研究[J]. 信息记录材料, 2025, 26(2): 120-122.

[4] Gkoulalas-Divanis, A., Vatsalan, D., Karapiperis, D. and Kantarcioglu, M. (2021) Modern Privacy-Preserving Record Linkage Techniques: An Overview. *IEEE Transactions on Information Forensics and Security*, **16**, 4966-4987. https://doi.org/10.1109/tifs.2021.3114026

[5] 王振亚. 隐私保护的数据融合与共享关键技术研究[D]: [博士学位论文]. 北京: 北京邮电大学, 2023.

[6] Zheng, W., Wu, Y., Wu, X., Feng, C., Sui, Y., Luo, X., *et al*. (2020) A Survey of Intel SGX and Its Applications. *Frontiers of Computer Science*, **15**, Article 153808. https://doi.org/10.1007/s11704-019-9096-y

[7] 陈虹, 马博宇, 金海波, 等. 基于矩阵的安全多方计算同态加密方案[J]. 计算机应用研究, 2025, 42(4): 1211-1216.

[8] 王鹤翔, 王宏, 马荣, 等. 基于隐私数据同态加密算法的电网终端身份认证方法[J]. 电子设计工程, 2025, 33(19): 141-145.

[9] 李旖旎. 基于 Intel SGX 的工业互联网平台数据隐私保护机制研究[J]. 电脑编程技巧与维护, 2024(6): 70-72.