计算机科学
COMPUTER SCIENCE

支持模糊匹配的带标签隐私集合交集计算协议

程恩泽, 张蕾, 魏立斐

引用本文

程恩泽, 张蕾, 魏立斐. 支持模糊匹配的带标签隐私集合交集计算协议[J]. 计算机科学, 2024, 51(12): 343-351.

CHENG Enze, ZHANG Lei, WEI Lifei. Fuzzy Labeled Private Set Intersection Protocol[J]. Computer Science, 2024, 51(12): 343-351.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.231000131

支持模糊匹配的带标签隐私集合交集计算协议

程恩泽[1]　张　蕾[1]　魏立斐[2]
1 上海海洋大学信息学院　上海 201306
2 上海海事大学信息工程学院　上海 201306
(chengenze_0220@163.com)

摘　要　支持模糊匹配的带标签隐私集合交集计算协议(Fuzzy Labeled Private Set Intersection,FLPSI)是PSI协议的变体,其特点在于发送方与接收方的集合元素并不完全相等,而是存在相似性,且发送方集合中的每个元素均关联一个标签,接收方仅得到相似匹配元素的标签,而不会泄露其他信息。现有的FLPSI协议大多使用汉明距离来判断二进制向量之间的匹配程度,协议基于昂贵的公钥密码来构建,计算开销大导致协议运行缓慢。对此,提出了一种基于对称密码构造的更加高效的FLPSI协议,通过模拟范例证明了协议在半诚实模型下是安全的,参与方均无法窃取额外的隐私信息。与现有方案相比,协议将整体通信复杂度与发送方的计算复杂度由$O(n^2)$降低为$O(n)$。实验仿真结果表明,所提方法在平衡场景下比现有FLPSI协议快3～10倍,通信量降低89%～95%;在非平衡场景下比现有FLPSI协议快7～10倍,与类似的模糊匹配协议相比具有明显优势。此外,还设计了FLPSI协议在隐私保护条件下人脸识别的应用,通过调整参数可以满足不同场景的要求。

关键词:隐私集合交集;模糊匹配;标签匹配;秘密共享;隐私计算
中图分类号　TP309

Fuzzy Labeled Private Set Intersection Protocol

CHENG Enze[1],ZHANG Lei[1] and WEI Lifei[2]
1 College of Information Technology,Shanghai Ocean University,Shanghai 201306,China
2 College of Information Engineering,Shanghai Maritime University,Shanghai 201306,China

Abstract　Fuzzy labeled private set intersection(FLPSI) is a variant of PSI where the elements in the sender's and receiver's sets are not the same but rather have some similarities. Each element in the sender's set is associated with a label,and the receiver only receives the labels of the matched elements and without revealing other information. Most existing FLPSI protocols use Hamming distance to determine the degree of matching between binary vectors. These protocols are built based on expensive public key ciphers,which requiring high computation overhead and resulting in slow running time. This paper proposes an efficient FLPSI protocol based on symmetric cryptography. It proves the security of the PSI protocol in the semi-honest model,ensuring that participants cannot obtain additional data. Compared to the existing schemes,the protocol reduces the overall communication complexity and the computational complexity of the sender from $O(n^2)$ to $O(n)$. Through experimental simulation,in balanced scenarios,the proposed protocol is 3～10x faster than the existing FLPSI protocol,and the communication is reduced by 89% to 95%. In unbalanced scenarios,the proposed protocol is 7～10x faster than the existing FLPSI protocol,and it also exhibits obvious advantages over similar fuzzy matching protocols. In addition,the application of FLPSI protocol in face recognition under privacy protection conditions is designed,which can meet the requirements of different scenarios by adjusting parameters.

Keywords　Private set intersection,Fuzzy matching,Labeled matching,Secret sharing,Privacy preserving computation

1　引言

隐私集合交集(Private Set Intersection,PSI)是安全多方计算的一种特殊情况,经典的PSI协议中参与方可以计算其私有集合的交集,最终由一方或多方获得集合交集,且不会泄露任何额外信息[1-2]。PSI作为安全多方计算中的重要密码学工具,被广泛应用于数据挖掘和人工智能领域,如隐私保护的数据挖掘[3-4]、安全的人类基因检测[5]和隐私通信录查找[6]等。

在传统的PSI协议中,参与比较的集合元素完全相同时,

到稿日期:2023-10-20　返修日期:2024-04-02
基金项目:国家自然科学基金面上项目(61972241);上海市自然科学基金面上项目(22ZR1427100);上海市软科学研究项目(23692106700)
This work was supported by the National Natural Science Foundtion of China (61972241), Natural Science Foundation of Shanghai (22ZR1427100) and Soft Science Project of Shanghai(23692106700).
通信作者:张蕾(Lzhang@shou.edu.cn)

才会输出准确的结果。但是在现实应用场景中,需要进行隐私计算的元素通常不是完全相同的,如安保系统中的人脸识别[7]功能,深度学习模型对同一人的面部在不同情况下提取的特征向量往往是不完全相同的。模糊匹配的隐私集合交集(Fuzzy PSI,FPSI)协议[8-14]作为 PSI 协议的变体,指在隐私保护的前提下,对参与者输入的元素集合之间进行模糊近似匹配,并输出匹配的结果。FPSI 技术可以在不泄露数据集元素的情况下,找出两个数据集中相似的元素,同时解决了隐私保护与数据模糊匹配的问题,为数据处理和隐私保护提供了可行的方案。针对不同的度量方式,研究人员提出了不同的 FPSI 方案,由于二进制向量具有高存储效率和计算效率的优势,因此二进制向量间汉明距离的相似匹配研究最为广泛。

支持模糊匹配的带标签隐私集合交集计算协议(Fuzzy Labeled Private Set Intersection,FLPSI)进一步扩大了 FPSI 协议的应用范围,其指发送方对于自身元素额外持有一个标签,接收方输出交集项目中对应的标签,对于其他信息一无所知,以实现私有信息检索的功能。与带模糊关键词的可搜索加密不同,该技术通常将数据加密,并构建索引数据结构,以便允许用户在加密数据集中进行搜索操作,而 FLPSI 协议中参与方数据无需本地加密存储,且能在保护参与方隐私的同时实现集合元素间的模糊匹配操作。FLPSI 协议具有广泛的实际应用场景,例如在生物识别系统中,不仅要实现生物特征相似度的比对,同时希望输出该生物特征对应的身份标签。

目前,在协议运行时间上有优势的 FLPSI 方案[13]采用客户端-服务器模型,适用于以汉明距离大小为相似度标准进行二进制向量之间的模糊匹配。客户端为了避免隐私泄露,先采用同态加密将明文向量加密成密文,再由服务器将密文编码为多项式进行相等性测试。虽然此方案中双方在离线阶段对数据进行了预处理,但是整体协议依然基于公钥密码学构建,导致方案的计算开销非常大,难以满足实际需求。因此,本文基于秘密共享、不经意伪随机函数和键值对打包技术,提出了一个更加高效的 FLPSI 协议,该协议同样适用于二进制向量间汉明距离的相似匹配。在该协议中,假设各参与方均为半诚实参与方,双方不会恶意地破坏协议的执行;参与双方在协议的在线阶段只需进行一次交互,由客户端发送查询请求并获得查询结果,而服务器在计算过程中无法获得客户端的任何隐私数据。该协议的构建主要基于对称加密框架,降低了计算开销。本文的主要贡献如下:

1)提出了一个高效的支持模糊匹配的带标签隐私集合交集计算协议,避免大量使用公钥操作,实现了交集的模糊匹配计算协议,保护了双方集合元素的隐私,也使接收方仅获得模糊交集中标签的信息,并使用模拟范式证明了该协议在半诚实安全模型下的安全性。

2)提出的协议的计算复杂度仅与集合的大小线性相关,通过仿真实验测试协议的计算开销与通信开销。结果表明,本文提出的协议相比最近的 FLPSI 方案[13]计算速度提高了 3～10 倍,并且在平衡场景下具有更优秀的通信负载。

3)与相关的模糊匹配协议相比,本文提出的协议在运行时间与通信量方面均具有明显的优势,实现了 FLPSI 协议在人脸识别中的应用,可以通过改变相关参数来满足不同场景对安全性和效率的要求。

2 相关工作

PSI 作为安全多方计算领域的热点问题,主要针对计算效率高、通信负载低两方面进行研究,一般来说主要有两种构造方法。第一种是基于通用电路的构造,参与方使用混乱电路进行逻辑或算数运算,尽管基于电路的协议可以灵活地适应 PSI 功能的变体,但构造的复杂度会随电路深度的增加而增加,而交集功能只能通过评估较大电路实现,因此基于此技术的 PSI 协议效率较低。另一种方法主要依赖于密码原语和密码学假设构造的专用 PSI 协议,基于公钥加密的 PSI 协议通常具有较低的通信成本[15-16],但其采用繁琐的公钥操作,产生了巨大的计算成本,导致运行效率非常低。首个真正实现 PSI 功能的协议[17]是基于 Differ-Hellman 假设[18]提出的,目前很多协议依然基于该假设设计,这些协议一般具有通信量较低的优势。

目前多数的 PSI 协议基于不经意传输(Oblivious Transfer,OT)框架实现,基于 OT 的 PSI 协议实现了计算和通信成本之间的良好平衡,目前局域网条件下速度最快的两方 PSI 协议来自于文献[19],其使用纠错码对选择向量进行编码,使元素相等性判定执行的 OT 次数不依赖于元素的长度,实现了大集合下实用的 PSI 协议。基于双方 OT 的 PSI 协议往往借助不经意伪随机函数设计,文献[20]基于 1-out-of-2 OT 设计出第一个单边恶意的 OPRF,文献[19]基于随机 OT 构造出单点 OPRF,文献[21]基于稀疏 OT 的扩展实现了多点 OPRF 操作,进一步降低了 PSI 协议的通信复杂度。同时,由于 OT 扩展协议可以并行化执行,因此基于 OT 的 PSI 协议也可以进行并行计算,提高了效率。

不经意键值对存储(Oblivious Key-Value Stores,OKVS)技术通过选择不同的数据结构优化存储空间、编码时间、解码时间,用于提升 PSI 协议的性能。文献[22]基于多项式插值的键值对打包技术实现具有信息论安全的隐私集合交集方案,通信开销较低,但使用多项式插值求解系统需要的时间复杂度为 $O(n\log^2 n)$。文献[23]分别介绍了基于混淆布隆过滤器的 OKVS 技术和基于表的 OKVS 技术,前者的通信复杂度为 $O(n)$,但其常系数较高,后者具有较小的通信开销与计算成本。文献[24]构造了一种新的 OKVS 数据结构——PaXoS(Probe and Xor of String,PaXoS)数据结构,其计算开销与存储空间是目前 PSI 协议中最优秀的。

现有的 PSI 协议大多针对精确匹配的场景,而 FPSI(Fuzzy Private Set Intersection,FPSI)协议更关注模糊匹配的场景。需要强调的是,FPSI 不同于阈值 PSI,前者关注两个元素是否近似,若满足条件,则输出对应元素,后者关注两个集合中相同元素的数量与阈值的关系,若相同元素数量大于阈值,则输出集合中的相同元素,本文讨论的是 FPSI。文献[8]借助多项式插值首次实现 FPSI 协议,文献[9]指出文献[8]方案中,半诚实的客户端可能会获得额外的隐私信息,并借助秘密分享(Secret Sharing,SS)构建了保护隐私的 FPSI 协议。文献[11]借助混乱电路提出了汉明距离度量方式下的 FPSI 协议,其通信成本与二进制向量长度线性相关。文献[12]

使用同态加密与不经意线性评估（Vector Oblivious Linear Evaluation，VOLE）实现了类似功能，作者指出通信成本与向量长度无关，而与汉明距离阈值的平方线性相关，因此提出了曼哈顿距离下的 FPSI 协议。文献[14]结合 OKVS 与函数秘密分享技术实现了结构感知的 FPSI 协议。

FLPSI 协议[13]是在 FPSI 协议的基础上更进一步的研究，其关注集合元素携带的对应信息，目前的研究较少。文献[13]首次提出 FLPSI 的概念，参与双方分别拥有二进制向量的数据集合，发送方额外拥有标签集合。在在线阶段，客户端发送自身集合元素的不同次幂，服务器为数据和标签分别构建多项式插值，并将其编码为密文向量发送给客户端，客户端本地解码多项式获得对应的标签信息。通过对元素汉明距离的度量，最终客户端获得近似元素对应的标签。协议总体的计算复杂度与通信复杂度均为 $O(n^2)$，对于计算能力和通信较弱的客户端来说是很大的计算负担。

3 预备知识

3.1 半诚实安全模型

半诚实安全模型（Semi-honest Security Model）是密码学中的一种安全分析模型，半诚实参与者在协议执行的过程中会遵守协议要求，但会保存协议的中间计算状态及所能收集到的一切信息，并试图利用这些信息推测额外的隐私信息，但无法干涉其他参与方诚实的执行协议。

在安全多方计算领域，常见的方式是采用理想-现实模型[25]（Ideal-real Model）证明协议的安全性。在这个理想的世界中，存在一个完全可信的实体，可以接收各方的秘密输入，并进行计算，然后返回计算结果给各方，这种理想世界的执行方式可以满足协议所需的安全性。因此，在半诚实安全协议的设计中，通常假设在理想世界中存在与现实世界能力相当的任何半诚实敌手。如果在实际协议的执行过程中，参与方在真实协议的视图与理想协议的视图是不可区分的，即在现实世界中执行协议并未泄露更多的消息，那么就可以称该协议是半诚实安全的[26]。

本文提出的协议设计由接收方和发送方构成，其中接收方持有集合 X，发送方持有集合 Y 和 L。为证明协议的安全性，需证明真实协议的视图与理想协议的视图是不可区分的。设 f 为协议计算函数，I 可表示协议的参与双方，参与方在执行输入为 $(X;Y,L)$ 的协议 π 时，其视图表示为 $VIEW_I^{\pi}(X;Y,L)=(\bar{I};M_I^1,M_I^2,\cdots,M_I^T)$，其中 $\bar{I}$ 表示参与方 I 的输入，M_I^T 表示 I 收到的第 T 条消息。

定义 1 在半诚实模型下，$f(X;Y,L)$ 表示协议的输出，如果存在概率多项式时间算法的模拟器 Sim_I，对于任意参与方 I，均有下述等式成立：

$$Sim_I(\bar{I};f(X;Y,L))\overset{c}{\equiv}VIEW_I^{\pi}(X;Y,L)$$

则能证明 π 安全地计算 f，其中 $\overset{c}{\equiv}$ 表示在计算上不可区分。

3.2 秘密分享

秘密分享作为一种基础的密码学技术，分为秘密分享与秘密重构两个阶段。在 (t,n) 秘密分享方案中，秘密 s 在 n 方之间分配，秘密分享阶段将一个秘密 s 分享给不同的参与者；在秘密重构阶段，只有至少 $t(t\leq n)$ 个参与方的信息组合在一起时，才能还原出初始秘密 s。本文协议采用基于拉格朗日插值法的 Shamir 秘密共享[27]。

秘密分享阶段：秘密的分发者为秘密 s 随机构造多项式 $f(x)=s+a_1x+a_2x^2+\cdots+a_{t-1}x^{t-1}$，随机选取 n 个元素 $x_1,x_2,\cdots,x_n$ 并计算秘密分享值 $y_i=f(x_i)$。秘密分享者将 (x_i,y_i) 分别秘密地发送给 n 个参与者。秘密重构阶段：任意 t 个参与方可使用下述公式重构出秘密信息 s：

$$s=\sum_{i=1}^{t}\left[y_i\prod_{j=1,j\neq i}^{t}\frac{-x_j}{x_j-x_i}\right]$$

3.3 不经意键值对存储

不经意键值对存储[28]是键值对存储（Key-Value Stores，KVS）的变体，指能够在隐藏 key 和 value 的前提下，保留 key-value 映射关系的数据结构。键值对存储由键的集合 K、值的集合 V 和 $Encode()$ 算法、$Decode()$ 算法组成，$Encode()$ 算法接收键值对 (k_i,v_i) 集合作为输入，并返回编码后数据结构 S；Decode() 算法接收数据结构 S 与密钥 k 作为输入，并返回相应的值 v。

定义 2 对于两组不同的密钥 $\{k_1^0,k_2^0,\cdots,k_n^0\}$ 和 $\{k_1^1,k_2^1,\cdots,k_n^1\}$ 分别选取 n 个随机值 v_i，并使用 Encode() 算法编码。如果返回的数据结构 S_0 和 S_1 在计算上是不可区分的，那么称该数据结构为 OKVS。

OKVS 的正确性：对于所有 $A\in K\times V$，均有 $Encode(A)=S$，并且所有的 $k\in K,v\in V$ 均有 $Decode(S,k)=v$ 成立，则可认为该 OKVS 是正确的。

OKVS 的完备性：如果 OKVS 对随机值进行编码，那么对于任意两组密钥 K^0 和 K^1，攻击者无法区分对 K^0 密钥进行 OKVS 编码的结果和对 K^1 密钥进行 OKVS 编码的结果。

本文提出的协议采用 PaXoS 实例化 OKVS，它将 n 个二进制字符串映射到 m 个二进制字符串中，对于原始字符串中的每个元素，都可以通过 m 个字符串的特定子集进行异或来检索。为了提高编码与解码效率，采用 Cuckoo hash[29]算法，将 n 个元素散列到 Cuckoo Table 中。在这个过程中，一个元素会被映射在多个位置上，使它们存储的值异或运算后等于该元素。

3.4 不经意伪随机函数

不经意伪随机函数（Oblivious Pesudorandom Function，OPRF）[19]是一种特殊的伪随机函数（Pesudorandom Function，PRF）。OPRF 旨在保护隐私的同时，允许接收方执行特定的计算。发送方拥有 OPRF 函数的密钥 k，接收方输入查询元素 q，允许接收方在不知道密钥 k 的情况下计算 $F(k,q)$ 并不能从计算结果推断出任何关于密钥 k 的信息，其中 F 表示伪随机函数。在 PSI 协议中，需要考虑接收方可以在静态选择的输入集上获得对应的计算结果。OPRF 协议的理想功能如图 1 所示。

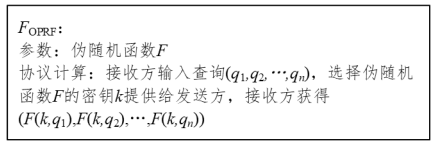


图 1 OPRF 的理想功能

Fig. 1 Ideal functionality of OPRF

本文采用的 OPRF 基于 VOLE 构建，VOLE 是一种允许各方生成随机向量并进行保密的向量乘法协议。OPRF 函数协议利用 VOLE 生成随机向量，并将其与查询元素进行向量乘法计算，实现了 PRF 计算。具体而言，VOLE 协议生成的随机向量 $\vec{A},\vec{B},\vec{C}$ 与随机值 Δ 满足 $\vec{C}=\Delta\vec{A}+\vec{B}$，其中发送方持有随机向量 $\vec{A}$ 和 $\vec{C}$，接收方持有随机向量 $\vec{B}$ 与随机值 Δ。然后分别将集合元素与持有的随机元素进行向量乘法计算，得到 OPRF 的结果。由于 VOLE 协议的保密性，发送方和接收方都无法从计算结果中推断出彼此的输入。

3.5 不经意可编程伪随机函数

不经意可编程伪随机函数(Oblivious Programmable Pesudorandom Function，OPPRF)[23]是伪随机函数(OPRF)的变体，其附加功能是允许在发送方选择的一组点上对伪随机函数的输出进行编程。其核心可编程伪随机函数(Programmable Pesudorandom Function，PPRF)由 $Keygen$ 与 F 两种算法组成，对于给定的一组点 $P=\{(x_1,y_1),(x_2,y_2),\cdots,(x_n,y_n)\}$，$Keygen$ 算法生成伪随机函数的密钥 k 和数据结构 $hint$；对于给定的查询点 q，F 计算 q 上的伪随机函数值。OPPRF 函数的理想功能如图 2 所示。

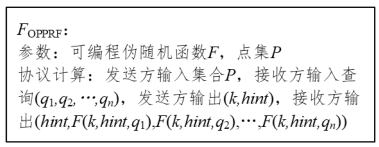


图 2 OPPRF 的理想功能

Fig. 2 Ideal functionality of OPPRF

本文采用 OPRF 与 OKVS 实例化 OPPRF 函数，接收方拥有查询集合 $Q=\{q_1,q_2,\cdots,q_n\}$，接收方拥有点集 $P=\{(x_1,y_1),(x_2,y_2),\cdots,(x_n,y_n)\}$。双方调用 OPRF 实例，其中接收方输入$(q_1,q_2,\cdots,q_n)$以获得由 OPRF 函数计算得到的伪随机函数值 $PRF(k,q_i)$，发送方获得伪随机函数密钥 k，并对自身集合 X 计算伪随机函数值 $PRF(k,X)$。之后调用 Encode()算法以 $PRF(k,X)$为键、Y 为值编码得到数据结构 $hint$ 发送给接收方。接收方使用 $Decode(hint,PRF(k,q_i))$算法进行查询实现 $F(k,hint,q_i)$的功能。

4 支持模糊匹配的带标签的隐私集合交集计算协议

4.1 问题描述

本节将介绍本文提出的半诚实模型下的 FLPSI 协议。协议的场景假定接收方 R 持有二进制向量集合 X，发送方 S 持有二进制向量集合 Y 以及标签集合 L，集合 L 为集合 Y 元素对应的标签。接收方希望获得双方相似元素对应的标签值。为了简化协议方案的描述，假定数据集合中存放着 N 个长度相同的向量。FLPSI 协议用于判断两个向量之间的汉明距离，若汉明距离小于阈值，则认为向量是相似的，否则是不相似的。协议需要解决的重点是保护参与方的隐私，确保接收方的查询对于发送方是不可见的，同时接收方无法获得接收方额外的隐私信息。

FLPSI 协议的理想功能如图 3 所示，其中汉明距离的匹配判定阈值 d 由协议双方共同协商确定。

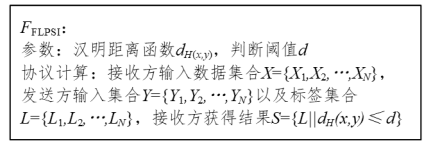


图 3 FLPSI 的理想功能

Fig. 3 Ideal functionality of FLPSI

4.2 协议设计

本节将在半诚实模型下提出具体的解决方案，为了保护数据隐私和缩短计算时间，采用对称加密构建各个密码学组件，提出基于 OPPRF 组件的 FLPSI 协议。方案的系统流程图如图 4 所示。

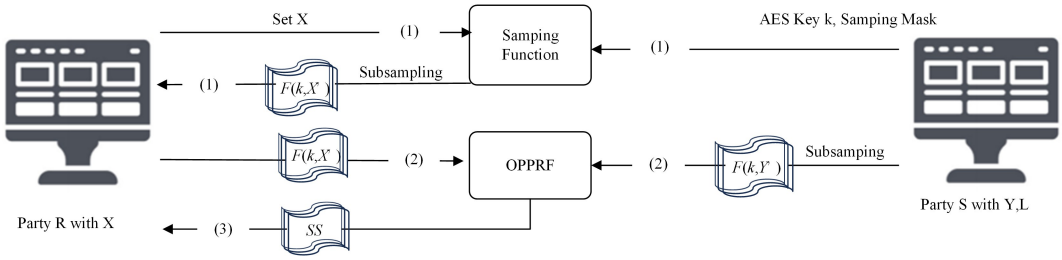


图 4 系统流程图

Fig. 4 System flowchart of FLPSI

1)双方调用采样函数，接收方输入集合 X，发送方输入选择的 n 个采样掩码 $mask$ 与 AES 密钥。采样函数将采样掩码与集合 X 分别执行按位与操作，再将结果加密。集合 X 中的元素 x 需执行 n 次采样操作并发送给接收方，第 j 次采样得到 $x_j=AES_j\{mask_j\wedge x\}$。2)发送方对集合 Y 进行本地采样作为 OPPRF 函数的输入，接收方将获得的采样数据作为 OPPRF 函数的输入。3)接收方得到秘密分享碎片后，重构出相应的标签值。

协议的主体部分分为预处理和在线两个阶段。

在预处理阶段，接收方首先向发送方发送查询请求，双方执行集合的采样操作。为了防止发送方获得接收方元素的信息，双方需共同调用安全计算函数，对接收方的元素进行 n 次采样并加密；为了防止接收方可能发起的重构攻击，每次执行协议时需更换采样掩码与 AES 密钥。本文采用了文献[13]中使用的安全计算函数，其借助 EMP 工具包将该函数实现为姚式混乱电路。

在在线阶段，双方依据不同的采样掩码执行 n 次 OPPRF 实例。在第 j 个实例中，接收方输入$\{x_{1,j},x_{2,j},\cdots,x_{N,j}\}$，输出 $F_{k_j}(x_{i,j})$，发送方将标签的秘密分享碎片编码进数据结构 Π_j 并发送给接收方。接收方使用 $F_{k_j}(x_{i,j})$解码数据结构 Π_j，

得到对应秘密分享碎片。

若元素 x 与元素 y 的汉明距离为 d，那么 x 与 y 仅有 d 位是不同的，若采样掩码中对应的 d 位均为 0，则 x 和 y 采样的结果相同，采样掩码为独立随机选取的，每个位置为 0 或 1 的概率均为 1/2，说明对 x 与 y 采样 n 次后的集合 $\{x_1, x_2, \cdots, x_n\}$ 与集合 $\{y_1, y_2, \cdots, y_n\}$ 中有 $n/2^d$ 对是相同的。因此设置 (t,n) 秘密分享时重构阈值 t 应当小于 $n/2^d$，以保证接收方可以重构出正确的标签值。协议的最后，接收方在本地进行秘密分享的重构即可获得双方汉明距离小于判断阈值的元素标签，而不会泄露其他任何的信息。

本文完整的协议设计如下：

协议 1　支持模糊匹配的带标签隐私集合交集计算协议

参数：n 为采样数量，t 为秘密分享的重构阈值，OPRF 函数 $F_k: \{0,1\}^\kappa \times \{0,1\}^* \to \{0,1\}^\zeta$，AES 加密 $\{0,1\}^\kappa \times \{0,1\}^{128} \to \{0,1\}^{128}$，$d_H(x,y)$ 为汉明距离函数，d 为匹配判定阈值，λ 为计算安全参数。

输入：接收方 R 的数据集合 $X=\{X_1, X_2, \cdots, X_N\}$，发送方 S 的数据集合 $Y=\{Y_1, Y_2, \cdots, Y_N\}$，标签集合 $L=\{L_1, L_2, \cdots, L_N\}$。

1）预处理阶段：R 输入数据集合 X，S 随机选择 n 个不同掩码和 AES 密钥。双方调用安全计算函数对 X 采样并加密得到新的集合 X' 输出给 R，$X_i'=\{x_{i,1}, x_{i,2}, \cdots, x_{i,n}\}$，$x_{i,j}=AES_j(mask_j \wedge X_i)$，$i \in [N]$，$j \in [n]$。集合 X 对于 S 是保密的，掩码 $mask$ 和 AES 密钥对于 R 不可见。对于所有的 $Y_i \in Y$，S 本地采样得到新集合 Y'，其中 $Y_i'=\{y_{i,1}, y_{i,2}, \cdots, y_{i,n}\}$，$y_{i,j}=AES_j(mask_j \wedge Y_i)$。

2）生成秘密分享：S 首先为标签 L_i 增添 λ 比特位，将其扩展为 $0^\lambda \| L_i$，然后生成 (t,n) 的秘密分享得到集合 $L_i'=\{l_{i,1}, l_{i,2}, \cdots, l_{i,n}\}$，其中 $l_{i,j}$ 长度为 ζ，最终 S 获得所有标签的秘密分享集合 $L'=\{L_1', L_2', \cdots, L_N'\}$。

3）OPPRF 函数计算：发送方 S 与接收方 R 并行执行 n 个 OPPRF 实例，每个实例中均包含 OPRF 函数计算、OKVS 编码和 OKVS 解码。

(1) OPRF 函数计算：接收方 R 输入 $x_{i,j}$，输出 OPRF 的函数值 $F_{k_j}(x_{i,j})$，发送方 S 输出密钥 k_j，并计算 $y_{i,j}$ 的 OPRF 函数值 $F_{k_j}(y_{i,j})$。

(2) OKVS 编码：S 计算编码值 $F_{k_j}(y_{i,j}) \oplus l_{i,j}$，并将 $F_{k_j}(y_{i,j})$ 与 $F_{k_j}(y_{i,j}) \oplus l_{i,j}$ 组成键值对集合 $S_j=\{F_{k_j}(y_{i,j}), F_{k_j}(y_{i,j}) \oplus l_{i,j}\}$。$S$ 将键值对集合编码为 $\Pi_j=Encode(S_j)$ 发送给 R。

(3) OKVS 解码：R 首先解码 Π_j，然后计算 $ss_j=Decode(\Pi_j, F_{k_j}(x_{i,j})) \oplus F_{k_j}(x_{i,j})$，输出集合 $SS_j=\{ss_{1,j}, ss_{2,j}, \cdots, ss_{N,j}\}$。

(4) 标签重构：R 得到集合 $\{SS_1, SS_2, \cdots, SS_n\}$ 后，对于所有的 $X_i \in X$，使用 $\{ss_{i,1}, ss_{i,2}, \cdots, ss_{i,n}\}$ 作为秘密分享重构算法的输入，尝试重构出标签 L_i，并依据前 λ 位比特是否为 0 判断是否为正确标签值。

输出：最终输出结果 $S=\{L \mid d_H(X,Y) \leqslant d\}$。

4.3　正确性分析

R 与 S 分别拥有集合 $X'=\{X_1', X_2', \cdots, X_N'\}$，$Y'=\{Y_1', Y_2', \cdots, Y_N'\}$，在第 j 个 OPPRF 实例中，R 选择 $\{x_{i,1}, x_{i,2}, \cdots, x_{i,n}\}$ 作为 OPRF 的输入，得到 OPRF 计算值 $F_{k_j}(x_{i,j})$。S 得到 OPRF 计算密钥 k_j，然后将 $F_{k_j}(y_{i,j})$ 作为 OKVS 编码的键，$F_{k_j}(y_{i,j})$ 与 $l_{i,j}$ 的异或作为 OKVS 编码的值，编码后得到数据结构 Π_j 发送给 R。若双方拥有相同元素 $x_{i,j}=y_{i,j}$，则 $F_{k_j}(x_{i,j})=F_{k_j}(y_{i,j})$，$R$ 使用 $F_{k_j}(x_{i,j})$ 解码 Π_j。依据 OKVS 正确性可知 $Decode(\Pi_j, F_{k_j}(x_{i,j})) \oplus F_{k_j}(x_{i,j})=F_{k_j}(y_{i,j}) \oplus l_{i,j} \oplus F_{k_j}(x_{i,j})=l_{i,j}$。当双方元素并不相同时，即 $x_{i,j} \neq y_{i,j}$，则 $F_{k_j}(x_{i,j}) \neq F_{k_j}(y_{i,j})$，由 OKVS 的安全性有 $Decode(\Pi_j, F_{k_j}(x_{i,j})) \oplus F_{k_j}(x_{i,j}) \neq l_{i,j}$。

在秘密重构阶段，由上文分析可知当 X_i' 与 Y_i' 中存在 $x_{i,j}=y_{i,j}$ 时，R 会获得 $0^\lambda \| L_i$ 的一个秘密分享碎片值，否则会得到一个伪随机值，即 $|X_i' \cap Y_i'|=\theta$，集合 $R_i=\{r \mid ss_{i,j}=l_{i,j}\}$ 的大小为 θ。在为标签 L_i 构造秘密分享时，为其增添了 λ 位的安全比特，R 通过观察重构值前 λ 位是否为 0 来判断是否重建了有效的标签值。由秘密分享的重构算法可知，若 $\theta \geqslant t$，则使用 $\{ss_{i,1}, ss_{i,2}, \cdots, ss_{i,n}\}$ 可以重构出 $0^\lambda \| L_i$；若 $\theta < t$，则 $\{ss_{i,1}, ss_{i,2}, \cdots, ss_{i,n}\}$ 重构出正确秘密值的概率为 $2^{-\lambda}$，即发生错误重构的概率是可以忽略的。

4.4　安全性证明

定理 1　以 OPPRF 的不可区分性和秘密分享的安全性为基础，协议 1 在半诚实模型下保密地完成了模糊匹配下带标签的隐私集合交集计算。

证明：采用理想-现实模型进行形式化证明，分别模拟半诚实发送方 S 和半诚实接收方 R，以证明协议 1 在半诚实模型下安全地完成了模糊匹配的带标签隐私集合交集计算。

1）半诚实接收方 R

R 的视图 $View_R^\pi(X;Y,L)=(X;X', F_{k_j}(x_{i,j}), \Pi_j, f(X;Y,L))$，$i \in [N]$，$j \in [n]$。模拟器 Sim_R 按照协议 1 模拟发送方 S 与接收方 R 交互。

Sim_R 随机生成发送方 S 的模拟数据集 $\bar{Y}$ 与 $\bar{L}$，并按照协议 1 的要求，依据采样规则对 $\bar{Y}$ 采样得到数据集 $\bar{Y}'$。之后模拟器 Sim_R 与接收方 R 模拟执行 OPPRF 函数，PRF 函数值 $F_{k_j}'(x_{i,j})$ 由 OPPRF 函数产生并发送给接收方 R，由 Sim_R 编码键值对集合 Π_j' 并发送给接收方 R。接收方 R 通过计算 $ss_j'=Decode(\Pi_j', F_{k_j}'(x_{i,j})) \oplus F_{k_j}'(x_{i,j})$，最终得到协议结果。

因为 OPRF 函数的输出与 OKVS 结果均具有不可区分性，同时模拟数据集 $\bar{Y}$ 与真实数据集 Y 均为任选的等长集合，即 Π_j 与 Π_j' 不可区分，同时无法区分 $ss_j=Decode(\Pi_j, F_{k_j}(x_{i,j})) \oplus F_{k_j}(x_{i,j})$ 与 $ss_j'=Decode(\Pi_j', F_{k_j}'(x_{i,j})) \oplus F_{k_j}'(x_{i,j})$。最终接收方 R 无法区分协议最终的输出结果，故 $Sim_R(X; f(X;Y,L)) \overset{c}{\equiv} View_R^\pi(X;Y,L)$。

2）半诚实发送方 S

S 的视图 $View_S^\pi(X;Y,L)=(Y,L;mask_j, ks_j, Y', k_j)$，$j \in [n]$。模拟器 Sim_S 按照协议 1 模拟接收方 R 与发送方 S 的交互。

Sim_S 随机生成接收方 R 的模拟数据集 $\bar{X}$，并按照相同的采样规则对集合元素进行采样得到数据集 $\bar{X}'$。此时发送方

S 随机生成 $mask_j'$ 与 ks_j'，由于半诚实发送方 S 每次协议执行前均会随机重新选取掩码与 AES 密钥，因此无法从中推断出更多信息，同时依据协议 1 的要求发送方 S 没有获得协议的最终输出。

之后模拟器 Sim_S 与发送方 S 模拟执行 OPPRF 函数，按照协议 1 的要求，PRF 密钥 k_j' 由 OPPRF 函数产生并发送给发送方 S。OPRF 函数的安全性可以保证密钥 k 的不可区分，即 k_j' 与 k_j 不可区分，故 $Sim_S(Y,L;f(X;Y,L)) \overset{c}{\equiv} View_S^{\pi}(X;Y,L)$。

综上所述，协议 1 在半诚实模型下安全地实现了模糊匹配下带标签的隐私集合交集计算。

5 实施与分析

5.1 性能分析与对比

协议的效率由通信复杂度和计算复杂度决定，本节分析协议 1 的通信复杂度与计算复杂度，并与现有的相关协议进行对比，以证明本文协议的优势。

1)通信复杂度：包括 PRF 采样计算与 OPRF 协议的输入输出以及键值对打包结果的传输。在 PRF 采样中，接收方 R 输入 N 个元素，接收 Nn 个采样值，通信量为 $N+Nn$；OPRF 协议中接收方 R 输入 Nn 个元素值，接收 Nn 个 OPRF 值，通信量为 $2Nn$。执行 OKVS 阶段，协议接收方 R 无发送，本文采用的 PaXoS 中 Cuckoo Table 的长度设置为输入长度的 2.4 倍，即接收方 R 接收的数据量约为 $2.4Nn$，因此总体的通信复杂度为 $O(Nn)$。

2)计算复杂度：协议的计算复杂度由接收方 R 和发送方 S 的计算量构成。接收方 R 的计算量主要包括 OPPRF 计算与秘密分享的重构。执行 OPRF 协议，接收方 R 的计算复杂度为 $O(Nn)$，OKVS 解码的计算复杂度为 $O(Nn)$，一个元素的秘密分享的重构计算复杂度为 $O(C_n^t t^2)$。发送方 S 的计算复杂度主要包括 PRF 的采样计算、OPRF 值的计算与 OKVS 的编码。首先 S 为 R 的集合采样，随后本地为自身集合采样，计算复杂度均为 $O(Nn)$，利用 OPRF 密钥 k 计算出 Nn 个 OPRF 值并编码，计算复杂度为 $O(2Nn)$，因此总体的计算复杂度为 $O(NnC_n^t t^2)$。

表 1 列出了本文方案与相关工作的综合对比，其中 N 表示集合元素的个数，d 表示汉明距离的大小，l 为数据向量的长度，n 和 t 分别表示秘密分享份数与重构阈值大小，AHE 与 FHE 分别表示加法同态加密操作和全同态加密操作，SYS 表示对称密码原语操作。

表 1 相关 FPSI 方案对比

Table 1 Comparison of related FPSI schemes

协议	通信复杂度	计算复杂度		带标签	加密技术
		接收方	发送方		
文献[8]	$O(Nn)$	$O(N^2)$	$O(N^2)$	否	AHE
文献[9]-1	$O(NC_n^t)$	$O(N^2)$	$O(N^2)$	否	AHE
文献[9]-2	$O(NC_n^t)$	$O(N^2)$	$O(N^2)$	否	AHE
文献[10]	$O(N^2 l)$	$O(N^2)$	$O(N^2)$	否	AHE
文献[11]	$(N^2 l)$	$O(N^2)$	$O(N^2)$	否	SYS
文献[12]	$O(N^2 d^2)$	$O(N^2)$	$O(N^2)$	否	AHE
文献[13]	$O(N^2 n)$	$O(NnC_n^t t^2)$	$O(N^2 n)$	是	FHE
Ours	$O(Nn)$	$O(NnC_n^t t^2)$	$O(Nn)$	是	SYS

文献[8-10,12]都采用了基于加法同态加密的方法来实现集合的模糊匹配操作，接收方与发送方的计算复杂度均与集合大小的平方成线性相关。文献[8]中的通信复杂度较低，但其协议中存在隐私泄露风险。文献[11]借助混乱电路提供了通用的两方安全评估函数，虽然主要使用对称加密操作，但构造复杂度会随着电路深度增加而大幅增加，且通信复杂度与集合大小的平方和数据向量的长度成正比，因此其并不适合于集合的模糊匹配操作。

文献[13]是目前最新的支持模糊匹配的带标签隐私集合交集计算协议，其总体的通信复杂度与计算复杂度均为 $O(n^2)$。与之相比，本方案在将参与方的通信复杂度降低为 $O(n)$的同时将发送方的计算复杂度也降低为 $O(n)$，且文献[13]采用全同态加密算法计算，相较于本文主要基于对称加密的方式构建，执行效率更低。

5.2 实验分析

为了客观地评估各协议效率，本文对比了文献[13]中的协议，为了确保实验的准确性和可靠性，本文采用相同的环境设置进行了实现。编程语言选用 C++，参与双方均用配置 Inter© Xeon© Gold 6130 CPU @ 2.10 GHz，251 GB RAM 的 Ubuntu 16.04.4 LTS 服务器执行协议，在实施过程中使用了 Crypto++密码学库。本文与文献[13]在运行时间与通信量两方面进行对比，选择了与其相同的参数，包括集合元素大小为 256bit，AES 加密的密钥长度与输出长度均为 128 位，标签元素被拆分为 64 份秘密分享，重构阈值设置为 2，安全参数 $\lambda=40$。秘密分享算法选用基于多项式插值原理的 Shamir 秘密分享算法，重构阈值沿用了文献[13]中的设定值。重构阈值设定为 2 符合实际应用场景，可以保证相似元素对应标签的重构正确率。

本文首先在双方集合大小相同的平衡场景与文献[13]进行协议性能的比较。在离线阶段，文献[13]协议主要进行同态加解密、多项式编码与秘密分享的重构，本文提出的协议主要进行 OKVS 的解码与秘密分享的重构，在线阶段双方完成信息交互。实验结果如表 2 所列。

表 2 平衡场景下的性能评估

Table 2 Performance evaluation in balanced scenarios

集合大小 N	协议	离线时间/s		在线时间/s	通信量
		接收方	发送方		
2^8	文献[13]	0.103	0.137	0.025	51.69×10^6
	Ours	0.004	0.017	0.069	4.60×10^6
2^{10}	文献[13]	0.247	0.383	0.083	155.94×10^6
	Ours	0.016	0.054	0.071	7.53×10^6
2^{12}	文献[13]	1.068	1.337	0.326	626.44×10^6
	Ours	0.052	0.195	0.075	41.63×10^6
2^{14}	文献[13]	4.134	5.305	1.300	2439.86×10^6
	Ours	0.190	0.792	0.093	153.92×10^6
2^{16}	文献[13]	15.626	19.098	4.921	9222.67×10^6
	Ours	0.672	2.691	0.475	544.12×10^6
2^{18}	文献[13]	56.528	68.606	17.960	34769.47×10^6
	Ours	2.307	8.617	1.561	1897.84×10^6

从表中数据可知，本文方案的运行时间具有巨大的优势。当数据集合较时，文献[13]总的运行时间是本文协议的 10 倍以上。在离线阶段，接收方与发送方的离线时间均随着集合元素数量的增加而增加，但本文协议始终优于文献[13]的

协议，并且随着集合大小的增大，性能优势更加明显。在在线阶段，相较于文献[13]协议的运行时间与集合大小成线性关系，本文协议的在线时间更加稳定。数据量较大时，本文协议比文献[13]的在线运行速度快 14 倍。当数据量较小时，本文的运行时间略高于文献[13]，这是由于本文采用 VOLE 构造 OPRF 函数，需初始化 VOLE 随机种子。现实场景中数据量往往很大，故本文在实际应用中完全可以接受。

同时，由表 2 可知，本文方案的通信负载具有绝对优势，在大集合场景下，文献[13]的通信负载是本文协议的 16～18 倍。当数据集合大小为 2^{14} 时，文献[13]的方案需要 2.38 GB 的通信量，而本文方案仅需要 0.15 GB，大大降低了 FLPSI 协议的通信量，对于网络带宽受限的场景依然适用。

本文还模拟了双方数据量不平衡的情况，假设发送方是拥有更多数据的服务器，而接收方为数据量较少的客户端，分别设定发送方的集合 Y 大小为 $\{2^{16},2^{18},2^{20},2^{22},2^{24}\}$，而接收方集合大小为 $\{2^{10},2^{12}\}$。表 3 列出了协议运行时间的对比结果。

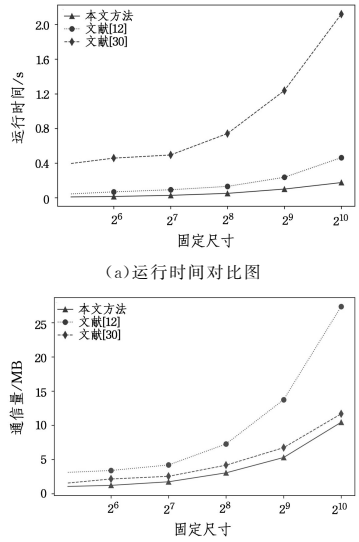
表 3 非平衡场景下的运行时间对比

Table 3 Comparison of runtime in unbalanced scenarios

$\lvert Y\rvert$	$\lvert X\rvert$	离线时间/s		在线时间/s	
		文献[13]	Ours	文献[13]	Ours
2^{16}	2^{10}	3.15	0.23	0.47	0.28
	2^{12}	4.06	0.29	0.64	0.28
2^{18}	2^{10}	9.79	0.63	1.21	0.53
	2^{12}	11.26	0.67	1.47	0.53
2^{20}	2^{10}	42.90	3.29	4.41	2.56
	2^{12}	49.73	3.33	4.97	2.56
2^{22}	2^{10}	187.11	17.18	14.99	12.24
	2^{12}	206.73	17.53	15.63	12.24
2^{24}	2^{10}	653.49	73.75	52.91	44.76
	2^{12}	722.08	74.59	55.27	44.76

由实验数据可知，文献[13]方案的离线时间与在线时间都很长，这与其采用全同态加密与多项式编码有关，这对于计算能力弱的客户端是很不友好的。具体而言，文献[13]方案的离线运行时间是本文方案的 10～16 倍，在线运行时间方面，本文方案为文献[13]方案的 36%～80%，因此本文方案也更适合于弱客户端且双方数据量大小不平衡的场景。

此外，本文还与相关的隐私模糊匹配方案[12,30]进行对比。文献[12]实现了不带标签的模糊向量的隐私匹配，其度量方式同样为汉明距离；文献[30]实现了基于阈值匹配的人脸面部识别协议。本文方案与二者运行时间和通信量的对比如图 5 所示。在拥有相同数据的情况下，相比于文献[12]和文献[30]，本文方案的运行时间与通信量均有显著降低。文献[30]的运行时间是本文的 10 倍以上，因为文献[30]采用了同态加密与乱码电路，增加了计算负担。文献[12]的通信量远高于本文方案且增长较快，这是由于其通信负载与向量间汉明距离的平方成正比。



(a)运行时间对比图

(b)通信量对比图

图 5 与隐私保护模糊匹配方案的对比

Fig. 5 Comparison between the proposed scheme and privacy preserving fuzzy matching schemes

6 应用

近年来，人脸、虹膜、指纹和 DNA 等生物识别技术受到了极大的关注，与传统的基于密码的身份认证相比，生物识别具有唯一性、移动性、不可转移性等优点[31]，但这些生物特征包含了大量的隐私。因此，如果是在未加密的情况下上传生物特征，则可能出现在未经用户许可的情况下收集、分析、滥用隐私信息等问题。

本文提出的协议主要应用于隐私保护下的支持模糊匹配的带标签场景，客户端在进行生物特征比对时，往往不仅仅需要比对的结果，还需要得到匹配用户的标签信息。本文以隐私保护下的人脸识别场景为例，介绍 FLPSI 协议的应用，流程图如图 6 所示。

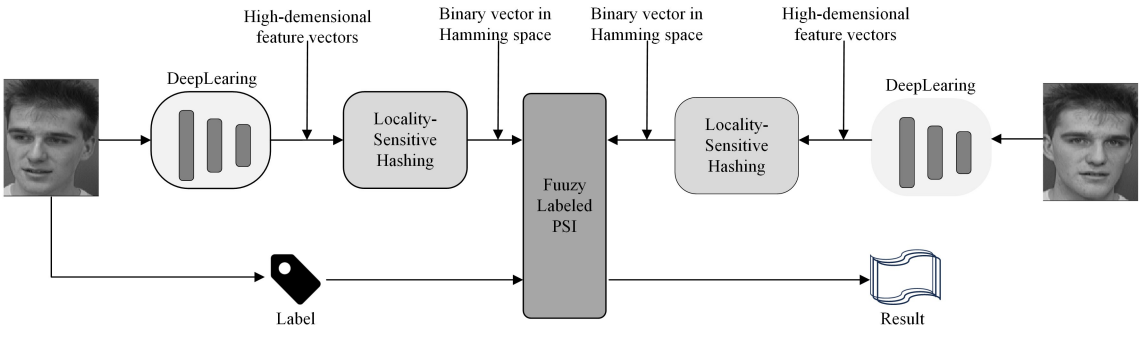


图 6 FLPSI 协议的应用

Fig. 6 Application of FLPSI protocol

服务器与客户端首先借助深度学习网络(Deep Learing，DL)[32]提取人脸的特征向量，由于深度学习网络提取的特征

向量往往是欧几里得空间中的高维向量，因此需要借助位置敏感哈希函数(Locality-Sensitive Hashing，LSH)[33]将高维向量映射到低维汉明空间获得二进制向量，作为 FLPSI 协议的输入，此时服务器额外输入标签值，经 FLPSI 协议运算之后由客户端获得最终的结果。

在评估生物识别系统性能时，往往需要考虑系统的准确性与安全性。下面分析系统的误识率 FAR(False Acceptance Rate)与误拒率 FRR(False Rejection Rate)，它们的定义如下。

定义 3(误识率 FAR) 非法用户被错误地认为是合法用户的概率。本系统中指协议输出了不满足条件的人脸标签秘密分享碎片，而非法客户端重构出了额外正确标签的比率。

$$FAR=\frac{|L|d_H(X,Y)>u|}{|X|}$$

定义 4(误拒率 FRR) 合法用户被错误地认为是非法用户的概率。本系统中指协议没有输出满足条件的人脸标签秘密分享碎片，而合法客户端无法重构出正确标签值的比率。

$$FRR=1-\frac{|L|d_H(X,Y)\leqslant u|}{|d_H(X,Y)\leqslant u|}$$

当非法用户被误识别为合法用户时，会泄露服务器的少量隐私，但由于 FLPSI 协议在预处理阶段会刷新采样掩码与密钥，因此客户端无法发动重构攻击，即客户端无法从单次泄露的少量隐私中获得有用的信息。当合法用户被错误地认为是非法用户时，不会泄露任何隐私消息，但高的误拒率会影响系统的运行效率。

本文将 AT&T 数据库中的人脸照片采用 FaceNet 算法提取出 128 维的特征向量，然后用位置敏感哈希函数映射后作为 FLPSI 协议的输入向量。在单个标签被秘密分享为 64 份的情况下，计算出不同重构阈值 t 下误识率 FAR 与误拒率 FRR 的大小，如图 7 所示。

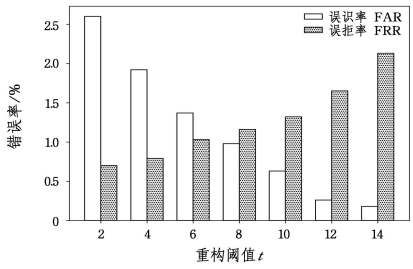


图 7 重构阈值对错误率的影响

Fig. 7 Impact of reconstruction threshold on error rate

由图中数据可知，随着重构阈值 t 的增加，误识率 FAR 在逐渐减小，而误拒率 FRR 在逐渐增大。因此，通过选择不同的重构阈值 t 可以满足不同场景下的系统要求。

本文将测试样本中合法用户被成功识别的数量占测试样本中合法用户总数的比例定义为识别率。FaceNet 算法测量特征向量间的欧氏距离，若两特征向量距离小于判断阈值，则认为照片来自于同一个人。随着判断阈值的减小，算法识别率会随之下降。依据本文的实验结果，当 FaceNet 算法的判断阈值为 1.1 时，识别率为 96.1%。为了进一步说明 FLPSI 协议在隐私保护条件下的可用性，本文依据不同的重构阈值计算识别率，并统计实验运行时间，实验结果如表 4 所列。

表 4 不同重构阈值下的性能评估

Table 4 Performance evaluation with different reconstruction thresholds

重构阈值	运行时间/s	识别率/%
2	0.318	97.3
4	0.474	95.8
6	0.708	93.1
8	0.996	89.3

由表中数据可知，随着重构阈值的增加，算法的运行时间有小幅的增长，增长的时间开销主要用于秘密分享算法中的多项式插值计算；算法识别率有小幅的下降，当重构阈值为 2 时，算法的识别率为 97.3%，可以满足实际应用场景的需要。因此，在需要提高访问效率的场景中，可以设置较低重构阈值 t 来降低误拒率并缩短运行时间，确保协议高效地运行。在需要提高系统安全性的场景中，可以提高重构阈值 t 降低误识率，避免未授权用户的非法访问，提高系统安全性。

结束语 隐私集合交集计算协议是安全多方计算领域一个重要的研究方向。本文面向模糊匹配的场景，设计了一种带标签的隐私集合交集协议，并在半诚实模型下证明了协议的安全性。所提出的协议允许接收者通过 OPPRF 函数和秘密共享，而不是通过昂贵的公钥操作来获得近似元素的标签，且不会泄露额外的隐私。对比实验结果表明，本文方法在平衡与非平衡场景下都具有更好的性能。此外，本文还给出了 FLPSI 协议在生物特征识别中的应用。在未来的工作中，将研究恶意模型下的 FLPSI 协议，如采用消息验证码来确保双方输入数据的完整性并提高匹配的精度。

参 考 文 献

[1] SHEN L Y, CHEN X J, SHI J Q, et al. Survey on Private Preserving Set Intersection Technology[J]. Jisuanji Yanjiu yu Fazhan/Computer Research and Development, 2017, 54(10): 2153-2169.

[2] WEI L F, LIU J H, ZHANG L, et al. Survey of Privacy Preserving Oriented Set Intersection Computation[J]. Jisuanji Yanjiu yu Fazhan/Computer Research and Development, 2022, 59(8): 1782-1799.

[3] YUNG M. From Mental Poker to Core Business: Why and How to Deploy Secure Computation Protocols? [C]// Proceeding of the 22nd ACM SIGSAC Conference on Computer and Communications Security. New York: Association for Computing Machinery, 2015: 1-2.

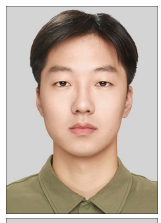
[4] AGGARWAL C C, YU P S. A General Survey of Privacy-Preserving Data Mining Models and Algorithms [J]. Privacy-Preserving Data Mining Models and Algorithms. 2008, 34: 11-52.

[5] BALDI P, BARONIO R, DE CRISTOFARO E, et al. Countering GATTACA: efficient and secure testing of fully-sequenced human genomes[C]// Proceedings of the 18ACM Conference on Computer and Communications Security. New York: Association for Computing Machinery, 2011: 691-702.

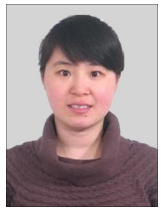
[6] DEMMLER D, RINDAL P, ROSULEK M, et al. PIR-PSI: Scaling Private Contact Discovery[J]. Proceedings on Privacy Enhancing Technologies, 2018(4): 259-178.

[7] EL-SHAFAI W, MOHAMED F A H E, ELKAMCHOUCHI H M A, et al. Efficient and Secure Cancelable Biometric Authenti-

cation Framework Based on Genetic Encryption Algorithm [J]. IEEE Access 2021(9):77675-77692.
[8] FREEDMAN M J,NISSIM K,PINKAS B. Efficient Private Matching and Set Intersection[C]// International Conference on the Theory and Application of Cryptographic Techniques. Berlin:Springer,2004:1-19.
[9] CHMIELEWSKI Ł,HOEPMAN J H. Fuzzy Private Matching [C]// 2008 Third International Conference on Availability,Reliability and Security. Oakland:IEEE,2008:327-334.
[10] OSADCHY M,PINKAS B,JARROUS A,et al. SCiFI- A System for Secure Face Identification[C]// 2010 IEEE Symposium on Security and Privacy. IEEE,2010:239-254.
[11] HUANG Y,EVANS D,KATZ J,et al. Faster Secure Two-Party Computation Using Garbled Circuits[C]// Proceedings of the 20th USENIX Security Symposium. Sanfrancisco:USENIX Association,2011:1-16.
[12] CHAKRABORTI A,FANTI G,REITER M. Distance-Aware Private Set Intersection[C]// Proceedings of the 32nd USENIX Security Symposium. USENIXA Association,2023:319-336.
[13] UZUN E,CHUNG S P H,KOLESNIKOV V,et al. Fuzzy Labeled Private Set Intersection with Applications to Private Real-Time Biometric Search[C]// Proceedings of the 30th USENIX Security Symposium. USENIX Association,2021:911-928.
[14] GARIMELLA G,ROSULEK M,SINGH J. Structure-Aware Private Set Intersection,with Applications to Fuzzy Matching [C]// Annual International Cryptology Conference. Cham:Springer,2022:323-352.
[15] BAY A,ERKIN Z,ALISHAHI M S,et al. Multi-Party Private Set Intersection Protocols for Practical Applications[C]// Proceedings of the 18th International Conference on Security and Cryptography. Springer,2021:515-522.
[16] BAY A,ERKIN Z,HOEPMAN J H,et al. Practical Multi-Party Private Set Intersection Protocols [J]. IEEE Transactions on Information Forensics and Security,2021,17:1-15.
[17] MEADOWS C. A More Efficient Cryptographic Matchmaking Protocol for Use in the Absence of a Continuously Available Third Party[C]// Proceedings of the 7th IEEE Symposium on Security and Privacy. IEEE,1986:134-134.
[18] SHAMIR A. On the power of commutativity in cryptography [C]// International Colloquium on Automata,Languages and Programming. Berlin:Springer,1980:582-595.
[19] KOLESNIKOV V,KUMARESAN R,ROSULEK M,et al. Efficient Batched Oblivious PRF with Applications to Private Set Intersection [C]// Proceedings of the 23rd ACM SIGSAC Conference on Computer and Communications Security. New York:Association for Computing Machinery,2016:818-829.
[20] PINKAS B,ROSULEK M,TRIEU N,et al. SpOT-Light:Lightweight Private Set Intersection from Sparse OT Extension[C]// Annual International Cryptology Conference. Cham:Springer,2019:401-431.
[21] CHASE M,MIAO P. Private Set Intersection in the Internet Setting From Lightweight Oblivious PRF [C]// Annual International Cryptology Conference. Cham:Springer,2020:34-63.
[22] ZHOU S F,LI S D,GUO Y M,et al. Efficient Secure Set Intersection Problem Computation [J]. Jisuanji Xuebao,2018,41(2):464-480.
[23] KOLESNIKOV V,MATANIA N,PINKAS B,et al. Practical Multi-party Private Set Intersection from Symmetric-Key Techniques[C]// Proceedings of the 2017 CM SIGSAC Conference on Computer and Communications Security. New York:Association for Computing Machinery,2017:1257-1272.
[24] PINKAS B,ROSULEK M,TRIEU N,et al. PSI from PaXoS:Fast,Malicious Private Set Intersection[C]// Annual International Conference on the Theory and Applications of Cryptographic Techniques. Cham:Springer,2020:739-767.
[25] EVANS D E,KOLESNIKOV V,ROSULEK M. A Pragmatic Introduction to Secure Multi-Party Computation [J]. Foundations and Trends© in Privacy and Security,2018,2(2/3):70-246.
[26] GOLDREICH O. Foundations of cryptography:volume2,basic applications[M]. Cambridge University Press,2009.
[27] SHAMIR A. How to share a secret [J]. Communications of the ACM,1979,22(11),612-613.
[28] GARIMELLA G,PINKAS B,ROSULEK M,et al. Oblivious Key-Value Stores and Amplification for Private Set Intersection [C]// Annual International Cryptology Conference. Cham:Springer,2021:395-425.
[29] PINKAS B,SCHNEIDER T,SEGEV G,et al. Phasing:Private Set Intersection Using Permutation-based Hashing:proceedings of the USENIX Security Symposium,F,2015 [C]// Proceedings of the 24th USENIX Security Symposium. USENIX Association,2015:515-530.
[30] SADEGHI A,SCHNEIDER T,WEHRENBERG I. Efficient Privacy-Preserving Face Recognition[C]// International Conference on the Theory and Application of Cryptographic Techniques. Berlin:Springer,2009:229-244.
[31] OGBANUFE O,KIM D. Comparing fingerprint-based biometrics authentication versus traditional authentication methods for e-payment [J]. Decision Support Systems,2018,106:1-14.
[32] SCHROFF F,KALENICHENKO D,PHILBIN J. FaceNet:A Unified Embedding for Face Recognition and Clustering[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recogintion,IEEE,2015:815-823.
[33] MARCAIS G,DEBLASIO D,PANDEY P,et al. Locality-sensitive hashing for the edit distance [J]. Bioinformatics,2019,35(14):127-135.



CHENG Enze,born in 1999,master,is a student member of CCF(No. H8993G). His main research interests include information security and secure computation.



ZHANG Lei,born in 1983,Ph.D,associate profersor,is a member of CCF (No. L0931M). Her main research interests include cryptography,data security and access control.

（责任编辑：何杨）