保护用户数量信息的安全虹膜识别方案

周 宇[1], 向剑文[1,2], 郑倩荣[1], 赵冬冬[1,2]

[1]武汉理工大学计算机与人工智能学院 交通物联网技术湖北省重点实验室 武汉 中国 430070
[2]武汉理工大学重庆研究院 重庆 中国 401135

摘要 由于传统密码认证方式的不便, 生物特征识别技术凭借其便捷、可靠、安全可溯源等特性脱颖而出。在不同的生物特征识别技术中, 虹膜识别已被证明能提供较高的识别性能和稳定性, 常被用于一些安全性要求较高的领域(如机密组织的认证管理等)。在这些领域中, 合法用户数量信息往往也属于机密信息, 是不能泄露的, 近年来针对虹膜识别的攻击手段也越加先进, 通过获得的数量信息可能推测出更多的其他信息, 造成更大的安全隐患。但是现有的安全虹膜识别方案仅考虑满足可撤销性、不可逆性和不可连接性, 未考虑保护用户数量信息。

本文提出一种保护用户数量信息的安全虹膜识别方案, 每个用户通过自身虹膜特征随机选择的结果及系统参数共同决定该用户的注册模板数量, 攻击者难以根据服务器中存储的虹膜模板数量推测出合法用户数量。该方案能够有效地与现有的安全虹膜识别方案进行结合。理论分析结果表明, 本文方案能够保护合法用户数量信息、保护新增用户数量信息、预防关联攻击、并且除了能够保持原始安全虹膜识别方案的可撤销性和不可连接性之外, 还能进一步提升原始安全虹膜识别方案的不可逆性。实验结果表明, 攻击者准确猜对合法用户数量信息的概率不足 15%, 且相对误差以及相对期望误差均超过 10%, 因此本文方案能有效保护用户数量信息, 并且不会对原始安全虹膜识别方案的识别精度的影响造成较大影响, 差异在 0.55%之内。

关键词 隐私保护; 虹膜识别; 用户数量信息
中图法分类号 TP309.2 DOI 号 10.19363/J.cnki.cn10-1380/tn.2023.05.05

Secure Iris Recognition with the Protection of the Number of Users

ZHOU Yu[1], XIANG Jianwen[1,2], ZHENG Qianrong[1], ZHAO Dongdong[1,2]

[1] School of Computer Science and Artificial Intelligence, Hubei Key Laboratory of Transportation of Internet of Things, Wuhan University of Technology, Wuhan 430070, China
[2] Chongqing Research Institute, Wuhan University of Technology, Chongqing 401135, China

Abstract Due to the inconvenience of traditional password authentication methods, biometric identification technology stands out due to its convenience, reliability and traceability. Among the different biometric technologies, iris recognition has been proven to provide high recognition performance and stability, and it often used in areas with high security requirements (e.g., authentication management of confidential organizations). In these fields, the number of users is often confidential and cannot be disclosed. In recent years, attacks on iris recognition have become more sophisticated, and the number of users obtained may lead to additional information and greater security risks. However, the existing secure iris recognition schemes only consider reversibility, irreversibility and unlinkability, and do not consider protecting the number of users.

In this paper, we propose a secure iris recognition scheme that protects the number of users, where each user determines the number of registration templates based on the result of random selection from their own iris feature data and the system parameters together. it is difficult for an attacker to infer the number of legitimate users of the system based on the number of iris templates stored in the server. The scheme proposed in this paper can be effectively combined with the existing secure iris recognition scheme. The theoretical analysis results show that the proposed scheme can protect the number of legitimate users, protect the number of new users, and prevent associated attacks. At the same time, in addition to maintaining the revocability and unlinkability of the original secure iris recognition scheme, it can further improve the irreversibility of the original secure iris recognition scheme. The experimental results show that the probability of the attacker correctly guessing the number of legitimate users is less than 15%, and the relative error and the relative expected error are more than 10%. Therefore, the proposed scheme can effectively protect the number of users. And it will not have a large impact on the recognition accuracy of the original security iris recognition scheme, the difference is within 0.55%.

通讯作者: 赵冬冬, 博士, 副教授, Email: zdd@whut.edu.cn。

本课题得到国家自然科学基金(No. 61806151)、湖北省重点研发计划(No. 2022BAA050)、海南省重点研发计划(No. ZDYF2021GXJS014)、重庆市自然科学基金(No. cstc2021jcyj-msxmX1146, No. CSTC2021JCYJ-MSXMX0002)资助。
收稿日期: 2022-09-09; 修改日期: 2022-11-07; 定稿日期: 2023-03-31

Key words privacy protection; iris recognition; the number of users

1 引言

身份鉴别和日常生活密切相关, 传统使用密码识别的方式, 用户需要记忆密码, 使用十分不便, 且用户往往选取和个人相关的信息作为密码, 容易遭到字典攻击, 安全性较弱。生物特征识别可根据个人的生理特征(如人脸、指纹、虹膜等)和行为特征(如步态、声音、手写签名等)进行识别, 相较于通过密码认证的方式更可靠, 因为生物特征信息不会丢失或遗忘, 且安全可溯源, 即用户难以否认使用生物特征信息访问过某一内容。但是生物特征数据是敏感的个人数据, 不像密码、令牌等泄露后可以更改, 生物特征数据是唯一的, 一旦泄露将无法挽回, 这将带来很大的安全风险。

大量的生物特征模板保护研究就是为了有效应对生物特征数据泄露带来的各种安全风险而开展的。考虑到生物特征数据泄露带来的各种安全风险, ISO/IEC 24745 标准[1]中定义了生物特征模板保护的要求: i)可撤销性, 如果生物特征模板被泄露, 可以根据原始生物特征数据生成新的模板替换被泄露的模板, 且被泄露的模板将失效。ii)不可逆性, 通过生物特征模板无法逆推得到原始生物特征数据。iii)不可连接性, 原始生物特征数据不能和生物特征模板进行匹配, 且存储在不同应用程序或服务器中的两个生物特征模板不能交叉匹配关联到同一主体。此外, 生物特征识别还需要保证识别性能的优异性。

在各种生物特征识别技术中, 虹膜识别因其准确度高、可识别性强、难以模仿、采集和编码简单、稳定等特点, 被广泛认为是最强大的识别工具之一[2]。同时虹膜特征在个体的一生中几乎保持不变, 不受遗传或环境因素的影响, 因此一旦虹膜特征数据被泄露将会带来很大的安全风险, 攻击者可利用盗来的虹膜特征数据在其他应用程序中重用。因此, 虹膜模板保护研究, 也被称为安全虹膜识别研究, 是十分具有现实意义的。

尽管近年来有大量的方法被提出来用于安全虹膜识别, 比如安全梗概、模糊提取器、模糊承诺、模糊金库、加盐法、不可逆变换、同态加密、安全多方计算、负数据库等, 但是几乎所有的方案都未考虑到用户数量信息的保护。在一些应用场景中, 保护用户数量信息也非常重要。例如, 1)特工组织采用虹膜特征数据对特工进行身份识别和管理, 特工组织的人员数量、人员变动情况等都属于机密信息。2)系统的用户数量信息能反应系统的运行状态, 用户数量较少或者用户数量呈现下降趋势会降低公众的信任度, 尤其是在商业竞争中是极其不利的。3)通过用户数量信息能够辅助攻击过程, 有助于提高攻击成功率, 造成更大的隐私泄露问题。因此, 设计保护用户数量信息的安全虹膜识别方案是十分有必要的。

本文的主要贡献如下:

1. 提出保护用户数量信息的安全虹膜识别方案, 通过用户自身虹膜特征随机选择操作以及系统参数共同决定该用户的注册模板数量, 使攻击者无法从服务器中的模板数量获得合法用户数量信息。

2. 将本文方案与现有的安全虹膜识别方案进行结合, 通过理论分析结果表明, 本文方案能够保护合法用户数量信息、保护新增用户数量信息、预防关联攻击、并能在一定程度上提高原始安全虹膜识别方案的不可逆性。

3. 对不同合法用户数量下的数量信息的安全性进行衡量, 并进行了相关参数的调整, 实验结果证实本文方案能有效保护用户数量信息。

4. 在真实虹膜特征数据集上验证了本文方案对原始安全虹膜识别方案的识别精度造成的影响几乎可以忽略。

本文剩余部分组织如下: 第 2 节将介绍安全虹膜识别领域的相关工作; 第 3 节将进行问题描述, 并介绍一种现有的安全虹膜识别方案, 后文将在该方案的基础上实施本文方案; 第 4 节将介绍提出的保护用户数量信息的安全虹膜识别方案; 第 5 节将从合法用户数量信息保护、新增用户数量信息保护、关联攻击防御、三大安全性要求四个方面对本文方案的安全性进行理论分析; 第 6 节对本文方案的安全性以及识别精度展开实验; 第 7 节对全文进行总结和未来展望。

2 相关工作

目前的安全虹膜识别方案都是一个用户生成一个模板存于服务器中, 未考虑服务器端的合法用户数量信息的保护。本节主要讨论和分析现有的安全虹膜识别方案的相关工作。生物特征识别系统可分为生物特征加密系统和可撤销生物特征数据[3-5]。生物特征加密系统又可分为基于密钥生成的生物特征加密系统和基于密钥绑定的生物特征加密系统, 可撤销生物特征数据可分为加盐法和不可逆变换。

基于密钥生成的生物特征加密系统以安全梗概

和模糊提取器方案为代表, 这两个方案由 Dodis 等人[6]提出。安全梗概方案在注册阶段使用生成算法从生物特征数据中获得辅助信息, 识别阶段若待识别生物特征数据和原始生物特征数据近似, 可利用辅助信息重构得到原始生物特征数据。模糊提取器方案是从生物特征数据中提取近似均匀分布的随机信号作为密钥, 使用熵损失来衡量方案的安全性。Álvarezd 等人[7]将模糊提取器用于虹膜识别, 利用多项式产生大量随机点与纠错码相结合, 再将其哈希值与原始虹膜特征数据的差作为辅助数据用于识别阶段。Chang 等人[8]将模糊提取器与一种新颖的逐位加密方案相结合来生成可撤销的生物特征模板。

基于密钥绑定的生物特征加密系统以模糊承诺和模糊金库方案为代表。模糊承诺方案是由 Juels 等人[9]提出, 基本思想是把纠错码技术引入到生物特征加密过程中, 允许注册阶段与识别阶段的生物特征存在小范围的偏差。Hao 等人[10]使用 Hadamard 码和 Reed-Solomon 码双层纠错码提高纠错码的容错率, 设计了只有在虹膜特征数据和令牌都正确的情况下才能通过识别的双因子认证方案。Ouda 等人[11]将可撤销的生物特征模板和模糊承诺方案相结合, 在 BioEncoding[12]方案的基础上生成可变长的虹膜模板作为模糊承诺的输入以提高系统的纠错能力, 能达到更好的识别精度。模糊金库最早由 Juels 等人[13]提出, 基本思想是将生物特征投影到多项式上得到有限点集, 再添加大量的杂凑点构成金库, 通过这种方式来保护原始的生物特征信息。Zhang 等人[14]利用随机点与用户特征集合构造一个随机多项式, 使用该多项式系数表示金库。

加盐法最早由 Jin 等人[15]提出, 基本思想是用小波傅里叶梅琳变换特征和一组伪随机数对生物特征数据进行迭代内积, 再经过二值化后得到哈希序列存于服务器中。Zuo 等人[16]提出了 BIN-SALT 和 GRAY-SALT 两种盐化方法用于虹膜识别。Asaker 等人[17]提出了一种基于加盐法的可撤销虹膜识别方案, 将原始的虹膜特征数据与加密的用户特有的掩码图像进行异或作为最终的虹膜模板。

不可逆变换常借助查找表、布隆过滤器、随机映射等构建可撤销的生物特征模板。Dwivedi 等人[18]使用随机查找表生成可撤销的生物模板, 首先将虹膜特征块转化为对应的十进制值, 再使用随机查找表对应位置的二进制块进行替换。Jeong 等人[19]在文献[18]的基础上增加局部排序操作, 能够有效提高识别精度。Rathgeb 等人[20]提出了基于布隆过滤器的可撤销模板生成方法。Ajish 等人[21]基于双层布隆滤波器的特征变换提高了数据压缩率、模板保护、匹配的响应时间和匹配的准确性。Pillai 等人[22]提出了一种基于随机投影和稀疏表示的虹膜识别框架。Yang 等人[23]提出了一种基于特征自适应随机投影的方法, 该方法由一个基本矩阵和局部特征槽生成投影矩阵, 当特征槽不同时, 生成的投影矩阵也不同, 且使用后的投影矩阵将被丢弃。Singh 等人[24]使用卷积神经网络提取虹膜特征数据, 再将其投影到随机子空间, 随后使用 KNN 分类器和 SHA-3 散列后得到最终的虹膜模板。

不可逆变换方法具有丰富的变换形式, Zhao 等人[25,26]提出了基于随机响应技术和聚合块信息以及基于局部排序的两种安全虹膜识别方案。Lai 等人[27]基于最小哈希函数提出索引优先的安全虹膜方案, 通过将虹膜特征数据分块, 记录每块中第一次出现 1 的索引值来估算两个集合的相似程度。Sadhya 等人[28]引入了一种基于局部敏感哈希的随机位采样的安全虹膜识别方案。

此外, 还有很多其他方法用于安全虹膜识别, Kolberg 等人[29]结合同态加密技术对识别阶段进行隐私保护。Morampudi 等人[30]通过完全同态加密实现了一个基于机器学习的安全虹膜识别方案。Blanton 等人[31]考虑了在非交互式单服务器和不共谋的多服务器两种情况下构建基于安全多方计算的虹膜识别系统。Bauspieß 等人[32]将秘密共享和混淆电路技术进行结合提出了能有效阻止爬山法攻击和量子计算攻击的安全虹膜识别方案。Zhao 等人[33]结合负数据库提出安全虹膜识别方案, 注册阶段由虹膜特征数据生成负数据库, 识别阶段将待识别的虹膜特征数据与负数据库进行距离估算来判断能否通过认证。

3 预备知识

本节的主要内容分为两部分: 3.1 节对本文方案的应用场景以及拟解决的问题展开介绍; 由于本文提出的保护用户数量信息的安全虹膜识别方案需要与现有的安全虹膜识别方案相结合, 考虑到文献[28]中提出的随机位采样(Randomized Bit Sampling, RBS)方案具有细粒度的良好识别精度, 因此选择在该方案的基础上实施本文的方案。为了简化表达, 将随机位采样方案简写为 RBS 方案, 因此, 3.2 节主要介绍 RBS 方案的模板生成方法以及相似度比较方法。

3.1 问题描述

本文的方案是基于以下不安全场景提出的, 这也是目前安全虹膜识别领域常见的场景假设:

1. 假设传输链路是不可信的, 因此注册阶段和识别阶段都使用虹膜模板来实现。

2. 假设注册阶段是安全的, 在识别阶段, 攻击者可以通过一些攻击手段入侵服务器, 对服务器中存储的信息进行获取, 但并不具备长期监视服务器的能力。

目前的安全虹膜识别方案中服务器中存储的虹膜模板数量与该系统的合法用户数量相等, 如图 1 所示, 虹膜识别中将可能面临以下三个问题:

Q1: 攻击者入侵服务器后获得服务器中存储的模板数量, 根据模板数量获得合法用户数量, 此时, 该系统的合法用户数量遭到泄露。

Q2: 攻击者在不同时刻入侵服务器, 将会获得该时间间隔内的合法用户数量信息变化情况, 此时, 该系统在这段时间内的新增用户数量遭到泄露。

Q3: 攻击者在不同时刻对服务器进行监视, 由于同一个合法用户在正确匹配的前提下将会匹配上服务器中的同一个注册模板, 因此攻击者可能会关联到同一个合法用户的匹配记录, 此时, 该系统将面临关联攻击。

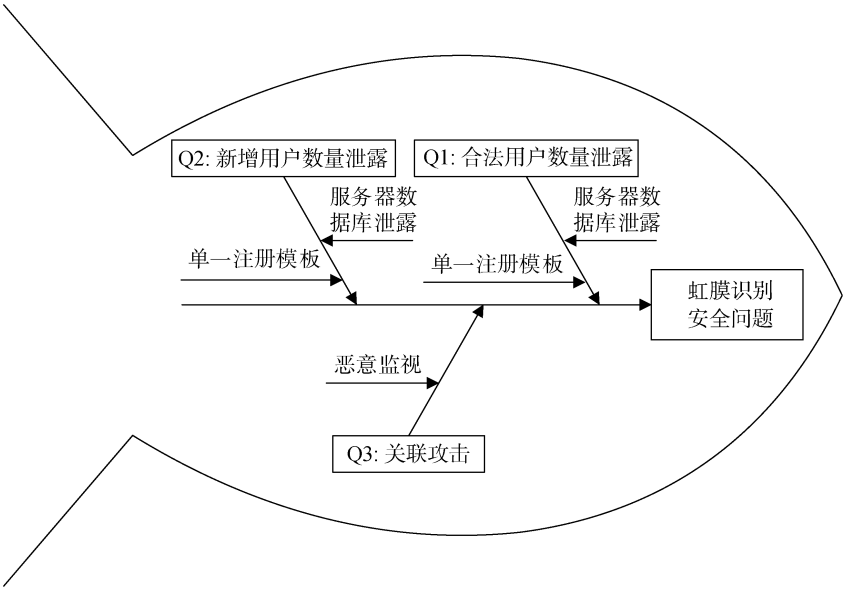

图 1 虹膜识别安全问题鱼骨图

Figure 1 Fishbone diagram of iris recognition security problems

综上所述, 目前的安全虹膜识别方案可能面临上述三种安全问题, 在具有高安全性需求的应用场景中, 比如特工组织、军工涉密人员等, 这些机密组织的内部成员数量信息属于机密信息, 一旦泄露将会加大身份信息暴露的风险, 因此本文提出一种保护用户数量信息的安全虹膜识别方案来解决上述场景下面临的安全问题。

3.2 RBS 方案

3.2.1 模板生成方法

RBS 方案的模板生成方法分为以下几个步骤, 示意图如图 2 所示。

1. 将虹膜特征数据 I (长度为 m 位)分为 n 块(每块长度为 b 位), 即 $b\times n=m$, 记每块为 $B_i \mid i=1,2\cdots n$。

2. 对块大小 b 进行索引采样, 每次从 b 个索引中随机选择 k 位, 重复 l 次, 得到 l 个大小为 k 的不同的哈希函数, 记为 $h_j \mid j=1,2\cdots l$, 其集合用 $H(B_i)$ 表示, 将该哈希函数生成过程用函数 $\mathcal{F}_1$ 表示。

$$\mathcal{F}_1 : B_i \xrightarrow{\mathbb{R}} \mathcal{H}(B_i) and \mid h_j \mid = k, \forall h_j \in \mathcal{H}(B_i) \quad (1)$$

3. 令 $[B]$ 为块 B_i 的随机采样索引, 对于 $[B]$ 的计算只进行一次, 即同一个虹膜特征数据每个块的 $[B]$ 是相同的, 但不同虹膜特征数据的 $[B]$ 是不同的, 相当于用户特有令牌, 将与虹膜模板一起存储在服务器中。

4. 根据 $[B]$ 从 B_i 取出对应索引的值得到标记位。

5. 将标记位转化为对应的十进制值得到十进制块。

$$c(i,j)=bin2dec(\lceil h_j \rceil) \quad (2)$$

6. 通过取模系数 T 对十进制块进行取模操作得到不可逆块。

$$c'(i,j)=c(i,j)\bmod(T\times 2^k) \quad (3)$$

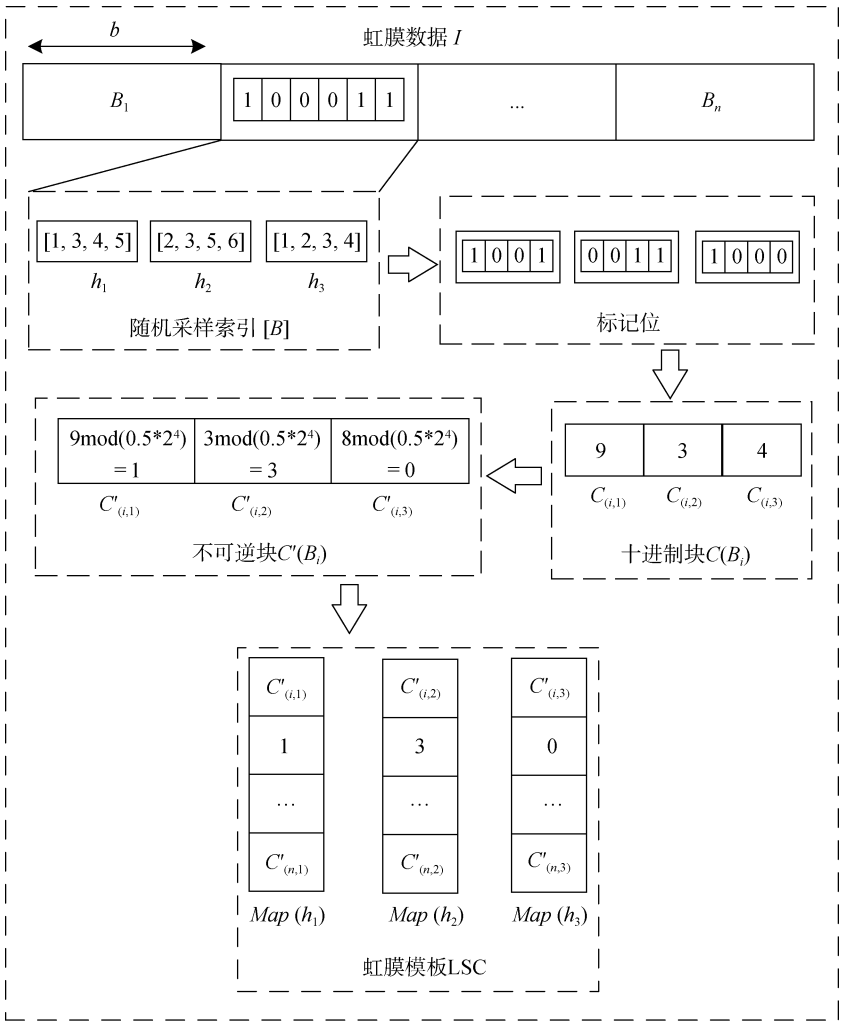

图 2　当 *l*=3, *k*=4 时, RBS 方案示意图

Figure 2　Schematic diagram of the RBS scheme when *l*=3, *k*=4

7. 将不可逆块与块索引 i 存储为 *Map* 键值对。

$$Map(h_j)=\{\underbrace{c'(i,j)}_{value},\ \underbrace{i}_{key}\} \tag{4}$$

8. 重复步骤 4-7, 直至所有的块都完成操作, 得到最终的虹膜模板。

$$\mathrm{LSC}=\left\{Map(h_j)\right\}\mid j=1,2\cdots l \tag{5}$$

3.2.2　相似度计算方法

两个虹膜模板的相似度计算通过对虹膜模板进行逐块比较, 再将每块的相似度累加后取平均值得到最终的相似度。记两个待比较的虹膜模板的某块分别为 $B_i(E)$, $B_i(Q)$, 记每块经过模板生成方法得到的不可逆块为 $E_i=\{e_1,e_2\cdots e_l\}$, $Q_i=\{q_1,q_2\cdots q_l\}$。通过公式(6)进行相似度计算:

$$S=\frac{\sum_{i=1}^{n}(\frac{|e_j=q_j|}{l})}{n}\mid j=\{1,2\cdots l\},\forall e_j\in E_i,\forall q_j\in Q_i \tag{6}$$

4　用户数量信息保护方案

本节主要分为两个部分: 4.1 节对本文的方案展开详细叙述; 4.2 节介绍本文的方案如何与 RBS 方案进行结合。

4.1　基础方案

保护用户数量信息的关键思想是为每个用户生成不定个数的虹膜模板存储在服务器中, 攻击者不能通过服务器中的虹膜模板数量获得合法用户数量信息, 因此, 注册模板数量的生成方式决定了用户数量信息的保护程度。现有的安全虹膜识别系统大多都依赖于可撤销令牌来生成可撤销的虹膜模板, 如异或串、置换序列等。本文就是通过每个用户使

用不同的可撤销令牌生成多个不同的注册模板。安全虹膜识别系统包括注册阶段和识别阶段, 下面将从这两个阶段展开叙述。

在注册阶段, 保护用户数量信息的安全虹膜识别方案具体可分为以下步骤:

1. 服务器通过设置参数 mb_1, mb_2, tn_{low}, D 来设置虹膜识别系统每个用户可注册模板数量的范围。

2. 服务器首先生成 tn_{low} 个不同的可撤销令牌存于服务器中, 服务器根据 D 来判断当前用户的系统最大位数, 将 tn_{low} 个可撤销令牌, mb_1 / mb_2 发送给当前用户, 其中 mb_1 和 mb_2 会根据规则只发送一个。

3. 记用户收到的系统最大位数为 mb_i, 显然 $mb_i = mb_1$ 或者 $mb_i = mb_2$, 用户将会从 1 至 mb_i 之间随机选择一个数作为用户最大位数, 记为 mb_x, 其中 $1 \leqslant mb_x \leqslant mb_i$。

4. 用户从自身的虹膜特征数据中随机选择连续的 mb_x 位, 将其转换为对应的十进制数值, 再与系统设置的最小模板数量 tn_{low} 相加得到该用户最终需要生成的注册模板数量, 记为 tn_i'。

5. 用户生成 $tn_i' - tn_{low}$ 个可撤销令牌, 结合服务器之前发送的 tn_{low} 个可撤销令牌, 此时用户将得到 tn_i' 个可撤销令牌。

6. 用户使用获得的 tn_i' 个可撤销令牌结合现有的某种安全虹膜识别方案生成 tn_i' 个注册模板, 并发送给服务器。

7. 服务器接收用户发送的 tn_i' 个注册模板, 判断模板是否有效, 如果是, 则执行步骤 8, 否则, 执行步骤 9。

8. 服务器发送“注册成功”信息给用户, 并存储用户发送的 tn_i' 个注册模板。

9. 服务器发送“注册失败”信息给用户, 并丢弃用户发送的 tn_i' 个注册模板。

在步骤 1 中, mb_1, mb_2 为两个不同的系统最大位数, 用于控制用户生成的最大模板数量, 且 $mb_1 < mb_2$; tn_{low} 为系统最小注册模板数量, 用于防止用户信息关联攻击; D 为系统设置的时间周期参数, 每隔时间 D 选择一个用户使用 mb_2 为其系统最大位数, 而其余用户使用 mb_1 为其系统最大位数, 这是出于安全性的考虑, 此外, 还定义 tn_{up} 为系统最大注册模板数量, 通过其他参数值计算获得, 用于调整系统安全性和识别效率之间的平衡。

在步骤 4 中, 虹膜特征数据为二进制串的形式, 根据自身虹膜特征数据选择注册模板数量的方式将增大攻击者的推测难度, 分析过程见 6.1 节。由于用户通过随机位数选择得到的十进制值可能为 0, 因此需要和最小模板数量 tn_{low} 相加, 确保每个用户至少会生成 tn_{low} 个注册模板。经过上述操作, 可以得到该用户的系统最大注册模板数量为 $tn_{up} = 2^{mb_2} - 1 + tn_{low}$。

此外, 当用户收到的系统最大位数为 mb_2 时, 可能需要生成较多的注册模板, 这将会对该用户造成较大的负担, 尽管这属于少数情况。此时, 服务器可以通过每隔 D 生成伪虹膜特征数据, 通过伪虹膜特征数据使用 mb_2 得到需要生成的注册模板数量, 再生成对应数量的虹膜模板存于服务器中, 这种方式能够避免被选中的用户生成较多数量的模板。但是当攻击者具有分辨真实虹膜特征数据和伪虹膜特征数据生成的虹膜模板不同的能力时, 这种方式可能不够安全, 因此本文仍采用由用户真实虹膜特征数据生成虹膜模板的方式。

在步骤 6 中, 用户使用服务器发送的 tn_{low} 个可撤销令牌生成模板是为了维持识别精度; 不同用户的注册模板数量 tn_i' 是不相同的, 因此攻击者通过服务器中存储的模板数量无法获得合法用户数量信息。

在识别阶段, 保护用户数量信息的安全虹膜识别方案具体可分为以下步骤:

1. 用户从存储的 tn_{low} 个可撤销令牌中随机选择一个生成待识别的虹膜模板, 发送给服务器。

2. 服务器接收用户发送的待识别模板, 判断模板是否有效, 如果是, 则执行步骤 3, 否则, 执行步骤 5。

3. 服务器将待识别模板和服务器中的所有注册模板进行比较, 具体的比较方法需要参考不同的安全虹膜识别方案, 将得到的相似度分数和阈值进行比较, 判断是否通过识别, 如果是, 执行步骤 4, 否则, 执行步骤 5。

4. 服务器发送“识别成功”信息给用户, 并丢弃用户发送的待识别模板, 用户获得进入系统的权限。

5. 服务器发送“识别失败”信息给用户, 并丢弃用户发送的待识别模板。

在步骤 1 中, 用户使用存储的 tn_{low} 个可撤销令牌生成待识别模板, 并非随机生成一个可撤销令牌生成待识别模板。原因在于即使是同一个虹膜特征数据使用不同的可撤销令牌生成的虹膜模板是完全不同的, 是无法匹配成功的。而服务器中存在该用户使用这 tn_{low} 个可撤销令牌生成的注册模板, 这样能确保对于每个合法用户的待识别模板, 在服务器中能找到使用同样可撤销令牌生成的注册模板, 这是维持原始安全虹膜识别方案的识别精度的主要原因。

4.2 RBS_NP 方案

为了验证本文方案的性能, 将在随机位采样方案的基础上实施本文的方案。为了简化表达, 使用 RBS_NP 方案表示在 RBS 方案的基础上实施了本文的方案, 本节主要介绍 RBS_NP 方案的注册阶段, 当 $mb_x=2$, $tn_{low}=2$ 时, RBS_NP 方案示意图如图 3 所示, 简单描述为以下步骤:

1. 注册模板数量生成过程:

1) 服务器设置安全虹膜识别系统的参数(包括 mb_1, mb_2, tn_{low}, D), 利用 RBS 方案中的 k, b, l 生成 tn_{low} 个不同的 $[B]$, 其中 $[B]$ 充当了可撤销令牌的角色, 根据 D 获得当前用户的系统最大位数 mb_i, 再将 mb_i 以及 tn_{low} 个 $[B]$ 发送给用户。

2) 用户接收服务器发送的 mb_i 以及 tn_{low} 个 $[B]$, 再通过 4.1 节的步骤 4 获得注册模板数量 tn_i', 同样利用 k, b, l 生成 $tn_i'-tn_{low}$ 个不同的 $[B]$。

2. 多个注册模板生成过程: 用户使用这 tn_i' 个 $[B]$ 按照 3.1 节的步骤生成 tn_i' 个注册模板。

通过以上步骤就能完成 RBS_NP 方案中用户注册模板的生成, 可以看出, 在 RBS 方案的基础上实施本文提出的方案只需要在获得当前用户的注册模板数量之后, 生成对应数量的 $[B]$, 再根据这些 $[B]$ 进行模板生成操作即可, 因此本文方案是容易实施的。

5 安全性分析

本节主要分为四个部分: 5.1 节对合法用户数量信息的安全性进行理论分析; 5.2 节对新增用户数量信息的安全性进行理论分析; 5.3 节对关联攻击的防御进行理论分析, 以上 3 小节是针对本文方案如何解决 3.1 节中提出的三个待解决问题进行的详细描述; 5.4 节简要分析本文方案对三大安全性的影响。

5.1 合法用户数量信息安全性分析

本节探讨攻击者通过服务器中存储的虹膜模板数量正确推测合法用户数量的可能性, 分为以下两步。第一步为正常的注册流程, 即指定数量的用户根据本文的方案进行注册, 获得服务器中存储的虹膜模板数量。第二步为攻击者通过服务器中存储的虹膜模板数量推测合法用户数量。

记 un 为合法用户数量, tn 为服务器中存储的虹膜模板数量, 其中 un, tn 是两个变量, 给定任意两个常量: un', tn', 需要获得当 $tn=tn'$ 时, $un=un'$ 的概率, 即攻击者获得服务器中存储的 tn' 个虹膜模板是由 un' 个用户生成的概率。当 $tn=tn'$ 时, 可以得到可能生成模板数量为 tn' 的用户数量范围, 记 un_{low} 为最小用户数量, un_{up} 为最大用户数量, 则 $un_{low}=\lceil tn'/tn_{up}\rceil$, $un_{up}=\lfloor tn'/tn_{low}\rfloor$。假设真实情况下出现不同的合法用户数量的概率是相等的, 即 $Pr(un=un_i)=Pr(un=un')=\dfrac{1}{un_{up}-un_{low}}$, $\forall un_i\in[un_{low},un_{up}]$, 可以得到, 当 $tn=tn'$ 时, $un=un'$ 的概率如公式(7)所示:

$$Pr(un=un'\mid tn=tn')=\frac{Pr(tn=tn'\mid un=un')}{\sum_{un_i=un_{low}}^{un_{up}}Pr(tn=tn'\mid un=un_i)} \tag{7}$$

推导过程如下:

$$\begin{aligned}
&Pr(un=un'\mid tn=tn')\\
&=\frac{Pr(tn=tn'\mid un=un')\times Pr(un=un')}{Pr(tn=tn')}\\
&=\frac{Pr(tn=tn'\mid un=un')\times Pr(un=un')}{\sum_{un_i=un_{low}}^{un_{up}}(Pr(tn=tn'\mid un=un_i)\times Pr(un=un_i))}\\
&=\frac{Pr(tn=tn'\mid un=un')\times\dfrac{1}{un_{up}-un_{low}}}{\dfrac{1}{un_{up}-un_{low}}\times\sum_{un_i=un_{low}}^{un_{up}}Pr(tn=tn'\mid un=un_i)}\\
&=\frac{Pr(tn=tn'\mid un=un')}{\sum_{un_i=un_{low}}^{un_{up}}Pr(tn=tn'\mid un=un_i)}
\end{aligned}$$

由于攻击者可能会计算出所有可能用户数量的概率, 再根据某种策略进行推测, 因此接下来介绍每一种可能的用户数量 un_i 是合法用户数量的概率

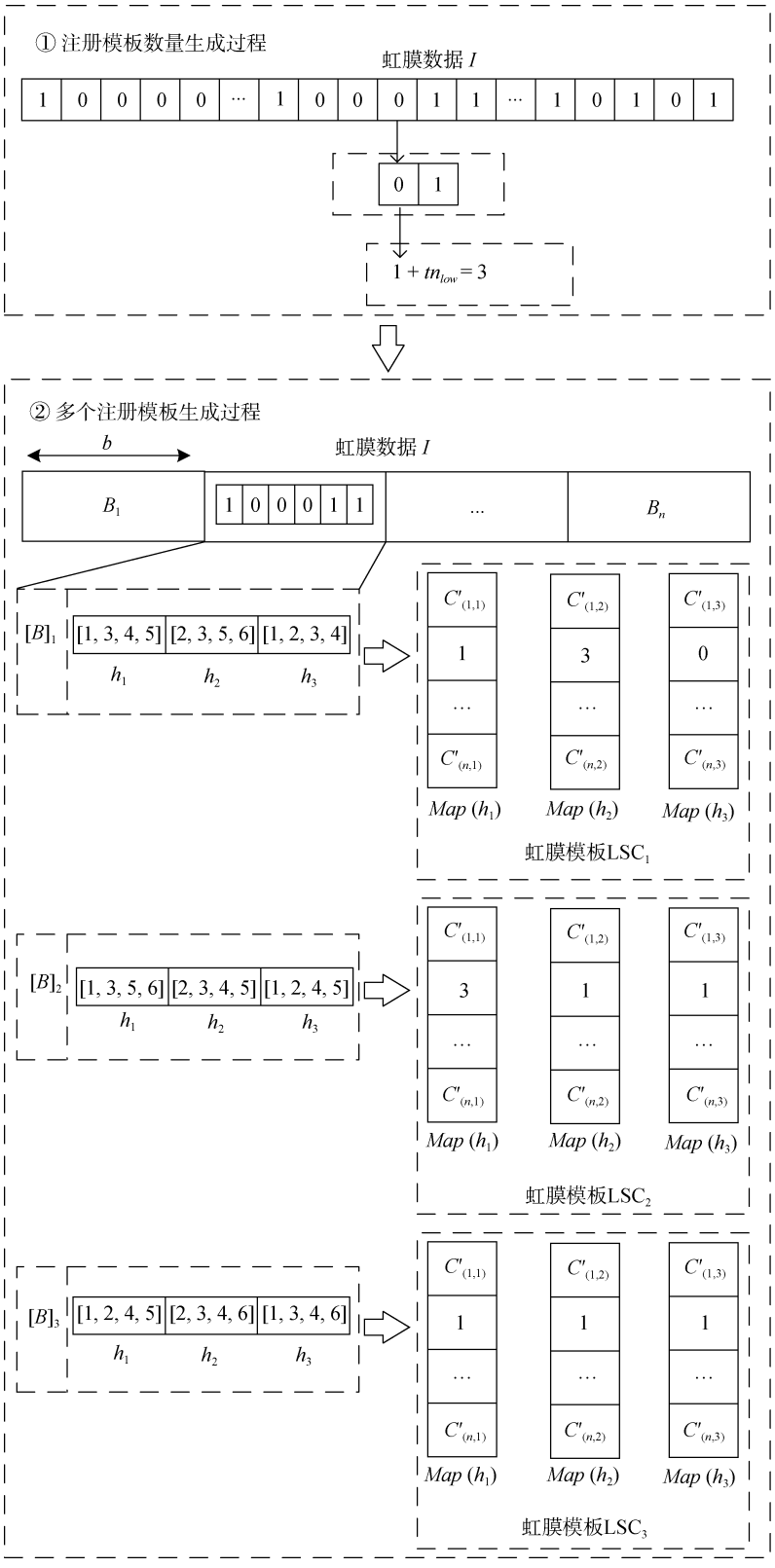

图 3　当 mb_x=2, tn_{low}=2 时, RBS_NP 方案示意图

Figure 3　Schematic diagram of the RBS_NP scheme when mb_x=2, tn_{low}=2

计算，即计算 $Pr(un=un_i \mid tn=tn'), \forall un_i \in [un_{low}, un_{up}]$。可以分为两步来实现，其一，攻击者计算出系统最大位数分别为 mb_1 和 mb_2 时，用户数量为 1 至 un_{up} 时生成的所有可能模板数量的概率矩阵，下面统一简称为概率矩阵，记 dp_i 为用户数量为 $un_i(1 \leqslant un_i \leqslant un_{up})$ 时的概率矩阵。主要思想为：首先

计算 dp_1, 也即初始概率矩阵 $init$, 假设此时用户的系统最大位数为 mb_x, 则该用户可能选择的系统最大位数为 $[1, mb_x]$, 每一种位数所占比例为 $1/mb_x$, 再将不同位数时获得的模板数量的概率相加即可得到 dp_1, 见算法 1 中的过程 1; dp_2 是在上一状态 dp_1 的基础上增加一个用户, 因此由 dp_1 和 $init$ 通过滑动窗口即可得到 dp_2, 以此类推, 最终将会得到用户数为 1 至 un_{up} 时的所有概率矩阵, 见算法 1 中的过程 4。

其二, 攻击者对于每一种可能的用户数量 un_i, 计算系统最大位数为 mb_1 和 mb_2 的用户数量组合中模板数量为 tn' 的最大概率。假设系统最大位数为 mb_1 的用户数量为 un_1', 系统最大位数为 mb_2 的用户数量为 un_2', 显然 $un' = un_1' + un_2'$。通过算法 2 可以获得 un_1' 和 un_2' 组合后模板数量为 tn' 的概率。主要思想为: 通过算法 1 中计算得到的 un_1' 个用户选择 mb_1 时生成模板数量的概率矩阵 res_1、un_2' 个用户选择 mb_2 时生成模板数量的概率矩阵 res_2, 利用滑动窗口来获得这种组合情况下 un' 个用户生成模板数量的概率矩阵。重复算法 2 遍历所有组合情况, 选择所有组合中模板数量为 tn' 的最大概率即可得到 $Pr(un = un_i \mid tn = tn'), \forall un_i \in [un_{low}, un_{up}]$。

算法 1. 相同系统最大位数概率矩阵算法

输入: 最大可能用户数量 un_{up}, 当前用户的系统最大位数 mb_x。

输出: 1 至 un_{up} 个用户生成所有可能模板数量的概率矩阵 res。

过程 1. 初始化用户数为 1 的所有可能模板数量的概率

FOR $i = 1 : mb_x$

  FOR $j = 0 : 2^i - 1$

    $dp_1[j] += (1.0/mb_x) \times (1.0/2^i)$

过程 2. 将 dp_1 赋值给 $init$

过程 3. 将 dp_1 加入 res

过程 4. 利用滑动窗口求出概率矩阵

FOR $i = 2 : un_{up}$

  FOR $j = 0 : len(dp_{i-1}) - 1$

    FOR $k = 0 : 2^{mb_x} - 1$

      $tmp[j+k] += dp_{i-1}[j] \times init[k]$

  $dp_i = tmp$

  将 dp_i 加入 res

算法 2. 组合系统最大位数概率矩阵算法

输入: mb_1 和 mb_2 的概率矩阵 res_1 和 res_2。

输出: un' 个用户生成所有可能模板数量的概率矩阵 res。

过程 1. 利用滑动窗口求出概率矩阵

FOR $i = 0 : len(res_1) - 1$

  FOR $j = 0 : len(res_2) - 1$

    $res[i+j] += res_1[i] \times res_2[j]$

值得注意的是, 攻击者通过公式(7)只能得到如何使推测成功的概率更大, 但无法判断自己是否推测成功。此外, 攻击者可以通过置信区间来衡量合法用户数量处于某一范围内的概率, 假设选择区间 $[un_1, un_2]$, 其中 $un_1 \geqslant un_{low}, un_2 \leqslant un_{up}$, 通过公式(8)可以得到当 $tn = tn'$ 时 $un \in [un_1, un_2]$ 的概率。

$$Pr(un_1 \leqslant un \leqslant un_2 \mid tn = tn') = \sum_{un_i = un_1}^{un_2} Pr(un = un_i \mid tn = tn') \tag{8}$$

5.2 新增用户数量信息安全性分析

本节探讨攻击者在不同时刻窃取服务器中虹膜模板数量, 通过获得的虹膜模板数量信息推测在这段时间内新增的用户数量。为了简化分析过程, 不考虑发生模板泄露时出现的删除注册模板现象, 即假设随着时间的增长, 用户注册数量的增多, 服务器中的虹膜模板数量是不断增长的。假设 t_a 时刻服务器中的虹膜模板数量为 tn_a', t_c 时刻服务器中的虹膜模板数量为 tn_c' (默认 $tn_c' \geqslant tn_a'$), 这将可能面临以下三种情况。

1. 当 $tn_c' = tn_a'$ 时, 服务器中的虹膜模板数量没有发生变化, 则攻击者可以知道在这段时间内没有发生用户注册。

2. 当 $0 < tn_c' - tn_a' < 2 \times tn_{low}$, 服务器中的模板数量的增量小于系统最小注册模板数量的两倍, 则攻击者可以知道在这段时间内有一个新增的注册用户。

3. 当 $tn_c' - tn_a' \geqslant 2 \times tn_{low}$, 服务器中的模板数量的增量大于等于系统最小注册模板数量的两倍, 此时攻击者无法确定这段时间内新增的合法用户数量, 只能令 $tn' = tn_c' - tn_a'$, 利用公式(7)进行计算。

显然, 第 3 种情况出现的概率最大, 当处于第 3 种情况时, 即使 t_c 时刻和 t_a 时刻误差较小, 新增用户数量较少, 但在合法用户数量较少时, 攻击者同样

有较大的概率推测失败, 在 6.1 节的实验部分将证实这一点, 因此可以认为本文的方案对于新增用户数量的保护在大多数情况下是有效的。

5.3 关联攻击防御分析

关联攻击产生的原因在于服务器中只存储了每个合法用户的一个注册模板, 因此通过多次监视服务器获得虹膜模板的匹配情况就能获得用户的登录情况, 若多次监视到同一虹膜模板就能对同一用户的多次登录信息进行关联。

本文方案可以进一步降低关联攻击的发生频率。在识别阶段只需要一个待识别模板即可完成对用户的识别, 由于使用不同可撤销令牌生成的虹膜模板是无关的, 因此同一用户每次进行识别时生成的待识别模板并非是完全一样的。当用户使用的待识别模板不同时, 服务器中匹配成功的将是不同的注册模板, 所以攻击者无法关联用户使用不同待识别模板的登录记录, 能一定程度上缓解关联攻击。在实际应用中, 可以将 tn_{low} 设置为更大的值来加大对关联攻击的抵抗程度, 但相应的需要生成更多数量的注册模板, 会降低识别效率。

5.4 三大安全性要求分析

保护用户数量信息的安全虹膜识别方案需要与现有的安全虹膜识别方案进行结合, 因此所提方案不应该对原始的安全虹膜识别方案的安全性产生较大的影响, 仍需满足可撤销性、不可逆性以及不可连接性, 以下分析建立在原始安全虹膜识别方案满足三大安全性要求的前提下展开叙述。

1. 可撤销性: 本文的方案是通过多个可撤销令牌生成多个虹膜模板的过程, 当服务器中存储的虹膜模板发生泄露时, 用户需要重新获得注册模板数量 tn_i', 生成 tn_i' 个可撤销令牌, 即可得到新的 tn_i' 个注册模板。此外, 用户还需要使用注册阶段服务器发送的 tn_{low} 个不同的可撤销令牌生成 tn_{low} 个虹膜模板, 将这 tn_{low} 个虹膜模板发送给服务器, 服务器将存储的所有模板与这 tn_{low} 个虹膜模板进行匹配, 删除匹配成功的虹膜模板。当然, 本文方案会使用更多数量的可撤销令牌, 但是现有的安全虹膜识别方案中可供选择的可撤销令牌数量是远远大于本文方案需要的数量。综上所述, 本文方案是满足可撤销性的。

2. 不可逆性: 本文的方案中尽管每个用户在服务器中存储了多个注册模板, 但这些模板是使用不同的可撤销令牌生成的, 攻击者无法判断哪些模板属于同一个用户, 因此攻击者仍然只能尝试从单个虹膜模板逆推回原始虹膜特征数据。此外, 攻击者只能获得服务器中存储的 tn_{low} 个可撤销令牌, 对于用户生成的 $tn' - tn_{low}$ 个可撤销令牌, 攻击者并不知晓, 因此攻击者无法确定服务器中存储的虹膜模板具体是使用哪一个可撤销令牌生成的, 只能使用已知的 tn_{low} 个可撤销令牌对服务器中存储的虹膜模板进行逆推, 这将对攻击者获得原始虹膜特征数据造成很大的困扰。综上所述, 本文方案能提升原始安全虹膜识别方案的不可逆性。

3. 不可连接性: 若原始安全虹膜识别方案满足不可连接性, 这说明同一个用户在服务器存储的多个虹膜模板是无法关联到该用户的。同样的, 不同服务器中使用的是不同的可撤销令牌, 也是无法关联到同一个用户的。综上所述, 本文方案是满足不可连接性的。

6 实验

本文使用 CASIA-IrisV3[34]虹膜特征数据集进行实验, 包含 CASIA-Iris-Interval, CASIA-Iris-Twins, CASIA-Iris-Lamp 三个子数据集, CASIA-Iris-Interval 包含来自 249 个人的 2639 张虹膜图片, CASIA-Iris-Twins 包含来自 200 个人的 3183 张虹膜图片, CASIA-Iris-Lamp 包含来自 411 个人的 16212 张虹膜图片, 再使用 OSIRISv4[35]系统将这些虹膜图片转化为 1536 位的二进制串。由于左眼数据和右眼数据相当于是两个不同的虹膜特征数据, 而仅使用左眼特征数据已经足够验证本文方法的性能, 因此与文献[20]一致, 本文实验中只使用左眼数据, CASIA-Iris-Interval 共计 198 类, CASIA-Iris-Twins 共计 200 类, 为了更利于不同用户数量信息的相关实验对比, CASIA-Iris-Lamp 只取前 200 类展开实验, 并且实验部分默认使用 CASIA-Iris-Interval 数据集, 只有不同用户数量信息的相关实验会使用上述的三个数据集。

本节的内容分为两个部分: 6.1 节对用户数量信息的安全性展开实验, 6.2 节对本文方案的识别精度展开实验。

6.1 安全性实验

6.1.1 实验设置

本节通过实验来测试本文的方案对用户数量信息的保护程度, 重复实验次数设置为 3000 次。假设保护用户数量信息的安全虹膜识别系统目前已经有 un' 个用户完成注册, 在服务器中得到 tn' 个模板, 攻击者将已知服务器中存储的模板数量 tn' 以及本文方案中的所有参数设置(包括 tn_{low}, mb_1, mb_2, D, 以

及RBS方案中的n，b，l，k，T)，并根据这些已知信息推测合法用户数量信息。通过以下三个评价指标来衡量本文方案对用户数量信息的保护程度。

(1) 在已知的参数设定下，攻击者首先获得可能生成tn'的用户数量范围，利用公式(7)获得不同用户数量的概率，攻击者推测其中最大概率的用户数量为合法用户数量，显然，推测成功的概率为$max(Pr(un=un_i \mid tn=tn')), \forall un_i \in [un_{low}, un_{up}]$，推测失败(speculation failed, sf)的概率为sf,如公式(9)所示:

$$sf=(1-max(Pr(un=un_i \mid tn=tn')))\times 100\%, \quad \forall un_i \in [un_{low}, un_{up}] \tag{9}$$

(2) 在已知的参数设定下，攻击者同样推测最大概率的用户数量，令该用户数量为un_i，un_i与un'的相对误差为δ，如公式(10)所示。

$$\delta=\frac{|un_i-un'|}{un'} \tag{10}$$

(3) 在已知参数设定下，攻击者通过不同用户数量的概率通过公式(11)计算出期望值E，推测E为合法用户数量，记E与un'的相对误差为δE，如公式(12)所示。

$$E=\sum_{un_i=un_{low}}^{un_{up}} Pr(un=un_i \mid tn=tn')\times un_i \tag{11}$$

$$\delta E=\frac{|E-un'|}{un'} \tag{12}$$

6.1.2 实验结果与分析

为了验证本文方案的安全性，本小节进行以下三组实验，验证不同参数对方案安全性的影响，分别为: 1)合法用户数量un对安全性的影响; 2)使用mb_2注册的用户数量un_2'对安全性的影响; 3)系统最大位数mb_1和mb_2对安全性的影响。另外，由于在本文的方案中，使用mb_2进行注册的用户数量un_2'是与时间周期D相关的，但是考虑到无法获得真实情况下不同时间的注册用户数量，因此通过对un_2'进行调整的方式代替D展开实验。考虑到$tn_{low}=2$时已能够预防关联攻击，且tn_{low}仅决定模板数量的下界，因此下列实验都固定$tn_{low}=2$。

1) 实验一: 合法用户数量un对安全性的影响

首先固定$mb_1=2$，$mb_2=6$，$un_2'=4$，调整合法用户数量un展开实验，指定数量的用户按照本文方案进行注册，根据注册阶段得到的模板数量，通过上述三个指标来展示对用户数量信息的保护程度。

结果如表 1 所示，可以观察到随着用户数量的增加，攻击者推测最大概率的用户数量成功的概率将逐渐降低，原因在于用户数量越多，可能生成的模板数量就越多，面临的组合情况将更多，因此第一个指标sf将越高。由于un_2'固定为 4，随着用户数量的增加，un_2'与un的比值逐渐降低，而mb_2一般设置的较大，将会使服务器中存储的模板数量出现较大的波动，比值降低的情况下造成的波动就会相对较小，因此第二个指标δ呈下降趋势。而期望值与真实值的相对误差反应了一种较为平均的状态，第三个指标δE并不存在明显规律的原因在于该值并不仅仅和un相关，同时和un_2'有关。并且在三个数据集上三个指标都呈现出一致的变化规律，因此本文方案能适用于不同的数据集。

表 1 $un_2'=4$时，不同的un对安全性的影响

Table 1 When $un_2'=4$, the impact on security of different un

数据集	评价指标	un					
		10	20	40	60	80	100
CASIA-Iris-Interval	sf	83.76%	88.54%	92.52%	94.41%	95.52%	96.23%
	δ	123.20%	62.35%	31.38%	21.65%	16.54%	13.29%
	δE	83.90%	34.95%	14.93%	12.45%	13.14%	15.06%
CASIA-Iris-Twins	sf	84.05%	88.56%	92.54%	94.40%	95.50%	96.24%
	δ	125.45%	61.36%	31.37%	21.04%	16.25%	13.55%
	δE	85.12%	33.43%	14.21%	12.03%	13.17%	14.78%
CASIA-Iris-Lamp	sf	84.06%	88.69%	92.58%	94.44%	95.52%	96.25%
	δ	125.78%	63.64%	32.40%	21.99%	16.66%	13.82%
	δE	85.27%	34.64%	14.69%	11.83%	12.90%	14.51%

2) 实验二: 使用 mb_2 注册的用户数量 un_2' 对安全性的影响

固定 $un=60$, $mb_1=2$, $mb_2=6$, un_2' 对安全性的影响如表 2 所示, 对于前两个安全指标 sf 与 δ, 随着 un_2' 的增大逐渐增大, 因为 un_2' 的增大将会使服务器中模板数量增多以及 un_2' 与 un 的比值逐渐增大。对于第三个指标 δE, un_2' 的增大并不会带来线性的增长或是降低, 这同样是因为 un_2' 的选择和合法用户数量 un 相关, 但两者的关系在实际应用中是变化的。并且, 结果显示, 适当增大 un_2' 的值将会对系统的安全性有大幅提升。

3) 实验三: 系统最大位数 mb_1 和 mb_2 对安全性的影响。

固定 $un=60$, $un_2'=4$, 调整系统最大位数的安全性结果如表 3、表 4 所示。第一个指标 sf 随着 mb_1 和 mb_2 差值的缩小逐渐上升, 此时最大概率的用户数量较为分散, 推测成功的概率变小; 而随着 mb_1 和 mb_2 差值的缩小, 此时攻击者猜错用户使用的系统最大位数后造成的模板数量差异会变小, 因此第二个指标 δ 明显下降。第三个指标 δE 在表 3 与表 4 中整体近似上升, 原因在于服务器中存储的模板数量增多, 面临排列组合的情况更多, 计算得到的概率矩阵会更分散。

表 2 un=60 时, 不同的 un_2' 对安全性的影响

Table 2 When un=60, the impact on security of different un_2'

评价指标	un_2'					
	1	2	4	6	8	10
sf	93.77%	93.99%	94.39%	94.74%	95.08%	95.36%
δ	7.56%	11.83%	21.35%	30.88%	41.75%	51.74%
δE	18.60%	15.79%	12.42%	12.32%	14.91%	19.26%

表 3 mb_2=6 时, 不同的 mb_1 对安全性的影响

Table 3 When mb_2=6, the impact on security of different mb_1

mb_1	1	2	3
sf	91.28%	94.39%	96.04%
δ	26.44%	21.17%	16.65%
δE	12.53%	12.34%	13.35%

表 4 mb_2=2 时, 不同的 mb_2 对安全性的影响

Table 4 When mb_2=2, the impact on security of different mb_2

mb_2	4	5	6	7
sf	94.84%	94.81%	94.39%	93.88%
δ	6.86%	11.78%	21.17%	37.15%
δE	12.03%	12.14%	12.34%	17.48%

本文的安全性保证主要来自于虹膜特征数据以及两次随机选择操作, 每个用户自身的虹膜特征数据并不具有特定规律, 选择的模板数量并非是等概率的, 而攻击者无法获知用户的模板数量的概率分布, 只能按照等概率的方式进行推测, 而两次随机选择操作进一步加大攻击者的推测难度, 因此得到的用户数量与真实值往往相差较大。通过上述实验结果发现, 保护用户数量信息的安全性与很多参数都有较大的关联, 同样需要取决于攻击者选择的推测方式, 而且攻击者无法判断是否推测成功, 因此, 本文的方案能有效保护用户数量信息。

6.2 识别精度

6.2.1 实验设置

在虹膜采集过程中可能会由于光线、角度等原因导致同一人每次采集时的结果会有一些误差, 可使用移位策略来解决这个问题。移位策略指在识别阶段, 用户通过对虹膜特征数据进行左移或右移操作得到新的虹膜特征数据, 将得到的数据生成虹膜模板后与服务器中的注册模板进行相似度计算, 取相似度分数最高的情况作为最终的相似度分数。因此, 本文与 RBS 方案一致, 同样使用移位策略。

为了更好的验证本文方案的识别性能, 将会对整个数据集的用户进行循环选择, 每次从中选择一个作为注册用户, 从该用户的所有虹膜特征数据中随机选择一条进行注册。在识别阶段, 该用户的所有虹膜特征数据生成的模板与该用户的注册模板进行比较的过程称为类内匹配, 其他用户的所有虹膜特征数据生成的模板与该用户的注册模板进行比较的过程称为类间匹配。相关实验部分见 6.2.2 节的实验一, 识别精度是重复 10 次实验取平均值的结果。

此外, 由于本文的关键是对用户数量信息的保

护, 将会补充不同合法用户数量下的识别精度实验。具体方法为: 以 CASIA-Iris-Interval 数据集为例, 假设合法用户数量为 10, 将会从 198 个用户中随机选择 10 个用户进行注册, 这 10 个用户将视为合法用户, 其余 188 个用户将视为非法用户, 服务器中存储的是这 10 个用户的注册模板, 此时的类内匹配为这 10 个用户的所有虹膜特征数据生成的模板与服务器中注册模板的匹配过程, 类间匹配为其余 188 个用户的所有虹膜特征数据生成的模板与服务器中注册模板的匹配过程。相关实验部分见 6.2.2 节的实验二至实验四, 并且考虑到注册用户数量是子集, 因此将重复实验次数增加到 30 次。

本文选取了两个评价指标衡量识别精度, 分别是: 1)错误接受率(False Acceptance Rate, FAR), 如公式(13)所示, 指在类间匹配时, 非法的虹膜特征数据被识别为合法的虹膜特征数据的比率; 2)正确接受率(Genuine Acceptance Rate, GAR), 如公式(14)所示, 指在类内匹配时, 合法的虹膜特征数据被正确识别的比率。本文的识别精度统一使用当 FAR=0.01%时的 GAR 值来进行衡量。

$$\mathrm{FAR}=\frac{\text{类间匹配识别为合法用户的次数}}{\text{类间匹配的总次数}}\times 100\% \quad (13)$$

$$\mathrm{GAR}=\frac{\text{类内匹配识别为合法用户的次数}}{\text{类内匹配的总次数}}\times 100\% \quad (14)$$

6.2.2 实验结果与分析

为了验证本文方案的识别精度, 本小节进行以下四组实验, 分别为: 1)RBS 方案与 RBS_NP 方案识别精度对比; 2)合法用户数量 un 对识别精度的影响; 3)使用 mb_2 注册的用户数量 un_2' 对识别精度的影响; 4)系统最大位数 mb_1 和 mb_2 对识别精度的影响。下列实验同样固定 $tn_{low}=2$。

1) 实验一: RBS 方案与 RBS_NP 方案识别精度对比

由于本文采用的虹膜数据集为 1536 位的虹膜串, 而 RBS 方案中采用的虹膜数据集为 10240 位的虹膜串, 因此将 RBS 方案中的参数分别调整为: 块数 $n=\{8,16,32\}$, 块大小 $b=\{192,96,48\}$, 哈希个数 $l=\{50,100,150,200\}$, 随机采样索引 $k=\{5,10,15\}$, 取模系数 $T=\{0.25,0.5,0.75,1\}$。为了提高实验效率, 从 RBS 方案中选择三组参数展开实验, 第 1 组为效率较高的参数: $b=192$, $l=50$, $k=5$, $T=0.5$, 其中 T 对效率无影响, 因此取中间值; 第 2 组为安全性较高的参数: $b=48$, $l=200$, $k=5$, $T=0.25$; 第 3 组为中间值参数: $b=96$, $l=100$, $k=10$, $T=0.5$。

固定 $mb_1=2$, $mb_2=6$, $un_2'=10$, 在这三组参数下, RBS_NP 方案与 RBS 方案的识别精度对比如表 5 所示, 观察可得, 识别精度的最大差异仅为 0.34%, 这可能是由于RBS方案中某些带有随机属性的步骤造成的精度浮动, 在可接受范围内。此外, 观察第 3 组参数的实验结果可知, RBS_NP 方案并非一定造成识别精度的下降。因此, 本文方案并不会对原始安全虹膜识别方案识别精度造成太大的影响。

表 5 FAR=0.01%时, RBS 方案与 RBS_NP 方案的 GAR 值

Table 5 The GAR (when FAR=0.01%) of RBS_NP scheme and RBS scheme

组号	GAR	
	RBS	RBS_NP
1	99.30%	99.05%
2	99.73%	99.61%
3	99.06%	99.40%

2) 实验二: 合法用户数量 un 对识别精度的影响

选取 RBS 方案中的上述的三组参数, 即效率较高的、安全性较高的以及中间值参数, 固定 $mb_1=2$, $mb_2=6$, $un_2'=4$, 调整 un。结果如表 6 所示, 观察可得, RBS_NP 方案都未对 RBS 方案的识别精度造成较大影响, 最大差异为 0.55%, 因此本文方案在任何合法用户数量下都不会对原始安全虹膜识别方案的识别精度产生较大影响。并且, un 的改变也未对实验精度产生较大影响, 最大差异为 0.67%, 因此本文方案在任何合法用户数量下都能达到良好的识别精度。

接着选取 RBS 方案的第 3 组中间值参数, 在三个数据集下调整 un, 结果如表 7 所示, 观察可得, un 的改变都未对 RBS_NP 方案的识别精度造成较大影响, 最大差异为 0.71%, 因此本文的方案在识别精度方面适用于不同的数据集。

3) 实验三: 使用 mb_2 注册的用户数量 un_2' 对识别精度的影响

选取 RBS 方案的第 3 组中间值参数, 固定 $un=60$, $mb_1=2$, $mb_2=6$, 调整 un_2', 结果如表 8 所示, 观察可得, 识别精度的最大差异为 0.43%, 说明 un_2' 同样不会对识别精度造成较大影响, 这进一步说明选用不同的时间周期参数 D 不会对识别精度产生较大影响。

表 6 FAR=0.01%时, RBS_NP 方案不同 *un* 的 GAR 值

Table 6 The GAR (when FAR=0.01%) of RBS_NP scheme using different *un*

组号		*un*					
		10	20	40	60	80	100
1	RBS	98.93%	98.40%	98.73%	98.84%	99.02%	98.80%
	RBS_NP	98.71%	98.81%	98.94%	98.82%	98.68%	98.58%
2	RBS	99.48%	99.32%	99.05%	99.47%	99.38%	99.35%
	RBS_NP	98.84%	99.19%	99.51%	99.48%	99.19%	99.39%
3	RBS	98.87%	99.26%	99.40%	99.29%	98.87%	99.04%
	RBS_NP	98.93%	98.90%	99.15%	98.74%	99.26%	99.13%

表 7 FAR=0.01%时, 不同数据集时 RBS_NP 方案不同 *un* 的 GAR 值

Table 7 The GAR (when FAR=0.01%) of RBS_NP scheme using different *un* in different dataset

数据集	*un*					
	10	20	40	60	80	100
CASIA-Iris-Interval	98.93%	98.90%	99.15%	98.74%	99.26%	99.13%
CASIA-Iris-Twins	95.04%	94.33%	94.79%	94.54%	94.90%	94.53%
CASIA-Iris-Lamp	96.13%	96.47%	96.62%	96.53%	96.49%	96.14%

表 8 FAR=0.01%时, RBS_NP 方案不同 un_2' 的 GAR 值

Table 8 The GAR (when FAR=0.01%) of RBS_NP scheme using different un_2'

un_2'	1	2	4	6	8	10
GAR	99.02%	99.17%	98.74%	99.09%	99.09%	99.08%

4) 实验四: 系统最大位数 mb_1 和 mb_2 对识别精度的影响

选取 RBS 方案的第 3 组中间值参数, 首先固定 $mb_2=6$, $un_2'=4$, 对 mb_1 进行参数调整, 观察多数用户使用的 mb_1 对识别精度产生的影响。如表 9 所示, 识别精度的最大差异仅为 0.15%, 可以得到 mb_1 不会对识别精度产生较大的影响。

表 9 FAR=0.01%时, RBS_NP 方案不同 mb_1 的 GAR 值

Table 9 The GAR (when FAR=0.01%) of RBS_NP scheme using different mb_1

mb_1	1	2	3
GAR	99.55%	99.40%	99.48%

接着固定 $mb_1=2$, $un_2'=4$, 对 mb_2 进行参数调整, 观察少数用户使用的 mb_2 对识别精度产生的影响。如表 10 所示, 识别精度的最大差异仅为 0.13%, 可以得到 mb_2 不会对识别精度产生较大的影响。

综上所述, mb_1 和 mb_2 不会对识别精度产生较大影响, 结合实验三得到的结论 un_2' 也不会对识别精度造成较大影响, 这三个参数都会影响到用户需要生成的注册模板数量, 进而说明每个用户生成多个模板的方式不会对识别精度造成较大影响。

表 10 FAR=0.01%时, RBS_NP 方案不同 mb_2 的 GAR 值

Table 10 The GAR (when FAR=0.01%) of RBS_NP scheme using different mb_2

mb_1	4	5	6	7
GAR	99.44%	99.31%	99.40%	99.39%

通过上述实验可得, 本文的方案不会对原始安全虹膜识别方案产生较大影响, 并且相关参数的调整也不会对识别精度产生较大影响, 因此本文方案在识别精度方面是符合要求的。

7 总结与展望

本文针对目前安全虹膜识别领域的相关方案都未涉及用户数量信息保护的问题, 提出了一种保护用户数量信息的安全虹膜识别方案, 可以和现有的安全虹膜识别方案相结合, 实施较为便利。通过理论分析得出本文方案能保护用户数量信息、保护新增用户数量信息、预防关联攻击以及在一定程度上增加原始安全虹膜识别方案的不可逆性, 并通过实验证实了本文方案对用户数量信息的保护是有效的,

并且本文方案不会对原始安全虹膜识别方案的识别精度造成较大影响。

下一步, 将从两个方面继续研究工作, 一方面是目前本文的工作在新增用户注册的情况下, 仅能达到不知晓系统新增了多少用户的目标, 下一步将会考虑如何满足攻击者无法知晓系统是否发生新增用户注册, 进一步提高对用户数量信息保护的程度。另一方面是探索如何将保护用户数量信息和安全虹膜识别的三大安全性要求结合的更加紧密, 进一步提出提升系统安全性的方案。

参考文献

[1] Information Technology-Security Techniques-Biometric Information Protection. ISO/IEC JTC1 SC27 IS 24745 (2022). https://www.iso.org/standard/75302.html. Accessed 12 Mar 2022.

[2] Lim S, Lee K, Byeon O, et al. Efficient iris recognition through improvement of feature vector and classifier[J]. *Electronics and Telecommunications Research Institute journal*, 2001, 23(2): 61-70.

[3] Wu Lifang, Ma Yukun, Zhou Peng, et al. A biometric template protection survey[J]. *Chinese Journal of Scientific Instrument*, 2016, 37(11): 2407-2420.
(毋立芳, 马玉琨, 周鹏, 等. 生物特征模板保护综述[J]. *仪器仪表学报*, 2016, 37(11): 2407-2420.)

[4] Wang Huiyong, Tang Shijie, Ding Yong, et al. A biometric template protection survey[J]. *Journal of Computer Research and Development*, 2020, 57(05): 1003-1021.
(王会勇, 唐士杰, 丁勇, 等. 生物特征识别模板保护综述[J]. *计算机研究与发展*, 2020, 57(05): 1003-1021.)

[5] Kumar N. Cancelable biometrics: a comprehensive survey[J]. *Artificial Intelligence Review*, 2020, 53(5): 3403-3446.

[6] Dodis Y, Reyzin L, Smith A. Fuzzy extractors: How to generate strong keys from biometrics and other noisy data[C]. *International Conference on the Theory and Applications of Cryptographic Techniques*. Springer, Berlin, Heidelberg, 2004: 523-540.

[7] Hernández Álvarez F, Hernández Encinas L, Sánchez Ávila C. Biometric fuzzy extractor scheme for iris templates[J]. *Security and Management* 2009: 563-569.

[8] Chang D, Garg S, Hasan M, et al. Cancelable multi-biometric approach using fuzzy extractor and novel bit-wise encryption[J]. *IEEE Transactions on Information Forensics and Security*, 2020, 15: 3152-3167.

[9] Juels A, Wattenberg M. A fuzzy commitment scheme[C]. *Proceedings of the 6th ACM Conference on Computer and Communications Security*. 1999: 28-36.

[10] Hao F, Anderson R, Daugman J. Combining crypto with biometrics effectively[J]. *IEEE Transactions on Computers*, 2006, 55(9): 1081-1088.

[11] Ouda O, Tsumura N, Nakaguchi T. Effective combination of iris-based cancelable biometrics and biometric cryptosystems[J]. *International Journal of Advanced Computer Science and Applications*, 2019, 10(11): 658-668.

[12] Ouda O, Tsumura N, Nakaguchi T. Tokenless cancelable biometrics scheme for protecting iris codes[C]. *2010 20th International Conference on Pattern Recognition*. IEEE, 2010: 882-885.

[13] Juels A, Sudan M. A fuzzy vault scheme[J]. *Designs, Codes and Cryptography*, 2006, 38(2): 237-257.

[14] Zhang Shumiao, Zhang Shuye, Feng Quan, et al. A polynomial representation of fuzzy vault [J]. *Computer Engineering*, 2011, 37(23): 147-148+151.
(张淑苗, 张书晔, 冯全, 等. 模糊金库的多项式表示方法[J]. *计算机工程*, 2011, 37(23): 147-148+151.)

[15] Jin A T B, Ling D N C, Goh A. Biohashing: two factor authentication featuring fingerprint data and tokenised random number[J]. *Pattern Recognition*, 2004, 37(11): 2245-2255.

[16] Zuo J, Ratha N K, Connell J H. Cancelable iris biometric[C]. *2008 19th International Conference on Pattern Recognition*. IEEE, 2008: 1-4.

[17] Asaker A A, Elsharkawy Z F, Nassar S, et al. A novel cancellable Iris template generation based on salting approach[J]. *Multimedia Tools and Applications*, 2021, 80(3): 3703-3727.

[18] Dwivedi R, Dey S, Singh R, et al. A privacy-preserving cancelable iris template generation scheme using decimal encoding and look-up table mapping[J]. *Computers & Security*, 2017, 65: 373-386.

[19] Jeong J Y, Jeong I R. Efficient cancelable iris template generation for wearable sensors[J]. *Security and Communication Networks*, 2019, 1:1-13.

[20] Rathgeb C, Breitinger F, Busch C. Alignment-free cancelable iris biometric templates based on adaptive bloom filters[C]. *2013 International Conference on Biometrics*. IEEE, 2013: 1-8.

[21] Ajish S, AnilKumar K S. Iris template protection using double bloom filter based feature transformation[J]. *Computers & Security*, 2020, 97: 101985.

[22] Pillai J K, Patel V M, Chellappa R, et al. Secure and robust iris recognition using random projections and sparse representations[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2011, 33(9): 1877-1893.

[23] Yang W, Wang S, Shahzad M, et al. A cancelable biometric authentication system based on feature-adaptive random projection[J]. *Journal of Information Security and Applications*, 2021, 58: 102704.

[24] Singh A, Vashist C, Gaurav P, et al. A generic framework for deep incremental cancelable template generation[J]. *Neurocomputing*, 2022, 467: 83-98.

[25] Zhao D, Hu X, Tian J, et al. Iris template protection based on randomized response technique and aggregated block information[C]. *2018 IEEE 29th International Symposium on Software Reliability Engineering*. IEEE, 2018: 248-258.

[26] Zhao D, Fang S, Xiang J, et al. Iris template protection based on local ranking[J]. *Security and Communication Networks*, 2018:1-9.

[27] Lai Y L, Jin Z, Teoh A B J, et al. Cancellable iris template generation based on Indexing-First-One hashing[J]. *Pattern Recognition*, 2017, 64: 105-117.

[28] Sadhya D, Raman B. Generation of cancelable iris templates via

randomized bit sampling[J]. *IEEE Transactions on Information Forensics and Security*, 2019, 14(11): 2972-2986.

[29] Kolberg J, Bauspieß P, Gomez-Barrero M, et al. Template protection based on homomorphic encryption: Computationally efficient application to iris-biometric verification and identification[C]. *2019 IEEE International Workshop on Information Forensics and Security*. IEEE, 2019: 1-6.

[30] Morampudi M K, Prasad M V N K, Raju U S N. Privacy-preserving iris authentication using fully homomorphic encryption[J]. *Multimedia Tools and Applications*, 2020, 79(27): 19215-19237.

[31] Blanton M, Aliasgari M. Secure outsourced computation of iris matching[J]. *Journal of Computer Security*, 2012, 20(2-3): 259-305.

[32] Bauspieß P, Kolberg J, Demmler D, et al. Post-Quantum Secure Two-Party Computation for Iris Biometric Template Protection[C]. *2020 IEEE International Workshop on Information Forensics and Security*. IEEE, 2020: 1-6.

[33] Zhao D, Zhou X, Xiang J, et al. NDBIris with better unlinkability[C]. *2020 IEEE Symposium Series on Computational Intelligence*. IEEE, 2020: 2948-2956.

[34] CASIA-IrisV3, http://biometrics.idealtest.org/. Accessed 15 Mar 2022.

[35] Othman N, Dorizzi B, Garcia-Salicetti S. OSIRIS: An open source iris recognition software[J]. *Pattern Recognition Letters*, 2016, 82: 124-131.

周宇 于 2020 年在安徽工程大学计算机科学与技术专业获得学士学位。现在武汉理工大学软件工程专业攻读硕士学位。研究领域为生物信息安全。研究方向包括: 隐私保护和安全虹膜识别。Email: zhouyu98@whut.edu.cn

向剑文 于 2004 年在武汉大学计算机软件与理论专业获得博士学位, 于2005年在日本北陆先端科学技术大学院大学情报科学专业获得博士学位。现在武汉理工大学计算机与人工智能学院副院长。研究领域为可靠性工程。研究方向包括: 网络安全、软件老化。Email: jwxiang@whut.edu.cn

郑倩荣 于 2020 年在中北大学软件工程专业获得学士学位。现在武汉理工大学计算机科学与技术专业攻读博士学位。研究领域为生物信息安全。研究方向包括: 隐私保护和安全虹膜识别。Email: qianrongzheng@whut.edu.cn

赵冬冬 于 2016 年在中国科学技术大学计算机应用技术专业获得博士学位, 现在武汉理工大学计算机与人工智能学院副教授。研究领域为隐私保护, 研究方向包括: 安全虹膜识别, 信息负表示。Email: zdd@whut.edu.cn