·安全技术·　　文章编号：1000-3428(2016)10-0164-05　　文献标志码：A　　中图分类号：TP309

基于在线手写签名的密钥生成方法

何　梦[1,2]，吴仲城[2]，李　芳[2]

（1. 中国科学院合肥物质科学研究院，合肥 230031；2. 中国科学院强磁场科学中心，合肥 230031）

摘　要：为实现生物密钥特征可靠性编码，基于在线手写签名，提出一种在线签名生成密钥的方法，即基于等概率质量量化的动态位分配方法。该方法采用 Fisher Ratio 准则进行用户依赖的特征选择，选取最能表征用户的特征，运用基于似然比的静态位分配方法中等概率量化思想，根据特征区间控制系数确定注册样本特征量化概率质量，以提高注册阶段的时间性能。考虑到计算量化概率质量时定积分运算的时间成本，提出以曲线段所接梯形面积来替代定积分的优化策略。在 SVC2004 签名数据库上进行认证性能验证实验及时间复杂度对比实验，结果表明，该方法获得的错误接受率为 2.54%，错误拒绝率为 28.63%，梯形改进的量化方法签名注册时间为 9 s，约为原高斯积分量化方法的 1/10。

关键词：生物密码；在线手写签名；身份认证；用户依赖；梯形改进

中文引用格式：何　梦，吴仲城，李　芳. 基于在线手写签名的密钥生成方法[J]. 计算机工程，2016，42(10)：164-168.

英文引用格式：He Meng，Wu Zhongcheng，Li Fang. Key Generation Method Based on On-line Handwritten Signature[J]. Computer Engineering，2016，42(10)：164-168.

Key Generation Method Based on On-line Handwritten Signature

HE Meng[1,2]，WU Zhongcheng[2]，LI Fang[2]

（1. Heifei Institutes of Physical Science，Chinese Academy of Sciences，Heifei 230031，China；
2. High Magnetic Field Laboratory，Chinese Academy of Science，Heifei 230031，China）

【Abstract】 In order to implement biological keys reliablility coding，on the basis of online hand-writing signature，this paper proposes a key generation method using on-line handwritten signature，which is based on the method of equal probability mass quantization. It utilizes the user-dependent fisher ratio criterion to get the user characteristics，which are the most representative of the user，and applies the thought of equal probability mass in the static bit allocation method based on likelihood ratio，but the quantitative probability mass of the registered sample is determined by the characteristic interval control coefficient，which improves the time performance of the registration phase. Factoring in time like definite integral in the computation of the quantitative probability mass，this paper raises an optimization strategy which replaces the definite integral with the trapezoidal area of curved section. The identity authentication experiment and the time complexity contrast experiment are carried out on the SVC2004. Results show that the method gains False Acceptance Rate(FAR) of 2.54% and False Rejection Rate(FRR) of 28.63%. The registration time of the trapezoidal improvement is 9 s，which is about 1/10 of the original Gauss integral method.

【Key words】 biological cryptography；on-line handwritten signature；identity authentication；user dependence；trapezoidal improvement

DOI：10.3969/j.issn.1000-3428.2016.10.029

1　概述

随着计算机网络和计算机通信技术的发展，计算机密码学得到前所未有的重视并迅速普及。在计算机环境中，用户的身份主要是通过“what you know”(静态口令)或者“what you have”(智能卡、动态口令、usb key 等)的方法进行认证。上述手段并不能保证用户身份的真实性，且极易丢失或被盗用。

基金项目：国家自然科学基金资助项目(61273323)。

作者简介：何　梦(1986—)，女，硕士研究生，主研方向为自然交互与计算、生物密码学；吴仲城，研究员、博士生导师；李　芳，博士研究生。

收稿日期：2015-10-20　　**修回日期：**2015-12-04　**E-mail：**993031732@qq.com

为解决数字身份与用户物理身份不统一的问题,研究者利用“who you are”(指纹、虹膜、人脸、手写签名等)方法开展了基于生物特征的认证研究。由于生物特征具有可随身携带、无需记忆、不会丢失、重复概率小、难以复制等诸多特点,基于生物特征的认证近年来逐渐成为密码学的主流发展方向,有望为数据安全保障提供一种更加自然和可靠的解决方案。

生物特征与密码学相结合即是生物密码,可以分为如下 3 类:密钥释放,密钥绑定,密钥生成[1-2]。密钥释放和密钥绑定方案都是采用生物特征和外部输入密钥相结合的方式,不合理的结合方法可能导致认证过程中外部输入密钥占主导地位,系统安全性仍然依赖于密钥本身而非生物特征[2]。密钥生成方案是基于生物特征离散化生成唯一密钥,冒名顶替者难以生成相同的生物密码。本文针对第 3 种方案进行研究,计算机密码学中密钥必须是确定的,生物特征则是由一些连续的数据组成,所以,二进制的直接表示必须依赖于有效的离散化方案。

生物离散化方案包括量化、编码 2 个阶段。静态位分配法是一种早期的离散化方法,它将每维特征映射到一段固定长度的码串,并将各维特征码串相级联即得到可以用于身份认证的生物密钥。文献[3]从击键特征中提取基于特定的阈值,将一维特征空间量化为 2 个区间,区间分别用 0 或 1 表示,用户的每维特征落在哪个区间就被映射为相应的二进制标志,尽管全局的量化方案可以避免特定用户信息的泄露,认证结果并不令人满意;文献[4-5]针对人脸、指纹特征提出了一种基于似然比的静态多比特离散化方案,根据概率质量确定似然比函数间隔,进而对左右区间进行分割,并将原始特征映射为二进制格雷码;文献[6]根据虹膜的第三层小波系数的正负提出单比特特征码;文献[7]针对指纹方向场提出了基于等度数间隔的量化方法。然而每维特征可区分性不同,即所含信息量不同,采用固定长度的码串表示在一定程度上模糊了有用信息。文献[8-9]对文献[10]的等间隔量化方法做了改进,生成特征码串长度和其本身特征分布有关,特征可分性越大,所分配码串长度越长,然而从攻击者角度来说,一旦用户边界和全局特征边界已知,则信息即已泄露;文献[11]根据人脸特征可靠性权重依次分配码串长度,调整权重阈值直到满足码串长度,可靠的特征具有分配更多码串的优势,然而近似全局的量化方案使得认证性能并不十分理想。

手写签名数据的采集具有非侵入性的特点,并且不存在诸如采集指纹、人脸、虹膜时带来的用户感情排斥或侵犯隐私的问题。从运动生物力学和神经生理学角度,文献[12]认为手写签名是一种快速而熟练的“弹道运动”(ballistic motion),反映了一个人长期形成的书写习惯,手写签名的个体差异性特点使其能够被用于身份鉴别;文献[13]分别从特征偏差、特征熵、特征稳定性与熵的相关性对手写签名特征进行了评价。

为实现特征可靠性编码,基于在线手写签名,本文提出一种依赖用户的动态位分配方法。在该方法中,不同用户相同特征分布可能不同,不同特征分布为真正的动态位分配提供了保障,且能够生成稳定且较长的码串。此外,等概率质量量化方法使得在特征边界泄露的情况下,攻击者仍不能完全破解特征码串。

2　基于 Fisher Ratio 的特征选择

生物特征提取过程中有许多不可分特征,并且每个用户的不可分特征不可能完全相同,本文提出基于用户依赖的特征选择方法进行特征选择。Fisher Ratio 是一种常用的特征选择方法,其理论基础是可分性较强的特征具有较小的类内距离和较大的类间距离。采用 Fisher Ratio 作为准则,根据得分进行特征排序,并从 D 维特征中选取可分性较强的前 n 维特征,从而达到获得稳定的生物密钥的目的。假设数据集中共有 I 个用户,每个用户有 J 个样本作为注册样本,D 维特征空间内用户 i 依赖的 Fisher Ratio 表示为:

$$r_i = [\, r_i^1 r_i^2 \cdots r_i^d \cdots r_i^D \,] = \left[\frac{s_b^1}{s_w^1} \ \frac{s_b^2}{s_w^2} \ \cdots \ \frac{s_b^d}{s_w^d} \ \cdots \ \frac{s_b^D}{s_w^D} \right] \quad (1)$$

其中,s_b^d,s_w^d 分别表示用户 i 的第 d 维特征在注册样本集上的类间距离、类内距离。某维特征的得分越高说明其具有较好的类别可分性。对 r_i 进行降序排序,选择得分高的前 n 维特征作为特征子集,以达到去除噪声特征及弱可分性特征的目的。

采用用户依赖的 Fisher Ratio 准则对 I 个用户分别做特征选择,可以获得 I 个 n 维特征向量,每个特征向量可能不同,为了身份认证时能够正确地选择相应的特征,需要将注册过程中选择的特征下标作为辅助数据存储。

3　基于等概率质量的动态位分配方法

当在线手写签名的注册样本足够多时,根据中心极限定理,可以认为特征分布服从高斯分布。本文采用注册样本特征区间控制系数来确定注册样本特征区间,具有不同统计参数的生物特征为动态位分配提供了保障。

3.1　特征量化

首先给出 2 个概念:背景概率密度函数和原始

概率密度函数。背景概率密度函数是指一维特征所有值的分布曲线,原始概率密度函数是指用户的该维特征所有值的分布曲线。不同于文献[3]的单比特量化,本文提出的依赖于注册用户的量化可以贡献更多的比特位,并且每维特征贡献比特位数可能不相等。在一个一维特征空间中,量化概率质量由注册用户特征的分布区间决定,如式(2)所示。

$$P_w(k_w)=\int_{u_w-k_w\sigma_w}^{u_w+k_w\sigma_w}G(v,u_0,\sigma_0)\mathrm{d}v \tag{2}$$

其中,u_w,σ_w 分别为单用户注册样本的期望、方差;u_0,σ_0 分别为所有用户注册样本的期望、方差;k_w 为特征区间控制系数;v 为特征;$G(v,u_0,\sigma_0)$为背景概率密度函数。

图1是一维特征空间中特征量化的示意图。根据等概率质量量化的原则分别对注册用户特征区间的左侧区域和右侧区域进行分段,左侧区域最多有 $L=\left\lfloor\frac{\int_{-\infty}^{u_w-k_w\sigma_w}G(v,u_0,\sigma_0)\mathrm{d}v}{P_w(k_w)}\right\rfloor$ 段区域具有与注册用户特征区间相等的概率质量,右侧区域最多有 $R=\left\lfloor\frac{\int_{u_w+k_w\sigma_w}^{+\infty}G(v,u_0,\sigma_0)\mathrm{d}v}{P_w(k_w)}\right\rfloor$ 段。除了上面被分段的区域外,还存在2个冗余的区域:最左边区域和最右边区域,背景概率函数在它们区域内的积分小于 $P_w(k_w)$,将这2个冗余区域记为1个环形区域。所以,该维特征的背景概率分布可以被分为 $S=L+R+2$ 段。根据特征区间控制系数 k_w 进行量化,每维特征的量化结果依赖于该维特征的原始概率密度函数在背景概率函数中的分布情况,所以不同用户同维度特征的分段结果 S 可能各不相同,表现为特征的动态位分配。

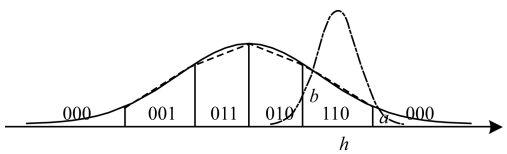

图1 基于等概率质量的一维特征量化

在图1中,实线为背景概率密度函数;点画线为注册用户原始概率密度函数 $G(v,u_w,\sigma_w)$;a,b,h 分别为对应线段长度,$h=2k_w\sigma_w$。

上述基于等概率质量的量化虽然能够较好地抵御攻击者从猜测离散化输出中取得优势,但在注册时的时间性能上完全输于基于等间隔的量化。这主要是由于概率质量的计算需要对背景概率密度函数进行积分,而曲线积分运算是一个耗时运算。曲线积分即是求曲线下方的面积,在量化分段时运用背景概率密度函数分段曲线下方的梯形面积来近似求解曲线积分,以达到降低运算时间、提高运算效率的目的,如图1所示。简化后的量化概率质量为:

$$Pw(k_w)=\frac{(a+b)h}{2} \tag{3}$$

3.2 特征编码

背景概率密度函数被分割的段数决定了特征编码的位数,则当函数被分割为 S 段时,至少需要 $\lceil\mathrm{lb}(S)\rceil$ 位才能满足编码需求,采用 $\lceil\mathrm{lb}(S)\rceil$ 位对生物特征进行编码。直接进行二进制编码容易引起海明域中较大的类内间距,而格雷码作为可靠性编码,是一种错误最小化的编码方式。在海明域中,格雷码的单步特性能够减小由于类内变化所引起的量化误差,从而降低错误拒绝率(False Rejection Rate, FRR),本文采用格雷码实现特征编码。根据 Fisher Ratio 准则选取的 n 维特征通过量化编码后,n 个子码串顺序连接成一个码串,即为生物密钥。

特征量化区间及生物密钥均须作为辅助数据存储,在用户身份认证时,通过判断生物特征在哪个区间内进而实现生成用户认证密钥,并将其与存储的密钥模板进行匹配,如果两者相匹配则为真实用户,否则为伪造用户。

3.3 算法流程

从签名数据中共提取 D 维特征,根据用户依赖的 Fisher Ratio 准则选择前 n 维特征作为位分配方法的输入,在用户注册阶段,基于概率的动态位分配方法具体算法过程如下:

步骤1 计算数据库所有用户注册样本的 D 维特征的期望、方差。

步骤2 依据用户 i 的注册样本计算其第 d 维特征的期望、方差,并根据统计特征构建原始概率密度函数。

步骤3 从步骤1中读取相应的第 d 维特征的期望、方差,构建背景概率密度函数。

步骤4 利用特征区间控制系数 k_w 确定原始概率密度函数所占背景概率密度函数的宽度,并根据该宽度计算量化概率质量 $P_w(k_w)$。

步骤5 依据 $P_w(k_w)$ 将背景概率密度函数量化分段,并根据量化分段结果对用户特征进行格雷编码,即为用户 i 第 d 维特征的码串。

步骤6 将用户 i 的 n 维特征均执行步骤2~步骤5,n 个特征码串顺次级联即为用户 i 的在线签名生物密钥。

4 实验结果与性能分析

为了验证基于等概率质量动态位分配方法的有效性,本文采用 SVC2004 签名数据库公开部分作为

实验数据集。公开部分包含英文签名和中文签名,共 40 个用户的签名数据,每个用户包含 20 个真实签名、20 个至少由 4 个其他用户提供的熟练伪造签名,20 个真实签名分 2 次采集完成,每次采集 10 个,2 次采集至少相隔一周[14]。

由于在线手写签名是生物行为特征,它在稳定性方面不如指纹、虹膜等生物生理特征。为减弱在线手写签名在稳定性方面的劣势,实验方案的设计是:从 2 次采集的真实签名中分别随机选取 5 个作为注册样本;将剩余的 10 个真实样本、20 个熟练伪造样本作为认证阶段的测试样本。根据上述描述重复 10 次,取平均值作为该系统的认证性能,真实样本和伪造样本如图 2 所示。

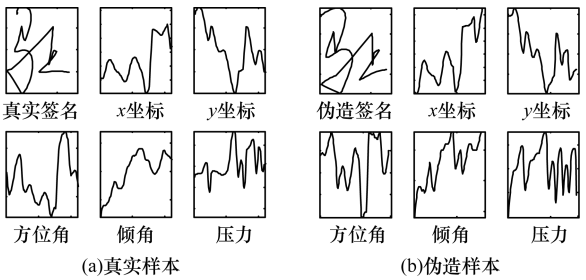

图 2　真实样本与伪造样本

手写签名认证的性能评价主要基于 FAR 与 FRR 指标。特征区间控制系数 k_w 的值需要进行折中选择:当 k_w 较大时,注册用户特征区间也较大,认证时容易将伪造签名认证为真实签名,即获得较大的 FAR,反之,当 k_w 较小时,注册用户特征区间也较小,认证时容易将真实签名认证为伪造签名,即获得较大的 FRR。

在给定 FAR 的约束下,通过求解最小 FRR 的最优化问题,可以使用户的不同维度特征获得不同的 k_w。为了简化问题,本文中注册用户的每维特征采用相同的 k_w,在不引入码串模糊参数的前提下,用户手写签名认证错误率 FRR 与 FAR 随 k_w($k_w \in [1,7]$)变化的曲线如图 3 所示。

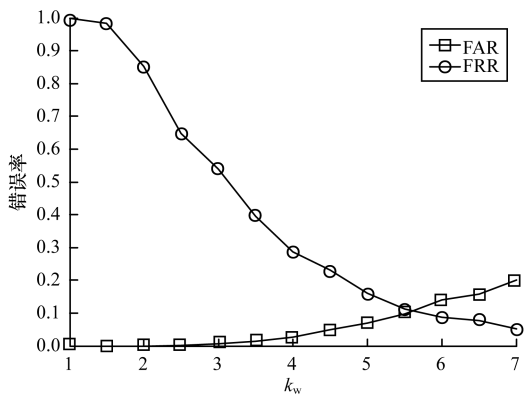

图 3　随 k_w 变化的错误率曲线

根据曲线可知 ERR 为 10%,文献[15]给出的认证结果中 ERR 为 10.90%,SVC2004 签名数据库公布的最好测试结果 ERR 为 6.90%,对比可知,虽然本文提出的方法不能获得最优的认证性能,但是测试结果也验证了基于在线手写签名生成密钥方法的有效性。FAR 表示了从伪造者可以生成合法密钥的比率,它越低,用户就越能够确信自身身份的安全性,故而选择 $k_w = 4$,此时 FRR 为 28.63%,FAR 为 2.54%,若选取码串长度模糊比为 99%,FRR 为 15.99%,FAR 为 6.28%。

表 1 为基于不同生物特征密钥生成方法的认证结果。从表中可以看出,基于行为特征的身份认证与基于生理特征的身份认证结果相比,普遍具有较差的认证性能,这也反映了行为特征具有比生理特征更大的类内距离,即行为特征稳定性要弱于生理特征的特点,然而,虽然获得的 FRR 相对生理特征的 FRR 要高出很多,但是用户在签名时平均只需要尝试 1.4 次即可克服因类内距离所引起的认证错误,这对大多数用户来说是可以接受的。随机伪造的数量严重影响系统的 FAR,然而本文提出的方法在不引入随机伪造时仍具有比早期基于行为特征的密钥生成方法较低的 FAR。总的来说,本文提出的方法具有较好的认证性能。

表 1　基于生物特征的密钥生成结果

生物特征	FRR/%	FAR/%	码长
声音[3]	–	20.00	46
签名[8]	23.50	5.90	190 ~ 360
指纹[4]	4.30	1.00	100
指纹[6]	–	–	128
人脸[11]	11.50	11.50	128
本文签名	28.63	2.54	123

猜测熵是衡量系统安全性的重要概念,它描述了攻击者获得可信密钥做出的猜测的期望数[3,8,16],当密钥长度为 S bit 时,穷举猜测次数的期望大约为 2^S,S 与猜测熵成指数关系,S 越大,系统越安全。本文在不引入随机变量的前提下获得了较长的密钥,密钥平均长度为 123 bit,足以抵抗穷举搜索攻击。此外,等概率质量的量化能够较好地抵御攻击者从猜测离散化输出中获得密钥信息。

为了验证改进方法在降低计算复杂度时的有效性,在相同的实验条件下,分别对 2 种等概率质量量化和等间隔量化进行实验,表 2 中记录了 $k_w = 4$ 用户签名注册的时间期望,可以看出,梯形改进的量化方法签名注册时间为 9 s,约为原高斯积分量化方法的 1/10。

表 2 签名注册时间期望 s

硬件环境	等概率质量量化		等间隔量化
	高斯积分	梯形改进	
Pentium Dual-Core CPU E5200(2.5 GHz) 2.00 GB RAM	91.5	9	2.7

5 结束语

本文根据用户手写签名具有代表用户身份的特性及可分特征存在差异性的特点,结合基于等概率质量量化的思想,提出一种新的在线签名认证方法。该方法采用 Fisher Ratio 准则进行用户依赖的特征选择,应用基于似然比的静态位分配方法中等概率量化思想,由特征区间控制系数确定注册样本特征量化概率质量,提高注册阶段的时间性能。此外提出的梯形改进方法在时间性能上也有较大的提升。本文下一步将从下面 3 个方面开展研究:(1)本文采用时域特征作为特征值,考虑增加频域特征提高认证性能;(2)不同用户经过动态位分配方法后获得的密钥长度可能各不相同,可以考虑通过存储平均密钥长度,增加系统的安全性;(3)对基于在线手写签名生成的生物密钥进行随机性检验,以进一步研究密钥的安全性。

参考文献

[1] Nandakumar K,Jain A K,Pankanti S. Fingerprint-based Fuzzy Vault:Implementation and Performance[J]. IEEE Transactions on Information Forensics and Security,2007,2(4):744-757.

[2] 游 林. 生物特征密码技术综述[J]. 杭州电子科技大学学报:自然科学版,2015,35(3):1-16.

[3] Monrose F,Reiter M K,Li Q,et al. Cryptographic Key Generation from Voice[C]//Proceedings of IEEE Symposium Security and Privacy. Sacramento,USA:IEEE Press,2001:202-213.

[4] Chen C,Veldhuis R N J,Kevenaar T A M,et al. Multi-bits Biometric String Generation Based on the Likelihood Ratio[C]//Proceedings of IEEE International Conference on Biometrics:Theory,Applications,and System. Washington D. C.,USA:IEEE Press,2007:1-6.

[5] Chen C,Veldhuis R N J,Kevenaar T A M,et al. Biometric Quantization Through Detection Rate Optimized Bit Allocation[J]. EURASIP Journal on Advances in Signal Processing,2009,3(1):1-16.

[6] 周 俊,曹 琦,王 帅. 基于虹膜特征的密钥生成研究[J]. 计算机工程与应用,2012,48(21):31-34.

[7] Sheng Weiguo,Howells G,Fairhurst M,et al. Reliable and Secure Encryption Key Generation from Fingerprints[J]. Information Management & Computer Security,2012,20(3):207-221.

[8] 常 郝. 基于动态手写签名的生物特征密钥生成研究[D]. 合肥:合肥工业大学,2007.

[9] 王培沛. 基于掌纹特征的密钥生成算法[D]. 哈尔滨:哈尔滨工业大学,2009.

[10] Chang Y J,Zhang W,Chen T. Biometric-based Cryptographic Key Generation[C]//Proceedings of IEEE International Conference on Multimedia and Expo. Copenhagen,Denmark:IEEE Press,2004:2203-2206.

[11] Lim M H,Teoh A B J,Toh K A. An Efficient Dynamic Reliability-dependent Bit Allocation for Biometric Discretization[J]. Pattern Recognition,2011,45(5):1960-1971.

[12] Denier J J,Gon V,Thuring J. The Guiding of Human Writing Movements[J]. Kybemetik,1965,4(3):145-148.

[13] Vielhauer C,Steinmetz R. Handwriting:Feature Correlation Analysis for Biometric Hashes[J]. EURASIP Journal on Applied Signal Processing,2004,8(4):542-558.

[14] Yeung D Y,Chang H,Xiong Y,et al. SVC2004:First International Signature Verification Competition[C]//Proceedings of Conference on Lecture Notes in Computer Science. Berlin,Germany:Springer,2004,5(3):16-22.

[15] 邱益鸣,胡华成,郑建彬,等. 基于曲线相似的在线签名认证方法[J]. 系统工程与电子技术,2014,36(5):1016-1020.

[16] 严霄凤. 基于熵的密码强度估计[J]. 网络安全技术与应用,2012,12(11):36-38.

编辑 索书志