

基于遗传算法和灰色关联分析的击键特征识别算法

王 暄^{1,2}, 陈伟伟¹, 马建峰²

(1. 陕西师范大学 物理学与信息技术学院, 陕西 西安 710062;

2. 西安电子科技大学 计算机网络与信息安全教育部重点实验室, 陕西 西安 710071)

(wxuan@snnu.edu.cn)

摘 要:基于用户击键特征的身份认证比传统的基于口令的身份认证方法有更高的安全性,现有研究方法中基于神经网络、数据挖掘等算法计算复杂度高,而基于特征向量、贝叶斯统计模型等算法识别精度较低。为了在提高识别精度的同时有效降低计算复杂度,在研究现有算法的基础上提出了一种基于遗传算法与灰色关联分析的击键特征识别算法。该算法利用遗传算法根据用户训练样本确定表征用户击键特征的标准特征序列,通过对当前用户击键特征序列与标准特征序列进行灰色关联分析实现用户身份认证。实验结果表明,该算法识别精度达到神经网络、支持向量机等算法的较高水平,错误拒绝率与错误接受率分别为 0% 与 1.5%。且计算复杂度低,与基于特征向量的算法相近。

关键词:用户身份认证;击键特征;遗传算法;灰色关联分析

中图分类号: TP391.4 **文献标识码:** A

User authentication algorithm with keystroke features based on genetic algorithms and grey relational analysis

WANG Xuan^{1,2}, CHEN Wei-wei¹, MA Jian-feng²

(1. School of Physics and Information Technology, Shaanxi Normal University, Xi'an Shaanxi 710062, China;

2. Key Laboratory of Ministry of Education for Computer Networks and Information Security, Xidian University, Xi'an Shaanxi 710071, China)

Abstract: User authentication based on keystroke dynamics features is more secure than conventional user authentication approach only based on passwords. The neural network and data mining-based methods present high authentication accuracy, but have a high computational cost. The statistical and vector-based methods have shown low computational complexity, but are less accurate in user authentication. In order to improve authentication accuracy and reduce computational complexity synchronously, a new user authentication approach based on keystroke patterns was proposed. In the proposed approach, Genetic algorithm was employed to generate the common keystroke pattern of each user from the training set consisting of the user's normal keystroke samples. Then Grey Relational analysis method was applied to calculate the degree of grey slope incidence between common keystroke pattern and current keystroke pattern, the resultant value was compared with a threshold value determined by experiment to implement user authentication. Experimental results show this approach represents the same user authentication accuracy as neural network and data mining-based methods in terms of False Acceptance Rate (FAR) and False Rejection Rate (FRR), false acceptance rate and false rejection rate of this method are 1.5% and 0% respectively. It is also shows that the computational complexity of the proposed method is lower than that of some other methods.

Key words: user authentication; keystroke feature; genetic algorithm; grey relational analysis

0 引言

采用生物特征识别的身份认证是根据用户独特的生理特征(如指纹、虹膜、语音、击键特征等)来认证用户的合法性。与传统的身份认证相比,它具有更好的安全性。在这些生理特征中,击键特征不需要增加新的硬件设备,在用户输入口令时,由系统自动提取其特征,对口令和击键特征同时进行认证,比其他的生物特征使用更加方便。

击键特征用于用户身份认证最先由 Gaines 等人^[1]于 20 世纪 80 年代提出,此后出现了许多这方面的研究成果。Leggett^[2]和 Bleha^[3]等人用统计分析理论,在假设样本为正态分布的前提下,将测试样本与训练样本进行比较。该方法识别精度低,需要一定的训练样本而且目前尚没有实验数据和

理论依据证明击键特征符合正态分布。Ru 等人^[4]采用了模糊数学的方法,将击键速度分为快、中、慢 3 个等级后,进行模糊处理并判断,随后 Tapiador 等人^[5]也尝试了模糊算法,但该类方法特征选取不充分,分类不精确,Brown 和 Rogers^[6]和 Cho 等人^[7]用神经网络的方法,Enzhe Yu^[8]将遗传算法和支持向量机理论结合,获得较好的效果,但所需训练样本空间大、训练时间长。Aykut Guven^[9]等人提出一种基于特征向量的快速识别算法。他只取该用户上一次成功登录的击键特征向量作为标准向量与测试样本进行比较来确定用户身份,这种简单的处理方法虽然有效降低了计算量,识别速度快,但也同时降低了系统的鲁棒性,导致较高的误报率。

本文基于遗传算法和灰色关联分析理论,提出一种计算复杂度低而识别精确度较高的击键特征用户身份认证算法,

收稿日期:2006-11-29;修订日期:2007-01-16 基金项目:国家 863 计划资助项目(2002AA143021)

作者简介:王暄(1966-),男,陕西定边人,副教授,博士研究生,主要研究方向:图像处理、网络安全; 陈伟伟(1981-),女,山西大同人,硕士研究生,主要研究方向:网络安全; 马建峰(1963-),男,陕西临潼人,教授,博士生导师,主要研究方向:信息安全、图像处理、密码学。

该算法采用全局搜索性强的遗传算法,基于训练样本,得到最佳描述用户击键特征的标准特征序列,然后通过计算当前用户的击键特征序列与合法用户的标准击键特征序列之间的灰色斜率关联度来认证用户的合法身份。实验结果表明,该算法识别精度可以达到神经网络、支持向量机等算法的较高水平,其正常用户的平均拒绝率(FRR)为0%,非法用户的平均通过率(FAR)为1.5%,而且计算量及识别速度与 Aykut Guven^[9]等人的算法相近。

1 遗传算法与灰色关联分析

1.1 遗传算法

遗传算法是一种模拟生物进化过程的随机方法,它的思想来源于生物遗传学和优胜劣汰的生存竞争规则。它将问题的可能解集通过基因编码构成初始种群(父代)开始迭代,计算父代个体的适应度,满足优化准则,跳出迭代;否则,通过选择、交叉或基因重组、变异等操作形成新的个体(子代),计算子代个体的适应度,子代插入到父代中,重新开始循环,直至满足优化准则,使种群的进化到包含近似最优解的状态。其流程图如图1所示。

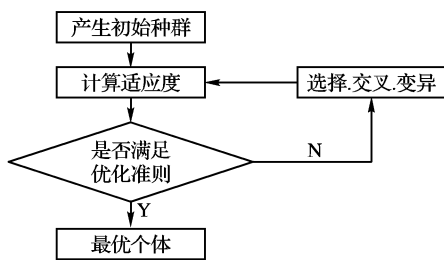


图1 遗传算法的流程

1.2 灰色关联分析

灰色系统理论由我国学者邓聚龙教授于1982年提出,用于处理信息不完备、不确定,数据较少的系统。灰色系统理论与概率和数理统计的方法相比,需要的样本空间小,数据不一定服从典型的分布,计算量小,不会出现量化结果和定性分析结果不符的现象;与模糊数学相比,不依赖经验,强调信息优化,研究现实规律^[10]。

距离空间只限于点点的距离比较,其距离作为比较的测度,点集拓扑是整体的比较,是领域的比较,但没有测度。距离空间与点集拓扑空间相结合,就构成了灰色关联分析空间^[11]。它是一种有参考系的、有测度的整体比较。灰色关联分析是根据序列曲线几何形状的相似程度来判断其联系是否紧密。曲线相似程度越大,相对应的序列联系越紧密,关联度越大,反之,关联度就小。灰色关联分析反映出灰色系统各因素之间发展趋势的相似或相异的程度。

灰色斜率关联度根据序列曲线的斜率的接近程度来判定曲线的接近程度,曲线斜率相差越小,序列的关联度就越大。灰色斜率关联度^[8]的相关定义如下:

设 $X(t), t = 1, \dots, n$ 为参考序列, $Y(t), t = 1, \dots, n$ 为比较序列,称

$$\xi(t) = \frac{1 + \left| \frac{1}{\bar{x}} \cdot \frac{\Delta x(t)}{\Delta t} \right|}{1 + \left| \frac{1}{\bar{x}} \cdot \frac{\Delta x(t)}{\Delta t} \right| + \left| \frac{1}{\bar{x}} \cdot \frac{\Delta x(t)}{\Delta t} - \frac{1}{\bar{y}} \cdot \frac{\Delta y(t)}{\Delta t} \right|} \quad (1)$$

为 $X(t)$ 与 $Y(t)$ 在 t 时刻的灰色斜率关联系数。

其中:

$$\bar{x} = \frac{1}{n} \sum_{t=1}^n x(t), \Delta x(t) = x(t + \Delta t) - x(t);$$

$$\frac{\Delta x(t)}{\Delta t} \text{ 为参考序列 } X(t) \text{ 在 } t \text{ 到 } t + \Delta t \text{ 的斜率};$$

$$\bar{y} = \frac{1}{n} \sum_{t=1}^n y(t), \Delta y(t) = y(t + \Delta t) - y(t);$$

$$\frac{\Delta y(t)}{\Delta t} \text{ 为比较序列 } Y(t) \text{ 在 } t \text{ 到 } t + \Delta t \text{ 的斜率};$$

当 $X(t), Y(t)$ 皆为1-时距的离散序列时, $X(t)$ 与 $Y(t)$ 在 t 到 $t + \Delta t$ 的灰色斜率关联系数公式简写为:

$$\xi(t) = \frac{1 + \left| \frac{\Delta x(t)}{\bar{x}} \right|}{1 + \left| \frac{\Delta x(t)}{\bar{x}} \right| + \left| \frac{\Delta x(t)}{\bar{x}} - \frac{\Delta y(t)}{\bar{y}} \right|} \quad (2)$$

$X(t)$ 与 $Y(t)$ 的灰色斜率关联度 ε 定义为:

$$\varepsilon = \frac{1}{n-1} \sum_{t=1}^{n-1} \xi(t) \quad (3)$$

2 本文算法

2.1 击键特征模型

击键特征是指人在敲击键盘时的力量和速度特性,不同的人,敲击键盘的节奏模式就不相同,属于人的生物特征之一。击键特征主要体现在:击键延迟时间(keystroke duration time)和击键间隔时间(keystroke latency time), $D(i)$ 表示第 i 键的延迟时间,指第 i 键按下与第 i 键弹起之间的时间, $L(i)$ 表示第 i 键与第 $i+1$ 键的间隔时间,指第 i 键弹起到第 $i+1$ 键按下之间的时间。

击键特征序列由 $D(i)$ 和 $L(i)$ 共同构成。对于 m 位长的密码,击键延迟时间和击键间隔时间分别为: $D(i), i = 1, 2, \dots, m$ 和 $L(i), i = 1, 2, \dots, m-1$ 。特征序列长为 $2m-1$, 其表示如下:

$$T = D(1), L(1), D(2), L(2), \dots, D(m-1), L(m-1), D(m)$$

2.2 遗传算法产生标准序列

本文的算法是通过计算合法用户的标准序列和登录用户的测试序列的灰色斜率关联度,来判断用户的合法性,其中标准序列能否真正的反映合法用户的击键特征成为一个重要的制约因素。由于遗传算法的适应性强,能搜索到全局最优解的良好搜索能力,我们对于训练样本采用遗传算法来搜索,寻找最佳的标准序列 T^* , 该标准序列与样本空间的所有序列的灰色斜率关联度都很大。

对于新用户,采集用户键入密码稳定时的特征序列 T_i ($i = 1, 2, \dots, S$) 构成训练样本,将训练样本 $T_i (i = 1, 2, \dots, S)$ 作为初始种群,采用二进制的编码方式。个体 i 的代价函数 Ψ_i 的定义:

$$\Psi_i = 1 - \sum_{j=1}^s \varepsilon_{ij} \quad (4)$$

其中 ε_{ij} 表示 i 个体与训练样本空间中的训练序列 j 的灰色斜率关联度。

Ψ_i 越小,反映出个体 i 与样本空间的序列越相似,越能代表用户的击键特征。种群的代价函数 $\Psi = \{\Psi_1, \Psi_2, \dots, \Psi_s\}$, 适应度计算我们采用了 rank-scale^[12] 的方法,对种群的代价函数值进行升序排列,排在第 j 位的个体适应度为:

$$\phi_j = \frac{S}{\sum_{i=1}^s \frac{1}{\sqrt{i}}} \quad (5)$$

代价函数 ϕ_j 越小,其适应度越大。

我们采用轮盘赌和跨世代精英的方法来进行选择。根据适应度进行轮盘赌的选择机制,被选中的个体进行单点交叉和变异的遗传操作,形成中间种群,该种群与父代种群合并,计算合并种群中每一个体的代价函数,适应度,将适应度从大到小排序,适应度大的前 S 个个体作为下一代种群,进行下一次的循环。

当种群中含有代价函数 Ψ_i 小于 0.02 的个体时,循环结束,即找到了最优个体;或者迭代结束时,代价函数最小的个体为最优个体。

2.3 基于灰关联的击键特征识别算法

对于熟练的计算机使用者,按照上述击键间隔时间的定义,可能会出现负值,而灰关联分析是对非负序列进行分析的,必须对数据进行预处理。灰色斜率关联度和序列的空间位置没有关系,仅与其几何形状有关。我们对训练样本的序列及登录时测试序列的各个因素都加上数值 K ,来保证数据的非负性。

基于灰关联的击键特征识别算法为:当用户登录时,生成比较序列 T ,而由训练样本通过遗传算法形成的标准序列 T^* 作为参考序列,根据式(3) 计算两序列的灰色斜率关联度 ε ,两序列的关联度越大,说明比较序列越相似于参考序列,根据实验,设置合适的阈值 M ,当 $\varepsilon > M$ 时,判定用户为合法登录;当 $\varepsilon \leq M$ 时,判定用户为非法入侵。

3 实验结果及分析

本文数据采集通过 VC++ 6.0 编程获得,精度为 ms 级。随机设置口令 4 个,长度为 4~10 个字符,对于每个口令,采集合法用户在一段时间内击键样本 280 个,模拟入侵样本 230 个,合法用户的测试样本由用户间隔几个小时输入密码一次和模仿人疲惫时,连续不断的输入密码两种形式构成,对于每一个口令均有 4 个以上的非法入侵者,模拟恶意入侵的情况连续输入密码 50 次,构成入侵样本。在击键样本中随机抽取 80 个,在模拟入侵样本中随机抽取 30 个做训练样本集 A,其余构成测试样本集 B。

3.1 实验结果

对每个口令,在训练样本集 A 中随机抽取 50 个合法用户的击键样本,根据遗传算法确定合法用户的标准特征向量,交叉率取 0.8,变异率取 0.005。计算训练样本集中剩余击键样本与标准特征向量的灰色斜率关联度,结果如图 2 所示,其中横轴为灰色斜率关联度,纵轴为样本数。

由图 2 可以看出,对于合法用户样本,灰色斜率关联度特征分布非常集中,类内离散度很小。由于入侵样本来自不同人的击键模式,所以分布较为分散,但其值均小于 0.9,因此阈值 M 选为 0.89。用测试样本集 B 对此算法的性能进行了测试,结果见表 1。

表 1 用测试样本集 B 对此算法的性能测试结果

	Password1	Password2	Password3	Password4
FAR	1.0	2.5	1.0	1.5
FRR	0	0	0	0

3.2 实验结果分析

由图 1 可以看出,用灰色斜率关联度作为分类特征,合法用户样本的类内离散度很小,入侵样本虽然分布分散,类内离

散大,但与合法用户的样本分布有一定距离,所以灰色斜率关联度是一种很好的分类特征,由表 1 可以看出非法用户的通过率(FAR)和正常用户的拒绝率(FRR)都非常小,平均为 1.5%与 0%,显著优于基于特征向量、贝叶斯统计模型等算

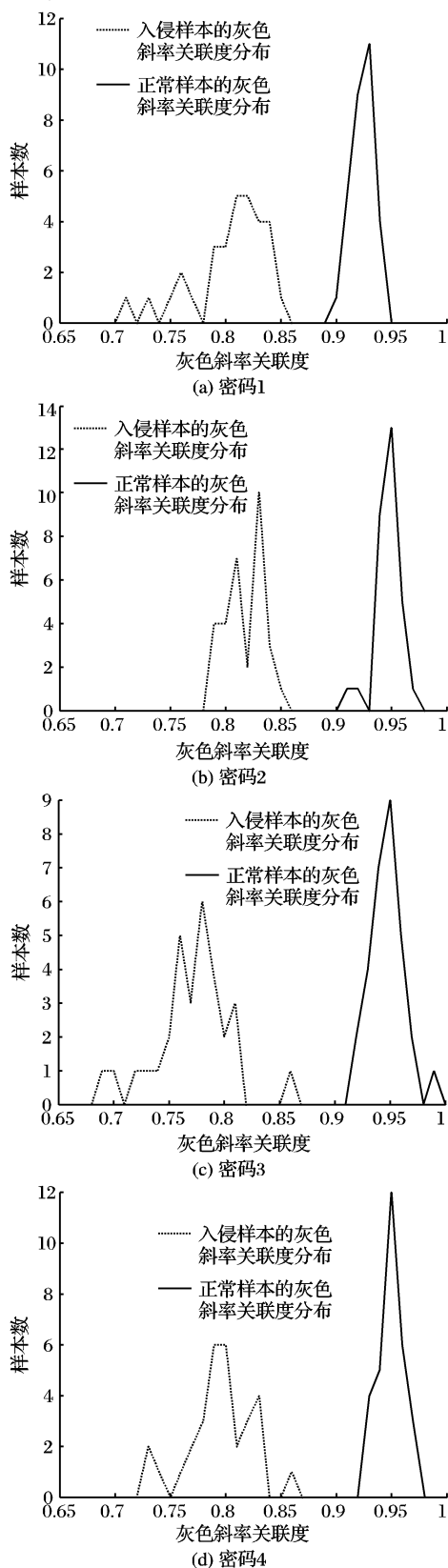


图 2 训练样本与标准特征向量的灰色斜率关联度分布

法,已经达到了神经网络方法、数据挖掘和支持向量机算法的最高水平。本算法虽然在利用遗传算法获得描述用户击键特征的标准特征向量有一定的算法复杂度,但其识别过程只需

要计算当前用户击键特征向量与合法用户标准击键特征之间的灰色斜率关联度,其时间复杂度与基于特征向量的快速识别算法相当,是一种计算复杂度低、识别精度高的基于击键特征的用户身份认证方法。

4 结语

本文提出一种基于击键特征的用户身份认证方法,该方法利用遗传算法在训练样本集中确定能表征用户击键特征的标准特征向量,然后计算待识别用户的当前击键特征序列与标准特征向量的灰色斜率关联度来识别合法用户和非法用户,实验表明,该算法性能显著优于基于特征向量、贝叶斯统计模型等算法,已经达到了神经网络方法、数据挖掘和支持向量机算法的最高水平。其识别过程只需要计算当前用户击键特征向量与合法用户标准击键特征之间的灰色斜率关联度,其时间复杂度与基于特征向量的快速识别算法相当,是一种计算复杂度低、识别精度高的基于用户击键特征的身份认证方法。

参考文献:

- [1] GAINES R, LISOWSKI W, PRESS S. Authentication by keystroke timing: some preliminary results[EB/OL]. <http://www.rand.org/pubs/reports/2006/R2526.pdf>, 2006-10-10.
- [2] LEGGETT J, WILLIAMS G, USNICK J. Dynamic identity verification via keystroke characteristics[J]. International Journal of Man-Machine Studies, 1991, 35(6): 859-870.
- [3] SALEH B, CHARLES S, BASSAM H. Computer-access security systems using keystroke dynamics[J]. IEEE Transaction on Pattern Analysis and Machine Intelligence, 1990, 12(12): 1217-1222.
- [4] DE RU WG, ELOFF JHP. Enhanced password authentication through fuzzy logic[J]. IEEE Expert, 1997, 12(6): 38-45.
- [5] TAPIADOR M. Fuzzy keystroke Biometrics on Web Security[A]. AutoID'99 Proceedings, Workshop on Automatic Identification Advanced Technologies[C]. IEEE, 1999. 133-136.
- [6] BROWN M, ROGERS SJ. User identification via keystroke characteristics of typed names using neural networks[J]. International Journal of Man-Machine Studies, 1993, 39(6): 999-1014.
- [7] CHO S, HAN C, HAN D, et al. Web-based keystroke dynamics identity verification using neural network[J]. Journal of Organizational Computing and Electronic Commerce, 2000, 10(4): 295-307.
- [8] YU E, CHO S. Keystroke dynamics identity verification-its problems and practical solutions[J]. Computer & Security, 2004, 23(5): 428-440.
- [9] GUVEN A, SOGUKPINAR I. Understanding users' keystroke patterns for computer access security[J]. Computer & Security, 2003, 22(8): 695-706.
- [10] 刘思峰. 灰色系统理论及其应用[M]. 第3版. 北京: 科学出版社, 2004.
- [11] 邓聚龙. 灰理论基础[M]. 武汉: 华中科技大学出版社, 2002.
- [12] SCHMITT LM. Theory of genetic algorithms[J]. Theoretical Computer Science, 2001, 259(1): 1-61.

(上接第1034页)

表1 组密钥协商方案的效率比较

	计算时间	模乘法运算量	轮数	通信开销
文献[9]	$(n^2 + 2n)T_E + 3n^2T_M + nT_L + (n^2 + n)T_C$	$(243n^2 + 480n)T_M + nT_L + (n^2 + n)T_C$	3	$3n^2 l + n l $
本文方案	$(n^2 + 2n)T_{ECM} + (3n^2 - n)T_{ECA}$	$(29.36n^2 + 57.88n)T_M$	2	$2n^2 l $

由表1可见:我们的方案在计算复杂度上只有 Bresson 和 Catalano^[9]方案的 1/8,在通信代价方面也只有它的近 1/2。因此,对于大的群组来说,我们的方案实现了更高的效率。

4 结语

基于椭圆曲线密码体制,提出了一个高效可认证的组密钥协商协议。协议仅需要两轮交互,就可以实现组密钥协商;利用类 ElGamal 密码系统,无需使用密钥分享技术;协议能够抵抗自适应选择消息攻击。与基于有限域上的离散对数相比,椭圆曲线上的离散对数计算更为困难。分析表明本文方案在计算复杂度与通信代价等方面比文献[9]具有更多的优势。

参考文献:

- [1] DIFFIE W, HELLMAN ME. New directions in Cryptography[J]. IEEE Transactions on Information Theory, 1976, 22(6): 644-654.
- [2] BELLARE M, ROGAWAY P. Entity Authentication and Key Distribution[A]. Crypto'93, LNCS 773[C]. Berlin: Springer-Verlag, 1994. 232-249.
- [3] BLAKE - WILSON S, MENEZES A. Authenticated Diffie - Hellman key agreement protocols [A]. SAC'98, LNCS1556 [C]. Berlin: Springer-Verlag, 1998. 339-361.
- [4] BRESSON E, CHEVASSUT O, POINTCHEVAL D. Provably authenticated group Diffie-Hellman key exchange - the dynamic case [A]. Advances in Cryptology ASIACRYPT'01 Proceedings, LNCS [C]. Berlin: Springer-Verlag, 2001. 290-309.
- [5] BRESSON E, CHEVASSUT O, POINTCHEVAL D. Dynamic group Diffie-Hellman key exchange under standard assumptions[A]. Advances in Cryptology EUORCPT'02 Proceedings, LNCS2332[C]. Berlin: Springer-Verlag, 2002. 321-336.
- [6] JOUX A. A one-round protocol for tripartite Diffie-Hellman[A]. Proceedings of ANTS-4 Conference, LNCS1838 [C]. Berlin: Springer Verlag, 2000. 385-394.
- [7] BURMESTER M, DESMETS YD. A secure and efficient conference key distribution system[A]. Advances in Cryptology EUORCPT'94 Proceedings, LNCS950[C]. Berlin: Springer-Verlag, 1995. 275-286.
- [8] 王志伟,谷大武. 基于树结构和门限思想的组密钥协商协议[J]. 软件学报, 2004, 15(6): 924-927.
- [9] BRESSON E, CATALANO D. Constant round authenticated group key agreement via distributed computation[A]. Public Key Cryptography-PKC2004, LNCS2947[C]. Berlin: Springer-Verlag, 2004. 115-129.
- [10] MIYAJI A. Elliptic curves over F_p suitable for cryptosystems[A]. Advances in Cryptology-AUSCRYPT'92 Proceedings, LNCS718[C]. Berlin: Springer-Verlag, 1993. 479-491.
- [11] LIN C, LEE C. Elliptic-Curve undeniable signature scheme[A]. Proceedings of the Eleventh National Conference on information Security[C]. 2001. 331-338.