Software Engineering and Applications 软件工程与应用, 2022, 11(1), 41-49
Published Online February 2022 in Hans. http://www.hanspub.org/journal/sea
https://doi.org/10.12677/sea.2022.111006

Hans汉斯

基于用户行为特征的低能耗身份认证

林梦琪，张晓梅

上海工程技术大学电子电气工程学院，上海

收稿日期：2021年12月24日；录用日期：2022年1月6日；发布日期：2022年1月30日

摘　要

目前，移动设备中传统的身份认证如密码、指纹等方式存在安全性低、易被破解等风险，不能完全保护使用者的隐私。本文提出基于用户行为特征的持续身份认证方案，通过调用移动终端传感器获得与用户行为相关的数据，采用最大互信息系数提取出能表征用户的特征，再利用机器学习算法进行模型训练，对用户身份进行识别。为了使持续的身份认证方式尽可能减少能耗，提出了低能耗的身份认证模型，对内置传感器的不同采样频率和识别算法进行能耗分析，实验结果表明，模型采用朴素贝叶斯算法，传感器采样率为25 Hz时，可以使认证精度达到97.94%，且显著降低认证模型的能耗。

关键词

行为特征，移动终端，传感器数据，身份认证，低能耗

Low Energy Identity Authentication Based on User Behavior Characteristics

Mengqi Lin, Xiaomei Zhang

School of Electronic and Electrical Engineering, Shanghai University of Engineering Science (SUES), Shanghai

Received: Dec. 24th, 2021; accepted: Jan. 6th, 2022; published: Jan. 30th, 2022

Abstract

At present, the traditional identity authentication in mobile devices, such as password and fingerprint, has low security, easy to be cracked and has other risks, and cannot completely protect the user's privacy. In this paper, a continuous identity authentication scheme based on user behavior characteristics is proposed. The data related to user behavior are obtained by calling mobile terminal sensors, and the features that can represent users are extracted by using maximum mutual information coefficient. Then the user identity is identified by machine learning algorithm for model training. In order to make the continuous authentication way as far as possible to reduce

文章引用: 林梦琪, 张晓梅. 基于用户行为特征的低能耗身份认证[J]. 软件工程与应用, 2022, 11(1): 41-49.
DOI: 10.12677/sea.2022.111006

energy consumption, low energy consumption of the authentication model is put forward, with built-in sensors of different sampling frequency and identification algorithms for analysis of energy consumption, the experimental results show that the model uses naive Bayesian algorithm, the sensor when the sampling rate is 40 Hz, can make the authentication accuracy reach 97.94%, and significantly reduce the energy consumption of authentication model.

Keywords

Behavioral Characteristics, Mobile Terminal, Sensor Data, Identity Authentication, Low Energy Consumption

Copyright © 2022 by author(s) and Hans Publishers Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/
Open Access

1. 引言

随着智能手机中存储的敏感信息(如银行账户)和个人信息(如短信、个人照片)的内容越来越多，隐私信息泄露正成为整个信息社会需要担忧的问题[1]。为了保护隐私和安全，目前移动端常用的身份识别方案有数字密码、图案密码、指纹识别、面部识别等[2] [3] [4]。然而，这类身份认证方式并不能够百分之百地保护，容易被攻破[5]并且认证频繁，每次唤醒设备都要进行验证。当前基于生物行为特征的连续身份验证越来越流行[6]，采集移动终端传感器以及用户行为等信息[7]，通过算法识别出当前使用者的身份，可以避免频繁使用口令、密码等显式认证方式，且能够对用户进行持续不间断的身份认证，因此能够兼顾用户的体验度与设备的安全性[8]。

生物行为识别技术基于用户的行为，如挥手的手势[9]、步态[10]、击键[11]、触摸屏[12]等，以及其他一般行为特征。Frank 等人[13]讨论了用户与触摸屏的交互作为连续认证的行为生物特征是否有效的问题，并证明了简单的触摸行为就足以验证用户。Hestbek 等人[14]提出了一种基于穿戴传感器和非循环特征提取步态的方法。然而，持续的身份认证会增加移动设备的验证开销，Riva 等人提出了渐进式身份认证的方式[15]，该方法将用户被要求输入密码的次数减少了 42%，使身份验证成为当前不使用安全锁用户的可行解决方案，并通过支持向量机算法得到 92.5%的召回率。Y. Yang 和 J. Sun 提出了基于行为的移动设备隐式认证的高效 w 层[16]，该方法消耗了设备总电池使用量的 14.5%，并得到了 96.73%的平均准确率。

在此基础上，本文提出了基于用户行为特征的身份认证方案，通过采集移动设备传感器数据，得到用户与移动设备触摸屏的交互信息，并提取出用户的行为特征，采用机器学习算法训练数据集从而实现对用户的身份验证。为了降低整个认证模型的开销，本文分别对采集数据阶段以及模型训练和认证阶段进行能耗分析，最终得到基于用户行为的低能耗认证模型，该模型不仅在一定程度上降低了能耗，并且能够得到较高的认证准确率。

2. 认证方案设计

基于用户行为特征的身份认证主要有以下几个阶段：数据采集、特征提取、模型训练、身份识别。数据采集部分主要是调用手机内置传感器获得与用户行为相关的交互数据；特征提取部分将采集到的原始数据进行一系列的计算得到更能够表征用户行为的特征(例如滑动轨迹长度等)，再将提取出的特征进行

分类；模型训练部分本文采用特征融合的方式将分类后的特征进行融合，采用机器学习算法进行模型的训练，并将模型保存；身份识别部分则将测试数据放入认证模型中，并得到认证结果。

2.1. 低能耗身份认证模型

在用户使用手机的过程中，会调用手机传感器来生成与用户行为相关的数据，再通过算法识别出用户的身份，从而达到持续的认证效果。然而这一过程会使认证后台持续运行，导致移动设备有能量的损耗。因此，本文提出了基于用户行为特征的低能耗的身份认证模型，在保证模型准确率的情况下，同时降低模型的能耗。

可以看出，整个身份认证的过程中，最体现能耗的部分即为传感器的调用和算法的识别，因此本文主要从内置传感器的数据采集和识别算法两个方面进行讨论，从降低数据采集过程中设备电量的消耗、减少训练模型过程中所耗费的时间和占用的内存来体现低能耗。认证模型如图 1 所示。

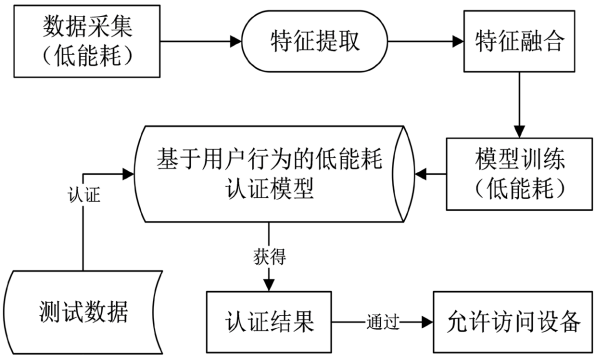

Figure 1. Low energy consumption identity authentication model based on user behavior characteristics
图 1. 基于用户行为特征的低能耗的身份认证模型

2.1.1. 内置传感器

本文主要调用加速度传感器、GPS 和陀螺仪传感器数据，以及触摸屏的 XY 坐标和速度。手机在调用传感器过程中，会导致 CPU、内存、网络和能量发生变化，而本文中的数据采集过程需要大量的调用传感器并返回数据，会频繁地占用内存以及消耗能量。为了探究低能耗的认证模型，本文决定采用不同的采样频率来采集数据，在 Android studio 平台将采集数据的 app 进行设置，采样率分别设为 50 Hz、25 Hz 和 5 Hz 进行实验。

2.1.2. 识别算法

模型训练和身份识别是模型中十分耗能的部分。由于不同的算法内部逻辑不同，训练一个模型所消耗的时间也会有所差异；调用训练好模型进行认证时，不同的算法得到的模型所占用的内存也各不相同；另外，模型认证所耗时间和准确率也有一定区别。

本文采用了贝叶斯、支持向量机、逻辑回归和决策树四种算法进行对比实验，并对算法的能耗进行分析，将准确率较高且能耗较低的算法选入低能耗的身份认证模型。

2.2. 数据采集

为了获得准确和适用的实验数据，本文通过 Android studio 平台自主开发阅读软件，使用 vivo Y38 手机运行软件，用户在阅读过程中滑动屏幕并调用内置传感器，软件后端记录每一滑动过程所产生的一系列数据。

实验数据由 10 名用户进行采集，为了保证数据的连贯性以及减少误差，要求每位用户进行 6 次数据采集，单次采集时长不少于半小时，最终各用户得到不少于 3000 条的轨迹信息。

由于在采集数据的过程中，用户可能会因为不当操作导致数据异常或冗余，因此需要对异常值进行剔除。本文采用 Z-score 对数据进行标准化，并判断出异常数据，主要步骤如下：

1) 将一组数据定义为 $x_1, x_2, \cdots, x_n$，计算均值 μ：

$$\mu = \frac{x_1 + x_2 + \cdots + x_n}{n} \tag{1}$$

2) 根据贝塞尔公式得到标准差 σ：

$$\sigma = \sqrt{\sum_{i=1}^{n} \left(x_i - \mu\right)^2 \Big/ n-1} \tag{2}$$

3) 得到标准化的数据 z_i：

$$z_i = \frac{x_i - \mu}{\sigma} \tag{3}$$

判断 z_i 的绝对值是否大于某个阈值，本文中阈值设置为 2，若 z_i 的绝对值大于 2，则视为异常的数据，进行剔除。其中阈值的设定是基于经验，在数据集较大的情况下，阈值为 2 时，表示有 2.5%的数据会被标记为异常。

2.3. 特征提取

本文通过数据采集共得到五类原始数据，分别为：加速度传感器(X、Y、Z)、陀螺仪传感器(X、Y、Z)、GPS (经度和纬度)、滑动坐标(X、Y)、滑动速度(X、Y)。

通过实验发现，原始数据并不能很好的表征用户，得到较高的认证准确率，因此需对数据进行处理，进一步得到这四类数据(除 GPS)的均值、方差、最大值、最小值、起始点和终止点，以及一次滑动的时间和 GPS 共计 59 个特征。

本文采用特征融合的方式，由于特征融合发生在特征之间，分类之前，也称作 early fusion。主要是利用特征选择技术，从融合的所有特征中找到最大限度的提高分类器性能的特征。

最大互信息系数

通过计算均值、方差、最大值、最小值、起始点和终止点的方式得到了 59 个特征，为了得到更能表征用户行为的特征，本文通过计算各特征与用户身份之间的最大互信息系数(Maximal Information Coefficient)来进行筛选。

最大互信息系数是用于来衡量两个变量之间的关联程度，与互信息(Mutual Information)相比，它具有更高的准确率。最大互信息数的计算利用了互信息的概念。

设有两个离散的随机变量 X 和 Y，它们之间的互信息量计算公式为：

$$I\left(X;Y\right) = \sum_{y \in Y} \sum_{x \in X} p\left(x,y\right) \log\left(\frac{p\left(x,y\right)}{p\left(x\right)p\left(y\right)}\right) \tag{4}$$

在上式中，$p\left(x,y\right)$ 是 X 和 Y 的联合概率分布函数，$p\left(x\right)$ 和 $p\left(y\right)$ 是 X 和 Y 的边缘概率分布函数。最大互信息系数是针对两个变量之间的关系离散在二维空间中，使用散点图来表示，将当前二维空间在 x、y 方向分别划分为一定的区间数，然后查看当前的散点在各个方格中落入的情况，这样就解决了在互信息中的联合概率难求的问题。最大互信息系数的计算公式为：

$$MIC(X;Y)=\max_{a*b<B}\frac{I(X;Y)}{\log_2\min(a,b)} \tag{5}$$

公式(5)中 a、b 表示 x、y 方向上的划分格子的个数，本质上就是网格分布，B 是变量，其中 B 的大小设置为数据量的 0.6 次方左右最佳。

通过计算 59 个特征与用户身份的最大互信息系数，本文选择将最大互信息系数大于 0.3 的特征筛选出来，共计 37 个特征作为特征集，如表 1 所示。

Table 1. Maximum mutual information coefficient between each feature and user identity
表 1. 各特征与用户身份的最大互信息系数

特征名	MIC	特征名	MIC	特征名	MIC
Latitude	0.939	X_ACC_Start	0.606	Z_ACC_Start	0.495
Longitude	0.935	XCoordinate_Std	0.602	XCoordinate_End	0.483
Y_ACC_Max	0.815	Z_ACC_Min	0.585	XCoordinate_Average	0.468
Y_ACC_Average	0.803	XCoordinate_Min	0.576	XCoordinate_Max	0.464
Y_ACC_Min	0.775	XCoordinate_Start	0.565	YVelocity_End	0.413
Y_ACC_End	0.764	Z_ACC_Max	0.558	YVelocity_Std	0.363
Y_ACC_Start	0.763	YCoordinate_Max	0.540	YCoordinate_Average	0.324
X_ACC_Max	0.704	YCoordinate_End	0.533	YVelocity_Average	0.314
X_ACC_Average	0.700	Z_ACC_End	0.528	X_GYRO_Max	0.313
X_ACC_End	0.638	YCoordinate_Min	0.527	XVelocity_Max	0.310
X_ACC_Min	0.636	Time	0.510	X_GYRO_Std	0.309
Z_ACC_Average	0.627	YCoordinate_Start	0.498	YVelocity_Max	0.303
YCoordinate_Std	0.619				

3. 实验结果与分析

本文从采样频率和识别算法两个方面对模型的能耗进行评估，并得到不同采样率下的认证准确率，使模型在保证能耗尽量低的情况下能有更高的精度。

在采样频率的设置中，本文根据 Android studio 平台中传感器的采样率参数，分别采用 SENSOR_DELAY_GAME、SENSOR_DELAY_UI 和 SENSOR_DELAY_NORMAL 进行实验，其分别代表采样率为 50 Hz、25 Hz 和 5 Hz。假设用户的一次滑动过程需要 1 s，那么在 GAME 状态下每次滑动可以采集到 50 条数据，而在 NORMAL 状态下只有 5 条数据。其中 SENSOR_DELAY_GAME、SENSOR_DELAY_UI 和 SENSOR_DELAY_NORMAL 是在 Android 编程中 SensorManager 的频率参数。

而在模型的特征筛选过程中，例如均值、方差和极值等，无论单条轨迹包含多少个点，均只会得到一个结果。因此，不同采样率下产生的用于训练和测试的数据量是相同的，但会导致数据的值各不相同。

3.1. 采样频率能耗分析

为了得到不同采样率下，身份认证模型的能量损耗，要求用户在不同的采样率下各使用 1 小时软件来采集数据，并通过电池容量监测管理软件监测手机电池使用情况。电池容量监测管理软件为长沙张量

方程信息科技有限公司所开发，版本为 3.1，可在安卓手机的应用商店中下载使用，界面如图 2 所示。在采集数据过程中，该软件会记录实时的 CPU 使用频率、CUP 温度以及电量消耗等信息。

Figure 2. Battery capacity monitoring management software
图 2. 电池容量监测管理软件

本文实验所用的手机设备均为 vivo Y38，其参数如表 2 所示。

Table 2. Vivo Y38 parameters
表 2. vivo Y38 参数

基本参数	
操作系统	Android 8.1.0
CPU	联发科 MT6762
CPU 核心数	2.0 GHz 八核处理器
GPU	IMG GE8320
运行内存	4 GB
机身容量	64 GB
电池容量	3260 mAh

采样频率实验结果

本文利用池容量监测管理软件进行实时的检测，并记录了不同采样率的情况下，用户使用采集数据软件 1 小时的过程中，电量的消耗、电量的下降速度、CPU 使用频率的最大值以及 CPU 在使用之初和使用结束时的温度范围，具体数值如表 3 所示。

Table 3. Energy consumption at different sampling rates
表 3. 不同采样率的能耗情况

	电量消耗	电量下降速度	实时 CPU 使用频率最大值	CPU 温度范围
GAME (50 Hz)	297.96 mAh	9.14%/h	90%	29℃~32℃
UI (25 Hz)	273.19 mAh	8.38%/h	87%	31℃~32℃
NORMAL (5 Hz)	263.41 mAh	8.08%/h	80%	31℃~32℃

由表可以看出，在电量消耗方面，GAME 模式明显高于 UI 模式和 NORMAL 模式，在 1 小时内，GAME 模式比 UI 模式多消耗 24.67 mAh，若使用 24 个小时，将多出 592.08 mAh 的电量，多出的电量相当于总电量的 1/5。除此之外，CPU 的温度在 GAME 模式下会上升 3℃，UI 模式和 NORMAL 模式几乎没变化。因此，选用 UI 模式或 NORMAL 模式来进行数据采集更佳。

3.2. 算法能耗分析

对于用户身份的识别，属于二分类问题，本文在识别算法中选择了朴素贝叶斯、支持向量机、逻辑回归和决策树这四种模型，通过该模型的运行时间和占用内存来评估其能耗，并根据准确度(Accuracy)与精密度(Precision)比较不同算法之间的认证精度。

其中 Accuracy 和 Precision 的计算方式如下：

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \tag{6}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{7}$$

TP 表示实际为正例且被分类为正例的个数，TN 表示实际为负例且被分类为负例的个数，FP 表示实际为负例但被分类为正例的个数，FN 表示实际为正例但被分类为负例的个数。Accuracy 表示的是分类器整体上的正确率，而 Precision 表示分类器预测为某个类别的正确率。

识别算法实验结果

本文将 1.2 节中采集到的数据进行特征的提取与模型的训练，并得到不同算法的运行时间，将训练好的模型保存到本地，记录该模型的占用空间，具体数据如表 4 所示。

Table 4. Time and space taken up by different algorithms
表 4. 不同算法所占用的时间和空间

	运行时间(ms)	占用空间(KB)
朴素贝叶斯	19.988	2.05
支持向量机	69.962	135
逻辑回归	45.970	1.09
决策树	20.019	3.06

识别算法中，支持向量机表现出的效果最差，不仅运行时间很长，而且占用空间也是其他算法的几十倍。在算法的选择中，若选择朴素贝叶斯或决策树，不仅更加节省识别算法所造成的时间和空间的消耗，而且可以提高整个身份认证过程的效率。

3.3. 认证准确率

除了考虑采样率和识别算法的能耗问题，身份认证中最主要的是身份识别的准确率，本文对不同采样率、不同算法进行了实验，并得到 Accuracy 和 Precision 来评估模型的整体性能。详细的实验结果如表 5 所示。

Table 5. The experimental results
表 5. 实验结果

算法	采样率	准确度(Accuracy)	精密度(Precision)
朴素贝叶斯	GAME (50 Hz)	97.96%	97.33%
	UI (25 Hz)	95.32%	97.94%
	NORMAL (5 Hz)	94.56%	93.33%
支持向量机	GAME (50 Hz)	84.69%	77.61%
	UI (25 Hz)	89.47%	84.62%
	NORMAL (5 Hz)	80.27%	72.55%
逻辑回归	GAME (50 Hz)	95.92%	92.86%
	UI (25 Hz)	95.91%	96.08%
	NORMAL (5 Hz)	95.24%	94.87%
决策树	GAME (50 Hz)	95.92%	94.59%
	UI (25 Hz)	94.56%	92.11%
	NORMAL (5 Hz)	93.88%	92.00%

从表 5 可以看出，朴素贝叶斯得到的准确率最高，在 GAME 模式下的 Precision 能够达到 97.33%，而支持向量机的准确率最低。

结合 2.1 节中对采样频率的能耗分析和 2.2 节中对算法的能耗分析，若选择在 UI 模式下进行数据采集，并选用朴素贝叶斯进行模型的训练和认证，一天内能减少手机 18%的电量，模型的训练时间与占用空间相较于其他算法也很少。将其作为低能耗的身份认证模型能得到 95.32%的准确度和 97.94%的精密度。

4. 结束语

本文提出基于用户行为特征的持续身份认证方案，通过调用移动设备传感器数据，利用这些原始数据提取出能够表征用户行为的特征，并采用机器学习算法对特征进行训练，实现对用户身份的持续认证。为了考虑认证模型的能耗问题，本文分别对传感器的采样频率和识别算法进行能耗分析，最终使用 25 Hz 的采样率和朴素贝叶斯算法得到基于用户行为特征的低能耗身份认证模型，实验结果表明，该模型不仅能在一定程度上降低能耗，且能达到 97.94%的认证精度。虽然本文的认证方案得到了较高的准确率，但是本文只考虑了用户静坐情况下使用移动设备的场景，下一步将优化应用场景，设置走路、跑步等运动状态，以实现同一用户在不同场合下的持续身份认证。

基金项目

国家自然科学基金(61802252)。

参考文献

[1] Shen, C., Zhang, Y., Cai, Z., *et al*. (2015) Touch-Interaction Behavior for Continuous User Authentication on Smartphones. *Proceedings of ICB*, Phuket, 19-22 May 2015, 157-162. https://doi.org/10.1109/ICB.2015.7139046

[2] Sarsavadia, R. and Patel, U. (2018) A Survey on Intelligent Face Recognition System. In: *International Conference on ISMAC in Computational Vision and Bio-Engineering*. Springer, Cham, 1209-1215. https://doi.org/10.1007/978-3-030-00665-5_114

[3] Bortolon, C., Capdevielle, D. and Raffard, S. (2015) Face Recognition in Schizophrenia Disorder: A Comprehensive Review of Behavioral, Neuroimaging and Neurophysiological Studies. *Neuroscience & Biobehavioral Reviews*, **53**, 79-107. https://doi.org/10.1016/j.neubiorev.2015.03.006

[4] Ali, M.M.H., Mahale, V.H., Yannawar, P., *et al*. (2016) Overview of Fingerprint Recognition System. 2016 *International Conference on Electrical, Electronics, and Optimization Techniques* (*ICEEOT*), Chennai, 3-5 March 2016, 1334-1338. https://doi.org/10.1109/ICEEOT.2016.7754900

[5] Liu, X., Li, Y., Deng, R.H., *et al*. (2018) When Human Cognitive Modeling Meets PINs: User-Independent Inter-Keystroke Timing Attacks. *Computers & Security*, **80**, 90-107. https://doi.org/10.1016/j.cose.2018.09.003

[6] 林梦琪, 张晓梅. 基于行为足迹的多模态融合身份认证[J]. 计算机工程, 2021, 47(10): 116-124.

[7] Yu, S.J., Younglee, J., Kim, M.H., *et al*. (2020) A Secure Biometric Based User Authentication Protocol in Wireless Sensor Networks. 2020 10*th Annual Computing and Communication Workshop and Conference* (*CCWC*), Las Vegas, 6-8 January 2020, 0830-0834. https://doi.org/10.1109/CCWC47524.2020.9031136

[8] Patel, V.M., Chellappa, R., Chandra, D., *et al*. (2016) Continuous User Authentication on Mobile Devices: Recent Progress and Remaining Challenges. *IEEE Signal Processing Magazine*, **33**, 49-61. https://doi.org/10.1109/MSP.2016.2555335

[9] Shrestha, B., Saxena, N. and Harrison, J. (2013) Wave-to-Access: Protecting Sensitive Mobile Device Services via a Hand Waving Gesture. Springer International Publishing, Berlin. https://doi.org/10.1007/978-3-319-02937-5_11

[10] 童随兵. 基于深度网络的步态识别技术研究[D]: [博士学位论文]. 上海: 上海交通大学, 2019.

[11] Crawford, H. (2010) Keystroke Dynamics: Characteristics and Opportunities. *IEEE* 8*th International Conference on Privacy Security & Trust*, Ottawa, 17-19 August 2010, 205-212. https://doi.org/10.1109/PST.2010.5593258

[12] Xu, H., Zhou, Y. and Lyu, M.R. (2014) Towards Continuous and Passive Authentication via Touch Biometrics: An Experimental Study on Smartphones. 10*th Symposium on Usable Privacy and Security* (*SOUPS* 2014), Menlo Park, 9-11 July 2014, 187-198.

[13] Frank, M., Biedert, R., Ma, E., *et al*. (2013) Touchalytics: On the Applicability of Touchscreen Input as a Behavioral Biometric for Continuous Authentication. *IEEE Transactions on Information Forensics and Security*, **8**, 136-148. https://doi.org/10.1109/TIFS.2012.2225048

[14] Hestbek, M.R., Nickel, C. and Busch, C. (2012) Biometric Gait Recognition for Mobile Devices Using Wavelet Transform and Support Vector Machines. 19*th International Conference on Systems, Signals and Image Processing* (*IWSSIP*), Vienna, 11-13 April 2012, 205-210.

[15] Riva, O., Qin, C., Strauss, K., *et al*. (2011) Progressive Authentication: Deciding When to Authenticate on Mobile Phones. *Proceedings of Usenix Security Symposium*, San Francisco, 8-12 August 2011, 15.

[16] Yang, Y. and Sun, J. (2017) Energy-Efficient W-Layer for Behavior-Based Implicit Authentication on Mobile Devices. *IEEE Conference on Computer Communications*, Atlanta, 1-4 May 2017, 1-9. https://doi.org/10.1109/INFOCOM.2017.8057222