COMPUTERS & SECURITY 45 (2014) 147–155

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/cose

Computers & Security

# Soft biometrics for keystroke dynamics: Profiling individuals while typing passwords

Syed Zulkarnain Syed Idrus [a,b,c,d,*], Estelle Cherrier [b,c,d], Christophe Rosenberger [b,c,d], Patrick Bours [e]

[a] Universiti Malaysia Perlis, 01000 Kangar, Perlis, Malaysia
[b] Université de Caen Basse-Normandie, UMR 6072 GREYC, F-14032 Caen, France
[c] ENSICAEN, UMR 6072 GREYC, F-14032 Caen, France
[d] CNRS, UMR 6072 GREYC, F-14032 Caen, France
[e] NISlab, Gjøvik University College, Gjøvik, Norway

ARTICLE INFO

*Article history:*
Received 31 October 2013
Received in revised form
6 April 2014
Accepted 25 May 2014
Available online 10 June 2014

*Keywords:*
Biometrics
Keystroke dynamics
Soft biometrics
Pattern recognition
Data fusion
Computer security

ABSTRACT

This paper presents a new profiling approach of individuals based on soft biometrics for keystroke dynamics. *Soft biometric traits* are unique representation of a person, which can be in a form of physical, behavioural or biological human characteristics that differentiate between him/her into a group people (*e.g.* gender, age, height, colour, race *etc.*). *Keystroke dynamics* is a behavioural biometric modality to recognise how a person types on a keyboard. In this paper, we consider the following soft traits: the hand category (*i.e.* if the user types with one or two hands), the gender category, the age category and the handedness category. For this purpose, we collected a new database. Two cases are studied: static passwords and free text. By combining machine learning and fusion process, the results are promising.



## 1. Introduction

It is accepted that the way a person types on a keyboard contains timing patterns, which can be used to label him/her and this is called *keystroke dynamics*. Keystroke dynamics is an interesting and a low cost biometric modality (Bours, 2012; Giot et al., 2011a), indeed no additional device is required. Keystroke dynamics belongs to the class of behavioural biometrics, in the sense that the template of a user reflects an aspect of his/her behaviour. Among the behavioural biometric modalities, we can mention signature dynamics analysis, gait recognition, voice recognition, or keystroke dynamics (Impedovo and Pirlo, 2007; Moustakas et al., 2010; Klevans and Rodman, 1997; Monrose and Rubin, 2000). In general, the global performances of behavioural biometric modalities (and especially keystroke dynamics) based authentication systems are lower than the popular morphologic biometric modalities based authentication systems (such as fingerprints, face or iris) (Maio and Jain, 2009; Wildes, 1997). The fact that the

* Corresponding author. Université de Caen Basse-Normandie, UMR 6072 GREYC, F-14032 Caen, France. Tel.: +33 645246908; fax: +33 231538110.
E-mail addresses: syed-zulkarnain.syed-idrus@ensicaen.fr, syzul@yahoo.com (S.Z. Syed Idrus), estelle.cherrier@ensicaen.fr (E. Cherrier), christophe.rosenberger@ensicaen.fr (C. Rosenberger), patrick.bours@hig.no (P. Bours).
http://dx.doi.org/10.1016/j.cose.2014.05.008
0167-4048/

performances of keystroke dynamics are lower than other biometric modalities can be explained by the *intra-class* variability of the users behaviour. This intra-class variability pertaining to computer users can be accounted for by a way of typing which is different when they are nervous, or angry, or even sad … (Epp et al., 2011).

One solution to cope with this variability is to study *soft biometrics*, which was first introduced by Jain et al. (2004). In that paper '*soft biometric traits*' are defined as "*characteristics that provide some information about the individual, but lack the distinctiveness and permanence to sufficiently differentiate any two individuals*". Jain *et al.* considered gender, ethnicity and height as complementary data for a usual fingerprint based biometric system. Thus, soft biometrics allow a refinement of the search of the genuine user in the database, resulting in a computing time reduction. For example, if the capture corresponds to a male according to a soft biometrics module, then, the standard biometric identification system can confine its search area to male users, without considering female ones.

Since the work of Jain *et al.*, several other articles related to soft biometrics can be found in the literature. In the paper (Ailisto et al., 2006), body weight and fat measurements are considered as soft criteria to enhance a standard fingerprint based biometric system. An overview can be found in Dantcheva et al. (2011) about soft biometrics, under a '*Bag of Soft Biometrics*', where Dantcheva *et al.* make a comparison with the pioneering work of Alphonse Bertillon, whose anthropometric criteria gave rise to soft biometrics (Rhodes, 1956). That paper proposes some facial soft biometrics and also body soft biometrics, namely weight and clothes colour detection. In Park and Jain (2010), Park and Jain present how gender or ethnicity and facial marks such as scars, moles and freckles can be used to enhance face recognition. In reference Dong and Woodard (2011), shape based eyebrow features are used for biometric recognition and soft biometric classification. In Denman et al. (2011), the authors use soft biometrics (height and colour model of head, torso and legs) to help identifying people in videos in surveillance networks. Marcialis et al. (2009) use hair colour and ethnicity as soft biometrics combined with face modality.

Regarding keystroke dynamics, Bixler and D'Mello (2013) look into the likelihood of 44 people's behaviour, whether they stay idle, involved or bored when asked to write on a given task. Their results are between 11% and 38% higher than random guessing. In our previous study (Idrus et al., 2013a), the results also show that it is possible to detect users' way of typing by using one/two hand(s) with over 90% recognition rate; gender between 65% and 90%; age between 65% and 82%; and handedness between 70% and 90% correct recognition accuracy with 110 users.

The objective of this paper is to propose an extended study of soft biometrics for keystroke dynamics from our previous study in Idrus et al. (2013a) on a new biometric benchmark database called 'GREYC-NISLAB Keystroke' (Idrus et al., 2013b) that we have created. We propose in this paper a thorough evaluation of the soft biometrics system and a comparison between static passwords and free text (digraphs). Thus, the novelty (compared to our papers mentioned) is to study to what extent soft biometrics can enhance the recognition performance of keystroke based authentication systems. Furthermore, we show how the performances can be increased significantly by data fusion for passwords. As soft criteria, we propose to test if it is possible to predict if the user:

| | |
|---|---|
| 1. types with one or two hands | 3. belongs to a particular age category |
| 2. is a male or a female | 4. is right-handed or left-handed |

This paper is organised as follows. Section 2 is devoted to the description of the proposed method. In Section 3, we describe the protocol that we applied and present the obtained results on the benchmark database in Section 4. Section 5 presents the conclusions and the different perspectives of this study.

## 2. Proposed methodology

In general, keystroke dynamics authentication systems involve a keyboard and an application for the capture and processing of the biometric information. Users are required to type on a keyboard running a dedicated application. Each capture is stored in a database within the application in the form of keystroke or timing features for all correct and incorrect entries. These features are composed of several timing values that are extracted, which is the *pattern vector* that is used for the analysis. For each soft criterion, two steps are involved in recognition evaluation: (i) a training step, and (ii) a test step, both relying on a machine learning algorithm. Here we have chosen SVM (Support Vector Machine) (Vapnik, 1998), on account of its efficiency. As a result, we compute the accuracy rate of the prediction of each soft category by the trained SVM. A graphical representation of the overall process is illustrated in Fig. 1. In order to enhance the overall recognition performance, data fusion is then applied.

### 2.1. Data capture

Different types of features can be extracted from a user while typing on a keyboard (Giot et al., 2011a): "*(i) code of the key; (ii) the type of event (press or release); and (iii) the time of the event*". All this timing information is stored in the form of raw data, which contains (see Fig. 2):

- *ppTime (PP)*: the latency of when the two buttons (keys) are pressed;
- *rrTime (RR)*: the latency of when the two buttons (keys) are released;
- *prTime (PR)*: the duration of when one button (key) is pressed and the other is released;
- *rpTime (RP)*: the latency of when one button (key) is released and the other is pressed.
- *vector (V)*: the concatenation of the previous four timing values.

Subsequently, the keystroke template *V* is utilised for the analysis for each soft category. For keystroke dynamics systems, we apply two approaches, namely: static passwords and free text. Concerning static passwords, we analyse all the typing features previously described. For free text, the