Pattern Recognition Letters 52 (2015) 80–86

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

Contributions to empirical analysis of keystroke dynamics in passwords ☆

Jugurta Montalvão [a,*], Eduardo O. Freire [a], Murilo A. Bezerra Jr. [a], Rodolfo Garcia [b]

[a] "Universidade Federal de Sergipe", 49100-000, São Cristóvão, Brazil
[b] "Pontifícia Universidade Católica do Paraná", 80215-901, Paraná, Brazil

ARTICLE INFO

Article history:
Received 3 April 2014
Available online 17 October 2014

Keywords:
Soft biometrics
Typing stabilization
Password length

ABSTRACT

Rhythmic patterns in passwords are addressed through biometric verification tests. Experimental results are obtained through three publicly available databases, whereas experiments are guided by questions **Q1:** How does a subject develop a stable *rhythmic signature* associated to a new password? and **Q2:** How does the number of symbols affect biometric performance? Measurements show that even if subjects are instructed to train themselves before sample acquisition, a clear habituation phenomenon is noticed at the beginning of the first sessions, both for the password *.tie5Roanl* and the passphrase *greyc laboratory*, with significant consequences in terms of biometric verification performances. As for **Q2**, all experiments show that error rates are consistently lowered as password length increases. Additionally, a marginal but potentially useful observation is the stabilization of patterns around a rhythmic profile which seems to be induced by the corresponding sequence of symbols, whose consequences are addressed.

1. Introduction

An important security issue in information systems is systematic verification of claimed user identity, which is typically done through the entering of a personal login password. Unfortunately, the weakness of most passwords chosen by users is in their simplicity, intended to facilitate memorization. This inherent weakness of human choices represents a threat to most systems, against which a pragmatic countermeasure is to impose criteria to user choices, through *ad-hoc password strength* measurements.

It is not surprising that *strong passwords*, according to these criteria, are frequently hard to memorize, which may push users to take note of them in notebooks or even in electronic files, to make them easy to grasp whenever necessary. Consequently, these notes end up becoming another kind of threat to system security.

A possible solution to this apparent deadlock comes from typing-dynamic based biometrics, or keystroke biometrics, a kind of behavioral biometrics which is based on the analysis of time intervals between events such as key pressing/releasing. In this kind of biometrics, it is assumed that every typist (or user of the system) produces unique dynamic patterns. Therefore, to some extent of precision, this dynamic uniqueness can be extracted and used to verify a claimed identity.

As early as 1980, researchers (e.g. [4,5,7,24]) have been studying the use of typing behavior for identification. Results from these works showed high correlation between measurements from the same subject. Following similar ideas, many approaches have been proposed since then, ranging from mean and covariance-based strategies [1] to artificial neural networks [16]. More recently, a study [2] was conducted concerning the discriminability of keystroke feature vectors for authentication from fixed texts. There, it was shown, first theoretically and then experimentally, that heterogeneous vectors can provide higher discriminability than aggregate ones.

In [19], a thorough overview of methods published before 2005 was presented, with comparisons between the main approaches. In 2009, Giot et al. [8] published an experimental package (Greyc) that includes a new database and algorithms for comparative tests. More recently, a broader overview was presented by Karnan et al. [11] in 2011, which summarizes most approaches used in keystroke dynamics during the last two decades. Just one year later, another wide survey was published by Banerjee and Woodard [3], followed by Pisani and Lorena [20], a detailed review of keystroke applications for intrusion detection, reinforcing the idea that keystroke dynamics are attracting the attention of an increasing number of researchers.

Banerjee and Woodard [3], in addition to a review of the psychological basis behind the use of keystroke dynamics, also discuss usual data acquisition methods, approaches and performances of most methods used by researchers when studying standard computer keyboards. These recent studies corroborate our feeling of an increase in importance of keystroke dynamics as a *serious* biometric solution, in spite of its simplicity of use.

In this paper, we focus on keystroke dynamics related to passwords as a supplementary behavioral signal that can be used to strengthen security of password-base systems. Passwords to be

☆ This paper has been recommended for acceptance by Nappi Michele.

* Corresponding author: Tel.: +5579-8876-1736.
E-mail address: jmontalvao@ufs.br, jugurta.montalvao@gmail.com (J. Montalvão).

http://dx.doi.org/10.1016/j.patrec.2014.09.016
0167-8655/

memorized by users typically are short sequences of characters, which could lead us to presuppose bad biometric verification performances for most practical purposes. However, we believe that keystroke dynamics in passwords carries enough biometric information to significantly improve security systems. Therefore, we empirically study rhythm profiles associated to password-like sequences of characters, through three publicly available databases.

By considering the acquisition protocols of these databases, explained in Section 2, we study how a subject develops a stable rhythmic signature associated to one of these password-like sequence of symbols. In Section 3, the preprocessing step (or time interval equalization) first proposed in 2006 [15] is briefly recalled and adjusted to process both down–down keystroke latency and key hold time. Improvements from this preprocessing step are shown through experiments adapted from papers by the owners of the databases presented in Section 2. These improvements justify the use of preprocessed data throughout experiments done in Section 4 to tackle the two questions presented there. Experimental results are further discussed in Section 5.

2. Databases

Three publicly available databases are used in this work. Although they were built by separate research groups in different years, they have in common the use of a single hypothetical password or passphrase for all subjects. By contrast, their sampling protocols are quite different, as explained in the next three subsections.

2.1. Database ".tie5Roanl"

The first database [12] we use in this study, publicly available for research purposes, features the following:

- The password *.tie5⟨Shift⟩Roanl* was chosen by the database owners as being representative of a *strong* (i.e. difficult to guess) 10-character password.
- 400 samples per subject were acquired with a laptop computer, throughout eight 50-sample sessions, at least one day apart.
- A total of 51 subjects took part in this effort.
- Along with the software application used to record typed keystrokes, an external reference timer for time stamping those keystrokes was also used for improved time precision.
- Every typed password was systematically checked for correctness.
- The acquisition software recorded each key-down or key-up event associated to the name of the key involved and a time stamp for the moment at which the keystroke event occurred.

2.2. Database "try4-mbs"

From the experimental part of the work reported by Loy et al. [14], another database for a password-like sequence of symbols was made publicly available. They called it *Keystroke100 benchmark dataset*. However, in this text, we call it *try4-mbs* as an alias which corresponds to the typed symbols.

Three relevant differences from the first database are the absence of capital letters, the acquisition of down–down keystroke latency only, and the fact that volunteers were *requested* to familiarize themselves with the password prior to data acquisition. Moreover, because this database is aimed at researches on both pressure-sensitive keystroke dynamics and typing biometrics, a device was developed by the database owners to simultaneously acquire keystroke latency and keystroke pressure from a total of 100 computer users. A total of 10 timing profiles plus 10 pressure measurement sequences were acquired from each participant. Thus, pressure signal is a relevant specificity of this database, although we discard pressure information in our study.

2.3. Database "Greyc"

According to Giot et al. [8], this database was initially created to test the evolution of the biometric model depending on the keyboard. During sample acquisition, volunteers were asked to enroll themselves in the system over a 1 month period, with 1 or 2 sessions per week. They could learn and test the password *greyc laboratory* as they wanted, before they typed it 6 times on each keyboard (i.e. keyboards 1 and 2). Thus, all available keyboard events were recorded, along with the corresponding keyboard label.

Thanks to the recording format of the keyboard that were used, we are able to estimate the average interval between every two consecutive acquisition sessions as presented in Fig. 4. The total number of volunteers in this database is 133, although only 100 took part in at least 5 sessions. The average number of captures per volunteer is 51.

3. Interval description and signal preprocessing

In 2006, a method was published proposing the use of time-interval equalization as a nonlinear memoryless preprocessing approach to improve performances of most keystroke-based biometric methods [15]. Databases of static and free text were acquired to show, in terms of error rates, the resulting gains that equalization could provide to most known methods published to that date. In this paper, we use the preprocessing approach as a preliminary step toward the main questions discussed in Section 4. To reinforce our belief in the usefulness of this preprocessing step, we do some experiments with the very same methods already used by each database owner. These experiments, whose results are shown in Table 1 and throughout this subsection, study the preprocessing gains in terms of Equal Error Rate (EER, the operational point of a biometric detector for which False Rejection Rate equals False Acceptance Rate). However, we emphasize that the preprocessing step is just an auxiliary tool in this work, whose details are to be found in [15].

Keystroke signals can be separated into subsignals such as (see illustration in Fig. 1):

- Down–Down (DD) interval is the time elapsed between the consecutive pressing of two (possibly different) keys.
- Hold (H) interval is the time elapsed between the pressing and releasing events of a given (single) key.
- Up–Down (UD) interval is the time elapsed between the releasing of one key and the pressing of another one, when they are consecutively pressed keys.

It is noteworthy that UD can also assume negative values. For instance, a subject can hold a key with a finger and type the next key with another finger, before releasing the first one.

Because we cannot predict these intervals, suitable models for them are random variables. In [15], a single random variable was used to model the probabilistic source of all DD intervals. Here, because intervals DD, H and UD are clearly different, statistically speaking, a straightforward model improvement is the use of three random variables instead, one for each kind of interval.

The starting point for preprocessing (also referred to as *equilization*) is the assumption that the random variable which models DD intervals, x^{DD}, follows an almost log-normal probability density function (pdf), so that $y^{DD} = \log_e(x^{DD})$ follows a normal pdf. Consequently, we should expect $w^{DD} = g(y^{DD})$ to follow an almost flat distribution between 0 and 1, where [15]:

$$g(y^{DD}; \mu_y^{DD}, \sigma_y^{DD}) = \frac{1}{1 + \exp\left(-\frac{1.7(y^{DD} - \mu_y^{DD})}{\sigma_y^{DD}}\right)} \quad (1)$$

and μ_y^{DD} and σ_y^{DD} stand for mean and standard deviation of the roughly normal (or Gaussian) variable y^{DD}, respectively.