COMPUTERS & SECURITY 30 (2011) 427–445

available at www.sciencedirect.com

ScienceDirect

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

Computers & Security

Unconstrained keystroke dynamics authentication with shared secret

*Romain Giot**, *Mohamad El-Abed*, *Baptiste Hemery*, *Christophe Rosenberger*

GREYC Laboratory, ENSICAEN, University of Caen, CNRS, 6 Boulevard Maréchal Juin, 14000 Caen Cedex, France

ARTICLE INFO

Article history:
Received 18 May 2010
Received in revised form
18 February 2011
Accepted 27 March 2011

Keywords:
Biometrics
Authentication
Keystroke dynamics
Support vector machine learning
Benchmark
Supervised template update

ABSTRACT

Among all the existing biometric modalities, authentication systems based on keystroke dynamics present interesting advantages. These solutions are well accepted by users and cheap as no additional sensor is required for authenticating the user before accessing to an application. In the last thirty years, many researchers have proposed, different algorithms aimed at increasing the performance of this approach. Their main drawback lies on the large number of data required for the enrollment step. As a consequence, the verification system is barely usable, because the enrollment is too restrictive. In this work, we propose a new method based on the Support Vector Machine (SVM) learning satisfying industrial conditions (*i.e.*, few samples per user are needed during the enrollment phase to create its template). In this method, users are authenticated through the keystroke dynamics of a shared secret (chosen by the system administrator). We use the GREYC keystroke database that is composed of a large number of users (100) for validation purposes. We compared the proposed method with six methods from the literature (selected based on their ability to work with few enrollment samples). Experimental results show that, even though the computation time to build the template can be longer with our method (54 s against 3 s for most of the others), its performance outperforms the other methods in an industrial context (Equal Error Rate of 15.28% against 16.79% and 17.02% for the two best methods of the state-of-the-art, on our dataset and five samples to create the template, with a better computation time than the second best method).

1. Introduction

Authentication systems allow entities to be recognized before using resources; these resources can be physical, like a building, or logical, like a computer application. Traditionally, individuals authenticate themselves on computers by using the classical couple of *username* and *password*. This scheme, which is based only on one factor: the knowledge of the username and the password, suffers from various security holes (Conklin et al., 2004). *Strong authentication* uses multiple authentication factors to improve security. In this case, individuals are authenticated with the help of at least two authentication methods using one or several different factors among: (1) something *we know*; (2) something *we have*; (3) something *we are*.

Biometric systems can take part in the strong authentication scheme by providing the factor what *we are* when used with one of the two other factors. We can provide strong authentication in the password authentication scheme (what *we know*) by combining it with *keystroke dynamics* (Peacock and Ke, 2004), which is a behavioral biometric modality monitoring the way individuals type on the keyboard (what *we are*).

* *Corresponding author.*
E-mail addresses: romain.giot@greyc.ensicaen.fr (R. Giot), mohamad.elabed@greyc.ensicaen.fr (M. El-Abed), baptiste.hemery@greyc.ensicaen.fr (B. Hemery), christophe.rosenberger@greyc.ensicaen.fr (C. Rosenberger).
0167-4048/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.cose.2011.03.004

Its main interest lies the fact that it is considered as *unobtrusive*, because users already use passwords for authentication on computers and keystroke timing captures do not affect the user's habit. Several types of keystroke dynamics systems exist in the literature and are generally based on *very long texts* (Gunetti and Picardi, 2005), *passwords* (Hocquet et al., 2007) or *shared secrets* (Obaidat and Sadoun, 1997) although several studies used a shared secret without referring to this term. The biometric sample can be captured *statically* (*i.e.*, at login phase) or *continuously* (*i.e.*, during the computer session). In this study, we focus on static authentication with shared secrets. Using a shared secret means that all users use the same password. The system always acts as an authentication system, because only a certain group of people is aware of this secret (what *we know*) while all the members of the group type it differently (what *we are*). This kind of authentication is interesting and can be used in different contexts: (i) several users use the same account, but it can be useful to track which user is really using the account (in this case, we talk about *identification* if the user does not specify his own username. This case will not treated in this paper), (ii) in analogy with password-protected buildings, an application requires the same password for all users and this password is changed at regular intervals, etc.

During the verification, the system checks if the password is the required one, if it differs from what is expected, the user is rejected, otherwise, the system checks if the keystroke dynamics match. If the keystroke dynamics correspond to the claimant's, the user is accepted, otherwise he is rejected. We argue on the fact that most of the results presented in studies in the literature cannot be compared easily due to various reasons which will be presented in this paper. In order to help solve this problem, we propose a dataset whose aim is to be used as a reference database in further keystroke dynamics studies. We also propose a new method based on Support Vector Machine (SVM) (Vapnik, 1998) for unconstrained shared secret keystroke dynamics.

The paper is organized as follows: this first section has presented the objective of this work. In the second section, we present the state-of-the-art of keystroke dynamics. In the third section, we detail the proposed method. In the fourth section, we present an experimental study for the validation of the proposed method. These results are discussed in the fifth section. The sixth section discusses the results. We conclude and present some perspectives in the last section.

2. Background

In this section, biometric systems are first presented. An overview of their evaluation aspects is then provided. Finally various discussions on the differences of keystroke dynamics studies are presented.

2.1. General biometric systems

2.1.1. Presentation

The aim of biometric systems is to verify the identity of an entity which access to a resource. In the case of *physical access*, this resource can be a building or a room, whereas in the case of *logical access*, this resource can be an application on a computer.

Different biometric modalities can be classified among three main families (even though we can find slightly different characteristics in the literature like the biological one that is often forgotten):

- *Biological*: recognition based on the analysis of biological data linked to an individual (e.g., DNA, EEG analysis, …).
- *Behavioral*: based on the analysis of an individual behavior while performing a specific task (e.g., keystroke dynamics, signature dynamics, gait, …).
- *Morphological*: based on the recognition of different physical patterns, which are, in general, permanent and unique (e.g., fingerprint, face recognition, …).

In this work, we are interested in a behavioral biometric modality: the *keystroke dynamics* for managing *logical access* (*i.e*, access to a computer application).

Biometric authentication systems are generally composed of two main modules: (a) the *enrollment module* which consists in creating a template (or reference) for the user with the help of one or several biometric captures (or samples), and (b) the *verification module* which consists in verifying if the provided sample belongs to the claimed user by comparing it with its template. After verification, a decision is taken to decide to accept or to reject the user depending on the result of the comparison. We can also use an optional (c) adaptive module which updates the template of a user after a successful authentication in order to reduce the intra-class variability (the biometric data are not stable which implies that different captures of the same user may be quite different).

2.1.2. Evaluation methodologies

Many works have already been done on the evaluation of biometric systems (Theofanos et al., 2008; Iso, 2006; Mansfield and Wayman, 2002). This evaluation may be realized within three different aspects:

- *performance*: the objective is to measure various statistical criteria on the performance of the system (*Capacity* (Bhatnagar and Kumar, 2009), *Equal Error Rate (EER)*, *Failure To Enroll (FTE)*, *Failure To Acquire (FTA)*, *computation time*, *Receiver Operating Characteristic (ROC) curves*, *False Acceptance Rate (FAR)*, *False Rejection Rate (FRR)* etc (Iso, 2006));
- *acceptability and user satisfaction*: this gives some information on the individuals' *perception*, *opinions* and *acceptance* with regard to the system (Theofanos et al., 2008; El-Abed et al., 2010);
- *security*: this quantifies how well a biometric system (algorithms and devices) can resist several types of logical and physical attacks such as Denial of Service (DoS) attack or spoofing or mimicking attacks (ISO, 2008).

In this work, we are mainly interested in performance evaluation, as our work deals with authentication algorithms and not a whole system and its working environment. The used metrics are the following ones:

FAR *False Acceptance Rate* which represents the ratio of impostors accepted by the system;