A two factor biometric framework for user authentication

MONICA CARFAGNI, MATTEO NUNZIATI, MATTEO PALAI

Dipartimento di Meccanica e Tecnologie Industriali

University of Florence

Via S.Marta 3, Florence

ITALY

{monica.carfagni, matteo.nunziati, matteo.palai}@unifi.it http://www.dmti.unifi.it

Abstract: - Recent efforts in enhancing security systems are oriented towards so called multi-factor authentication. Such kind of authentication methodologies relay on the usage of two or more sources of data for identity retrieval. In this paper, a two factor authentication framework is experimented, which is based on biometric parameters. The proposed framework involves the acquisition of online signatures and pass-phrases as two different and independent sources of data. The goal of this work is to explore possible multi-factor solutions, which avoid the current need of passwords and/or other authentication gadgets.

Key-Words: - Multifactor authentication, biometry, online signature recognition, speaker recognition, system fusion, generative models.

1 Introduction

Multi-factor authentication (MFA) is the most diffused solution to secure identification systems world wide. Examples of this approach are the two-factor verification systems introduced by banks, in order to authenticate users during online transactions [1]. Additionally, relevant ICT providers such as Google [2] are moving their authentication infrastructure towards this model.

In a multi-factor framework, people is authenticated by means of something they know (STK), something they have (STH) and something they are (STA). From a simplified perspective biometry could be considered as a way to obtain information of STA type. In this paper we want to test biometric solutions from a different perspective.

Biometry is generally referred to as the science (and technology) which provides information about someone's identity by means of his/her biological traits. Generally speaking, two different kinds of biometric traits exist: behavioral and physical. Behavioral biometry (BB) tries to assign unique habits to a given subject. Examples of BB are keystroke rhythm and signature pressure and speed. On the other side, physical biometry involves the analysis of anatomic traits such as fingerprints, face or voice (actually voice indirectly represents vocal tract anatomy [3]). In this paper we focus our research on online signature recognition and speaker (voice) recognition.

Signature recognition is the ability to assign/verify a unique id to a given signature act, analyzed during its realization by means of tablets. Speaker recognition, on its turn, is the ability to assign/verify a unique id to a given pass-phrase. Moving from the fact that applying a signature is a behavioral trait, this element can be considered as the unique ability a person has to impress a paper (rather that transmitting a set of impulses at a given speed and pressure [4]). On its turn, a pass-phrase can be considered as both something a person knows (the content of the pass-phrase) and something the person is (the frequency response induced by the vocal tract anatomy). In other terms we would like to explore the feasibility of a two factor biometric authentication system where signature is used as STK, while voice represents the STA component.

ISBN: 978-1-61804-109-8

2 Biometric models

The following section provides a brief description of methods employed in this paper for biometric authentication.

2.1 Outline of a biometric system

Biometric techniques requires that specific parameters, the *features*, are extracted from a row biometric signal (e.g. a voice recording or a signature on a tablet), later, a statistical model is enrolled against such features and stored in a specific facility along with an used ID. The model itself, also referred as a template, is used in order to provide a statistical representation of a given person. Recognition is then performed by means of *similarity* and *typicality* comparison.

Let $F=\{f_i with \;\; i=1,\ldots,n\}$ be the feature set acquired by a biometric device at a fixed sampling frequency, and Θ_0 be a template, a *similarity* score S_0 is defined as:

$$S_0=\frac{1}{n}\sum_{i=1}^{n} P(f_i|\Theta_0) \qquad (1)$$

where P is the probability operator. If an alternative template Θ_1 exists, it is possible to estimate a second *typicality* score S_1 and retrieve the normalized log-score (NLS) as:

$$NLS=\log\left(\frac{S_0}{S_1}\right) \qquad (2)$$

Fixed an acceptance threshold θ, if NLS$\geq\theta$ a subject is considered as the target of Θ_0, that is he/she is the person the template has been derived from. Otherwise, the person is rejected as unknown. Usually, Θ_1 is named Universal Background Model (UBM) and it is generated by pooling together feature sets obtained from a reference database R. This model is expected to provide a good estimation of the probability that certain features can occur among different people.

In other terms Θ_0 accounts for the similarity between a template and a person's biometric trait, while Θ_1 assesses for the typicality of a certain feature, that is, its frequency among a wide group of people.

The typicality allows a system to weight the actual similarity between a template and a feature set, evaluating how much original a component of a biometric trait is.

The acceptance threshold θ is usually fixed empirically during a test session and is defined on an application basis. We define here $\theta=\theta_{EER}$, that is, the value for which an Equal Error Rate (EER) is attained [5].

2.2 Models

The so named UBM-GMM model is widely employed in this paper for both signature and voice. This kind of model represents a special case of the Maximum A Posteriori (MAP) estimator for HMM parameters, described in [6].

In order to compute a proper template, Gaussian Mixture Models (GMM) are commonly used (compare [7] among others). Given a number h of multivariate Gaussian distributions $N(x,\mu_j,\Sigma_j)$, with x being a vector of iid random variables, a GMM based template is defined as:

$$\Theta=\sum_{j=1}^{h} \alpha_j N(x,\mu_j,\Sigma_j) \qquad (3)$$

where the covariance matrices are commonly constrained to diagonal form and weight coefficients (α_j) are constrained in order to satisfy:

$$\int_{-\infty}^{\infty}\sum_{j=1}^{h} \alpha_j N(x,\mu_j,\Sigma_j)\,dx=1 \qquad (4)$$

In order to properly compute a GMM based template, the unbiased estimators for each mean μ_j and covariance matrix Σ_j as well as the weights α_j must be retrieved. A straightforward solution is the application of the well known iterative Expectation-Maximization (EM) algorithm.

Anyway, the classical EM algorithm needs a relevant number of data for its estimates to be accurate enough. As a matter of fact, biometric traits do not provide such an amount of data. By applying MAP estimation, templates result more robust to random variation of biological traits. The MAP algorithm interpolates at each iteration of the EM algorithm between the UBM parameters and the parameters retrieved by the EM itself.

MAP is widely used in speaker recognition and the authors in [4] have successfully proposed the MAP for signature recognition tasks. Therefore, the templates used in this paper are based on the same computational model.

3 Features

This section describes both signature and speech features employed in this paper.

ISBN: 978-1-61804-109-8

3.1 Signature

On-line signature recognition requires the employment of digitizing tablets. Such tools allow to record several temporal patters, such as: the pen position on the tablet (x,y), its pressure (p).

In [4], the authors have reviewed the signature process from a physical perspective. Briefly, the whole act of signature making can be reduced to the motion of a point in space (the pen tip); therefore, the signature can be described by the classical problem of a material point moving in a bi-dimensional space. According to classical equations of mechanics, a material point moving on a generic path can be represented by a dynamic system, where the state is defined by the vector (x, y, δ, $\dot{x}$, $\dot{y}$, $\dot{\delta}$), that is, point's position and instantaneous velocity (being δ the angular velocity), while the input is defined by the acceleration provided to it by external forces: ($\ddot{x}$, $\ddot{y}$, $\ddot{\delta}$). Moving from this model and by adding the pressure information, authors proposed the following feature vector:

$$f' = [x, y, \delta, p, v, \dot{\delta}, \dot{p}, \dot{v}, \ddot{\delta}, \ddot{p}] \qquad (5)$$

This feature vector is employed again in this paper, computing a vector for each dataset acquired at 100Hz by a digitizing tablet.

3.2 Voice

One of the most commonly used features in speaker recognition are Mel Frequency Cepstral Coefficients (MFCC) [8]. In order to increase the discrimination capability of a speaker recognition system, such features are associated with their first and second order derivatives in a manner similar to the procedure described for signatures. Moreover the first and second order derivatives of signal energy are included to incorporate user habits related to loudness modulation in voice. MFCC are used to approximate the voice spectrum via discrete cosine transformation. In other terms by varying the number of MFCCs a more or less accurate representation of the speaker voice can be attained. State of art uses from 13 to 19 MFCC. Thus, each feature vector accounts for 41 to 59 parameters. Each feature vector is extracted from a 20 to 30 ms signal window (commonly overlapped Hamming windows are used). In this paper 20 ms Hamming windows are employed, extracting 13 MFCC plus their derivatives and energy derivatives, leading to 41 parameters for each signal window.
Additionally, the average energy value of each window is used in order to discriminate between silence and actual voice in recordings. Windows are grouped by means of a 2 component GMM. The windows belonging to the GMM component with the lower average energy are discarded as silence, while the others are employed for model training.

4 Experiment layout and results

The experiment proposed in this paper has been conducted by using chimeric data. A biometric chimera is a set of biodata obtained by grouping biometric traits belonging to different people. In such a way, even if a multi modal database doesn't exist, it is possible to generate fake authentication sessions. Researchers have warned [9] about the overestimation of discrimination capabilities induced by the usage of chimeric data. The authors are aware of this issue, nonetheless we consider that the provided results still provide insightful information about the strength obtainable by merging different biometric traits.
In the presented experiments, authors compare the reliability of signature based authentication methods with that of a fused system, where both voice and signature are used to discriminate people. In order to define the NLS of the fused system, independence of signature and voice is hypothesized, leading to the following NLS formulation:

$$NLS_{tot} = NLS_{signature} + NLS_{voice} \qquad (6)$$

the employed dataset involves the usage of 38 subjects. 19 of them are used to define the correct decision threshold, while the others are used for validation, that is, to simulate the effective performance of the system during its runtime. During training, 8 simulated accesses per user are simulated, for a total of 152 tempted accesses. Additionally 2682 fraudulent accesses are computed in order to evaluate system resistance to attackers. These accesses are used to define the EER. Validation is performed with the same procedure, by using the other half of the user set.
The following table resumes the results of experiments:

System	FA [%]	FR [%]	HTER [%]
Signature only	1.4	0.8	1,1
Voice only	14.5	21	17,75
Voice + signature	2	0	1

Table 1: system reliability in terms of false acceptance (FA), false rejection (FR) and half total error (HTER).

ISBN: 978-1-61804-109-8

4 Discussion and conclusion

The basic system (signature only) provides a quite good performance, with an half total error ((FA+FR)/2) of 1,1%. Even introducing a weak biometric model, as the speaker recognition one, the final HTER is decreased of 9%. Such an improvement is attained with a relevant improvement in user experience, as no rejection is emerged during tests for authorized people. The need to re-authenticate themselves due to system errors is a non neglectable issue in real life, therefore this attainment is as good as the HTER reduction. Nonetheless, this improvement is connected to an increased amount of access grants provided to non authorized people. Being this a major security risk, the presented framework do not represents a definitive solution for real world applications.

Eventually, the decrement in the global error (HTER) shows that purely biometry systems, adequately fused, can lead to interesting results in multi-factor authentication, therefore, major study will be dedicated in this direction, in order to overcome the current need of passwords and/or other authentication gadgets.

References:

1] Federal Financial Institutions Examination Council, *Guidance on Authentication in Internet Banking Environment*, http://www.ffiec.gov/pdf/authentication_guidance.pdf, 2005.

2] Nishit Shah, http://googleblog.blogspot.com/2011/02/advanced-sign-in-security-for-your.html, 2011.

3] Campbell J. P. jr., *Speaker recognition: a tutorial*, proceedings of the IEEE, vol. 85 (9), pp. 1437-1462 , 1997

4] M. Carfagni, M. Nunziati ,*An Improved Model and Feature Set for Signature Recognition*, Proceedings of the International Conference on COMPUTERS and COMPUTING, vol .1, pp.75-80, May 2011.

5] Doddington G.R., Przybocki M.A., Martin A.F., Reynolds D.A., *The NIST speaker recognition evaluation- overview, methodology, systems, results, perspective*, Speech Communication, vol 31, pp. 225- 254 , 2005.

6] J. L. Gauvain, C.-H. Lee, *Maximum a posteriori estimation for multivariate Gaussian mixture observations of Markov chains*, IEEE Trans. Speech Audio Process., Vol.2, pp. 291–298, 1994.

7] J. Richiardi, A. Drygajlo, *Gaussian Mixture Models for On-line Signature Verification*, Proceedings of the 2003 ACM SIGMM workshop on Biometrics, 2003.

8] Shurer T, *An experimental comparison of different feature extraction and classification methods for telephone speech*, 2nd IEEE Workshop on Interactive Voice Technology for Telecommunications Applications, pp.93-96 , 1994

9] N. Poh and S. Bengio, *Can Chimeric Persons Be Used in Multimodal Biometric Authentication Experiments?*, Machine learning for multimodal interaction, Springer, pp.87-100, 2006