International Journal of Business, Human and Social Sciences
ISSN: 2517-9411
Vol:2, No:8, 2008

Intelligent Speaker Verification based Biometric System for Electronic Commerce Applications

Anastasis Kounoudes, and Stephanos Mavromoustakos

***Abstract*—**Electronic commerce is growing rapidly with on-line sales already heading for hundreds of billion dollars per year. Due to the huge amount of money transferred everyday, an increased security level is required. In this work we present the architecture of an intelligent speaker verification system, which is able to accurately verify the registered users of an e-commerce service using only their voices as an input. According to the proposed architecture, a transaction-based e-commerce application should be complemented by a biometric server where customer's unique set of speech models (voiceprint) is stored. The verification procedure requests from the user to pronounce a personalized sequence of digits and after capturing speech and extracting voice features at the client side are sent back to the biometric server. The biometric server uses pattern recognition to decide whether the received features match the stored voiceprint of the customer who claims to be, and accordingly grants verification. The proposed architecture can provide e-commerce applications with a higher degree of certainty regarding the identity of a customer, and prevent impostors to execute fraudulent transactions.

***Keywords*—**Speaker Recognition, Biometrics, E-commerce security.

I. Introduction

TRANSACTION-BASED Internet applications, such as e-commerce, e-business and e-finance are continuously growing and gaining a considerable part of the marketplace. Providing access to information has been handled so far by implementation of personal identification numbers (PINs), cards or tokens [1]. The problem with these is that they do not ensure high security in cases such as transactions encountered in e-commerce. To further increase e-commerce growth and public respect, higher security protection should be provided to customers. By integrating advanced biometric verification in an Internet application, secure, low-risk and convenient transactions can be executed. Biometric verification means the automated use of physiological or behavioral characteristics, such as iris, face, signature, finger, or voice, to verify one's claimed identity [2].

Manuscript received May 31, 2006. This work was supported in part by the Cyprus Research Promotion Foundation and the General Secretariat for Research and Technology of the Ministry of Development of Greece under the Grant KY-EL/0603/77 and PLHRO/0603/01.

A. Kounoudes is with 'The Philips College', Lamias 4-6 Nicosia, 2001 Cyprus (e-mail: tkounoudes@cytanet.com.cy).

S. Mavromoustakos, is with 'The Philips College', Lamias 4-6 Nicosia, 2001 Cyprus (phone: +357.99.556939; e-mail: stephano@logosnet.cy.net.cy).

Verification methods using biometrics can replace or complement conventional authorization mechanisms, namely passwords and personal identification numbers (PINs), for higher security applications. The main risk of traditional authorization methods is that passwords and PINs are sensitive to be stolen, guessed or retrieved by a person. Moreover, considering the amount of e-commerce services a customer uses that require a password, it is difficult for him/her to possess securely multiple and difficult to be guessed passwords. On the other hand, biometrics utilize intrinsic characteristics of a person and are not susceptible to fraud. Another advantage of biometric methods over the traditional ones is that the verification is not restricted to a binary decision, thus multiple levels of security can be posed.

Comparing the biometric methods mentioned above, voice biometric systems, which are based on speaker verification, could be regarded as the most promising technique for being widely utilized in e-commerce applications. The increased presence of microphones devices and their low cost compared to other biometric acquisition devices makes voice biometrics the least expensive to deploy. Furthermore, voice biometrics do not invade customer's privacy and users are more willing to cooperate (voice has not been used for individual tracking and monitoring).

This paper, proposes an intelligent speaker verification process for enhancing security on e-commerce transactions using biometrics. The proposed method suggests the transmission of voice features instead of the whole speech signal to ensure maximum security and privacy and also save on bandwidth. The structure of this paper is as follows: Section II discusses biometric verification and its advantages and disadvantages. Section III provides an overview of the system. Section IV explains the proposed speaker verification process including the enrolment process and the verification process. This section also describes the tests performed to evaluate the performance of the system in real environment conditions. Finally, Section V sums up the findings of the paper and provides some concluding remarks.

II. Biometrics: Advantages and Disadvantages

Biometrics is the science of using digital technology to identify the identity of individuals based on behavioral or physiological characteristics. By basing a security system on the physiological features rather than a few keystrokes or a password, the possibilities of fraud are drastically reduced.

The terrorist attacks of September 11th 2001, and the desire to tighten security in every way possible, particularly in the USA, resulted enormous funds being made available to the research and development of biometric systems. As a result, the biometric industry is now emerging and is rapidly gaining acceptance from governments, companies and individuals.

Already, there are many industries employing biometrics, including the U.S. Immigration and Naturalization Service, major western countries armies, international banks, governments and healthcare organizations. The European Union also moves towards creating standards for biometric passports which will be deployed in the near future, while Britain plans to issue new identity cards which include biometrics. 'Athens 2004' also used biometrics to enhance the security of athletes and buildings during the Olympic Games of 2004.

There are many types of biometrics, but among the most common are scanning fingerprints, voices, faces, retinas or irises. Computer hardware and software programs have been developed to scan a thumb print, for example, and then compare it with a stored databank of other prints for an exact match, or a voice is compared to a bank of voice-print samples using pattern classification algorithms. Face recognition is the measurement of certain characteristics, such as the distance between eyes. Retina scanning has the computer camera inspecting the pattern of veins in a human eye. And, finally, iris scanning takes retina scanning one step further by concentrating on the color pattern surrounding one's pupils.

Key features of voice biometric that differentiate it from other types of biometric procedure are that it is non-invasive and that it can be performed remotely by telephone or via Internet. Approaches such as fingerprint analysis and retina scanning are much less acceptable to users. In addition, the cost and complexity of the systems required for fingerprint or retinal scanning far exceed that of the single microphone of a voice-based system that is, in any case, already provided in typical PC systems, the telephone and the mobile networks. Voice biometric systems generally include classical pattern recognition components; that is data acquisition (recording of speech signals), pre-processing, feature extraction and classification. These components are used in the two primary functional biometric system components, the enrolment and the verification processes discussed in Section III.

The main advantage that biometrics can offer is security and convenience. Among the various types of biometric technologies available, voice recognition is one of the cheapest to implement, especially for volume deployment for the e-commerce industry [1]. Iris scanning provides high security and is convenient in that it allows the users to keep their glasses on throughout the scan [3]. A biometric system is not based on a standard true or false system [1], but by utilizing a threshold of acceptance closeness to the user's characteristic different levels of physical security, authenticity, integrity and confidentiality can be established [4].

While biometric verification includes several advantages it does have some drawbacks as well. Even though, difficult but not impossible, fingerprints and pictures can be copied from anywhere and voice can be recorded [4]. Another major drawback is the cost associated with these technologies with iris scanning as being more expensive [1]. Finally, users of these systems concern of their privacy data. However, educating these people will curb their misguided fears [1].

III. System Overview

The system consists of the client, the E-commerce server and the Secure Voice Biometric Server (SVBS). The client could be any computer with Internet connection in which the user can access the e-commerce service. The role of the E-commerce server can be attributed to multiple, transaction-based e-commerce applications. The SVBS is a secure server that could be located away from the e-commerce server as a third party service. SVBS generates trains and updates the user's unique set of speech models (voiceprint), stores them securely in a database, and performs the matching process to authenticate a user.

Consider the case when a user needs to purchase an expensive product from an e-commerce site utilizing the proposed speaker verification approach for enhanced security. After registering to the e-commerce service, the user is asked whether he requires biometric user verification on his transactions. If the user selects this feature, then he is redirected to the SVBS where he follows the enrolment procedure to create his/her voiceprint, which is stored in the secure server.

Every time the user wants to purchase a product, he/she is redirected by the application sever to the SVBS where biometric verification is performed to verify the user's identity. If the user is the one who claims to be, then authorization is granted, and the user is free to proceed with the transaction. The verification procedure is shown in Fig. 1.

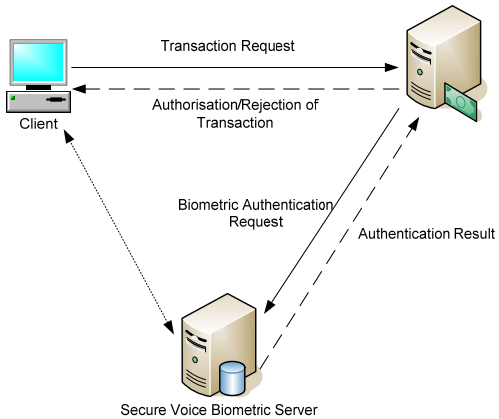

Fig. 1 Voice biometric verification for e-commerce transactions

IV. Voice Biometric Verification

During a speaker verification procedure, the user provides an identity claim together with speech samples corresponding

to prompts from the SVBS. The processing of the raw speech data results in distinctive and representative voice features (Feature Extraction), which contain information of the physiological characteristics of the user. The extracted features are then compared with the voiceprint of the claimed user, which was created during the enrolment phase, and a matching score is calculated (Verification). If the matching score is over a predefined threshold, then the authorization is considered successful; otherwise a call back procedure is followed. The following subsections describe in detail the enrolment and verification processes.

A. Enrollment Process

When a biometric verification is needed for transactions between a user and an E-commerce Server, the interested user should enroll in the SVBS, as illustrated in Fig. 2. Thus, the first time the user requests the service from the e-commerce service, its client is redirected to the SVBS. To eliminate the probability of a fraudulent enrolment, SVBS sends a password to the interested user via email.

The user can log into the SVBS by using this password to establish a secure connection with the SVBS. The SVBS sends a random sequence of digits 0-9 to the client, and the client prompts it to the user. While the user is pronouncing the sequence, the speech signal is recorded, and the client performs the feature extraction task. When the user has prompted the whole digit sequence (a procedure that can last two to five minutes), and a specifically downloaded from the SVBS client's software has extracted all the appropriate speech features, these features are sent back to the SVBS. The SVBS processes the received features and trains whole-digit HMMs (Hidden Markov Models) [5] for the specific user. The user's voiceprint, which consists of all digits HMM models, is safely stored at the SVBS database. Since the enrolment procedure is unsupervised, there is an increased risk of a low-quality but still valid enrolment. Such an enrolment can increase the probability of False Rejection, as well as the probability of False Acceptance for a user. In order to avoid such a problem, after the voiceprint of the user has been created, the SVBS starts immediately a verification process. If the verification is successful, the user's voiceprint is considered accurate and the enrolment ends. Otherwise the SVBS deletes the problematic voiceprint from its database, terminates the enrolment process, and suggests the user a second trial.

The strict protocol followed during the enrolment process is obliged by the fact that user's voiceprint is created for the first time. Early unsuccessful verification indicates inadequate hardware, misspelled training phrases, noisy environment, or suspicious enrolment trial, and thus it should be rejected.

B. Verification Process

When the user's client starts a high-security transaction with the E-commerce server, he is redirected to the SVBS (Fig. 3). After a secure network connection has been established between the client and the SVBS, the latter asks

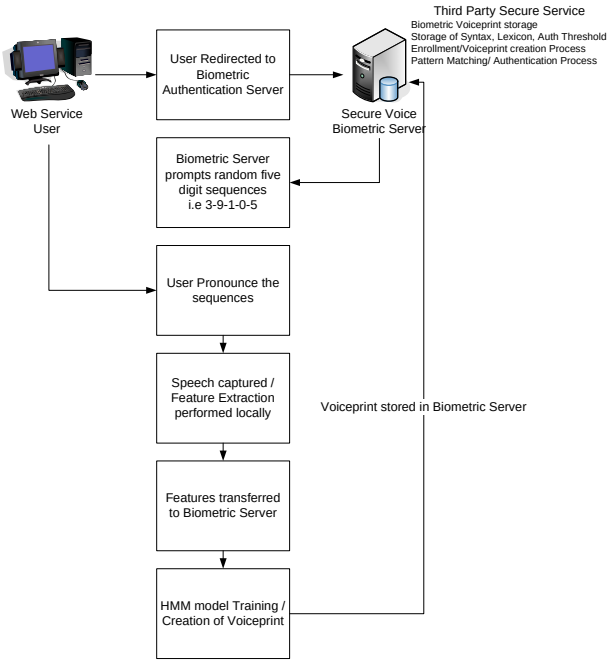

Fig. 2 The enrolment process

for an identity claim (such as a username) of the interested user. After receiving the username, the SVBS checks the user existence in its database. If such an identity exists, then the SVBS sends a sequence of five random digits to the client. The client's application prompts the user to pronounce the sequence, records the speech signal, extracts the required features, and sends them back to the SVBS. The SVBS performs the pattern matching operation between the received features and the voiceprint of the claimed user stored in its database.

If the matching score is above the threshold obliged by the security level of the application, then authorization is granted, and the result is forwarded to the E-commerce server. After a successful authorization, the SVBS updates the current voiceprint using the recently received features before storing it back to its database. In this way, the HMM models of each user are enriched to include more characteristics of the hardware configurations, environmental noise, and emotional conditions. Such a statistical generalization increases accuracy of the system. If the score does not meet the desired threshold, the authorization is repeated using a new digit sequence. In case the maximum number of three trials is exceeded, authorization is rejected. [8].

C. Internet based Evaluation Results

Speaker verification can be performed using various voice characteristics, while many approaches can be followed at the verification stage [3], [4]. Among the features one can extract from a speech signal for speaker verification purposes, the proposed method utilizes Mel-Frequency Cepstrum Coefficients (MFCC) [5], [6]. Tests using an in-house single digit database recorded over the Internet were performed to

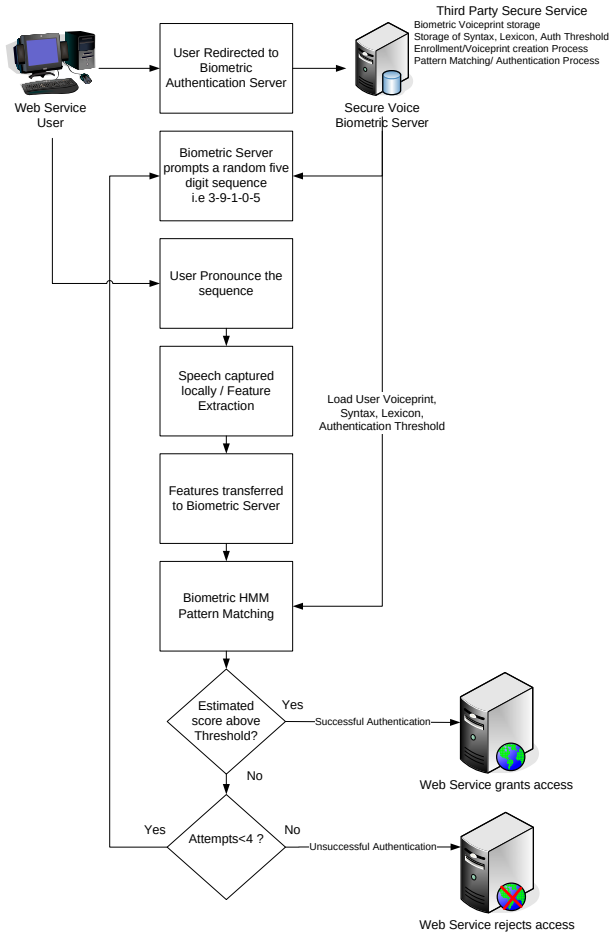

Fig. 3 The verification process

evaluate the performance of the proposed system. Specifically, recorded speech (8 KHz, 16 bits, mono) from ten users, were segmented in 25msec frames overlapping with other by 60%, thus a feature vector was output every 10msec. After pre-emphasis of the speech signal, 12 MFCC were computed. To capture time dynamics of the signal, the energy and MFCC first and second time derivatives (called Delta and Delta-Delta or Acceleration Coefficients) [10] were also computed, leading to 36-dimensional feature vector. Notice that Delta and Acceleration Coefficients were not computed at the client-side and transmitted back to the SVBS, since they could be directly computed through MFCC coefficients. Speaker verification is based on continuous density HMM (Hidden Markov Models). More precisely, a five-state left-to-right HMM with four mixtures is used for each digit, as well as for the silence interval [11]. An additional silence model was trained so as to model the beginning and ending of an utterance, as well as the intermediate pauses. The HMM are trained through the Baum-Welch algorithm [8], while speaker verification is performed using the Viterbi algorithm [8]. Data from ten users were used to evaluate the speaker verification performance against False Acceptance Rate (FAR), False Rejection Rate (FRR) and Equal Error rate (EER) [12]. Tests performed using the above conditions resulted an EER of 5%. In an internet-based e-commerce application, it is expected that different microphone configurations and/or environmental noise conditions will appear, and affect the speech signal in a different way. This problem, known as the 'mismatched condition' can severely degrade system's accuracy [6]. To maintain verification accuracy, a technique such as Cepstral Mean Subtraction (CMS) [13] was employed, and identical tests were repeated. It was found that CMS reduces the effect of the channel appearing in the recordings over the Internet, and increases verification performance by reducing the EER just below 1%. Moreover, the problem of mismatched condition can be eliminated through the dynamic update of user's voiceprint after a series of successful verifications.

V. Conclusion

Transaction-based e-commerce and e-business applications as continuously growing require higher security protection. Simple security mechanisms such as, username and password do not provide high security. Integrating advanced biometric authentication in such applications, ensure low-risk and convenient transactions.

This paper proposed a novel voice biometric authentication process for enhancing e-commerce service security. A system was developed, and demonstrated very good verification performance based on this approach. The system consists of the client, the e-commerce server and the Secure Voice Biometric Server (SVBS). The SVBS generates, trains, and updates users' voiceprints, stores them securely in a database, and performs the matching algorithm to authenticate a user.

The proposed architecture is advantageous since it is easily upgraded. Moreover, some heavy-duty functions, i.e. pattern matching and HMM training, have been loaded to the SVBS, and the main responsibility of the client is speech capturing and feature extraction. SVBS can apply different levels of security during the authentication procedure according to the security policy of the current application.

Acknowledgment

This work was supported in part by the Cyprus Research Promotion Foundation and the General Secretariat for Research and Technology of the Ministry of Development of Greece under the Grant KY-EL/0603/77 and PLHRO/0603/01. The views and conclusions contained in this document are those of the authors, and should not be interpreted as representing the official policies of Cyprus or Greek Governments.

References

[1] A. J. Harris and D. C. Yen, Biometric authentication: assuring access to information, Information Management & Security 10/1, pp. 12-19, 2002.
[2] J.L. Dugelay, J.C. Junqua, C. Kotropoulos, and R. Kuhn, Recent Advantages in Biometric Person Authentication, ICASSP 2002, International Conference on Acoustics, Speech and Signal Processing, May 13, 2002, Orlando, Florida, USA.

[3] J. Ashbourn, Biometrics: advanced identity verification: The complete guide, Springer-Verlag, London, 2000.
[4] A. Klosterman and G. Ganger, Secure continuous biometric-enhanced authentication, Carnegie Mellon University, Pittsburgh, PA.
[5] L. R. Rabiner, A Tutorial on Hidden Markov Models and selected applications in Speech Recognition, Proc. IEEE, vol. 77, pp. 257-286, Feb. 1989.
[6] R. J. Mammone, X. Zhang and R. P. Ramachandran, Robust Speaker Recognition, A Feature-Based Approach, IEEE Signal Processing Magazine, 13 (5), September 1996, 55-71.
[7] J. P. Campbell, Speaker Recognition: A Tutorial, Proceedings of the IEEE, 85(9), September 1997, 1437-1462.
[8] L. Rabiner, BH Juang, Fundamentals of Speech Recognition, (Prentice Hall, 1993).
[9] S. Furui, Cepstral Analysis technique for automatic speaker verification, IEEE Transactions on Acoustics, Speech and Signal Processing, vol. ASSP-29, 1981.
[10] J.R. Deller, J.G.Proakis, and J.H.L.Hansen, Discrete-Time Processing of Speech Signals, Macmillan 1993.
[11] D. Reynolds, Speaker Identification and Verification using Gaussian Mixture speaker models, Speech Communications, vol 17, pp. 91-108, 1995.
[12] S. Navanati, M. Thieme, and R. Navanati, Biometrics: Identify verification in a networked world (John Wiley & Sons, Inc. 2002.
[13] Hynek Hermansky, Exploring Temporal Domain for Robustness in Speech Recognition, 15th International Congress on Acoustics, 1995.