A Novel Method for Securing Online Databases

Stephanos Mavromoustakos, Cyprus College, 6 Diogenes Str., Engomi, P.O. Box 22006, 1516 Nicosia, Cyprus; E-mail: mstefanos@cycollege.ac.cy

ABSTRACT

In the past few years, we have experienced a rapid increase in the development and usage of Web-based systems. However, a main problem of these systems is the lack of security mechanisms to protect their data especially in cases where high security is required. In this work, we present the architecture of a secure system using voice biometrics. Among the existing security and biometric methods, voice biometrics can be an affordable technology providing high security. The proposed architecture includes a biometric server where the users' unique set of speech models (voiceprint) is stored. The authentication procedure requests from the user to pronounce a random 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 decides whether the received features match the stored voiceprint of the user who claims to be, and accordingly grants authentication. By implementing the proposed architecture, online databases are provided with a higher degree of certainty based on the user's identity.

1. INTRODUCTION

Web-based systems, such as web-banking and e-commerce, 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 provide a robust solution for e-commence and they are vulnerable to fraud. To further increase e-commerce growth and public respect, higher security protection should be provided to users. By integrating advanced biometric authentication in an Internet application, secure, low-risk and convenient transactions can be executed. Biometric authentication means the automated use of physiological or behavioral characteristics, such as iris, face, signature, finger, or voice, to verify one's claimed identity [2].

Authentication 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 Internet applications 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 utilizes intrinsic characteristics of a person and is not susceptible to fraud. Another advantage of biometric methods over the traditional ones is that the authentication is not restricted to a binary decision, thus multiple levels of security can be posed.

Comparing the biometric methods mentioned above, voice biometrics could be regarded as the most promising one for being widely utilized in Internet applications for securing online databases. The increased presence of microphone 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). Speaker authentication can be combined with other speech – enabled applications over the Internet, i.e. verbal passwords, speech recognition.

This paper, proposes a novel method for securing online databases using voice biometrics. The proposed method suggests the transmission of voice features instead of the whole speech signal to ensure maximum security and privacy and also to save on bandwidth. The structure of this paper is as follows: Section 2 discusses biometric authentication and its advantages and disadvantages. Section 3 provides an overview of the system. Section 4 explains the proposed voice biometric authentication including the enrolment process and the authentication process. This section also describes the tests performed to evaluate the performance of the system in real environment conditions. Finally, Section 5 sums up the findings of the paper and provides some concluding remarks.

2. BIOMETRICS

Biometrics is the science of using digital technology to identify the identity of individuals based on behavioural 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 America, 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. During the Olympic games in Athens in 2004, biometrics were also widely deployed to enhance the security of athletes and buildings.

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 [3].

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 handsets. 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 3.

The main advantage that biometric can offer is security and convenience. Among the various types of biometric technologies available, voice recognition is one of the cheapest to implement [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 authentication includes several advantages it does have some drawbacks as well. Even though it is 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].

3. SYSTEM OVERVIEW

The system consists of the client, the Application Server and the Secure Voice Biometric Server (SVBS). The client could be any computer with Internet connection in which the user can access an Internet service. The role of the Application Server can be attributed to multiple, online database applications. The SVBS is a secure server that could be located away from the application server as a third party service. The 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 voice biometric approach for enhanced security. After registering to the e-commerce service, the user is asked whether he requires biometric user authentication 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 voiceprint, which is stored in the secure server. Figure 1 illustrates the enrolment process while Figure 2 shows the e-commerce transaction process in an abstract form.

Every time the user wants to purchase a product, he is redirected by the application sever to the SVBS where biometric authentication is performed to verify (or not) 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.

4. VOICE BIOMETRIC AUTHENTICATION

During a speaker authentication 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 sections describe in detail the feature extraction, enrolment and authentication processes.

4.1 Feature Extraction

Speech is produced by the flow of air through the various articulators such as the vocal tract, lips, tongue, and nose. Air is forced out of the lungs through the trachea and the glottis, where it passes through the vocal cords. The vocal cords, if tense, vibrate like an oscillator, but if relaxed, do not vibrate and simply let the air pass through. The air stream then passes through the pharynx cavity and, depending on the position of a movable flap called the velum, exits either through the oral cavity (mouth), or the nasal cavity (nostrils). In the former case, the tongue and the teeth may modify the flow of the air stream as well. Different positions of these articulators give rise to different types of sounds. The different sounds produced by human beings are strongly related to the physiological characteristics of the vocal tract of each speaker. The fact that different speakers have different vocal-tract configurations for the same utterance is the basis for using vocal-tract filter parameters (feature coefficients) to good effect for speaker identification. These unique characteristics can be identified through a parameterization procedure called feature extraction.

Feature extraction is the process of measuring certain attributes of speech needed by the voice biometric system to differentiate people from their voice. The most often used technique that we also use in our system, is the mel frequency cepstral coding (MFCC) [6] which uses the Mel scale which is based on the human ear scale. The proposed system suggests the feature extraction process to be performed locally on the client's hardware and the calculated features to be securely transmitted

Figure 1. Enrolment to the biometric server

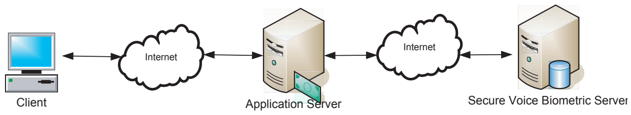


Figure 2. Voice biometric authentication for accessing an online database

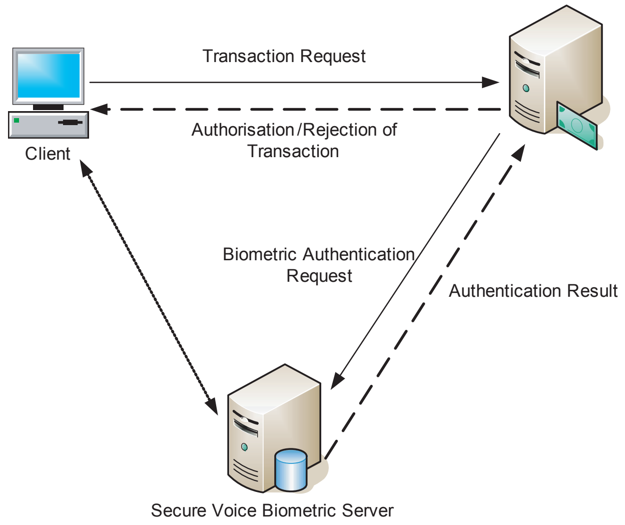
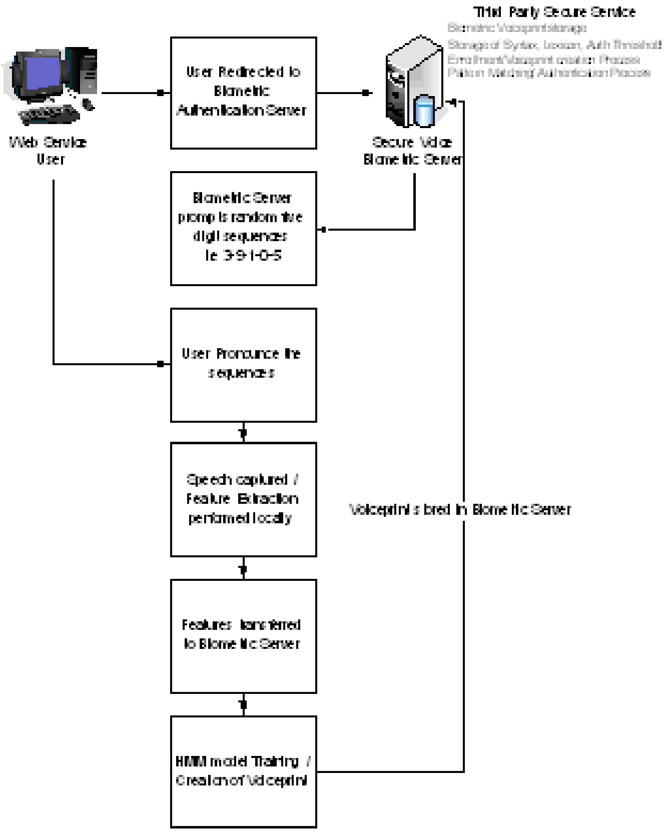

Transaction Request
Authorisation/Rejection of Transaction
Client
Biometric Authentication Request
Authentication Result
Secure Voice Biometric Server

Figure 3. The enrolment process


Third Party Secure Service
Biometric Voiceprint storage
Storage of Syntax, Lexicon, Auth Threshold
Enrollment/Voiceprint creation Process
Pattern Matching/ Authentication Process
Web Service User
User Redirected to Biometric Authentication Server
Secure Voice Biometric Server
Biometric Server prompts random five digit sequences ie 3-9-1-0-5
User Pronounce the sequences
Speech captured / Feature Extraction performed locally
Voiceprint stored in Biometric Server
Features transferred to Biometric Server
HMM model Training / Creation of Voiceprint

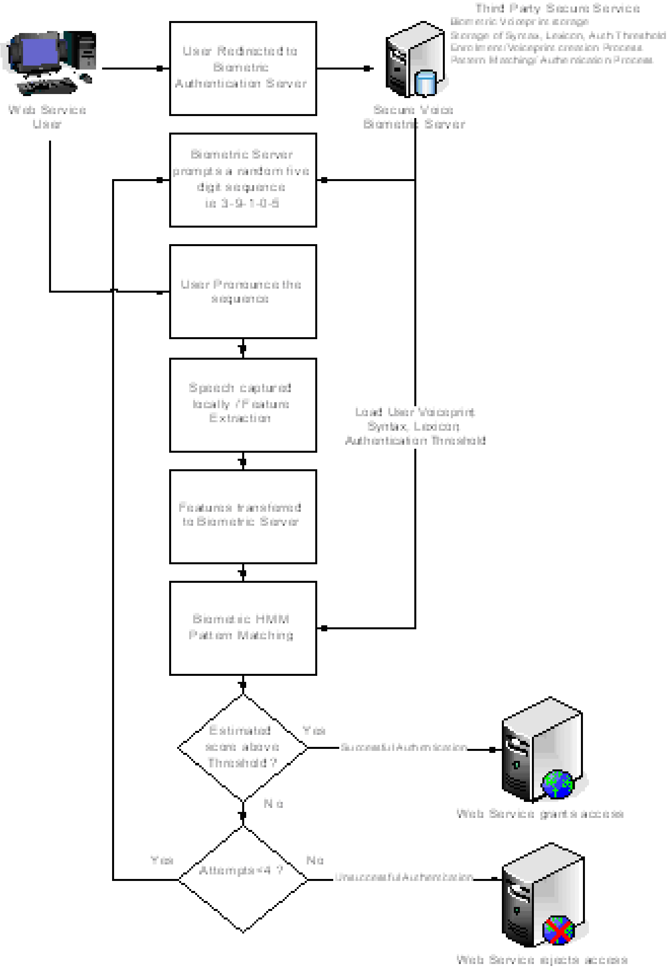
to the SVBS. The reason is that due to their size, the features take significantly less bandwidth when transmitted over the Internet making the whole interactive procedure of verification much secure and quick.

4.2 The Enrollment Process

When a biometric authentication is needed for transactions between a user and an Application Server, the interested user should enroll in the SVBS (Figure 3). Thus, the first time the user requests the service from the Application Server 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 which lasts 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 (0…9) 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 (FR) as well as the probability of False Acceptance (FA) for a user. In order to avoid such a problem, after the voiceprint of the user has been created, the SVBS starts immediately an authentication process. If the authentication 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.

Figure 4. The authentication process


Web Service User
User Redirected to Biometric Authentication Server
Secure Voice Biometric Server
Third Party Secure Service
Biometric Voiceprint storage
Storage of Syntax, Lexicon, Auth Threshold
Enrolment/Voiceprint creation Process
Pattern Matching/ Authentication Process
Biometric Server prompts a random five digit sequence ie 3-9-1-0-6
User Pronounce the sequence
Speech captured locally / Feature Extraction
Load User Voiceprint Syntax, Lexicon Authentication Threshold
Features transfered to Biometric Server
Biometric HMM Pattern Matching
Estimated score above Threshold ?
Yes
Successful Authentication
Web Service grants access
No
Yes
Attempts<4 ?
No
Unsuccessful Authentication
Web Service rejects access

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 authentication indicates inadequate hardware, misspelled training phrases, noisy environment, or suspicious enrolment trial, and thus it should be rejected.

4.3 The Authentication Process

When the user's client starts a high-security transaction with the Application Server, it is redirected to the SVBS (Figure 4). After a secure network connection has been established between the client and the SVBS, the latter asks 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 MFCC 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 and produces a matching score.

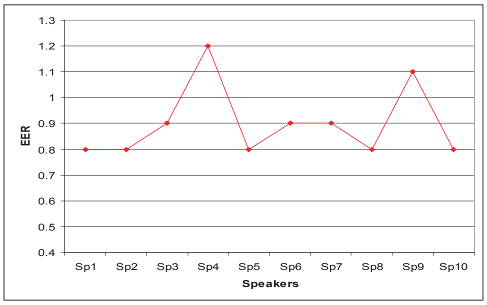
If the matching score is above the threshold obliged by the security level of the application, then authorization is granted and the result is forwarded to the Application Server to authorize the transaction. After a series of successful authorization, the SVBS updates the current voiceprint using the recently received features. In this way, the HMM models of each user are enriched to include more characteristics of the hardware configurations, and 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.

4.4 Evaluation Results

Speaker authentication 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 authentication 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 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 at the SVBS. Speaker authentication is based on continuous density HMM

Figure 5. Tests with CMS



(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 and also 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 authentication performance against False Acceptance Rate (FAR), False Rejection Rate (FRR) and Equal Error rate (EER) [12]. Tests performed using the above conditions resulted an average EER equal to 5%.

In an Internet-based 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 a system's accuracy [6]. To maintain authentication accuracy, a technique such as Cepstral Mean Subtraction (CMS) [13] was used and identical tests were repeated.

It was found that CMS can reduce the effect of the channel appearing in the recordings over the Internet and increase authentication performance by reducing the EER just below 1% as can be seen in Figure 6. Moreover, the problem of mismatched condition can be eliminated through the dynamic update of user's voiceprint after every successful authentication.

5. CONCLUSIONS

Transaction-based Internet applications as continuously grow require higher security mechanisms to protect their data. Simple security mechanisms such as, username and password do not provide high security. Integrating advanced biometric authentication in Internet applications, secure low-risk and convenient transactions.

This paper proposed a novel method for securing online databases using voice biometrics. A system was developed and demonstrated very good verification performance based on this approach. The system consists of the client, the Application Server and the Secure Voice Biometric Server (SVBS). The SVBS generates trains and updates the user's 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, 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.

6. 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.