Authenticating Sensitive Speech-Recitation in Distance-Learning Applications using Real-Time Audio Watermarking

Omar Tayan[1,2], Lamri Laouamer[3,4]
Dept. of Comp. Eng., College of Computer Sci. & Eng.,
Taibah University, KSA[1]
IT Research Center for Quran and its Sciences (NOOR)[2]
Department of Information Systems, CBE, Qassim
University, Buraidah, KSA[3]
Lab-STICC (UMR CNRS 6285), University of Bretagne
Occidentale, Brest, France[4]

Tarek Moulahi[5], Yasser M. Alginahi[2,6]
MIS Dept., CBE, Qassim University,
Buraydah, KSA[5]
Deanship of Academic Services,
Department Computer Science
Taibah Univ., KSA[6]

***Abstract*—Thispaper focuses on audio-watermarking authentication and integrity-protection within the context of the speech-data transmitted over the Internet in a real-time learning environment.The Arabic Quran recitation through distance learning is used as a case-study example that is characteristic of sensitive data requiring robust authentication and integrity measures. Thisworkproposes an approach for the purpose ofauthenticating and validatingaudio-data transmitted by a publisher or during communications between an instructor and students reciting via Internet communications.The watermarking approach proposed here is based on detection of the key patternswithin the audio signal as an input to the algorithm before the embedding phase is performed.The developed application could be easily used at both sides of the communication for ensuring authenticity and integrity of the transmitted speech signal and is proved effective for many distance-learning applications that require low-complexity processing in real-time.**

Keywords—Audio; Watermarking; Quran-recitation; Integrity; Authentication

I. INTRODUCTION

Recent advancements in information and communication technologies combined with the widespread growth of the Internet have enabled the ease of digital content distribution, communication and reproduction. Consequently, millions of users from the digital community are able to benefit from the advantages of fast and simple digital information exchange. However, such benefits come together in-hand with the problems and threats associated with ensuring digital copyright protection, preventing digital counterfeiting, proof-of-authentication and content-originality verification. Essentially, all such digital content in the Internet can be classified into images, text, audio and video, with the challenge being to ensure secure and reliable communications for each media type. In the literature, the techniques employed to provide the necessary security for such digital audio-content is known as digital-watermarking. Furthermore, cryptographic techniquesare commonly combined with the watermarking schemes in order to improve the security of the data within the communications channel.

The audio signals in their digital form are easily reproducible without any distortion; effective protection techniques have become essential. This need was reinforced by the new distribution supports such as the Internet and by compression methods such as MP-3 and MP-4 standards. Watermarking has been proposed to solve this problem. The watermark is a signal that is embedded into the original audio signal. The watermarked signal then contains information (hidden data that can be used for many purposes); for example, the hidden-data can be used to identify the owner (copyright protection to identify the source of illegal copies, verify the integrity of a document, track a signal in a network and include additional information such as the title, author, serial numbers, etc.Applications of digital audio watermarking are numerous and include: copyright protection, copy-protection, tamper proffering, access control, fingerprinting, digital right management, content authentication, annotation and privacy control, media forensics, communication enhancement, content protection, content filtering, improved auditing, content location, … etc. [1 - 2].

The constraints that should satisfy an audio watermarking scheme depend on the application. The main constraints are: 1) Inaudibility: the watermark signal must not be perceived by the listener, 2) Robustness: the watermark must resist any change in the signal, since this change does not result in degradation quality, 3) Capacity: the capacity corresponds to the quantity of bits to hide in the host signal, 4) Complexity: in practice, most watermarking operations must be done in real time (especially in the case of the watermark detection / extraction processes), this factor should be as low as possible by maintaining a high robustness. Hence, any watermarking scheme should find an ideal compromise between inaudibility, capacity, complexity, robustness and security which is not easy to achieve.

The audio watermarking techniques can be classified into two categories: spatial domain and frequency domain. The

spatial domain is the classic process where the watermark embedding /extraction will take place directly on the signal values and does not require any transform processing. The frequency domain is the space in which the signal will be considered as a sum of frequencies of different amplitudes by applying some transforms such as Discrete Cosine Transform (DCT), Discrete Wavelet Transform (DWT), Fast Fourier Transform (FFT), Singular Value Decomposition (SVD), etc. In this work, a technique based on DWT is proposed to identify the Quran reciter.

The objective of this work is to design and develop a digital audio-based encoding algorithm for sensitive speech (such as the Quran recitation case study used in this work) that embeds watermark data into the digital content whilst preserving the exact wordings, diacritics and Tajweed (quality of the pronunciation) sounds of the audio transmission.

The remainder of this paper is organized as follows: Section II provides the related work on audio-watermarking schemes and their classification. Section III provides the methodology and implementation for the proposed approach, and Section IV explains the integration of the various components and overall framework for the recognition system. Section V contains the analysis and results of the proposed framework, while Section VI provides a comparative study with other related works. Finally, Section VII concludes the paper.

II. Related Work

The research in audio watermarking started well after many techniques have been developed for watermarking on different multimedia files such as images and text. Embedding data in audio is usually more difficult compared to images since the Human Auditory System is more complicated than the Human visual system. Recently, many watermarking techniques have been developed to address different research problems related to audio files; however, techniques for highly-sensitive audio-content still need to be developed for sensitive applications such as the identification of recitations of the Holy Quran, in which even slight modifications to the audio-data can render the recitation/file as invalid [2]. The objective of this work is to develop an audio identification system for Quran reciters within the context of a distance-learning environment.

The patchwork technique first presented by Bender et al. [3]in 1996 was used on images. Statistical methods based on hypothesis testing had relied on large data sets. This method is usuallyapplied in a transform domain such as Fourier, Wavelet, ,… etc. to spread the watermarking in the time domain in order to increase robustness against any modifications [3-5].

Yee and Wei, implemented a non-blind two-channel time-frequency digital bits audio watermarking scheme with error-correcting code. The watermark bits are encoded with cyclic code before embedding it into the audio signal using time-frequency compression expansion technique with psychoacoustic model which decides on the coefficients to be deleted or added. Both channels of the stereo audio signal are used for watermark embedding. This combination of cyclic code and two-channel approach using the robust time-frequency technique of coding watermark bits has resulted in perfect recovery of watermark under attacks [6].

The work of Zhang et al. [7] dealt with the implementation of real-time audio watermarking techniques based on Digital Signal Processing (DSP). The implementation was illustrated using DSK5402.It uses qualitative watermark methods and fast Modulated Complex Lapped Transform(MCLT). The experimental results show the robustness and transparency of this technique [7]. Other works based on different transforms such as: Gao et al. in [8] proposes an audio zero-watermarking algorithm based on FFT. The proposed algorithm provided a solution for the contradictions of imperceptibility and robustness.The algorithm shows effective resistance to different types of attacks and appears to meet the requirements of watermarking security.Xiong-hua and Wei-zhen [9], proposed an adaptive digital audio blind detection fragile watermarking algorithm based on a modified Discrete Fourier Transform (DFT) transform.

The work developed by Furon and Pierre proposes an asymmetric watermarking method which provides higher security level against malicious attacks used for copy protection purposes.This method is versatile, as it can be adapted to a large number of watermarking techniques based on Direct Sequence Spread Spectrum (DSSS). The method studied was applied to copy protection framework by analyzing the possible threats and estimating the complexity of each class of attacks.The proposed method shows that watermarked content only attack is not possible with this method which is seen to be a real threat to other techniques such as DSSS and Watermarking Costa's Schemes(WCS). The disadvantages of this method are that asymmetric detectors need more complexity, memory and accumulate large amount of content in order to make a reliable decision [10].

The work by Tavakoli proposes a watermarking technique for cover communication through the telephone system.This technique is suitable for theIntegrated Services Digital Network(ISDN) and the Public Switched Telephone Network(PSTN) networks that can be modified for mobile systems. It uses a direct sequence spread spectrum algorithm with perceptual modeling of the Human Auditory System for embedding watermark into audio signals.Experimental results show the watermark is robust against attacks such as, Additive White Gaussian Noise(AWGN), Low Pass Filtering (LPF), D/A and A/D conversion, A-Law or u-Law conversion and down sampling to 64 kbps. In addition, it is also robust against audio format conversions such as wave to mp3 [11].

The authors in [12] proposed an audio watermarking technique based on chaotic mapping and used DWT to extract the wavelet coefficients of the audio signal. Here, the detail wavelet domain is chosen to embed the watermark so that to achieve transparency and fragility; Principle Component Analysis(PCA) was used to help reduce the watermark information needed to be embedded. Therefore, the signal reconstruction was achieved by the extracted watermark and can accurately locate the tampered region in the time domain. The experimental results demonstrate the efficiency of the

proposed method in terms of fragility, transparency and tamper localization.

Dutta et al. proposed an audio watermarking method based on Biometrics, in which the biometric pattern of an iris is used to generate the watermark that has a stamp of ownership. The watermark is then embedded in the high-energy regions selectively, which makes the embedding process robust against cropping and synchronization attacks [13].

In Chen et al. [14], a fragile watermarking scheme was proposed that embeds watermark data into the principle-components of the detailed wavelet coefficients with blind extraction based on the fast independent component-analysis system. The proposed fragile authentication scheme had demonstrated excellent transparency and tamper-detection capabilities under a number of simulated attack-scenarios.

Zhao and Shen [15] presented a semi-fragile audio watermark algorithm for authentication that includes tamper detection capabilities. The experimental results had also confirmed its robustness against various signal-processing operations. The contribution in [16] describes an inaudible speech and watermarking algorithm which embeds copyright information into audio files as proof of ownership. In this study, the watermarking process was achieved using a cascade of the SVD and DWT transforms. A set of attack scenarios specified by the Stirmark benchmark for audio files were simulated. It was demonstrated that the embedding logo could be successfully extracted whilst remaining robust against most attacks being simulated.

Baranwal and Datta, presented a comparative study of spread spectrum based audio watermarking techniques [17]. Er and Gul [18] presented a comparison of audio watermark techniques that can be used for source-origin authentication in real-time session initiation protocol (SIP) such as, Voice over IP (VoIP). The least significant-bit (LSB), DC-level shift (DCSHIFT), frequency-hopping spread spectrum (FHSS) andDSSS approaches were compared in terms of robustness and evaluation-times, complexity and capacity metrics. The results demonstrated the effectiveness of the FHSS and DSSS schemes for VoIPapplications that require source-authentication.

Other recent works found in literature include: The work of Kang et al. whoproposed a multi-bit spread-spectrum audio watermarking technique based on geometric invariant log coordinate mapping feature [19]. Chen et al. proposed an optimization based audio watermarking technique based on DWT [20]. Wang and Zhao proposed a synchronized invariant audio watermarking scheme based on DWT and DCT [21], Petrovic and Yang developed an audio watermarking in the compressed domain [22]. Zhao and Shen developed an audio watermarking algorithm for audio authentication [23]. Al-Haj et al. proposed a hybrid DWT-SVD audio watermarking [24].For further details on some of the above techniques or others the reader may refer to the following surveys published on this topic: [1], [25 – 27].

In an attempt to classify audio watermarking techniques found in literature, but not limited to [1 – 27], the authors developed the classification tree, Figure 1, of potential Audio-Speech Watermarking Techniques considered for the Holy Quran. Table 1 provides the general classification of audio watermarking techniques with their applicability to Quran computing.In addition, it provides the limitations and considerations needed.Finally, from studying the methods available in the literature, Figure 2 provides an overview of key issues of particular importance in Audio-Watermarking for Quran recitations.

TABLE I. AUDIO WATERMARKING TECHNIQUES APPLICABLE TO QURAN COMPUTING (SUMMARIZED FROM [2])

Application Domain	Technique	Applicable to Quran IT	Limitations and Considerations
Audio-Based Watermarking	Time-domain techniques	Not Ideal	Modifies fine details in recitation/speech; poor robustness to attacks
	Transform-domain techniques	Yes-conditional	Effective when used with spread-spectrum approach, provides good robustness. Ensure high inaudibility and no distortion.
	Spread-Spectrum Techniques	Yes-conditional	Embedding may exploit those less active segments during recitation for minimum distortion; good robustness to attacks.

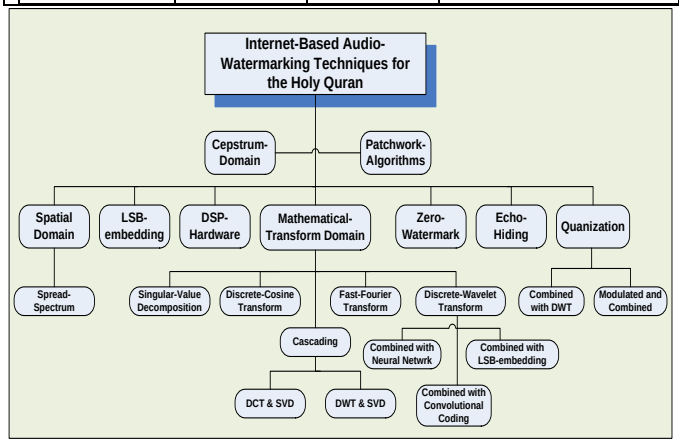

Fig. 1. Classification of potential Audio-Speech Watermarking Techniques for the Holy Quran

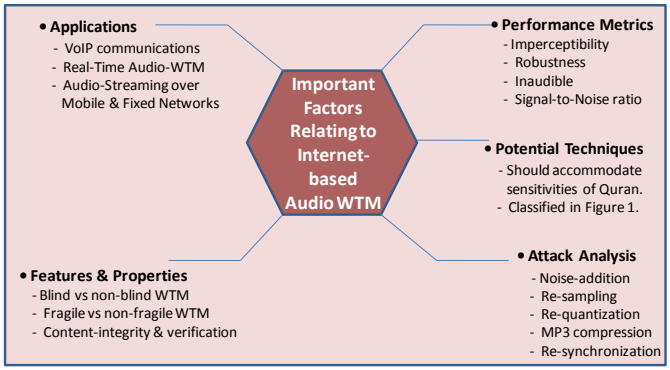

Fig. 2. Factors related to Internet-based audio Watermarking

In this work, the problem being addressed isthat of the security aspect of integrity and authentication with regard todigital Quran recitations and audio resources. Hence, a critical requirement is to ensure that all digital Quran audio/content that had originated from a known reference or reciter would be secure from being tampered with or modified in any way. That is, any modification or tampering of the digital Quran audio signal by an original publisher or source-recitation would be easily detected by the detection software and rendered as an invalid signal.The proposed methodology is provided in the next section.

III. Methodology and Implementation

The proposed work involves identifying the Holy Quran reciter by providing an improvedapproach based on DWTfor one level to achieve very low complexity (low watermark size). The frequency domain based research works in the literature deals with copyright issues and individual/public property. It should be noted that the audio watermarking approach based on transforms such as FFT, DCT, DWT,SVD, etc., provide remarkable robustness, but unfortunately are associated with high complexity which makes it difficult to adapt for real time applications. In contrast, the results obtained in our approach were remarkable in terms of robustness and complexity. We explain our improvement through three implementation approaches considered in this work, which include: (i) the embedding and extraction process for the case of enhanced robustness in section A, (ii) the approach of enhanced robustness and security by applying the Rivest, Shamir, Adleman (RSA) algorithm on the watermark as in section B, and finally, (iii) the approach for enhanced robustness with the use of secure Hyper Text Transfer Protocol (HTTPS), which employs the secure-socket layer (SSL) technique (section C). The use of security measures in section B and C significantly helps to preserve the outstanding safety against falsification of the reciter identity. However, due to the need for avoiding high processing-complexity and running-times, the RSA-based encryption/decryption approach (section B) was then replaced with the use of HTTPS, and SSL (section C) in order to achieve reduced online complexity for real-time applications.

A. Encoding Based on Robustness Requirement Only

In this section, we describe the first of three implementation approaches considered in which only the robustness metric is considered in the embedding scheme prior to signal transmission from the user-end. The mathematic and algorithmic steps involved in the embedding and extraction schemes are detailed in section A.1 and A.2, respectively.

1) Watermark Embedding Scheme

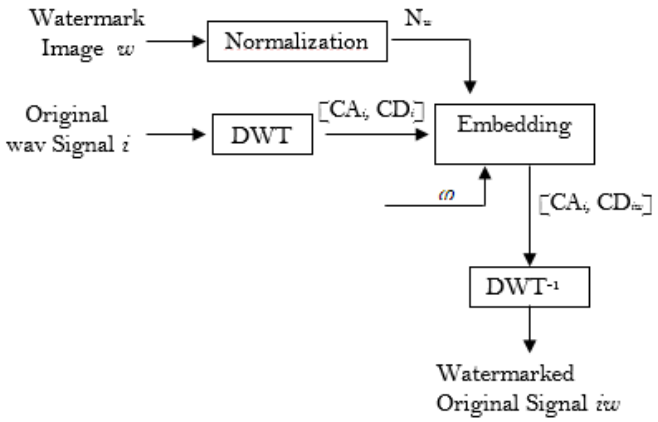

Fig. 3. Linear Interpolation-Based Watermarking Embedding Scheme

The linear interpolation based watermarking embedding process from Figure 3is defined as:

$$i_w^{(p,q)} = (1 - \varphi)w^{(p,q)} + \varphi i^{(p,q)}$$

where:

$i_w^{(p,q)}$ is the watermarked original audio signal

$w^{(p,q)}$ is the used watermark

$i^{(p,q)}$ is the original audio signal

(p, q) is the position of the sampled point

φ is the watermarking key, $\varphi \in]0, 1[$

CA: is the audio-segment with the most important information (from the input signal)

CD: are the detailed-segments within the input audio-signal (embedding is applied on this segment).

CD_{iw} is the embedded signal

Normalization: is the process of transforming the input matrix into a linear-vector.

In linear interpolation based watermarking, two cases will be analysed:

Case1: audible watermarking

When: $\varphi \to \mathbf{0}$ means $i_w^{(p,q)} \cong w^{(p,q)}$

Case2: inaudible watermarking

When: $\varphi \to \mathbf{1}$ means $i_w^{(p,q)} \cong i^{(p,q)}$

Since the embedding process will only concern the CD_icomponent, we have:

$$CD_{iw}^{(p,q)} = (1-\varphi)N_w^{(1,k)} + CD_i^{(p,q)}$$

Where:

$$N_w^{(1,k)} = \frac{w^{(S1,S2)}}{255}$$

s_1, s_2 are the size of the watermark image *w*

k is the length of the normalized watermark image $N_w^{(1,k)}$

Hence, after applying the wavelet inverse on the components $\left(CD_{iw}^{(p,q)}, CD_i^{(p,q)}\right)$ as follows, we will obtain the audible watermarked original audio-signal $i_w^{(p,q)}$:

$$i_w^{(p,q)} = DWT^{-1}\left(CD_{iw}^{(p,q)}, CD_i^{(p,q)}\right)$$

Algorithmic Steps:

1. Read the original wav signal *i*.
2. Compute the wavelet components of *i* (CA*i*, CD*i*)
3. Create a watermark image *w* (representing the reciter name)
4. Normalize *w* and obtain N*w* (where N*w*=*w*/255)
5. Perform the watermark embedding: $CD_{iw}=(1-t)N_w+tCD_i$
6. Apply the wavelet inverse (DWT^{-1}) to obtain *iw* (the watermarked original signal), which is sent over the SSL transmission line.

2) Watermark Extracting Scheme

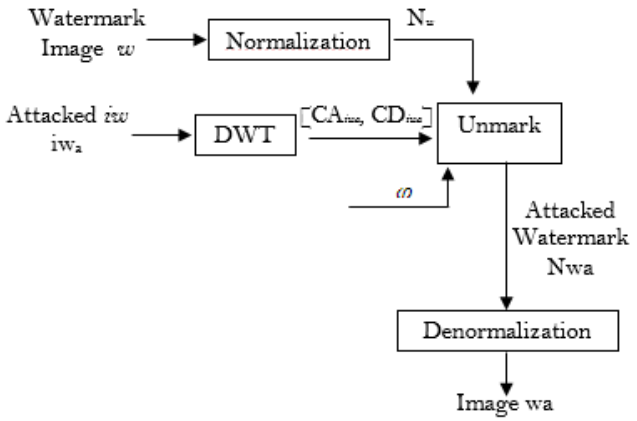

Fig. 4. Linear Interpolation-based watermarking embedding scheme

The extraction process, Figure 4,is defined by:

$$w_a^{(p,q)} = \frac{1}{\varphi} CD_{iwa}^{(p,q)} - \frac{1-\varphi}{\varphi} w^{(p,q)}$$

Where:

$N_{wa}^{(p,q)}$ is the extracted normalized watermark after an attack $CD_{iwa}$is the extracted watermark following an attack during transmission

$CD_{iwa}^{(p,q)}$ is the attacked component $CD_{iw}^{(p,q)}$

To obtain the attacked watermark image, we must denormalize the vector $N_w^{(1,k)}$ by:

$$w_a = N_w^{(1,k)} \times 255$$

Algorithmic Steps:

1. Read the attacked signal (iw_a), derived from the original-signal (*iw*) following an attack.
2. Compute the wavelet components of iw_a (CAiw_a, CDiw_a)
3. Read the watermark image *w* (e.g. a text-string or bit-stream of the reciter's name)
4. Normalize *w* and obtain N*w* (where N*w*=*w*/255)
5. Perform watermark extraction by: $Nw_a=1/(1-t)\ CDiw_a - (t/1-t\ )Nw$
6. DenormalizeNw_a to obtain w_a, such as w_a= Nw_a *255

B. Encoding Based on Robustness and Security (RSA Encryption)

In this section, we describe the second implementation approach developed, in which the robust embedding scheme follows with RSA encryption prior to signal transmission from the user-end in order to ensure robust and secure transmission. The mathematic and algorithmic steps involved in the embedding and extraction schemes are detailed in section B.1 and B.2, respectively.

1) Watermark Embedding Scheme

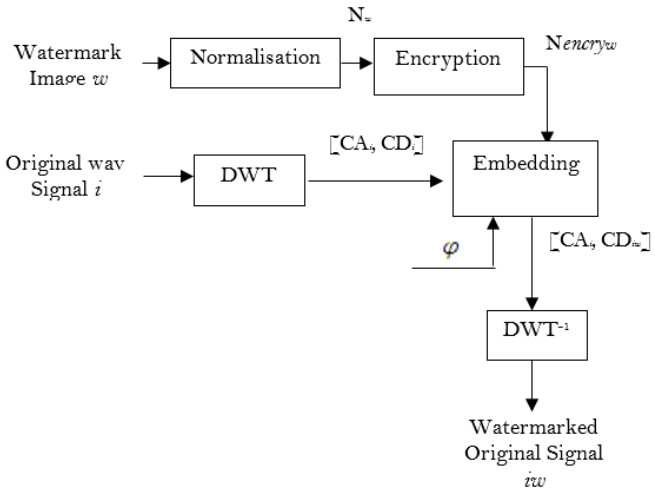

Fig. 5. Encoding Scheme based on Robustness and Security

The algorithmic steps for embedding and encryption are summarized as follows:

Algorithm steps:

1. Read the original wav signal *i*.
2. Compute the wavelet components of *i* (CA*i*, CD*i*)
3. Create a watermark image *w* (representing the reciter name)
4. Normalize *w* and obtain N*w* (where N*w*=*w*/255)
5. Encrypt N*w* (based on RSA Algorithm) and obtain $Nencry_w$following encryption.

6. Achieve the watermark embedding by: CD_{iw}=(1-t) $Nencry_w$ +tCD_i
7. Apply the wavelet inverse (DWT^{-1}) to obtain *iw* (the watermarked original-signal).

2) Watermark Extraction Scheme

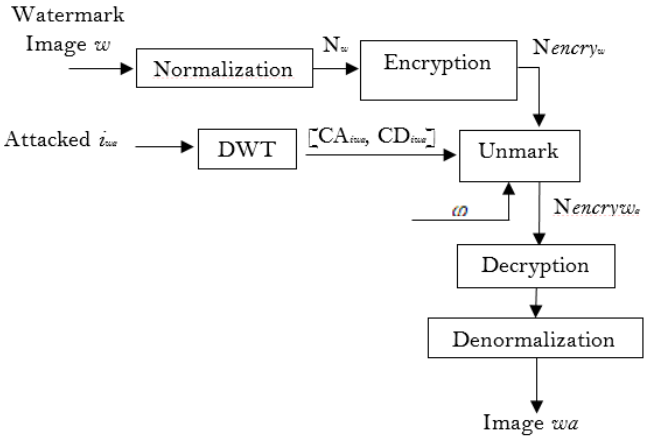

Fig. 6. Extraction Scheme based on Robustness and Security

The algorithmic steps for extracting an RSA-encrypted signal are summarized as follows:

Algorithm steps:

1. Read the attacked signal (iw_a), derived from the original-signal (*iw*) following an attack.
2. Compute the wavelet components of iw_a (CAiw_a, CDiw_a)
3. Read the watermark image *w* (e.g. the reciter name as a bit-stream/text-string)
4. Normalize *w* and obtain N*w* (where N*w*=*w*/255)
5. Encrypt N*w* (based on the RSA Algorithm) and obtain $Nencry_w$
6. Perform watermark extraction using: $Nencry_{wa}$ =1/(1-t) CDiw_a –(t/1-t) $Nencry_w$
7. Decrypt $Nencry_{wa}$ and denormalize it to obtain w_a

C. Enhanced Security Mechanism For Real-Time Support by Combining Robust Watermarking with HTTPS/SSL

This section describes the third implementation approach developed, in which the robust embedding scheme follows with transmission from the user-end using HTTPS and SSL in order to ensure robust and secure transmission with lower complexity as compared with the RSA approach (from section B).

To enforce the security of the watermarked *.wav* signal and also the authenticity of the reciter, it was decided to assign a session-based protocol to each reciter, identified using the username and password followed by a user-verification code sent to the user's email. This was the first layer of security used in this work and was used in order to prevent (internal) attacks/non-authentic access from the client-side. The second layer of security used was to secure all data transmitted on the network from external attacks. As previously mentioned, the RSA algorithm was initially considered and applied to provide the required encryption. However, the RSA encryption scheme was then replaced using the HTTPS protocol, using SSL libraries in order to provide a lightweight security scheme that creates a secure session between the client and the server before proceeding with the audio transmission. The main advantage of this alternative security-mechanism is that it is able to secure the transmitted audio-signal using a lightweight secure-socket layer (SSL) approach that is more suitable for real-time requirements since the RSA-based approach (section B) proved to be too complex when encrypting the whole signal to be transmitted or even when encrypting the watermark-signal only.

IV. Proposed Quran Recitation Recognition Framework

The proposed Quran recitation recognition framework is shown in Figure 7, and comprises of two main parts; the encoder at the sender-side and the decoder at the receiver-side. Biometric fingerprints were simulated using textual bit-streams of the user's name for demonstration of the initial prototype. In reality the text-string would be replaced by a bit-string of the fingerprint applied by the client.

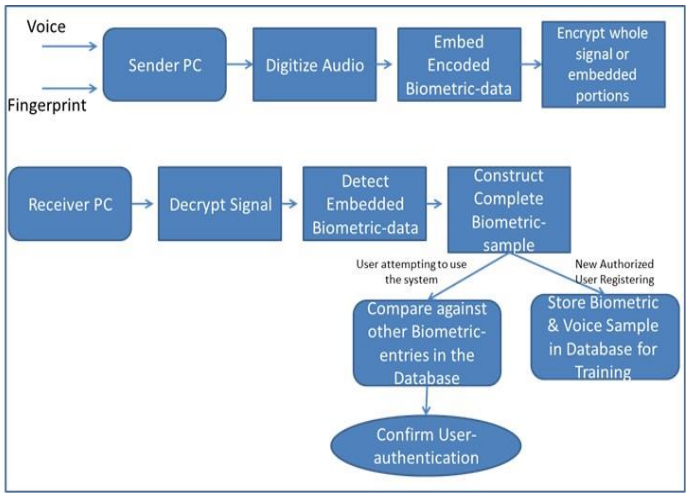

Fig. 7. Framework Diagram of Functional Blocks at the Sender and Receiver

The encoding part is at the sender's side where we have voice input by a given student (reciter) and at the same time we have a database for all students in the course where a fingerprint/watermark is stored for each student/trainer. The fingerprint could be any type of user-specific signal/signature from a biometric database (or a new biometric signature), which is also input into the application. In Figure 7, the input voice-signal is initially in analog format and has active and inactive periods whereby the signal is sampled and quantized into a digital signal. Embedding in done on the central region of the detail-components (e.g. the *CD* component from section III) of the signal, whilst avoiding the most important information-components (e.g. the *CA* component from section III) that contains the main recitation signal components. The process of watermark embedding is illustrated in section V. Following the embedding phase, the Quran-signal would not be altered, since due to its sensitivity requirement, a small change would render the signal as invalid. On the other hand, the experiments showed that the embedded watermark had resulted with low noise effect without altering any

fundamental characters, Tajweed or diacritic pronunciations in the Quran recitation (which are mainly found in the unaltered *CA* components). Additionally, the encoded-signal may undergo security operations as described in section III.B and section III.C, for the case of the RSA and HTTPS/SSL approaches, respectively. This work initially considered and applied the RSA cryptographic algorithm, which was then replaced by the HTTPS approach with SSL in order to avoid the high complexity and long processing delays for the RSA algorithm to encrypt and decrypt the audio-signal. Through our experiments, the SSL approach was found to be significantly better than the RSA algorithm for real-time audio-streaming applications due to the lower complexity when using SSL. Figure 8 summarizes the stages of data-flow at the sender's side prior to transmission.

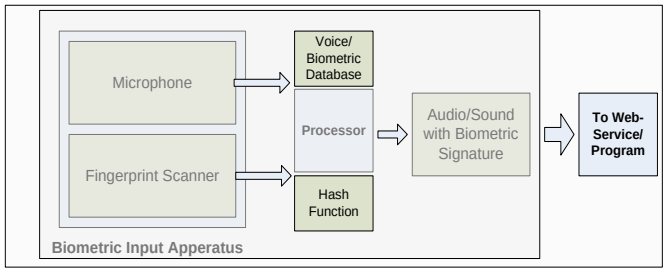

Fig. 8. Detailed Analysis of Data-Flow at the Senders-Side

At the receiver's side, when this information is received, the decoding process is completed by going through the following steps: decrypting the signal (optional; as in the case when RSA was used), identifying the watermark portions in the signal, extracting the watermark to generate the original watermark code, then comparing this watermark to the stored watermark in the database: if there is a match, then the reciter is successfully identified and validated in the system, or otherwise discarded as invalid. In the case of a new reciter, the new signature/watermark is stored, verified by the receiver-institution and usable thereafter. The principle system architecture diagram, combining all functional components at the sender and receiver sides is now summarized in Figure 9.

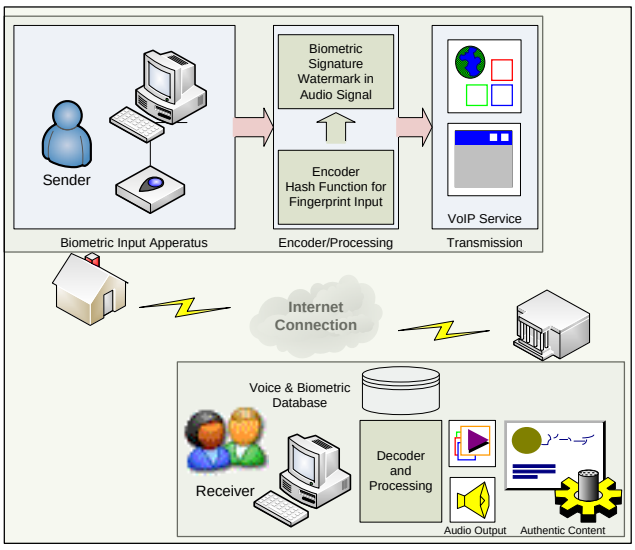

Fig. 9. Principle System Architecture Diagram

V. Analysis and Results

The example of the watermark embedding/extracting processis illustrated as follows:

1) Reading the original wave signal i,Figure 10.

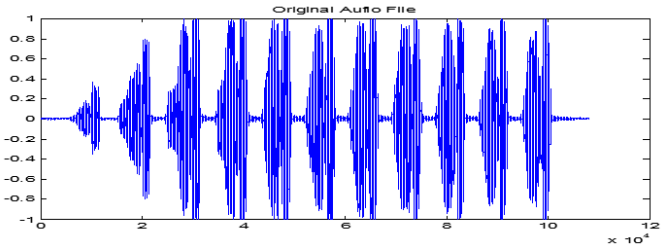

Fig. 10. Original Audio File

2) Calculate the wavelets coefficients of i (CAi and CDi) on one level, Figure 11.

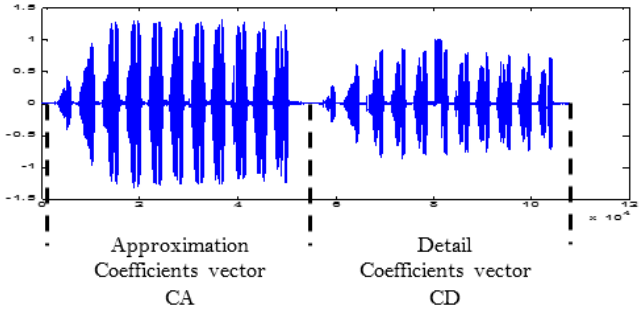

Fig. 11. Calculate the wavelets coefficients of *i* (CA*i* and CD*i*) on one level

3) Embedding the normalized watermark in CD coefficients of the signal i by the watermark w in the center of CDi, Figure 12.

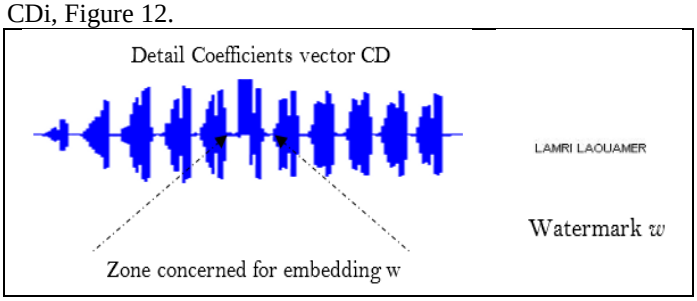

Fig. 12. Embedding the normalized watermark in CD coefficients

4) Obtaining the watermarked signal iw, Figure 13.

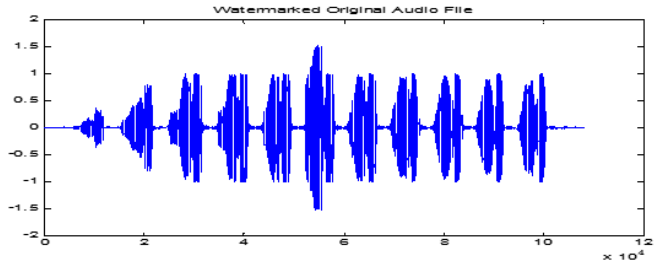

Fig. 13. Finding the watermarked Signal,*iw*

For demonstration purposes, the RSA-approach will be illustrated when being applied on the watermark. Here, the

security will be checked and the embedding/extracting processes will concern the encrypted watermark and not the watermark itself. An example of encrypting and decrypting the normalized watermark is shown in Figure 14.

Fig. 14. Illustration of RSA-based Encryption and Decryption on the Watermarked-Signal

A website was developed which consists of two-tier architecture as follows:-

1) The client-end can access the application through any web-browser (e.g. Internet Explorer, Mozilla, Chrome, etc. ....).

2) The server-end is where the whole application is accessed.

- PHP programming language was used for developing basic services in the developed application, such as the uploading function.
- HTTPS protocol was used for securing all communications between the client-side and server-side.
- The operations of the watermark embedding, extraction and the checking of possible attacks on the watermarked signal were achieved using three Matlab programs.

These three programs have been used as executable files within our PHP web page. These programs are the core of this proposed audio prototype. The complexity of our scheme is polynomial, which makes our approach realizable with acceptable running-time performance. It remains to note that the checking tasks have worked as required with an accuracy rate of 100%.

The client/server prototype provides the user with the choice either focusing only on robustness (watermark embedding and extracting) or either to introduce the security aspect (encrypted watermark embedding and extracting) as an additional requirement during communications. The prototype essentially requires download and installation on a server-machine, which can be accessed for transmitting audio from the client side and receiving audio at the instructor-side. Hence, the sender and receiver both access the same application/interface; however, have different uses and access-privileges. Figure15 illustrates a snapshot of the application.Steps to use the application include: first, the client records the recitation, then the reciter embeds the recitation, following this the reciter submits the recorded recitation and sends the results.Finally, the audio-signal is encrypted and watermarked ready for transmission to the instructor/receiver-side.

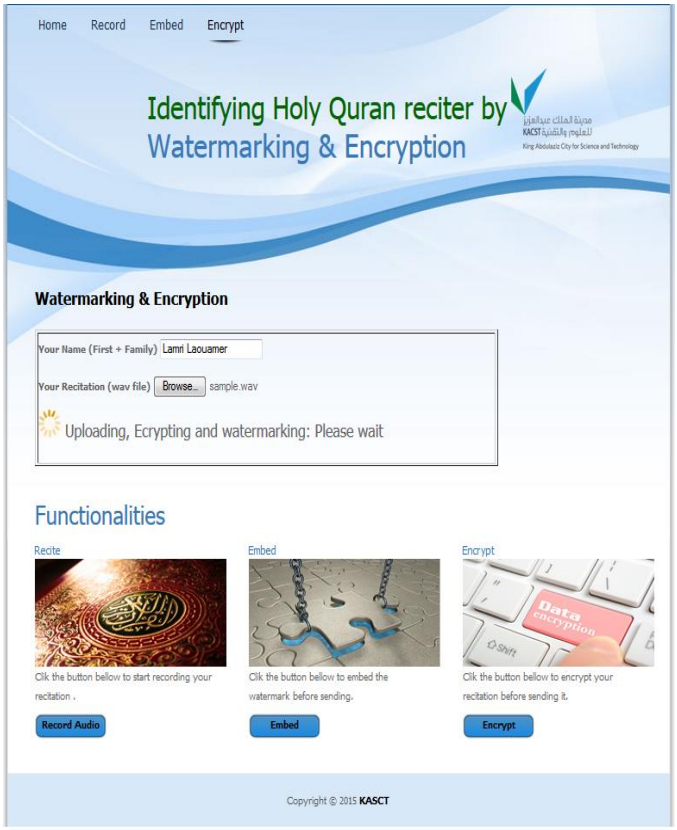

Fig. 15. Audio signal – Watermarking and Encryption

The receiver side operation proceeds with login as an admin i.e.Quran-instructor/evaluator, who is presented with the updated recitations received.

VI. Comparison with the State of the Art Audio Watermarking Literature

The scheme presented in this approach provides a number of practical advantagesand enhancements compared to some other existing schemes and can be summarized as follows:

- First, the proposed approachonly requires embedding little information(e.g. only a few data-bits to hide),thereby offering good capacityin terms of complexity representing therecite data,and is adaptable for the identification processas required in real time applications.

- Second, the *detailed* coefficients vector in wavelet transformwere exploitedduring the watermarking process, since it had resulted with no considerable alteration to the quality of the audible sound.Hence, one further enhancement to existing approaches in the literature was that all embedded data was completely encoded into the *detailed* coefficients vectors. In contrast,modifying the *approximation*coefficients vector in the wavelet transform(as done in other schemes) had a directly impact on the signal quality.

- Third, the proposed approachwas found to bemore robust against various attacks applied as compared to other related works in [23] and [28].

- A secure process is achieved in the proposed scheme thatconfirms the identification of the reciter and detects any false/unauthenticreciters.

- Finally, it is worth noting that the use of the improved audio-watermarking scheme proposed here for application in highly-sensitive Quran voice-signals (and thus the constraints consequently imposed on the embedding scheme) is itself novel and not found anywhere in the related literature.

Many works found in the audio-watermarking literature have focused onrobustness and inaudibility/capacity performance for various approaches using either the Stirmark Benchmark attacks or self-simulated attack scenarios by simulation after varying a number of influential parameters. In this study, we present our results in comparison with two other closely-related studies after matching the attack-scenarios, types and parameter-values used in the other studies (Table 2). The results obtained from our approach were found to be better than the results presented in [23] in terms of the normalized correlation coefficients (NCs), since all used NC-values were closer to 1 as compared with [23]. Furthermore, the proposed approach provides highly significant results in terms of Bit Error Rate (BER) values, which are close to 0 inalmost all attack scenarios. This suggests that the proposed approach is more robust against the attacks considered in the tests when compared to the results obtained in [28] (Table 2).

TABLE II. COMPARATIVE RESULTS AGAINST RELATED WORK UNDER SIMILAR SIMULATION PARAMETERS

Normalized Correlation Coefficient NC		
	Our approach	Proposed approach in [23]
Resamp22050	0.9832	0.9401
Re-quan	0.9956	0.9934
AWGN	0.9607	0.885
Low pass	0.9813	0.9318
MP3- 48kb	0.972	0.9279
MP3-64kb	0.983	0.9363
MP3-96	-	-
MP3-128kb	-	-
Bit Error Ratio BER		
	Our approach	Proposed approach in [28]
Resamp22050	0	0
Re-quan	0.56	1.25
AWGN	0.82	5.6
Low pass	0.93	7
MP3- 48kb	0.75	17
MP3-64kb	0.3	8
MP3-96	0.1999	3.7
MP3-128kb	0.171	3.7

VII. CONCLUSION

The authentication scheme presented in this work provided a robust, secure and practical approach in terms of achieving low complexity as required for ensuring real-time authenticity of sensitive-speech data. Arabic Quran recitations were taken as a case study during experimentations due to the sensitive nature of the recitation, which had resulted with additional complexities to overcome, in contrast to ordinary speech-data mainly addressed in the literature. The main novelty in this work was found in our application and enhancement of existing audio-watermarking techniques under the constraints of the sensitive voice data, which should not be altered, with the further requirement that any embedded data remain inaudible. Themechanism executes a number of functional stages at the sender and receiver sides and avoids distortion of the intelligible and audible Quran input-signal, and had therefore successfully addressed the sensitivities of the digital-Quran audio-signal. Following experiments with several protocol variations, our final solution had employed the DWTthrough an HTTPS protocol in order to achieve reduced complexity and online authentication in real-time. Notably, our contribution compared very well with the other related approaches in the literature and had provided enhanced results for our key metrics of interest that had included robustness, inaudibility and capacity, enabling us to achieve real-time authentication.

The Quran recitation recognition framework and prototype produced in this work facilitates Quran Learning Institutions to authenticate the student-identity/reciter over an unreliable network, such as the Internet in cases where remote/distance-learning is required. The work in this paper is also very useful for student-evaluation purposes in a distance-learning environment, particularly where certificates are issued by an institution following student/client-verification. The prototype was successfully tested and was able to confirm authentic and non-authentic client identities. Finally, such a system could also be employed for more general verification purposes or other similar online-based learning centers/institutions requiring user voice-authentication before issuing academic or other certificates.

ACKNOWLEDGMENT

The researchers would like to thank and acknowledge the King Abdulaziz City for Science and Technology (KACST) for their financial support for this research work during the academic year 2014/2015 under research grant no. “28-34”.

REFERENCES

[1] Singh, Prabhishek, and R. S. Chadha. "A Survey of Digital Watermarking Techniques, Applications and Attacks." International Journal of Engineering and Innovative Technology (IJEIT) 2, no. 9 (2013).

[2] Tayan, O.; Alginahi, Y.M., "A review of recent advances on multimedia watermarking security and design implications for digital Quran

computing," Biometrics and Security Technologies (ISBAST), 2014 International Symposium on , vol., no., pp.304,309, 26-27 Aug. 2014

[3] Bender, Walter, et al. "Techniques for data hiding." IBM systems journal 35.3.4 (1996): 313-336.

[4] Yeo, In-Kwon, and HyoungJoong Kim. "Modified patchwork algorithm: A novel audio watermarking scheme." Speech and Audio Processing, IEEE Transactions on 11, no. 4 (2003): 381-386.

[5] Cvejic, Nedeljko, and TapioSeppanen. "Robust audio watermarking in wavelet domain using frequency hopping and patchwork method." In Image and Signal Processing and Analysis, 2003. ISPA 2003. Proceedings of the 3rd International Symposium on, vol. 1, pp. 251-255. IEEE, 2003.

[6] HtayHtay Yee; Foo Say Wei, "Audio watermarking with error-correcting code," TENCON 2009 - 2009 IEEE Region 10 Conference, vol., no., pp.1,5, 23-26 Jan. 2009

[7] Qiuyu Zhang; Jiabin Deng; Zhanting Yuan, "Implementation of Real-Time Audio Watermarking Based on DSP," Intelligent Information Technology Application Workshops, 2009. IITAW '09. Third International Symposium on , vol., no., pp.145,148, 21-22 Nov. 2009

[8] Liting Gao; Wei Zhao; Xiumei Wen; Lixia Wang, "An audio zero-watermarking algorithm based on FFT," Networking and Digital Society (ICNDS), 2010 2nd International Conference on , vol.1, no., pp.274,277, 30-31 May 2010.

[9] Huang Xiong-hua; Jiang Wei-zhen, An adaptive fragile audio watermarking based on MDFT, International Conference on Intelligent Computing and Integrated Systems (ICISS), 2010.

[10] Furon, Teddy, and Pierre Duhamel. "An asymmetric watermarking method."Signal Processing, IEEE Transactions on 51, no. 4 (2003): 981-995.

[11] Tavakoli, E.; Vahdat, B.V.; Shamsollahi, M.B.; Sameni, R., Audio Watermarking for Covert Communication through Telephone System, IEEE International Symposium on Signal Processing and Information Technology, 2006.

[12] Ning Chen; Meng-yao Zhu; Sen Liu, "A new fragile audio watermarking scheme," Audio Language and Image Processing (ICALIP), 2010 International Conference on , vol., no., pp.367,372, 23-25 Nov. 2010

[13] M. K. Dutta, P. Gupta, V.K. Pathak, "Biometric Based Watermarking in Audio Signals", proceedings of the International Conference on Multimedia Information Networking and Security, 2009.

[14] N. Chen, M-Y. Zhu, S. Liu, "A New Fragile Audio Watermarking Scheme", International Conference on Audio Language and Image Processing, China, 2010

[15] H.Zhao, D. Shen, "An Audio Watermarking Algorithm for Audio Authentication", 2010.

[16] M. A. Nematollahi, S.A.R. Al-Haddad, S. Doraisamy, M. I. Bin Saripan, "Digital Audio and Speech Watermarking Based on the Multiple Discrete Wavelets Transform and Singular Value Decomposition", proceedings of the Sixth Asia Modelling Symposium, 2012.

[17] N. Baranwal, K. Datta, "Comparative Study of Spread Spectrum Based Audio Watermarking Techniques", IEEE-International Conference on Recent Trends in Inforrmation Technology, Chennai, India, 3-5 June, 2011.

[18] F. Citak Er, E. Gul, "Comparison of Digital Audio Watermarking Techniques for the Security of VOIPCommunications", 2011.

[19] Xiangui Kang; Rui Yang; Jiwu Huang, "Geometric Invariant Audio Watermarking Based on an LCM Feature," *Multimedia, IEEE Transactions on* , vol.13, no.2, pp.181,190, April 2011

[20] Chen, S.-T.; Wu, G.-D.; Huang, H.-N., "Wavelet-domain audio watermarking scheme using optimisation-based quantisation," *Signal Processing, IET* , vol.4, no.6, pp.720,727, Dec. 2010

[21] Xiang-Yang Wang; Hong Zhao, A Novel Synchronization Invariant Audio Watermarking Scheme Based on DWT and DCT, IEEE Transactions on Signal Processing, Volume: 54 , Issue: 12, Page(s): 4835 – 484,0 2006.

[22] Petrovic, R.; Yang, D.T., Audio watermarking in compressed domain, 9th International Conference on Telecommunication in Modern Satellite, Cable, and Broadcasting Services, 2009. TELSIKS '09.

[23] Hong Zhao; Dongsheng Shen, An audio watermarking algorithm for audio authentication, IEEE International Conference on Information Theory and Information Security (ICITIS),pages: 807 – 809, Beijing, China, 2010.

[24] Al-Haj, A.; Twal, C.; Mohammad, A., Hybrid DWT-SVD audio watermarking, Fifth International Conference on Digital Information Management (ICDIM), 2010.

[25] Sharma, Shweta, JitendraRajpurohit, and Sunil Dhankar. "Survey on different level of audio watermarking techniques." *International Journal of Computer Applications* 49, no. 10 (2012): 41-48.

[26] Komal, Ms, V. Goenka, MsPallavi, and K. Patil. "Overview of Audio Watermarking Techniques.", International Journal of Emerging Technology and Advanced Engineering, Vol. 2, Issue 2, 2012.

[27] Olanrewaju, R.F.; Khalifa, O., Digital audio watermarking; techniques and applications, International Conference on Computer and Communication Engineering (ICCCE), 2012.

[28] Mahmood Movassagh, Ali AsgharBeheshtiShirazi, "A Novel Method For Real-Time Audio Watermarking Using Wavelet Transform", Canadian Conference on Electrical and Computer Engineering, CCECE 2008, pages: 83-88, Niagara Falls, ON, Canada, 2008.