Multimed Tools Appl (2013) 64:649–669
DOI 10.1007/s11042-011-0969-9

Partial encryption and watermarking scheme for audio files with controlled degradation of quality

Kamalika Datta · Indranil Sen Gupta

Published online: 3 January 2012

Abstract With the rapid progress in communication and multimedia technology, protection of multimedia assets has become a major concern. *Encryption* and *watermarking* are two complementary techniques that are used for safeguarding multimedia data like audio, video and images. These methods serve two different purposes; encryption helps in making the media unintelligible, while watermarking helps in providing copyright information in it. In this paper a combination of both encryption and watermarking is incorporated on audio files with this aim in view. Discrete Wavelet Transformation (DWT) is performed on audio files up to third level which gives rise to a binary tree like structure. The areas for watermark embedding and encryption are selected among the wavelet coefficients in the leaf nodes of the tree. Detailed experimentation is carried out with analysis of *SNR* values, that leads to the determination of degree of degradation of the audio quality after encryption. Such controlled degradation can be used for safe distribution of audio contents over public networks, whereby only the authorized users can have access to the high quality contents, while other users can only access a lower quality version.

Keywords Perceptual encryption · Watermarking · Discrete wavelet transformation · Commutative watermarking and encryption (CWE)

K. Datta (✉)
Department of Information Technology, Bengal Engineering and Science University,
Shibpur, Howrah 711103, India
e-mail: kdatta.iitkgp@gmail.com

I. Sen Gupta
Department of Computer Science & Engineering,
Indian Institute of Technology Kharagpur,
Kharagpur 721301, India
e-mail: isg@iitkgp.ac.in

1 Introduction

With the fast proliferation of Internet and the underlying applications, handling multimedia traffic and services have gained in importance over the last couple of decades. In particular, safe distribution and management of multimedia contents have become the real challenge. This is required to protect piracy and copyright issues in distributing multimedia contents, particularly audio and video. In this paper a technique for partial encryption and digital watermarking of audio files has been presented, which can be very effective to address the problem as mentioned. A typical scenario is an on-line music portal, which sells and distributes audio files to customers over the Internet. To protect against piracy, users have to register to the site through proper authorization mechanism. While registered users can gain access to high-quality audio contents, unregistered users can only have access to lower-quality preview. The method uses a combination of partial encryption, perceptual encryption, and commutative watermarking and encryption [1, 13] on audio files.

Partial Encryption [13] is a process that encrypts only a selected part of the multimedia contents, while leaving other parts unchanged. The decryption process is similar, where only the encrypted portion is decrypted to get back the original contents. Since the whole file is not encrypted, the encryption process is faster and can lead to efficient on-line implementation. Figure 1 captures the basic idea behind this concept.

Perceptual Encryption [13] is a type of controlled encryption process which degrades the quality of a multimedia file depending upon the requirements. By providing encryption with various levels of degradation, media contents with various quality levels can be generated, like preview mode, medium quality mode, high quality mode, etc. To carry out this kind of encryption, the sensitive parameters of the media file have to be extracted out, and some of them encrypted, as illustrated in Fig. 2. For example, in an audio file, we can perform discrete wavelet transformation (DWT), and select some subset of the wavelet coefficients for encryption.

In Commutative Watermarking and Encryption (CWE) [13, 23, 24], both encryption and watermarking are applied on disjoint parts of the multimedia contents. Encryption is used to protect the confidentiality, while watermarking [7] can be used for key distribution and protection of media copyright. The encryption and watermarking steps are normally applied one after the other in a specified sequence, both at the sender side and the receiver side. In CWE, since the parts are disjoint, the steps can be applied in any order (i.e. commutative). Thus, encryption on the watermarked version, or watermarking on the encrypted version, will both generate

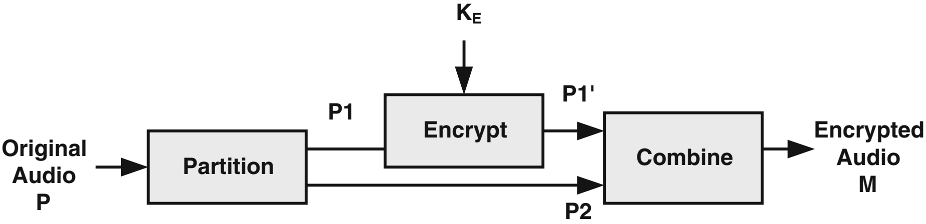


Fig. 1 Partial encryption

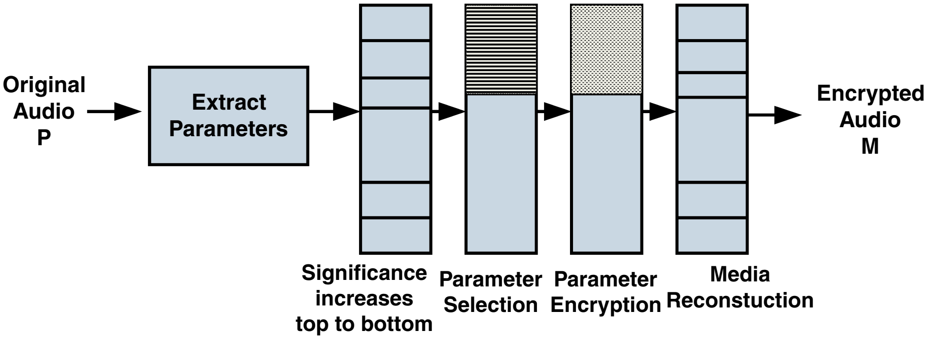


Fig. 2 Perceptual encryption

very similar final media content, as illustrated in Fig. 3. The main advantage of having a commutative system [3] is flexibility, where the same watermarked version can be variously encrypted depending on varying requirements, or the same encrypted version can be provided with a different watermark if required. This saves computation time as compared to carrying out both watermarking and encryption every time.

In this paper, we have proposed an audio encryption and watermarking [17] scheme that is commutative in nature. The DWT coefficients of a given audio file are partitioned, and encryption and watermarking are performed on disjoint partitions.

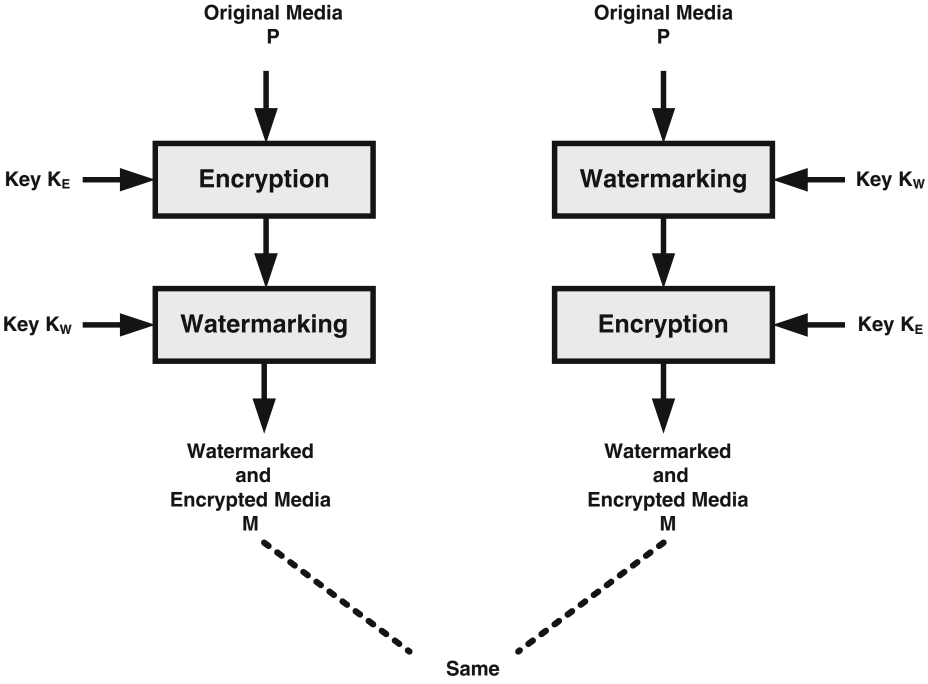


Fig. 3 Commutative watermarking and encryption

Moreover, the quality of the watermarked audio can be controlled by a user specified parameter. This feature helps in controlling distribution of media (audio) files over a public network.

The rest of the paper is organized as follows. A brief review of similar existing works is presented in Section 2, followed by the details of the proposed scheme in Section 3 along with experimental results. Section 4 gives some comparisons of the results with existing works, and finally Section 5 provides with concluding remarks.

2 Review of the existing works

There exist applications of digital watermarking where copyright protection is not the major concern. Rather, issues like content authentication, secure distribution, etc. assume greater importance. There exists a set of methods in the literature which combine conventional watermarking with encryption to achieve this objective. Some of these methods are reviewed in this section. Although quite a few research works have been reported on this field, there is a scope for improvement as far as the security and flexibility of the schemes are concerned.

Juan et al. [9] proposed a perception based scalable encryption model for Audio Video Coding Standard (AVS). Depending on the type of applications available, scalable encryption technique provides different protection levels to these applications. All audio bits are not equally important, and so a perceptual classification of the bit streams is incorporated in this paper. The main design goal of this scheme is to account for the degree of degradation of the audio content. The security of this scheme, however, depends on the encryption algorithm used.

Servetti et al. [19] proposed a low complexity perceptual based partial encryption scheme. This scheme provides very good content protection. Here instead of encrypting the whole multimedia data, only a selected fractional part is encrypted which intuitively reduces the processing load. This scheme is applied to clean speech taken from NTT Multilingual Speech Database. The bitstream is partitioned as per their scheme and is then encrypted. Two partial encryption techniques are reported in the paper, a low protection scheme, which helps in protecting most kinds of eavesdropping, and a high protection scheme, based on encrypting perceptually important bits.

Servetti et al. [20] proposed a low-complexity scheme based on partial encryption for content protection of MP3 audio. The main motivation of this work is to provide the users with degradable quality of audio which can be improved to original quality by attaining a key. Decryption process is applied only to a selected number of bits (1–10% of the total bitstream). Here from the MP3 audio file, MDCT coefficients are selected and divided into several frequency regions, and these spectral subdivisions are mainly exploited to degrade the perceptual quality of the compressed audio by using low-pass filtering. For introducing annoying artifacts to the compressed audio, limiting frequency content is an effective way which is used in this technique. Moreover in this technique the cut-off frequency is modified by increasing or decreasing the number of coefficients to get the desired degree of perceptual quality. The result shows that low-pass filtering at 5.5 kHz preserves audio contents. But this paper lacks some formal tests, which should be performed for overall effectiveness of the proposed scheme. Also in this paper there is no mention of the set of MP3 audio files taken for experimentation.

Lemma et al. [11] proposed a secure embedding scheme that incorporates traditional watermarking and partial encryption. Two new techniques are proposed; one for MASK watermarking on baseband audio, and the other for spread spectrum watermarking on MPEG-2 encoded video streams. In the first technique MASK watermarking system is used. Here the embedding process is performed by modifying the envelope of the host signal. Encryption is performed by modulating the signal with a piece wise stationary random sequence so that the encrypted audio is perceptually degraded. For each individual clients, server generates the watermark information. Firstly encryption is performed on the audio signal and then the generated watermark is embedded to the encrypted audio. In the second method a simple additive spread spectrum watermarking scheme on MPEG-2 compressed video stream is presented. The security issue of the partial encryption method used is not addressed here, only efficiency of the secure watermark embedding process is analyzed.

In [2] a protocol is proposed which uses cryptographic techniques to address the piracy issue. Here commutative encryption technique is used to protect the piracy of the watermarked data. Here the fingerprint to be embedded is determined by the content provider and the customer. In this case it will be fairly judged by the public authority as who is guilty in the unauthorized distribution. So this scheme is found to be secure against any attack from content provider or customer. Although this method proposes a very elegant technique for secure distribution but there is very little mention about the security of the process.

In [14] a commutative watermarking and encryption technique is used for MPEG2 video. Usually in different scenarios, watermarking and encryption is performed on different parts of the media files, but this has got the disadvantage that it cannot protect against replacement attacks. So to overcome from this difficulty here watermarking and encryption are performed on same part of media data. Although replacement attack is very important, but it is more relevant for images [6]. Different researches have been performed on still images.

Lian et al. [16] proposed a commutative watermarking and encryption scheme for image files based on frequency characteristics of wavelet codec. They analyzed the variations of PSNR values with changes in the frequency bands used for encryption.

Though a number of works have tried to combine encryption with watermarking to achieve some degree of degradation in a given audio file, the process is not continuous in terms of the amount of degradation possible. There is, therefore, a need to device a scheme where based on the value of some user specified parameter, any arbitrary level of degradation in quality can be achieved.

3 The proposed scheme

Media encryption and media watermarking are two different techniques, which can be coalesced together to protect both confidentiality and identity. The proposed approach deals with a commutative watermarking and encryption (CWE) scheme based on partial audio encryption, which provides controlled level of degradation in the quality of the audio files. In order to have the commutative property, the discrete wavelet transform coefficients of the given audio file are partitioned, and watermarking and encryption are applied on disjoint partitions. The basic features of the scheme are discussed in the following subsections.

3.1 Partial (perceptual) encryption

This subsection provides an insight into the concept of partial encryption as used in the context of the proposed scheme. The original media file P is initially partitioned into two parts, $P1$ and $P2$. The part $P2$ is significant to perception and a change in the coefficients of this part renders the audio file unintelligible, whereas the human auditory senses are not sensitive to the part $P1$. The part $P1$ is encrypted using a key K_E to form the perception-sensitive part $P1'$. Then the encrypted part $P1'$ is recombined with the untouched part $P2$ to form the encrypted media M.

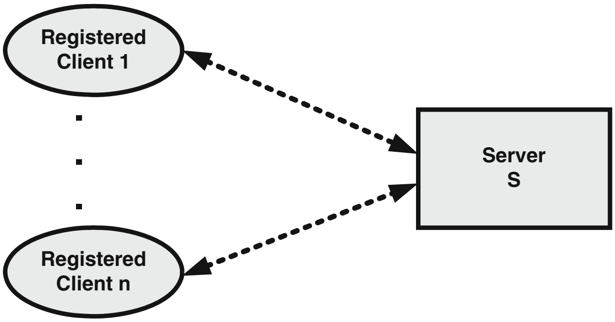
$$\begin{aligned}(P1, P2) &= Partition(P)\\ P1' &= Encrypt(P1, K_E)\\ M &= Combine(P1', P2)\end{aligned}$$

Actually, the partitioning is carried out in the DWT domain, and not in the time domain. Hence this method may also be regarded as a perceptual encryption scheme. As explained in the subsequent subsections, encryption is carried out on some of the higher frequency DWT coefficients, using a scheme explained in the following subsection.

3.2 The key distribution and encryption framework

The proposed requirements and the suggested solution for distributing the encryption key and partial encryption of the media file can be explained with respect to the general model as shown in Fig. 4. The server S allows users (clients) to register themselves, and have access to a database of media files hosted by the server. Unregistered users will have access to partially encrypted versions of media files which will result in degraded quality of playback, whereas registered users will be provided with a decryption key using which the high-quality version can be reconstructed. In the proposed work, the key is distributed along with the media using watermarking; however, it may also be sent over the secure channel that exists

Fig. 4 The client-server environment



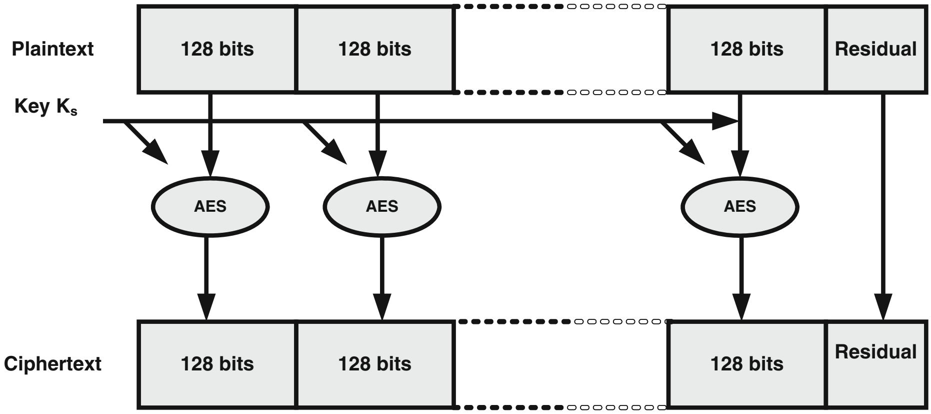


Fig. 5 AES encryption scheme

between a user and the server. The process of key distribution, and the partial encryption of the media files are explained below.

When clients register with the server S through a secure channel, they are provided with a (public key, private key) pair for being used with some standard public-key system like RSA [18]. For a client m, the public and private keys are denoted as KU_{cm} and KR_{cm} respectively.

A media file will be partially encrypted by the server using a randomly generated symmetric key K_s, using AES algorithm [4] as shown in Fig. 5. The block of data to be encrypted (referred to as plaintext in the figure) is divided into 128-bit sub-blocks, and each sub-block is encrypted using AES with the key K_s to obtain the ciphertext block. If the last sub-block is not a multiple of 128, it is not encrypted.

The server S distributes the encryption key K_s to a registered client m by encrypting the key using the public-key KU_{cm} of the client, and watermarking the encrypted key in the media file. This process is illustrated in Fig. 6. It may be noted that any other method that does not rely on watermarking may also be used for distributing the key. However, since in the context of the present work we are not considering signal processing attacks that can disturb the watermark, we have chosen to send the encrypted key watermarked along with the audio file.

3.3 DWT decomposition tree of coefficients

It is known that DWT transforms an audio signal at any level into approximate and detail coefficients. The approximate coefficients refer to the low frequency components, which the human auditory senses are sensitive to. The detail coefficients are the representation of high frequency components, which largely go undetected by human auditory senses.

In the first level of DWT transformation, the audio samples are decomposed into the low frequency components (approximate coefficients A) and the high frequency components (detailed coefficients D). With the DWT transformation to the second level, the approximate and detail coefficients again undergo transformation to form the four coefficients of AA, AD, DA, and DD. As we proceed down the levels, the

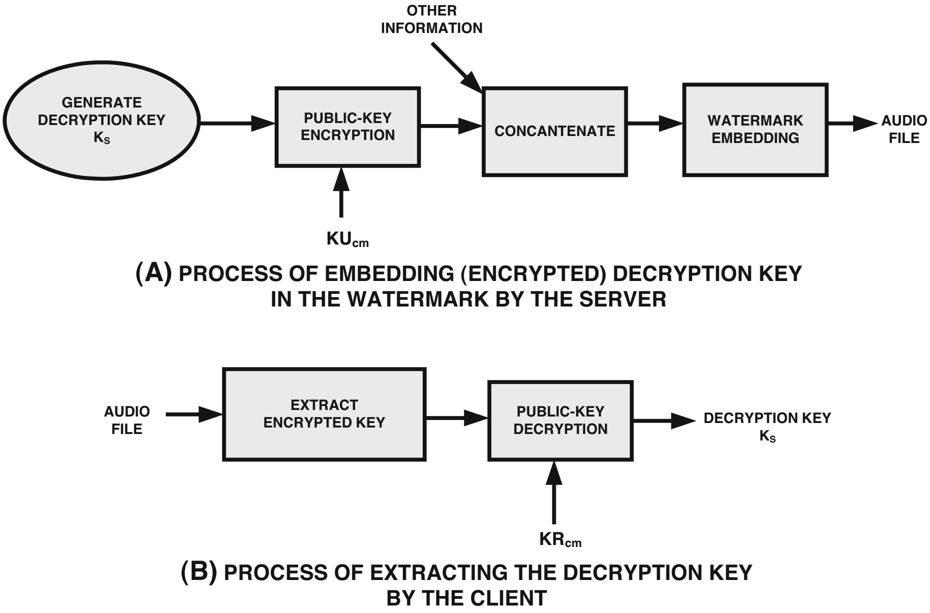

OTHER INFORMATION
GENERATE DECRYPTION KEY K_S
PUBLIC-KEY ENCRYPTION
CONCANTENATE
WATERMARK EMBEDDING
AUDIO FILE
KU_{cm}
(A) PROCESS OF EMBEDDING (ENCRYPTED) DECRYPTION KEY IN THE WATERMARK BY THE SERVER
AUDIO FILE
EXTRACT ENCRYPTED KEY
PUBLIC-KEY DECRYPTION
DECRYPTION KEY K_S
KR_{cm}
(B) PROCESS OF EXTRACTING THE DECRYPTION KEY BY THE CLIENT

Fig. 6 Key distribution process

audio file gets further transformed, giving rise to a binary tree like structure. At a level n, the total number of leaf nodes which represent the approximate and detail coefficients of the binary tree structure is 2^n. Figure 7 depicts the decomposition tree for $n = 3$.

Inverse discrete wavelet transform (IDWT) is used to recombine the decomposed low and high frequency coefficients to get back the audio file. In the proposed schemes, the given audio undergoes DWT up to three levels. For ease of explanation, we will use the following alternate notations in some of the following subsections: $x1$ (instead of AAA), $x2$ (instead of AAD), $x3$ (instead of ADA), $x4$ (instead of ADD), $x5$ (instead of DAA), $x6$ (instead of DAD), $x7$ (instead of DDA), $x8$ (instead of DDD).

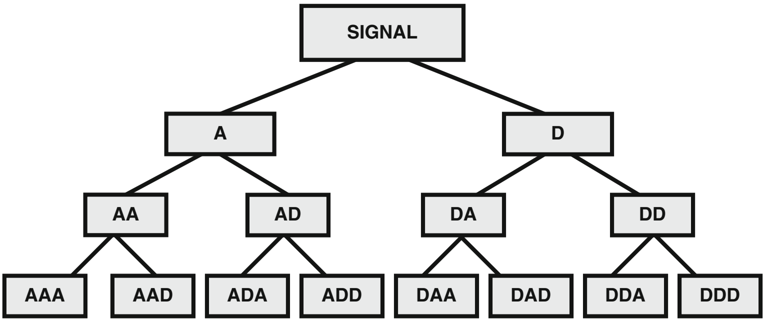


Fig. 7 DWT of an audio signal up to 3 levels

3.4 Proposed scheme 1

After decomposition of the original audio file up to three levels using DWT, parameter extraction is performed. The eight leaf nodes, $x1$ to $x8$, correspond to transform coefficients, arranged in order of increasing frequency. In the first scheme that is being proposed, coefficients in the lowest frequency components ($x1$) are used for watermarking, while all the coefficients in one of the frequency blocks $x2$ to $x8$ are encrypted. Through experimentation, the impact of the frequency block used for encryption on the quality of the resulting audio signal is evaluated in terms of the SNR values of the encrypted media files. The algorithms for watermark embedding and encryption, and watermark detection and decryption are given below. In Algorithm 1, the concatenation operation (denoted by ||) is carried out considering all its operands as bit strings. It may be noted that during the process of watermarking in $x1$, the audio signal will undergo some degradation which, however, is small as compared to that resulting due to encryption. In terms of SNR values, this degradation has been found to be within 5% for all the audio files that have been experimented with.

Algorithm 1 *Scheme1_Embed*

```
/* Embed watermark $w$ into *file*, and encrypt the *i*th third-level DWT
   coefficient block
   $x_i$ using key $k$, to generate *newfile*
   Input parameters: *file*, $i$, $w$, $k$
   Output parameter: *newfile* */
begin
   $x_1 \ldots x_8 =$ DWT of *file* upto 3 levels;
   $w = w||i||p\_encrypt(k, KU_{cm})$;
   $embed(x_1, w)$;
   $encrypt(x_i, k)$;
end
```

Algorithm 2 *Scheme1_Extract*

```
/* Extract watermark from *newfile*, and decrypt the encrypted portion
   to generate *file*
   Input parameter: *newfile*
   Output parameter: *file* */
begin
   $x_1 \ldots x_8 =$ DWT of *newfile* upto 3 levels;
   $new\_w = extract\ (x_1)$;
   $(w', i', k') = new\_w$;
   $k" = p\_decrypt\ (k', KR_{cm})$;
   $decrypt\ (x_{i'}, k")$;
   $file =$ IDWT of $x_1 \ldots x_8$;
end
```

3.4.1 Watermarking process

As mentioned above, the lowest frequency coefficients present in $x1$ are used for embedding the watermark. The watermarking scheme employs a mean quantization technique [10, 25], wherein the low frequency coefficients in $x1$ are used for embedding. In the process of embedding, coefficients in $x1$ are first divided into frames (number of frames being equal to the number of watermark bits to be embedded), the frame mean subtracted from each of the coefficient values, and an offset added or subtracted depending on the watermark bit (0 or 1). This watermarking scheme has been shown to be robust against many signal processing attacks like MP3 compression, quantization, etc. However, it may be noted that in the context of the present work, robustness in the watermarking process [22] is not a very important consideration.

Detail of the embedding process is depicted in Fig. 8. The following are the steps performed during the embedding process.

Step 1: At first DWT is performed up to three levels and we obtain AAA and AAD.

$$[A, D] = DWT(s, wavelet) \tag{1}$$

Here s is the audio sample values and *wavelet* defines the mother wavelet transformation for performing the analysis (*Haar* or *Daubechies*). Here A and D are first level approximate and detail coefficients. To obtain the third level approximate coefficients, we have to move down to third level by further decomposing approximate coefficients.

Step 2: The approximate coefficients of the third level (AAA) are divided into frames of fixed size $\lfloor n/m \rfloor$, and denoted as $f_1, f_2, \ldots, f_m$. Here n is the total number of coefficients in AAA, and m is the length of the watermark in bits.

Step 3: After the framing process is done, the means of all the frames are calculated.

$$m_i = Mean(f_i), \quad \text{where } i = 1, 2, ..., m \tag{2}$$

The calculated means of each frames are then subtracted from all the coefficient values of that frame.

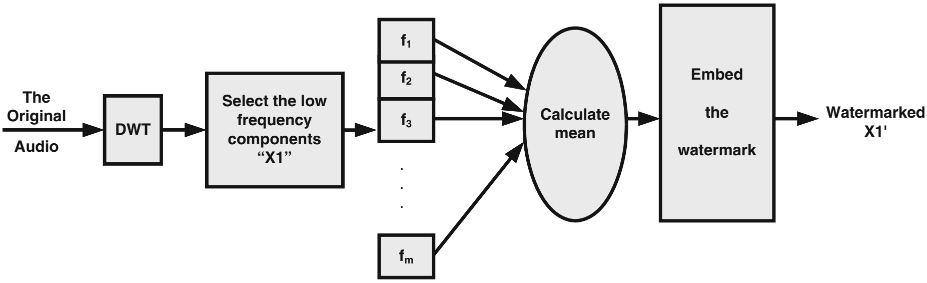


Fig. 8 Watermark embedding process

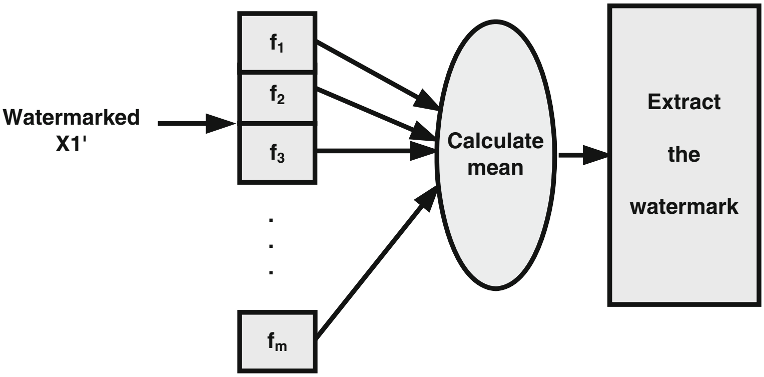


Fig. 9 Watermark extraction process

Step 4: The watermark bit W_i is embedded into the frame f_i as follows:

$$\begin{aligned} f'_{ij} &= f_{ij} + \alpha m_i, \text{ if } W_i = 1 \\ &= f_{ij} - \alpha m_i, \text{ if } W_i = 0 \end{aligned} \tag{3}$$

where f_{ij} denotes the jth coefficient of frame f_i, and α is a constant called embedding intensity.

Detail of the extraction process is depicted in Fig. 9. The following are the steps performed during the extraction.

Step 1 Here in the same way like the embedding process, DWT is performed and we obtain third level approximate coefficients.

$$[A', D'] = DWT(s', wavelet) \tag{4}$$

In a similar way we move down to third level and obtain AAA.

Step 2 Thereafter framing is performed and the mean is calculated for each frame.

$$m'_i = Mean(f_i), \text{ where } i = 1, 2, ..., m \tag{5}$$

Step 3 The ith watermark bit W_i is extracted using the following formula.

$$\begin{aligned} W_i &= 1, \text{ if } m'_i > 0 \\ &= 0, \text{ if } m'_i < 0 \end{aligned} \tag{6}$$

The watermarking process discussed here uses simple mean quantization technique. Various mean quantization techniques [5, 10, 12] are there in literature. It may be recalled that we do not need robustness as a necessary condition for successful implementation of our scheme, and hence the choice of the watermarking algorithm is not very critical.

3.4.2 Integration of encryption and watermarking

As mentioned, the encryption and watermarking operations are performed on the coefficients in the independent leaf nodes (among $x1$, $x2$, ..., $x8$). The commutation of the watermarking and encryption process is ensured due to the mutually independent sequencing of these processes. Since the choice of the leaf nodes (or the sections

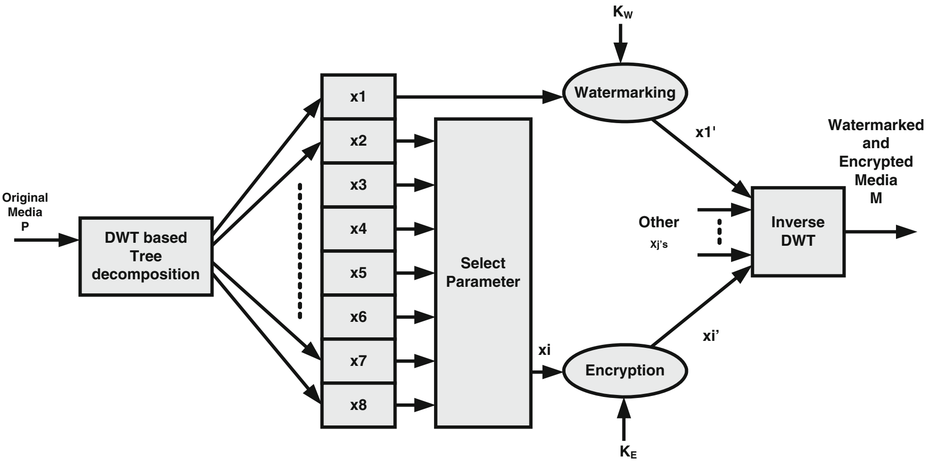


Fig. 10 Combined watermarking and encryption for proposed scheme 1

of the coefficients) for encryption and watermarking are disjoint, the processes can be implemented in any sequence. The original audio file whether encrypted first and watermarked second, or watermarked first and encrypted second generates the same media file. In this process, embedding of watermark into the encrypted audio file is independent of the knowledge of the decryption key which helps in controlled access and safe distribution of audio content. The process is depicted in Fig. 10.

3.4.3 Experimental results

In the implementation, we have used mean quantization method on the low frequency DWT coefficients for watermarking, and AES algorithm on some subset of higher frequency DWT coefficients for encryption. Experiments have been carried out using MATLAB on 100 audio files, by watermarking on $x1$ and encrypting one of the leaf nodes $x2$ through $x8$.

Representative results for four files are shown in Table 1, which depicts the variations in SNR values after encrypting individual leaf nodes at third level.

It may be observed from the table that SNR values after encryption do not vary monotonically with frequency. Though we have shown the results for four files only, similar results are found to hold for the other files also. We may conclude from the experimental results that the variation in SNR with the block x_i being encrypted depends quite heavily on the audio file under consideration, and we cannot make a

Table 1 Variation of SNR values with encryption of various x_i's

File	SNR values when encryption is done on						
	x_8	x_7	x_6	x_5	x_4	x_3	x_2
a1.wav	31.27	37.15	37.51	36.67	26.74	26.49	15.22
a2.wav	28.96	17.13	30.31	38.25	9.39	9.80	13.60
a3.wav	22.13	29.56	32.08	42.99	21.83	22.42	15.15
a4.wav	37.93	36.72	38.16	27.85	22.96	15.89	15.18

prediction as to which block will result in the desired level of degradation. For this reason, we have not explored this method any further.

Algorithm 3 *Scheme2_Embed*

```
/* Embed watermark $w$ into *file*, and encrypt $S$% of the high-frequency
   third-level DWT coefficient using key $k$, to generate $newfile$
   Input parameters: *file*, $S$, $w$, $k$
   Output parameter: $newfile$ */
begin
   $V$ = DWT of *file* upto 3 levels;
   $w = w||S||p\_encrypt(k, KU_{cm})$;
   $zone1$ = one-eighth of the coefficients of $V$ from beginning;
   $embed(zone1, w)$;
   $zone2$ = $S$% of the coefficients of $V$ from the end;
   $encrypt(zone2, k)$;
   $newfile$ = IDWT of $V$;
end
```

Algorithm 4 *Scheme2_Extract*

```
/* Extract watermark from $newfile$, and obtain $S$ and the encryption key;
   hence decrypt the encrypted portion, to generate $file$
   Input parameter: $newfile$
   Output parameter: *file* */
begin
   $V$ = DWT of $newfile$ upto 3 levels;
   $zone1$ = one-eighth of the coefficients of $V$ from beginning;
   $new\_w = extract(zone1)$;
   $(w', S', k') = new\_w$;
   $k'' = p\_decrypt\ (k', KR_{cm})$;
   $zone2$ = $S$% of the coefficients of $V$ from the end;
   $decrypt(zone2, k'')$;
   $file$ = IDWT of $V$;
end
```

3.5 Proposed scheme 2

With the previous scheme failing to provide us with a mechanism to provide a monotonous degradation mechanism based on some user-defined parameter, an alternate scheme has been proposed. As before, watermarking is carried on the leftmost DWT transformed coefficients block (namely, $x1$). However, instead of encrypting one of the higher frequency blocks, we encrypt a variable number of consecutive coefficients starting from the highest frequency side. Essentially, we treat all the coefficients in all the leaf nodes $x1, x2, \ldots, x8$ as a single vector V. Encryption is performed on S% of the coefficients in the vector V from the rightmost (high frequency) side. Watermarking is performed as before using mean quantization method on $x1$, which corresponds to 12% of the lowest frequency coefficients at level 3. With increase in the value of S, since the number of DWT coefficients that are encrypted increases, the amount of degradation is also expected to increase monotonously. The algorithms for watermark embedding and encryption, and watermark detection and decryption are given below.

3.5.1 Watermark embedding

As in the previous approach, we embed the watermark bits in the DWT transformed lowest frequency coefficients in $x1$ at the third level, using mean quantization technique.

3.5.2 Integration of encryption and watermarking

In this scheme, watermarking and encryption are carried out as illustrated in Fig. 11. Clearly, watermarking and encryption are carried out on disjoint sections of the DWT transformed coefficients, and hence the process is commutative. The value of S can be chosen seamlessly to encrypt any percentage of the higher frequency components as desired.

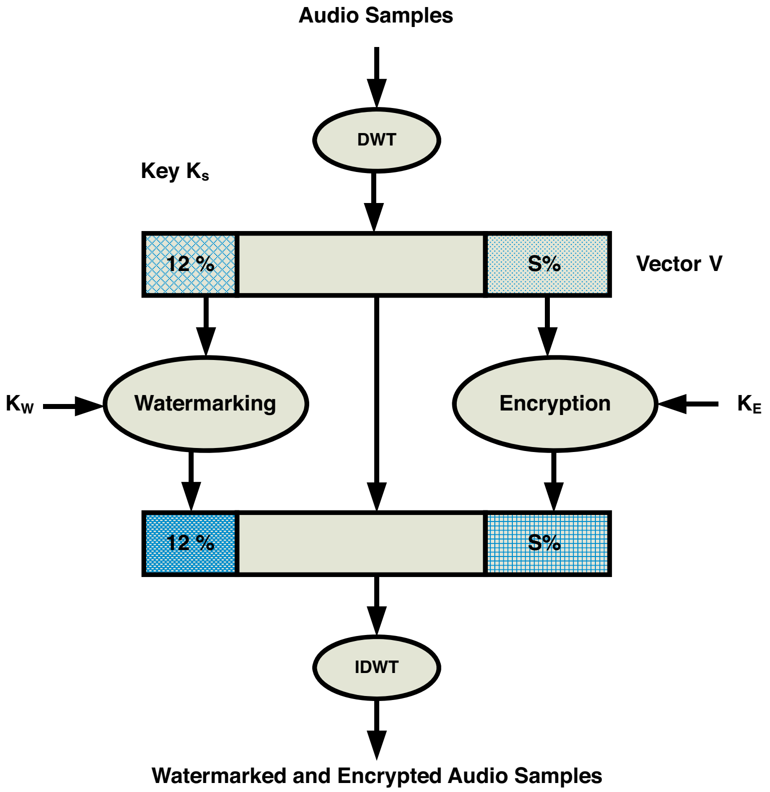


Fig. 11 Combined watermarking and encryption for proposed scheme 2

Table 2 Variation of SNR with *S*

File	Value of S in %																
name	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85
pop1	27.5	24.5	22.0	20.3	19.1	19.0	18.9	18.8	18.7	18.6	16.5	15.1	14.2	13.9	13.5	13.0	12.5
pop2	28.6	25.3	22.9	20.6	19.3	19.1	18.9	18.8	18.8	18.7	15.0	12.8	12.1	11.8	11.6	11.2	10.8
pop3	27.0	24.2	21.7	20.0	18.1	17.4	16.9	16.6	16.6	16.5	13.8	12.7	11.6	11.2	10.6	8.6	8.3
pop4	25.5	23.5	19.8	18.3	17.4	17.0	16.7	16.5	16.5	16.5	13.2	11.7	10.8	10.6	10.5	9.6	9.0
pop5	38.5	34.4	27.4	24.9	22.7	22.6	22.5	22.3	22.2	22.2	18.3	16.2	13.5	11.7	10.4	10.0	9.6
popb1	29.6	26.7	25.0	24.4	23.7	23.5	23.2	23.1	23.1	23.0	21.5	20.7	19.8	19.3	19.0	18.0	17.1
popb2	26.0	23.1	21.5	20.3	19.3	19.0	18.8	18.7	18.6	18.6	16.6	15.3	14.4	13.8	13.4	10.9	9.5
popb3	28.8	26.6	24.1	22.1	20.3	19.0	18.5	18.5	18.5	14.6	13.0	12.0	11.7	11.4	8.7	7.2	6.4
popb4	21.7	18.9	17.6	16.9	16.6	16.5	16.5	16.5	16.5	16.5	10.6	8.2	7.3	7.1	7.0	6.5	6.2
popb5	32.1	29.6	24.1	20.4	18.6	18.5	18.4	18.4	18.3	18.3	11.1	8.6	7.6	7.5	7.4	7.1	6.8
sam1	38.6	35.4	34.3	34.2	34.1	31.2	29.4	28.6	28.5	28.3	24.6	22.3	21.3	20.9	20.5	17.0	14.7
sam2	42.3	39.5	38.6	38.4	38.4	36.3	35.3	34.5	33.7	33.4	26.4	24.4	23.2	22.3	21.9	13.5	11.2
sam3	40.5	37.6	36.7	36.5	36.4	34.7	33.4	32.6	31.7	31.1	27.3	25.3	24.0	22.5	21.7	13.3	10.6
sam4	51.0	47.0	45.9	45.6	45.3	38.9	36.4	34.9	34.0	33.2	29.8	27.9	26.0	24.5	23.2	14.0	11.2
sam5	47.0	45.0	44.0	43.3	42.9	34.0	31.3	30.2	29.9	29.5	25.8	23.9	22.7	22.3	21.7	18.6	16.8
se1	29.6	26.5	24.4	23.5	22.1	21.9	21.7	21.5	21.4	21.3	19.7	18.6	17.5	17.3	16.7	12.9	11.0
se2	32.0	29.9	20.6	16.1	14.1	14.1	14.1	14.1	14.1	14.1	13.2	12.6	12.0	11.4	11.0	10.2	9.5
se3	32.5	30.8	28.8	27.7	26.9	26.5	26.4	26.2	26.0	25.9	18.0	15.8	14.8	14.6	14.5	12.3	11.2
se4	32.9	30.1	28.0	26.2	25.0	24.7	24.5	23.8	22.6	21.7	20.0	19.0	16.4	13.9	12.3	10.3	9.1
se5	40.2	34.1	23.3	18.9	16.7	15.4	14.4	13.9	13.9	13.9	9.4	7.2	6.4	6.3	6.3	6.2	5.8
ins1	24.7	23.0	21.0	18.8	17.7	17.3	17.2	17.1	17.1	17.0	13.3	12.2	11.1	9.8	8.8	7.0	6.4
ins2	30.0	27.2	26.6	26.3	26.1	25.3	24.8	24.7	24.6	24.6	18.4	16.4	15.7	15.3	15.0	12.1	11.1
ins3	41.9	38.9	37.9	37.8	37.7	37.2	36.7	36.3	36.1	35.7	29.2	26.5	25.5	25.1	24.7	22.0	19.3
ins4	31.3	29.3	23.7	20.3	18.4	13.6	11.4	10.6	10.5	10.5	10.2	10.1	9.9	9.7	9.5	7.7	6.9
ins5	34.7	32.0	30.1	28.3	27.2	26.4	25.7	25.4	25.1	24.9	20.5	18.2	17.4	16.7	16.1	11.7	9.1

Table 2 (continued)

File	Value of S in %																
name	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85
hindi1	38.0	34.9	33.9	33.9	33.8	33.5	33.2	32.8	32.4	32.0	29.8	28.2	26.9	26.0	25.1	19.6	17.2
hindi2	43.4	40.7	39.3	38.3	37.6	34.9	33.8	33.1	31.7	31.4	25.4	23.8	22.5	20.6	20.1	12.2	10.0
hindi3	34.2	31.5	30.4	30.2	30.1	29.4	29.0	28.6	28.3	28.1	20.1	17.5	16.6	16.4	16.1	13.7	12.0
hindi4	46.4	42.4	41.5	41.3	40.9	37.9	35.0	34.1	33.3	32.4	27.3	23.7	22.7	22.0	20.9	15.9	12.1
hindi5	35.9	33.8	30.3	27.5	25.8	25.2	25.1	24.8	24.6	24.6	14.9	12.4	11.3	10.9	10.5	6.7	5.4
rock1	27.3	25.2	22.5	20.9	20.1	19.9	19.8	19.7	19.7	19.7	17.3	16.6	15.6	14.7	14.2	12.5	11.3
rock2	31.3	28.1	26.7	26.1	25.3	25.2	25.1	25.0	25.0	24.9	24.2	23.0	22.1	21.4	20.7	17.1	14.9
rock3	36.8	33.6	31.1	28.4	26.9	26.5	26.2	26.1	25.8	25.7	22.8	21.1	20.5	19.3	18.6	13.9	11.3
rock4	33.3	29.6	26.7	24.0	22.5	22.2	21.9	21.8	21.8	21.8	20.8	19.3	18.5	17.4	16.7	14.5	11.9
rock5	33.0	31.1	24.5	20.9	19.1	18.9	18.8	18.7	18.7	18.7	13.1	10.8	9.8	9.6	9.4	8.6	8.2
rock6	32.6	28.9	27.7	26.5	25.9	25.0	24.1	23.6	23.4	23.3	13.1	10.2	9.3	9.2	9.1	8.5	7.7
rock7	46.0	43.4	42.0	41.3	40.7	38.6	37.4	36.6	35.5	35.0	30.3	28.6	27.5	26.0	25.5	16.6	14.2
rock8	47.4	44.6	44.0	42.9	42.2	40.5	39.4	39.1	38.5	38.2	25.6	22.7	21.8	21.7	21.6	20.1	19.1
rock9	42.5	39.8	38.1	37.4	36.8	33.5	31.8	30.6	29.4	28.6	23.2	21.1	19.6	18.2	17.3	10.7	8.4
rock10	28.0	25.1	23.3	22.2	21.4	21.3	21.1	21.1	21.0	21.0	19.2	18.0	17.3	17.0	16.7	12.3	10.2

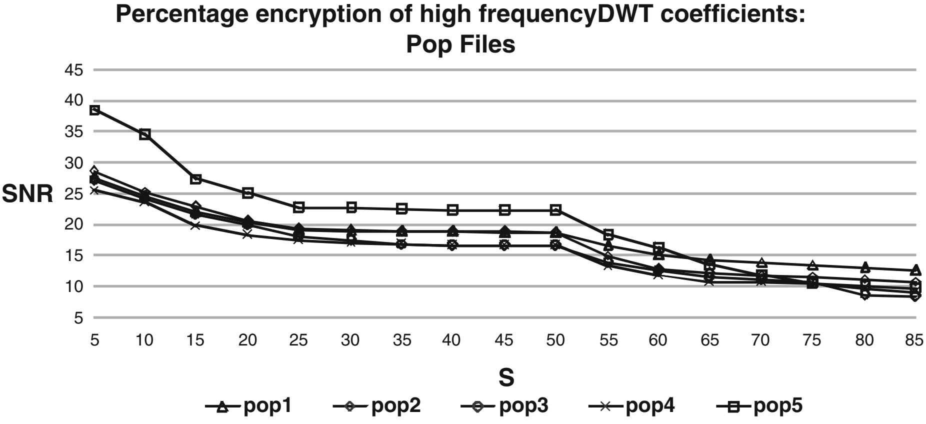


Fig. 12 Variations with *S* for pop files

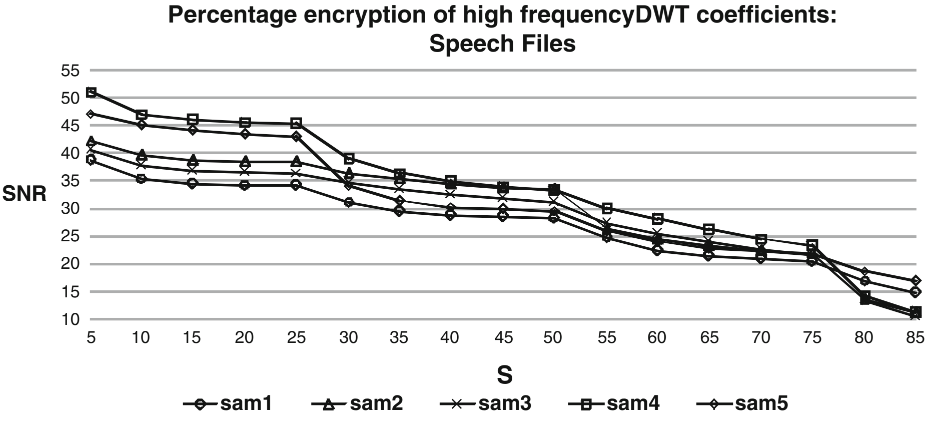


Fig. 13 Variations with *S* for speech files

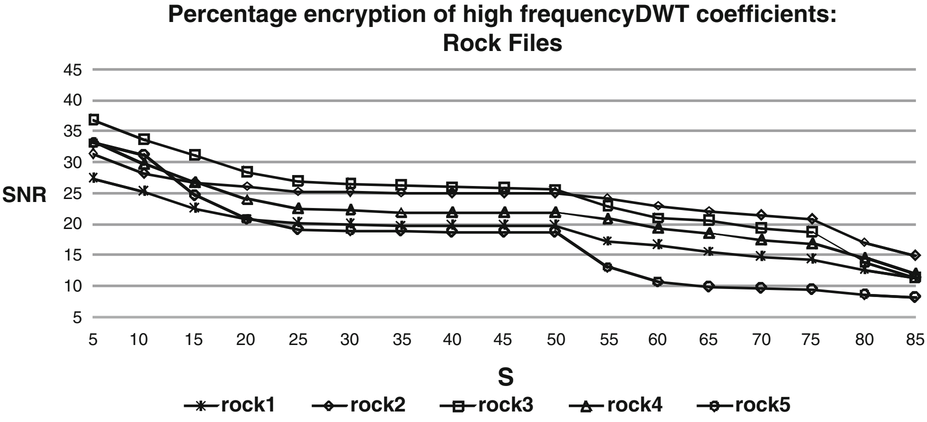


Fig. 14 Variations with *S* for rock files

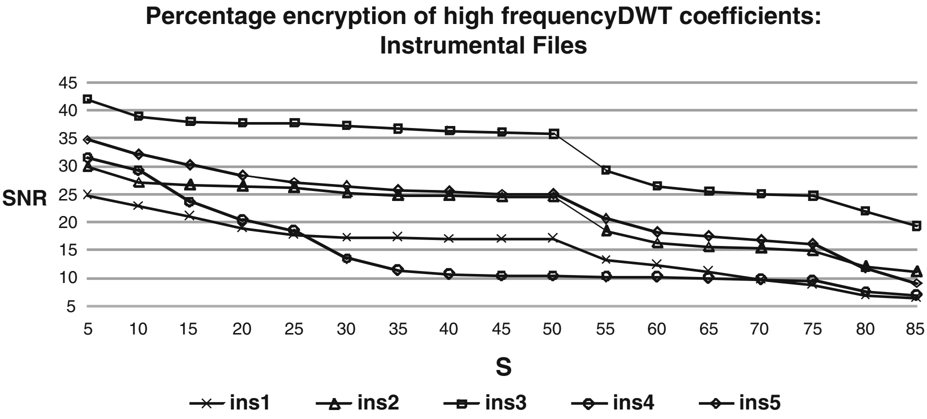


Fig. 15 Variations with *S* for instrumental files

3.5.3 Experimental results

The experiment has been carried out using MATLAB on a set of 100 audio files, which include pop, speech, rock and instrumental files, by varying the value of *S* from 5 to 85, in increments of 5. Table 2 shows the variations in SNR values with variations in *S* for 40 of the files, with some of the corresponding plots shown in Figs. 12, 13, 14, and 15. The following observations may be drawn from the results.

- The SNR values deteriorate when $5 \leq S \leq 25$.
- The SNR values remain more or less constant when $25 \leq S \leq 50$.
- The SNR values further degrade when $S > 50$.

Therefore, by varying the value of *S*, any desired level of degradation can be achieved. Through experimentation, it has been found that any value of *S* between 30 and 60 can be used to provide acceptable levels of degradation in the audio quality. Since the value of *S* is also watermarked as stated in Algorithm 1, the decoder at the receiver end can extract the value of *S* before carrying out decryption.

4 Analysis and comparison

In the first scheme as reported (Algorithms 1 and 2), in addition to some watermark to identify the origin of the file, the value of *i* (indicating which leaf node block x_i is encrypted), and the encryption key *k* are embedded into the audio file as watermark. Assuming that we use 128 bit encryption key, 3 bits for *i*, and 69 bits to store some information about the audio, the number of watermark bits is 200. Assuming the frame size for mean quantization method for watermarking is 16 (which gives good results), this requires a minimum of 3,200 number of coefficients in $x1$. In terms of the embedding capacity, this translates into one bit for every $8 \times 16 = 128$ audio sample values. This in turn implies that the minimum number of samples in the audio file is $8 \times 3{,}200 = 25{,}600$. This determines the minimum size of the audio file for which the proposed method can be implemented.

Similarly, in the second scheme (Algorithms 3 and 4), watermark will consist of information about the file, S, and k, which again will be around 200 bits if S is encoded in multiples of 5. The minimum size of the audio file in this case will also be around 25,600 samples.

In the context of the present work, the word *security* needs some explanation. We are not concerned about the robustness of the watermarking algorithm, since the attacker does not achieve its objective by destroying the watermark. What the attacker tries to achieve is to either extract the encryption key k, or have access to the high quality version of the audio. Clearly, the latter cannot be achieved without knowledge of the former, which is safeguarded by encrypting it using the public key of the client and watermarking in into the audio.

The results of the proposed approach cannot be directly compared to other techniques because in the context of the application scenario considered, there are no methods available in the literature. The partial audio encryption scheme proposed in [21] is somewhat similar, where the FFT parameters of speech data are encrypted thereby degrading the quality. However continuous control over the level of degradation as can be done in the proposed scheme is not possible. The scheme proposed in [19, 20] works on G.729 compressed speech, and is used to encrypt telephone voice signals. In [8] which works on MP3 audio, some bit allocation information or some Huffman codes are encrypted to provide degradation. Again, continuous control over the level of degradation is not easy. Similarly, in [15, 16], a CWE scheme for compressed MPEG4 video is proposed, where video parameters like inter or intra-prediction mode, motion vector difference and residue coefficient sign are encrypted, while the amplitudes of DC or AC are watermarked. Here again, continuous control is difficult, as there are too many parameters.

5 Conclusion

A partial encryption and watermarking scheme has been proposed in this paper, which is based on the DWT coefficients corresponding to a given audio file. Two different schemes have been reported. For watermarking purpose low frequency components are chosen whereas for encryption high and middle frequencies are selected. In the first scheme, SNR values after encryption do not vary monotonically with frequency, and so the selection of leaf node cannot be always made to have the expected degradation. But the results of the second scheme provide us with a mechanism for selecting the degree of encryption so as to degrade the original audio by a desired amount. The value of S can be tuned to obtain any desired level of degradation for a given audio file. This method can be used effectively for safe distribution of audio files over the Internet.

References

1. Asokan N, Shoup V, Waidner M (1998) Optimistic fair exchange of digital signatures. In: Advances in cryptology (EUROCRYPT-98), vol 1403. Springer-Verlag, pp 591–606
2. Cheung S, Leung H, Wang C (2004) A commutative encrypted protocol for the privacy protection of watermarks in digital contents. In: 37th Hawaii international conference on system sciences

3. Cox I, Miller M, Bloom J (2001) Digital watermarking principles and practice. The Morgan Kaufman Series in Multimedia and Information Systems
4. Daemen J, Rijmen V (2002) The design of Rijndael. Information security and cryptography. Springer
5. Datta K, Sengupta I (2009) A robust encrypted audio watermarking scheme using discrete wavelet transformation. In: Proceedings of 13th world multi conference on systemics, cybernetics, informatics (WMSCI 2009), pp 423–425
6. Dorr G, Dugelay J, Grang L (2004) Exploiting self-similarities to defeat digital watermarking systems: a case study on still images. In: International workshop on multimedia and security, pp 133–142
7. European network of excellence in cryptology: first summary report on hybrid systems (2002) http://www.ecrypt.eu.org/documents/
8. Gang L, Akansu AN, Ramkumar M, Xie X (2001) Online music protection and mp3 compression. In: International symposium on intelligent multimedia, video and speech processing, pp 13–16
9. Juan L, Tie-Jun H, Jun-Hua Q (2007) A perception-based scalable encryption model for AVS audio. In: IEEE international conference on multimedia and expo, (ICME 2007), pp 1778–1781
10. Lanxun W, Chao Y, Jiao P (2007) An audio watermark embedding algorithm based on mean-quantization in wavelet domain. In: Proceedings of 8th intl. conf. on electronic measurement and instruments, (ICEMI), pp 423–425
11. Lemma A, Katzenbeisser S, Celik M, Veen M (2006) Secure watermark embedding through partial encryption. In: International workshop on digital watermarking (IWDW 2006), pp 433–445
12. Li M, Lei Y, Liu J, Yan Y (2006) A novel audio watermarking in wavelet domain. In: Proceedings of the intl. conf. on intelligent information hiding and multimedia signal prcessing, pp 27–32
13. Lian S (2009) Multimedia content encryption: techniques and applications. CRC Press, Boca Raton, FL
14. Lian S (2009) Quasi-commutative watermarking and encryption for secure media content distribution. Multimedia Tools Appl (LNCS) 43(1):91–107
15. Lian S, Liu Z, Ren Z, Wang H (2006) Commutative encryption and watermarking in compressed video. IEEE Circuits Syst Video Technol 17(6):774–778
16. Lian S, Liu Z, Ren Z, Wang H (2006) Commutative watermarking and encryption for media data. Int J Opt Eng 45(8):1062–1078
17. Podilchuk CI, Delp EJ (2001) Digital watermarking: algorithms and applications. IEEE Signal Process Mag 18:33–46
18. Rivest R, Shamir A, Adleman L (1978) A method for obtaining digital structures and public-key cryptosystem. Commun ACM 21(2):120–126
19. Servetti A, Martin J (2002) Perception-based partial encryption of compressed speech. IEEE Trans Speech Audio Process 10(8):637–643
20. Servetti A, Testa C, Martin J (2003) Frequency-selective partial encryption of compressed audio. In: IEEE international conference on acoustic speech and signal processing (ICASSP 03), vol 5, pp 668–671
21. Sridharan S, Dawson E, Goldberg B (1991) Fast fourier transform based speech encryption system. In: Proceedings of IEE communications, speech and vision, vol 138, pp 215–223
22. Tang X, Niu Y, Yue H, Yin Z (2005) A digital audio watermark embedding algorithm. Int J Inf Technol 11(12):24–31
23. Wu T, Wu S (1997) Selective encryption and watermarking of mpeg videos. In: International conference on image science, systems and technology, (CISST '97)
24. Yong S, Lee S (2005) An efficient fingerprinting scheme with symmetric and commutative encryption. In: International workshop on digital watermarking (LNCS), vol 3710, pp 54–66
25. Yu F, Bao WB, Ru LC, Qiang QN (2004) A novel algorithm for robust audio watermarking in wavelet domain. Journal of Electronic Science and Technology of China (JEST) 2(2):70–78

Kamalika Datta completed her BSc. (Computer Science) from Ravenshaw College (cuttack). After graduating she did her Master in Computer Application from Biju Patnaik University of Technology. After which she joined a software firm where she worked as a software engineer for about 14 months. She joined Indian Institute of Technology Kharagpur as a Junior Project Assistant along with she completed her MS (Information Technology) form IIT Kharagpur. She also worked as a Research Consultant at IIT Kharagpur in a project entitled "Fault diagnosis of digital systems". She had also worked as an Assistant Professor at Kalinga Institute of Industrial Technology, KIIT University, Bhubaneswar for about 13 months. Currently she is pursuing her PhD work at Bengal Engineering and Science University, Shibpur in the broad area of reversible logic synthesis.

Indranil Sen Gupta has obtained his B.Tech., M.Tech. and PhD degrees in Computer Science and Engineering from the University of Calcutta. He joined Indian Institute of Technology, Kharagpur, as a faculty member in 1988, in the Department of Computer Science and Engineering, where he is presently a Professor. A Centre of Excellence in Information Assurance has been set up at IIT Kharagpur under his leadership. He has over 23 years of teaching and research experience. His research interests include cryptography and network security, side-channel attacks on cryptosystems, VLSI design and testing, and mobile computing.