TÜBİTAK

Turkish Journal of Electrical Engineering & Computer Sciences

http://journals.tubitak.gov.tr/elektrik/

Research Article

Turk J Elec Eng & Comp Sci
(2017) 25: 2136 – 2148
© TÜBİTAK
doi:10.3906/elk-162-267

A steganographic approach to hide secret data in digital audio based on XOR operands triplet property with high embedding rate and good quality audio

Krishna BHOWAL[*], **Debasree SARKAR, Susanta BISWAS, Partha Pratim SARKAR**
Department of Engineering & Technological Studies, University of Kalyani, Kolkata, India

Received: 19.02.2016 • **Accepted/Published Online:** 18.08.2016 • **Final Version:** 29.05.2017

Abstract: In this paper, a unique and transparent data hiding algorithm based on XOR operands triplet (XOT) is proposed. In XOT, an XOR operation applied on any two members of a triplet provides the third member of the same triplet. Taking advantage of the low computational complexity and fascinating properties of the XOR operator, it is possible to embed 4-ary secret digits with negligible changes in the host digital audio. The proposed scheme has been designed to ensure that a minimum number of bit alterations happen in the host digital audio during the data hiding process, which also increases the security of the scheme and provides high quality embedded audio without compromising the statistical property of host audio signals. The scheme confirms that the maximum change is less than 12.5% of the digital audio samples and the average error for the individual digital audio sample is less than 6.25%. The experimental results show that the scheme has a high capacity (88 kbps) without perceptual distortion (objective difference grades are –0.1 to –0.31) and provides robustness against intentional or unintentional attack detection. Comparative analysis shows that our method has better performance than data hiding techniques reported recently in terms of imperceptibility, capacity, and security.

Key words: Embedding capacity, imperceptibility, steganography, watermarking, computational complexity, XOR operands triplet

1. Introduction

In the current digital information age, transmissions of digital content are increasing rapidly day by day and with the development of different new communication techniques for these digital transmissions. Unauthorized access of information and illegal copying or distribution of digital content has also become easier. To prevent unauthorized access of information and illegal copying or distribution of digital media, the most promising solutions are data hiding techniques where data are embedded secretly and imperceptibly in host digital media. That means that information security becomes more and more relevant in the current scenario. Data hiding techniques have developed a strong basis for a growing number of applications including authentication, copyright protection, tamper detection, and covert communications.

A data hiding algorithm may have different properties based on its applicability, but must satisfy the following basic requirements in all applications [1]:

Imperceptibility: The data hiding algorithm should be designed without affecting the quality of the audio after embedding secret data. Imperceptibility can be measured using subjective difference grade (SDG), objective difference grade (ODG), and signal-to-noise ratio (SNR) measures.

[*]Correspondence: ykbhowal@yahoo.co.in

Robustness: The data hiding algorithm should be designed in such a way that the modification made due to conventional digital signal processing operations or any other intentional attacks should be detectable or would not affect the extraction of hidden data from embedded audio signals.

Security: Security is the main challenge in designing a data hiding algorithm and security requirements may vary with the application of the data hiding technique. A data hiding algorithm is truly protected if the presence of hidden data in host audio is not perceived by unauthorized people even after learning the exact algorithm applied during the data hiding process. For this, secret data may be encrypted before being embedding in a host audio signal.

Embedding capacity: Some applications of data embedding require small amounts of information to be incorporated. On the other hand, many applications of data embedding, like covert communication, require much data to be incorporated. The ability to embed large quantities of data in a host medium will depend on how the embedding algorithm has been designed and also the type of the cover digital media utilized.

Therefore, to improve the overall performance of a data hiding technique, the key challenge is to decrease the number of bit level alterations required to be incorporated into the digital cover media during the data hiding process. Keeping all these basic requirements in mind, we have considered a data hiding scheme where the possibility to alter bit(s) in a host digital audio sample is at maximum 2 bits and at minimum 0 bits in a 16-bit digital audio sample. A good quality of audio signal has been generated after the embedding process.

2. Related works

In order to hide secret information in digital audio effectively, a variety of embedding techniques were discussed and implemented in [1–6]. Most of the schemes exploit sophisticated signal processing techniques for hiding secret data. In [2], to increase the robustness in the data hiding process, high level LSB positions were considered to embed secret message. To decrease the distortion generated due to higher LSB insertion, GA operators are used.

Generally, the robustness and the capacity hardly coexist in the same steganographic system due to tradeoff imbalance between these two criteria where increased robustness levels result in decreasing data hiding capacity [3]. In [4], parity coding and XORing of LSB-based methods were proposed. In the second method, XOR operation is performed between the LSB and the next bit has to be embedded. The LSB remains unchanged if equal, or otherwise is flipped. From the experimental results it is found that the encryption with steganography provides better security. The various types of steganography and watermarking techniques and their basic requirements like imperceptibility, capacity, and robustness were discussed elaborately in [5]. Data embedding by exploiting modification directions requires that each secret digit in a (2n+1)-ary notational system be embedded on n cover media samples, where n is a system parameter [6].

A method that performs watermark embedding in the frequency domain in order to take advantage of the frequency masking characteristics of the human auditory system was presented in [7]. In [8], a part of the frequency of the FFT spectrum was separated into small frames and a single secret bit was embedded into each frame. The largest Fibonacci number that is lower than each single FFT magnitude in each frame was computed. Based on the matching secret bit to be embedded, all samples in each frame are altered. All FFT samples in a frame are altered to the closest Fibonacci number with an odd index. A new adaptive audio watermarking algorithm based on empirical mode decomposition (EMD) was introduced in [9]. Each audio signal's frame is decomposed adaptively using the EMD concept into intrinsic oscillatory components called intrinsic mode functions. In spread-spectrum watermarking, the data are embedded by adding a pseudorandom sequence to the audio signal or some features derived from it. A spread-spectrum watermarking in the time domain was presented in [10].

A scheme based on a new inaudibility control procedure that locally regulates the watermark transparency, an embedding function that maximizes system robustness to additive channel perturbation by maintaining the error probability at a fixed value, and an efficient and low computational cost mechanism was proposed in [11]. Audio watermarking methods that add their watermarks in the time domain and also have attracted attention as a prevention technique against copyright violation were reported in [12]. The conventional method maintains good sound quality and is highly robust to pirate attacks like MP3 compression as proposed in [13] with payload of 2 bps and robustness to MP3 of 64 kbps. The audio signal intervals were quantized and the secret information was embedded in the quantization indices in [14].

From the above literature survey, it is clear that robustness and capacity are the main requirements in watermarking, On the other hand, imperceptibility and capacity are the main requirements in steganography. In our proposed scheme, we mainly concentrate on the imperceptibility, capacity, and robustness.

3. Proposed work

This work presents an efficient approach to achieve high quality audio. The proposed scheme is based on the interesting property of three operands of the XOR ($\oplus$) operator, i.e. triplet of $\oplus$ operator.

It is well known that the XOR ($\oplus$) bitwise operator has several fascinating properties. One of these interesting properties has been applied in this work as explained below.

First, a list of triplets (x_i, y_i, z_i), where x_i, y_i, and z_i are some positive integer numbers holding the following property, has been generated:

$$x_i = y_i \oplus z_i$$

$$y_i = x_i \oplus z_i$$

$$z_i = y_i \oplus x_i$$

If we apply the $\oplus$ operator on any of these two members of the triplet, we will get the third member of the same triplet.

Consider Table 1, generated based on the $\oplus$ operation.

Table 1. XOR operator-based table for digits 0 to 3.

$\oplus$	0	1	2	3
0	0	1	2	3
1	1	0	3	2
2	2	3	0	1
3	3	2	1	0

Considering Table 1 and for selecting the range of values of x_i, y_i, and z_i, a 4-ary number system is a very compatible choice. Initially, secret digits are converted to the 4-ary number system as shown in Table 2.

Table 3 is generated from Table 1 and the following table generation algorithm:

For i = 0 to 3 and for j = 0 to 3

$$T_{i\,X\,4+j,\;0} = i$$

$$T_{i\,X\,4+j,\;1} = j$$

Table 2. 10-ary digit to 4-ary digit conversion table.

Secret digits	0	1	2	3	4	5	6	7	8	9
4-ary secret digits	00	01	02	03	10	11	12	13	20	21

Table 3. XOR operands triplet table in 4-ary number system.

$T_{i*4+j,0}$	$T_{i*4+j,1}$	$T_{i*4+j,2}$
0	0	0
0	1	1
0	2	2
0	3	3
1	0	1
1	1	0
1	2	3
1	3	2
2	0	2
2	1	3
2	2	0
2	3	1
3	0	3
3	1	2
3	2	1
3	3	0

$$T_{i\,X\,4+j,\;2} = i \oplus j$$

Here $T_{i\,X\,4+j,\;0}$, $T_{i\,X\,4+j,\;1}$, and $T_{i\,X\,4+j,\;2}$ are used to represent secret digits, the 1st audio sample, and the 2nd audio sample respectively.

3.1. Embedding procedure

a) Convert secret message to secret digits.

b) Convert secret digits to 4-ary secret digits using Table 2.

c) Read audio file and generate 16-bit audio samples, AS_j.

d) Extract two bits from the right-hand side (LSB side) of each audio sample AS_j. The possible bits are 00, 01, 10, and 11, and the corresponding decimal representations are 0, 1, 2, and 3. These digits are represented by S_j in the remaining sections.

e) Consider d_i as the secret digits and (S_j, S_{j+1}) as 4-ary sample pairs extracted from two consecutive audio samples $(AS_j,\ AS_{j+1})$.

f) The following embedding algorithm is designed to hide secret digits d_i, as given below:

Let ‘nosd’ be the number of secret digits to be embedded into ‘nodas’ digital audio samples where 2*nosd <=nodas. We represent secret digits by d_i where i = 0 to nosd – 1. There are four cases considered during the embedding process because values of d_i are in the range of 0 to 3.

```
Case 1: if $d_i = 0$
        if $T_{k,1} = S_j$ and $T_{k,2} = S_{j+1}$, for k = 0 to 3
              then $S_j^{'} = S_j$ and $S_{j+1}^{'} = S_{j+1}$
        else
        { for $l$ = 0 to 3 and k = 0 to 3
              $S1DIFF_l = T_{k,1} - S_j$ and $S2DIFF_l = T_{k,2} - S_{j+1}$
              Now choose the minimum difference pair from $S1DIFF_l$ and $S2DIFF_l$, for $l$ = 0 to 3.
              Let $S1DIFF_m$ and $S2DIFF_m$ pair be the minimum pair for $0 \leq m \leq 3$.
              Calculate $S_j^{'}$ and $S_{j+1}^{'}$ as below:
              $S_j^{'} = S_j + S1DIFF_m$ and $S_{j+1}^{'} = S_{j+1} + S2DIFF_m$
        }
Case 2: if $d_i = 1$
        if $T_{k,1} = S_j$ and $T_{k,2} = S_{j+1}$, for k = 4 to 7
              then $S_j^{'} = S_j$ and $S_{j+1}^{'} = S_{j+1}$
        else
        { for $l$ = 0 to 3 and k = 4 to 7
              $S1DIFF_l = T_{k,1} - S_j$ and $S2DIFF_l = T_{k,2} - S_{j+1}$
              Now choose the minimum difference pair from $S1DIFF_l$ and $S2DIFF_l$, for $l$ = 0 to 3.
              Let $S1DIFF_m$ and $S2DIFF_m$ pair be the minimum pair for $0 \leq m \leq 3$.
              Calculate $S_j^{'}$ and $S_{j+1}^{'}$ as below:
              $S_j^{'} = S_j + S1DIFF_m$ and $S_{j+1}^{'} = S_{j+1} + S2DIFF_m$
      }
Case 3: if $d_i = 2$
        if $T_{k,1} = S_j$ and $T_{k,2} = S_{j+1}$, for k = 8 to 11
              then $S_j^{'} = S_j$ and $S_{j+1}^{'} = S_{j+1}$
        else
        { for $l$ = 0 to 3 and k = 8 to 11
              $S1DIFF_l = T_{k,1} - S_j$ and $S2DIFF_l = T_{k,2} - S_{j+1}$
              Now choose the minimum difference pair from $S1DIFF_l$ and $S2DIFF_l$, for $l$ = 0 to 3.
              Let $S1DIFF_m$ and $S2DIFF_m$ pair be the minimum pair for $0 \leq m \leq 3$.
              Calculate $S_j^{'}$ and $S_{j+1}^{'}$ as below:
              $S_j^{'} = S_j + S1DIFF_m$ and $S_{j+1}^{'} = S_{j+1} + S2DIFF_m$
        }
Case 4: if $d_i = 3$
        if $T_{k,1} = S_j$ and $T_{k,2} = S_{j+1}$, for k = 12 to 15
              then $S_j^{'} = S_j$ and $S_{j+1}^{'} = S_{j+1}$
        else
```

```
{ for $l$ = 0 to 3 and k = 12 to 15
        $S1DIFF_l = T_{k,1} - S_j$ and $S2DIFF_l = T_{k,2} - S_{j+1}$
        Now choose the minimum difference pair from $S1DIFF_l$ and $S2DIFF_l$, for $l$ = 0 to 3.
        Let $S1DIFF_m$ and $S2DIFF_m$ pair be the minimum pair for $0 \leq m \leq 3$.
        Calculate $S_j^{'}$ and $S_{j+1}^{'}$ as below:
        $S_j^{'} = S_j + S1DIFF_m$ and $S_{j+1}^{'} = S_{j+1} + S2DIFF_m$
}
```

Finally, the 1st and 2nd LSB bits of the binary representation of AS_j have been replaced by the two-bit representation of $S_j^{'}$ to get modified audio samples $AS_j^{'}$.

For validating the correctness of information embedding scheme, we give Theorem 1.

Theorem 1 *For $d_i = T_{k,0}$ where i = 0 to nosd – 1, $T_{k,1} = S_j^{'}$ and $T_{k,2} = S_{j+1}^{'}$, where k = 0 to 15.*

Proof From Table 3, it is clear that the $T_{k,0}$ column contains digits 0 for k = 0 to 3, 1 for k = 4 to 7, 2 for k = 8 to 11, and 3 for k = 12 to 15. Again, secret digits d_i are in the 4-ary number system. We have considered four cases for the four secret digits above.

For $d_i = 0$, we calculated $S_j^{'} = S_j + S1DIFF_m$ and $S_{j+1}^{'} = S_{j+1} + S2DIFF_m$ for $0 \leq m \leq 3$.

$$\begin{aligned}\text{Now, } S_j^{'} &= S_j + S1DIFF_m \\ &= S_j + T_{k,1} - S_j \text{ for } 0 \leq k \leq 3 \\ &= T_{k,1}\end{aligned}$$

$$\begin{aligned}\text{Again, } S_{j+1}^{'} &= S_{j+1} + S2DIFF_m \\ &= S_{j+1} + T_{k,2} - S_{j+1} \text{ for } 0 \leq k \leq 3 \\ &= T_{k,2}\end{aligned}$$

For $d_i = 1$ and $4 \leq k \leq 7$, $d_i = 2$ and $8 \leq k \leq 11$, and $d_i = 3$ and $12 \leq k \leq 15$, in the same way, we can prove the remaining cases. □

An example of embedding process: Let secret digit $d_i = 2$ and the audio sample pair (AS_1, AS_2) be (32690, 32671). The binary representation of 32690 is 01111111 101100**10** and 32691 is 01111111 101100**11**. Extracting the 1st and 2nd LSB bits from each of the samples, we get (S_1, S_2) = (10, 11), i.e. (2, 3). For $d_1 = T_{k,0} = 2$, for $8 \leq k \leq 11$ and corresponding values of $T_{k,1}$ and $T_{k,2}$, for $8 \leq k \leq 11$ are (0, 2), (1, 3), (2, 0), and (3, 1). Now we choose the pair that is closest to (2, 3) to ensure minimum deviation in the audio sample after the embedding process. In this example, (1, 3) is the closest pair to (2, 3) and the difference pair is (–1, 0). Now we calculate $S_1^{'} = S_1 - 1$ and $S_2^{'} = S_2 + 0$ and $(S_1^{'}, S_2^{'})$ = (1, 3) = (01, 11). Finally, we replace the 1st and 2nd LSB bits of the binary representation of AS_1 and AS_2 by 01 and 11, respectively. Modified audio samples are 01111111 101100**01** and 01111111 101100**11**, i.e. $(AS_1^{'}, AS_2^{'})$ = (32689, 32691).

3.2. Extraction procedure

Following are the steps to extract the hidden message from embedded audio signals without using original audio signals, i.e. a blind approach is followed here.

1. Extract 4-ary secret digits from the embedded audio samples by applying the proposed scheme.
2. Convert 4-ary secret digits to 10-ary secret digits.
3. Convert secret digits to the message.

For each pair of embedded audio samples $(AS^{'}_{j}, AS^{'}_{j+1})$ extract the 1st and 2nd LSB bits from each pair of embedded audio sample ($AS^{'}_{j}, AS^{'}_{j+1}$). The possible bits are 00, 01, 10, and 11, and corresponding decimal representations are 0, 1, 2, and 3. Let the digit pair be $(S^{'}_{j}, S^{'}_{j+1})$.

To extract a secret digit d_i from digits $S^{'}_{j}\, and\, S^{'}_{j+1}$, Eq. (1) is used, as below:

$$d_i = S^{'}_{j} \oplus S^{'}_{j+1} \tag{1}$$

For validating the correctness of the information extracting scheme, we give Theorem 2.

Theorem 2 *For i = 0 to nosd – 1, j = 0 to (nosd – 1)*2, secret digits* $d_i = S^{'}_{j} \oplus S^{'}_{j+1}$

Proof According to the table construction algorithm $T_{k,1}$ and $T_{k,2}$ hold all possible two element combinations using 0, 1, 2, and 3 and $T_{k,0}$ holds the result of XOR ($\oplus$) operation performed on these two elements for $0 \leq k \leq 15$. During the embedding process, we ensure that the $T_{k,1}$ and $T_{k,2}$ columns contain $S^{'}_{j}\, and\, S^{'}_{j+1}$, respectively. □

From Theorem 1, we have $S^{'}_{j} = T_{k,1}$ and $S^{'}_{j+1} = T_{k,2}$ so $S^{'}_{j} \oplus S^{'}_{j+1} = T_{k,1} \oplus T_{k,2} = T_{k,0} = d_i$.

Continuing the previous example, let the embedded audio sample pair be $(AS^{'}_{1}, AS^{'}_{2})$ = (32689, 32691). The binary representations of these samples are 01111111 101100**01** and 01111111 101100**11**. The 1st and 2nd LSB bits of both samples are 01 and 11 and the corresponding 4-ary representation is 1 and 3, respectively. Now, to get the secret digit $d_1 = 1 \oplus 3 = 2$.

4. Experimental results

To calculate the performance of our proposed scheme in terms of imperceptibility, security, capacity, and robustness, corresponding experiments are performed on 10 digital audio sequences from different music types like classic, jazz, country, pop, rock, folk, country-blues, folk-rock, jazz-rock, and pop-rock. All the clips were 44.1 kHz sampled mono audio files, represented by 16 bits per sample, and the length of the clips ranged from 10 to 20 s.

4.1. Audio quality evaluation and measurements

4.1.1. Measurement of similarity between original audio and embedded audio through correlation

The most familiar measure of similarity between two quantities is the linear correlation coefficient. If there is a series of n original audio samples X and a series of n embedded audio samples Y and they have been written as

$x_{i,}$ and y_i where i = 1,2,3,. . . n, respectively, then the sample correlation coefficient can be used in correlation r between X and Y. The audio sample correlation coefficient is written in Eq. (2) as follows:

$$r_{xy} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{(n-1)\, S_x S_y} \tag{2}$$

where $\bar{x}$ is the mean of original audio samples X, $\bar{y}$ is the mean of embedded audio samples Y, and S_x and S_y are the sample standard deviations of X and Y, respectively. Correlation coefficients are calculated for ten categories of audio clips in MATLAB and the value of r is 1 in all categories of audio clips.

4.1.2. Objective quality measurements

Here imperceptibility quality assessment has been performed using both SNR and ODG measurements. ODG is an appropriate measurement of audio distortions, since it is assumed to provide a precise model of the SDG results that may be obtained by a group of expert listeners. In this section, we perform ODG measurements, where ODG = 0 means no degradation happened in digitally embedded audio signals and ODG = –4 means a very annoying distortion happened in embedded digital audio signals. The SNR values are calculated using the original digital audio and embedded digital audio files in a later section, whereas the ODG measurements are provided using the advanced ITU-R BS.1387 standard [15], which is implemented in Opera software [16]. The ITU-R BS.1387 standard specifies a method where particular recommendations are suggested for perceptual evaluation of audio quality (PEAQ). PEAQ is completely compliant with the ITU-R BS.1387 standard, covering the applicability to high quality audio signals with sampling rates of 44.1 to 48 kHz. ODG values of the ten embedded audio signals are reported in Table 4. All ODG values of the embedded audio signals are between –0.1 to –0.31, which determines their good qualities.

Table 4. ODG, SDG, BER, and SNR value comparisons between different audio types.

Audio types	Objective difference grade (ODG)	Subjective difference grade (ODG)	BER	SNR(dB)
Audio_1	–0.31	4.9	0.01	92.95
Audio_2	–0.20	5.0	0.01	93.26
Audio_3	–0.30	5.0	0.01	92.48
Audio_4	–0.10	5.0	0.01	92.65
Audio_5	–0.10	5.0	0.01	92.31
Audio_6	–0.21	4.9	0.01	93.13
Audio_7	–0.10	5.0	0.01	93.11
Audio_8	–0.17	5.0	0.01	92.72
Audio_9	–0.10	5.0	0.01	93.16
Audio_{10}	–0.14	5.0	0.01	93.04

4.1.3. Subjective quality evaluation

Subjective quality measurements [17,18] have been performed to evaluate the inaudibility of our proposed data hiding scheme. Ten participants were nominated for these subjective listening tests; five of them were experts in music and the rest were general listeners. All of the participants are presented with the original and the embedded digital audio signals and were asked to report any difference between these two signals using a five-point SDG: (5: imperceptible, 4: perceptible but not annoying, 3: slightly annoying, 2: annoying, 1: very annoying). The output of the subjective tests is the average of the quality ratings, called a mean opinion score (MOS). The SDG values for different audio types are reported in Table 4, which shows that the perceived quality of the embedded audio signal is imperceptible (about 5.0 in all cases). From the data presented in Table 4, we can confirm the convenient imperceptibility of the secret message in the digitally embedded audio signals.

4.1.4. Signal-to-noise ratio measurement

The SNR is a very effective tool to measure the difference between the original and embedded audio signals [19]. The SNR is used to judge the quality of the embedded audio. In general, if the SNR value is higher than the standard measurement of 50 dB, then the secret data embedded in the cover media are imperceptible to the human auditory system. The SNR value is measured using Eq. (3) and the Figure shows the SNR values of 10 categories of audio clips. The original signal (the cover audio) is denoted as x(i), where i = 1 to N, while the stego-signal (the stego-audio) is denoted as y(i), i = 1 to N.

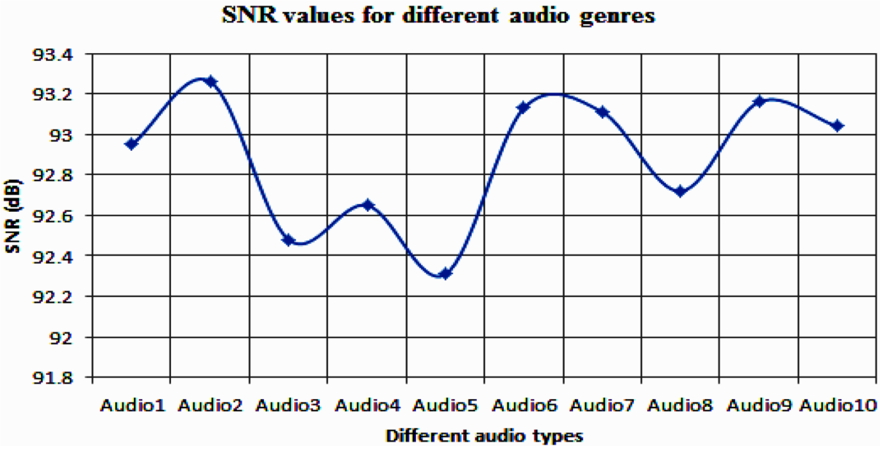

Figure. SNR value comparison among different audio types.

$$SNR = 10\,log_{10}\frac{\sum\limits_{i=1}^{N} x^2(i)}{\sum\limits_{i=1}^{N}\left(x\left(i\right)-y\left(i\right)\right)^2} \tag{3}$$

The bit error rate (BER) metric is used here to measure the quality of the embedded audio signals. The ratio of the number of the altered bits to the total number of embedded audio bits is defined as BER, which is expressed in Eq. (4).

$$BER = \frac{100}{len}\sum\nolimits_{i=0}^{l-1}\begin{cases}1, & AS_i^{'} = AS_i\\ 0, & AS_i^{'} \neq AS_i\end{cases} \quad \ldots \tag{4}$$

Here, len is the bit length of audio signals, AS_i is the ith bit of the original audio signals, and $AS_i^{'}$ is the ith bit of the embedded audio signals.

The BER values for different audio signals are reported in Table 4. The BER values obtained here confirm the good quality of embedded audio signals. Comparisons among the ODG, SDG, BER, and SNR values for different audio types are reported in Table 4. For simplicity, the 10 audio clips are denoted as Audio_1, Audio_2, Audio_3, Audio_4, Audio_5, Audio_6, Audio_7, Audio_8, Audio_9, and Audio_{10}.

4.2. Embedding and extracting complexity

Suppose there is n number of secret digits d_i to be embedded in digital audio signals. First, the availability of at least n × 2 number of audio samples is checked. To embed d_i where $i = 0 \, \text{to} \, n - 1$, a loop has to be iterated for n number of times. Four cases are considered for 4-ary secret digits 0, 1, 2, and 3 under this loop. The nearest pair $(T_{k,1}, \; T_{k,2}), \; k = 0 \, \text{to} \, 3$, is searched from the XOT table for each of these cases. Therefore, the total time complexity of the embedding process is n × 4 × 4 = 16 × n. The data embedding complexity is thus O(n). A loop has to be iterated for n number of times during the data extraction process. Again, the time complexity of the extraction process is O(n). Therefore, the overall time complexity is O(n), i.e. the time complexity is linear.

4.3. Security analysis

By minimizing the bit alteration during the embedding process it is normally guaranteed that the algorithm designed to identify the hidden data based on statistical analysis may be effectively disabled. Steganalysis of digital audio signals is comparatively unexplored compared to the steganalysis of digital image signals.

In this work, only two bits out of 16 bits are used to embed secret digits. The possibility of bit alteration during the embedding process is less than 12.5%. Both the number of secret digits embedded and the secret digits being in a 4-ary number system are key pieces of information for the receiver of our proposed scheme. Again, to make the system more secure and fulfill the data hiding requirement, the information can be encrypted before embedding. There are several cryptography techniques available and Advanced Encryption Standard (AES) encryption is a good selection in terms of computational complexity. AES is a symmetric three-block cipher. These ciphers encrypt and decrypt information in blocks of 128 bits using 128-bit, 192-bit, and 256-bit keys respectively with linear time complexity.

In [20–22], different steganalysis techniques were proposed and designed mainly based on statistical tools like analysis of variance, sequential floating search method, regression analysis classifier, and support vector machine classifier. Most of the techniques will thus not work on our proposed data hiding scheme because alteration of bits in audio signals is much less common and also random.

4.4. Robustness

Robustness of a data hiding technique is defined as the modification made due to conventional digital signal processing operations or any other intentional attacks on embedded audio signals; it should be detectable or would not affect the extraction of hidden data from embedded audio signals. The common attacks include AddNoise, BassBoost, echo addition, and LSB zero. Using original digital audio signals, the above attacks can be easily detected as follows. Let AS_j and $AS_j^{'}$ be original and embedded digital audio samples, respectively. The difference between AS_j and $AS_j^{'}$ is limited and the maximum value is 2 as per the algorithm proposed here, because only 2 LSBs of each audio sample are considered for embedding the secret digit. Modification happens between the 3rd bit and 16th bit of embedded digital audio samples due to common attacks as may be identified by $|AS_j - AS_j^{'}| > 2$. The common attack detection probability is about 87.5%.

4.5. Performance comparisons

The proposed scheme has been compared with some recent steganography and watermarking schemes in audio signals. Each data hiding scheme has different embedding algorithms and properties. For this reason, it is difficult to establish an impartial comparison of the proposed scheme with some other data hiding schemes in audio signals. In this section, a few recent and relevant audio data hiding techniques have been chosen for comparison. Table 5 provides a performance comparison between the proposed data hiding algorithm and several other recent data hiding techniques in audio signals.

A data hiding technique consisting of all the basic requirements practically is not possible to design. There is a tradeoff between certain parameters, i.e. it is not possible to embed a large message in digital audio to reach absolute undetectability and great robustness. Hence, there must be a tradeoff between undetectability and robustness.

The method in [8] provides a significant performance in the different properties of the data hiding technique. The method offers moderate embedding capacity solutions for data hiding in audio signals even though the imperceptibility in terms of SNR and ODG is not so good in some of the cases. The most important achievement of this scheme is robustness against attacks such as echo, filtering, and noises. The method in [9] achieves a low payload for the three audio files. The imperceptibility in terms of SNR is not so good, but the imperceptibility in terms of ODG is moderate in this work. This scheme has a good performance against MP3 (32 kb/s) compression and the maximum of BER against this is about 1%. The methods in [13,14] offer low embedding capacity, acceptable transparency, and reasonably robust against selected attacks. The method in [13] provides very a low embedding rate, high distortion, and very robust scheme, while that in [14] provides very low embedding capacity, highly distorted signals (SNR is 29.3 dB), and moderate robustness against some attacks.

The most important achievement of the proposed method is better imperceptibility in terms of SNR and ODG with higher embedding capacity. The comparison presented in Table 5 demonstrates the superiority in both capacity and imperceptibility of the proposed method with respect to the methods discussed in the literature. The proposed method can embed much more information by introducing less distortion in the stego-audio file. In brief, the proposed method achieves higher embedding capacity if we compare it to methods with similar imperceptibility. Furthermore, the proposed method is very robust in the case of detection of common attacks and attack detection probability is about 87.5%.

Table 5. Performance comparisons with recent and relevant data hiding techniques in audio signals.

Scheme	Capacity (bps)	Imperceptibility in SNR (dB)	Imperceptibility (ODG)
[8]	683 to 3 k	35 to 61	–0.30 to –1.10
[9]	46.9 to 50.3	26.38	–0.40 to –0.60
[13]	2	42.8 to 44.4	–1.66 to –1.88
[14]	4.3	29.3	Not reported
Proposed	88 kbps	93.26	–0.10 to –0.31

5. Conclusion and future work

This paper presents an effective data hiding scheme where secret digits are embedded in digital audio by the minimum number of bit alternations that happen during the secret digit embedding process. Secret digits are converted to 4-ary notational systems to accommodate the XOR operands triplet table's elements as explained above. From the experimental results it is clear that the scheme has a high embedding capacity (88 kbps) without perceptual distortion (ODG is –0.1 to –0.31). The values of ODG, SDG, and SNR ensure that the human auditory system will not be able to distinguish between the original audio and the stego-audio. The scheme is very effective in the case of detection of common attacks and attack detection probability is about 87.5%.

In this work, there is room to enhance the performance of the robustness of the proposed scheme by extending the algorithm, whereby hidden information can be extracted after common types of attacks including AddNoise, BassBoost, echo addition, and LSB zero.

References

[1] Swanson MD, Zhu B, Tewfik AH. Current state of the art, challenges and future directions for audio watermarking. In: IEEE 1999 Multimedia Computing and Systems Conference; 07–11 June 1999; Florence, Italy. New York, NY, USA: IEEE. pp. 19-24.

[2] Bhowal K, Bhattacharyya D, Pal AJ, Kim TH. A GA based audio steganography with enhanced security. Telecommun Syst 2013; 52: 2197-2204.

[3] Bender W, Gruhl D, Morimoto N, Lu A. Techniques for data hiding. IBM Syst J 1996; 35: 313-336.

[4] Kekre HB, Athawale A, Rao S, Athawale U. Information hiding in audio signals. International Journal of Computer Applications 2010; 7: 14-19.

[5] Cox I, Miller M, Bloom J, Fridrich J, Kalker T. Digital Watermarking and Steganography. 2nd ed. San Francisco, CA, USA: Morgan Kaufmann, 2008. pp. 425-490.

[6] Zhang X, Wang S. Efficient steganographic embedding by exploiting modification direction. IEEE Commun Lett 2006; 10: 781-783.

[7] Swanson MD, Zhu B, Tewfik AH, Boney L. Robust audio watermarking using perceptual masking. Signal Process 1998; 66: 337-355.

[8] Fallahpour M, Megías D. Audio watermarking based on Fibonacci numbers. IEEE T Audio Speech 2015; 23: 1273-1282.

[9] Khaldi K, Boudraa A. Audio watermarking via EMD. IEEE T Audio Speech 2013; 21: 675-680.

[10] Bassia P, Pitas I, Nikolaidis N. Robust audio watermarking in the time domain. IEEE T Multimedia 2001; 3: 232-241.

[11] Baras C, Moreau N, Dymarski P. Controlling the inaudibility and maximizing the robustness in an audio annotation watermarking system. IEEE T Audio Speech 2006; 14: 1772-1782.

[12] Lie W, Chang L. Robust and high-quality time-domain audio watermarking based on low frequency amplitude modification. IEEE T Multimedia 2006; 8: 46-59.

[13] Xiang S, Kim JH, Huang J. Audio watermarking robust against time-scale modification and MP3 compression. Signal Process 2008; 88: 2372-2387.

[14] Mansour M, Tewfik A. Data embedding in audio using time-scale modification. IEEE T Speech Audi P 2005; 13: 432-440.

[15] Thiede T, Treurniet WC, Bitto R, Schmidmer C, Sporer T, Beerens JG, Colomes C, Keyhl M, Stoll G, Brandenburg K et al. PEAQ - The ITU standard for objective measurement of perceived audio quality. J Audio Eng Soc 2000; 48: 3-29.

[16] Rydén T, Stoll G, Sporer T, Keyhl M. Perceptual Evaluation of Audio Quality (PEAQ) Software. Erlangen, Germany: OPTICOM GmbH, 1998.

[17] Unoki M, Imabeppu K, Hamada D, Haniu A, Miyauchi R. Embedding limitations with digital-audio watermarking method based on cochlear delay characteristics. Journal of Information Hiding and Multimedia Signal Processing 2011; 2: 1-23.

[18] Wang S, Unoki M. Speech watermarking method based on formant tuning. IEICE T Inf Syst 2015; E98-D: 29-37.

[19] Quackenbush SR, Barnwell TP 3rd, Clements MA. Objective Measures of Speech Quality. Englewood Cliffs, NJ, USA: Prentice Hall, 1988.

[20] Westfeld A, Pfitzmann A. Attacks on steganographic systems. Lect Notes Comp Sci 1999; 1768: 61-76.

[21] Ozer H, Avcibas I, Sankur B, Memon N. Steganalysis of audio based on audio quality metrics. In: SPIE 2003 Security and Watermarking of Multimedia Contents Conference; January 2003; Santa Clara, CA, USA. pp. 55-66.

[22] Avcibas I. Audio steganalysis with content-independent distortion measures. IEEE Signal Proc Let 2006; 13: 92-95.