A Method for Secure Communication Using a Discrete Wavelet Transform for Sound Data

Yuji Tsuda[1], Kouhei Nishimura[2], Haruka Oyaizu[3], Yasunari Yoshitomi[2], Taro Asada[2], and Masayoshi Tabuse[2]
1: Software Service, Inc.
Nishi-Miyahara, Yodogawa-Ku, Osaka 532-0004, Japan
2: Graduate School of Life and Environmental Sciences, Kyoto Prefectural University,
1-5 Nakaragi-cho, Shimogamo, Sakyo-ku, Kyoto 606-8522, Japan
E-mail: {yoshitomi, tabuse}@kpu.ac.jp, t_asada@mei.kpu.ac.jp}
http://www2.kpu.ac.jp/ningen/infsys/English_index.html
3: NHK Media Technology, Inc.
*Kamiyama-cho, Shibuya-ku, Tokyo 150-0047, Ja*pan

Abstract

We developed a method for secure communication using a discrete wavelet transform. Two users must have one piece of music before communicating each other. The music and the original sound data are transformed into a code using the scaling coefficients obtained from a discrete wavelet transform. The user can produce the sound similar to the original sound using an inverse discrete wavelet transform with the code made from the music, the information on the deference between these two codes.

Keywords: Secure communication, Sound data processing, Wavelet transform, Coding.

1. Introduction

Recently, some kinds of frauds have been critical issues. Especially, elder persons tend to be targets of fraud using a telephone. The fraud pretends to be a grandchild of the elder person in talking through the telephone, and appeals the elder person to send money, for example, through a bank budget account. When the elder person takes the fraud for a grandchild, the fraud can get money. Even if the voice of fraud is not similar to that of the grandchild, the elder person might send money to the fraud. This is because the fraud skillfully appeals a serious monetary trouble such as a traffic accident to the elder person who cherishes a real grandchild.

In the present study, we propose a method for secure communication using a discrete wavelet transform (DWT). The method can be used for the Internet protocol (IP) telephone, and has a potential for stopping frauds using a telephone.

2. Wavelet Transform

In this section, we briefly explain the DWT, according to the references.[1]

The original sound data $s_k^{(0)}$, which is used as the level-0 wavelet decomposition coefficient sequence, where k denotes the element number in the data, are decomposed into the elements of multi-resolution representation (MRR) and the elements of multi-resolution analysis (MRA) by repeatedly applying the DWT. The wavelet decomposition coefficient sequence $s_k^{(j)}$ at level j is decomposed into two wavelet decomposition coefficient sequences at level $j+1$ by using (1) and (2):

$$s_k^{(j+1)} = \sum_n \overline{p_{n-2k}} s_n^{(j)} \quad (1)$$

$$w_k^{(j+1)} = \sum_n \overline{q_{n-2k}} s_n^{(j)}, \quad (2)$$

where p_k and q_k denote the scaling and wavelet sequences, respectively, and $w_k^{(j+1)}$ denotes the development coefficient at level $j+1$. The development coefficients at level J are obtained using (1) and (2) iteratively from $j=0$ to $j=J-1$. Fig. 1 shows the process of multi-resolution analysis by DWT.

In the present study, we use the Daubechies wavelet for the DWT, according to the reference.[2] As a result, we obtain the following relation between p_k and q_k :

$$q_k = (-1)^k p_{1-k} \quad (3)$$

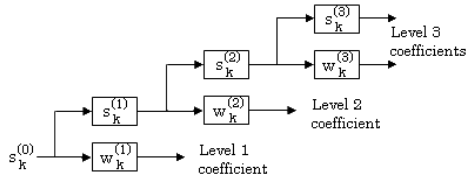

Fig. 1. Multi-resolution analysis by the DWT.[1]

3. Proposed Method

3.1. *Coding*

3.1.1. *Phenomenon exploited in coding algorithm for sound data*

It is known that the histogram of the wavelet coefficients of each domain of the MRR sequences has a distribution which is centered at approximately 0 when the DWT is performed on sound data.[1] In the present study, we have found that the histogram of the scaling coefficients of each domain of the MRA sequences also has a distribution which is centered at approximately 0 when the DWT is performed on sound data. Exploiting this phenomenon, we developed a secure communication method using sound data.

3.1.2. *Parameter setting*

As with the digital watermark (DW) techniques for images[2, 3] and digital sound[4], we set the following coding parameters.

The values of $Th(\text{minus})$ and $Th(\text{plus})$ in Fig.2 are chosen such that the non-positive scaling coefficients (S_m in total frequency) are equally divided into two groups by $Th(\text{minus})$, and the positive scaling coefficients (S_p in total frequency) are equally divided into two groups by $Th(\text{plus})$. Next, the values of $T1$, $T2$, $T3$, and $T4$, which are the parameters for controlling the authentication precision, are chosen to satisfy the following conditions:

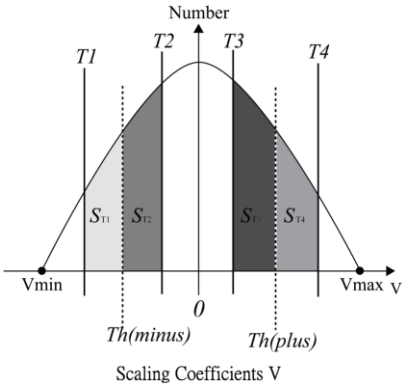

Fig. 2. Schematic diagram of the histogram of MRA scaling coefficients.

1) $T1 < Th(\text{minus}) < T2 < 0 < T3 < Th(\text{plus}) < T4$.
2) The value of S_{T1}, which is the number of scaling coefficients in $(T1, Th(\text{minus}))$, is equal to S_{T2} , which is the number of scaling coefficients in $[Th(\text{minus}), T2)$, i.e., $S_{T1} = S_{T2}$.
3) The value of S_{T3}, the number of scaling coefficients in $(T3, Th(\text{plus})]$, is equal to S_{T4}, the number of scaling coefficients in $(Th(\text{plus}), T4)$, i.e., $S_{T3} = S_{T4}$.
4) $S_{T1} / S_m = S_{T3} / S_p$.

In the present study, the values of both S_{T1} / S_m and S_{T3} / S_p are set to 0.3, which was determined experimentally.

3.1.3. *Coding*

To prepare the coding of sound data, the procedure separates the scaling coefficients V of an MRA sequence into five sets (hereinafter referred to as G_0, G_1, G_2, G_3, and G_4), as shown in Fig. 3, by using the following criteria:

- $G_0 = \{V \mid V \in V^{SC}, V \le T1\}$,
- $G_1 = \{V \mid V \in V^{SC}, T1 < V < T2\}$,
- $G_2 = \{V \mid V \in V^{SC}, T2 \le V \le T3\}$,
- $G_3 = \{V \mid V \in V^{SC}, T3 < V < T4\}$,
- $G_4 = \{V \mid V \in V^{SC}, T4 \le V\}$,

where V^{SC} is the set of scaling coefficients in the sound data file.

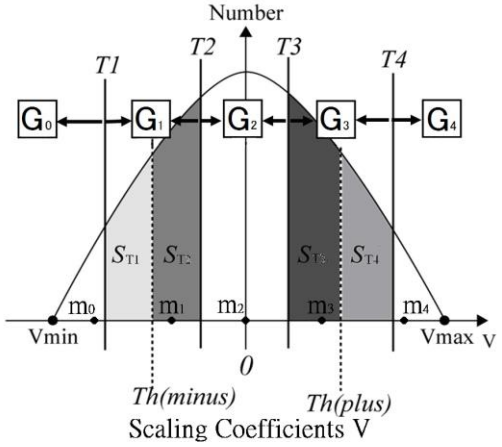

Fig. 3. Schematic diagram for demonstrating scaling coefficient selection in coding sound data.

The scaling coefficients of an MRA sequence are coded according to the following rules, in which V_i denotes one of scaling coefficients:

When $V_i \in G_0$, symbol c_i is set to be 0.
When $V_i \in G_1$, symbol c_i is set to be 1.
When $V_i \in G_2$, symbol c_i is set to be 2.
When $V_i \in G_3$, symbol c_i is set to be 3.
When $V_i \in G_4$, symbol c_i is set to be 4.

Then, the representative value for each set is set as its average value (m_0, m_1, m_2, m_3, m_4), respectively. For sound data formation, we use a code $\boldsymbol{C}$ (hereinafter referred to as an original code), which is the sequence of c_i, and m_j defined above.

3.2. *Sound data formation using code replacement*

The scaling coefficient sequence for sound data A is expressed by

$$S(A)_k = \{x_1, x_2, x_3, \ldots, x_k\},$$

where k is the number of scaling coefficient of A. Then, a sequence

$$C(A)_k = \{X_1, X_2, X_3, \ldots, X_k\}$$

is given, where $X_i (\in \{0,1,2,3,4\})$ is the index showing one of five sets of scaling coefficients to which x_i belongs.

Next, the sound data A' is defined as having the scaling coefficient sequence $S(A')_k$ and the value of zero for all wavelet coefficient values at every level. $S(A')_k$ is obtained as:

$$S(A')_k = \{a_1, a_2, a_3, \ldots, a_k\},$$

where $a_i \left(\in \{m_0^A, m_1^A, m_2^A, m_3^A, m_4^A\}\right)$ is an average at the range of scaling coefficients of A, which is indicated by $X_i (\in \{0,1,2,3,4\})$ obtained from A.

Then, the sound data B'_A is defined as having the scaling coefficient sequence $S(B'_A)_k$ and the value of zero for all wavelet coefficient values at every level. $S(B'_A)_k$ is obtained as:

$$S(B'_A)_k = \{b_{A,1}, b_{A,2}, b_{A,3}, \ldots, b_{A,k}\},$$

where $b_{A,i} \left(\in \{m_0^B, m_1^B, m_2^B, m_3^B, m_4^B\}\right)$ is an average at the range of scaling coefficients of B, which is indicated by $X_i (\in \{0,1,2,3,4\})$ obtained from A.

$S(B'_A)_k$ is obtained by replacing Y_i to X_i when $Y_i \neq X_i$, followed by replacing b_i to $b_{A,i}$, where b_i is an average at the range of scaling coefficients of B, which is indicated by Y_i. Therefore, $C(B'_A)_k = C(A)_k$. As a result, B'_A is expected to be similar to A.

3.3. *Data for communication*

A sequence $D1(B'_A)_n$ is defined as:

$$D1(B'_A)_n = \{z_1, z_2, \ldots, z_n\},$$

where n is the number is the total number of the cases of $Y_i \neq X_i$, and $z_p = \lfloor |y_i| \rfloor \bmod 256$, in which the integer p is increased from 1 to n one by one when $Y_i \neq X_i$ happens. Here, $\lfloor x \rfloor$ is the maximum integer that is not bigger than x. Then, a sequence $D2(B'_A)_n$ is defined as:

$$D2(B'_A)_n = \{Z_1, Z_2, \ldots, Z_n\},$$

where n is the number is the total number of the cases of $Y_i \neq X_i$, and $Z_p = X_i$, in which the integer p is increased from 1 to n one by one when $Y_i \neq X_i$ happens.

In the communication between two users, both a sender and a receiver have B as a secret key, and the sender sends $D1(B'_A)_n$ and $D2(B'_A)_n$ to the receiver. Then, using the processing described in the section 3.4, the receiver makes B'_A-like sound data named by B''_A, which is expected to be similar to A.

3.4. *Sound data composition*

The scaling coefficient sequence for sound data B is expressed by

$$S(B)_k = \{y_1, y_2, y_3, \ldots, y_k\},$$

where k is the number of scaling coefficient of B. Then, a sequence

$$C(B)_k = \{Y_1, Y_2, Y_3, \ldots, Y_k\}$$

is given, where $Y_i \left(\in \{0,1,2,3,4\}\right)$ is the index showing one of five sets of scaling coefficients to which a scaling coefficient y_i of B belongs. $S(B')_k$ is obtained as:

$$S(B')_k = \{b_1, b_2, b_3, \ldots, b_k\},$$

where $b_i \left(\in \{m_0^B, m_1^B, m_2^B, m_3^B, m_4^B\}\right)$ is an average at the range of scaling coefficients of B, which is indicated by $Y_i \left(\in \{0,1,2,3,4\}\right)$ obtained from B.

A sequence $D3(B)_k$ is defined as:

$$D3(B)_k = \{z_{B,1}, z_{B,2}, \ldots, z_{B,k}\},$$

where k is the number of scaling coefficient of B, and $z_{B,q} = \left[\left[y_q\right]\right] \bmod 256$. B_A'' is composed as follows; $S(B_A'')_k$ is composed from $S(B')_k$ by replacing b_q to $m_{Z_p}^B$ when $z_{B,q} = z_p$ happens from $p = 1$ to n; then, the sound data B_A'' is composed by IDWT using the scaling coefficient sequence $S(B_A'')_k$ and the value of zero for all wavelet coefficient values at every level. The receiver makes B_A'' using $D1(B_A')_n$ and $D2(B_A')_n$, which are made with both A and B and are sent by the sender, in addition to B which the receiver has beforehand. B_A'' is expected to be similar to A.

3.5. *Communication of sound data having arbitrary length*

In general, the amount of recording time of A is not the same as that of B. We use B having a unit recording time such as one second. Then, we apply the proposed method described in the above sections to A every unit time of recording time of A. When the recording time of A cannot be divisible by the unit time, additional scaling coefficients needed for application of the proposed method are set as 0.

4. Numerical Experiment

We applied the proposed method to several voice sounds as A using as B the music of (1) 'Classical', (2) 'Hiphop' having one second of recording time each**.** The music is clipped out from the database[5]. In all cases of the experiment, B_A'' was audible and similar to A. However, B_A'' had the noise mainly of low frequency. After erasing the low frequency noise, the sound became more audible and the tone of voice was changed as if the speaker had been a different person. In the next target, we try to develop a method for decreasing the noise with keeping the tone of voice of the speaker.

5. Conclusion

We developed a method for secure communication using a discrete wavelet transform for sound data. In this method, two users must have one piece of music, which has a length of one second and plays a role of secret key, before communicating each other. The music is beforehand transformed into a code using the scaling coefficients obtained from a discrete wavelet transform. The sound data are transformed into another code using the same method as that for the music. The information on the deference between these two codes is sent from one user to the other. The user who receives the information can produce the sound similar to the original sound using an inverse discrete wavelet transform with the code made from the music, the information on the deference between these two codes, and values of zero for all wavelet coefficients. The voice produced by the proposed method was audible.

References

1. Y. Yoshitomi, T. Asada, Y. Kinugawa, and M. Tabuse, An authentication method for digital audio using a discrete wavelet transform, *J. Inf. Sec.* **2**(2) (2011) 59-68.
2. D. Inoue and Y. Yoshitomi, Watermarking using wavelet transform and genetic algorithm for realizing high tolerance to image compression, *J. IIEEJ*, **38**(2) (2009) 136-144.
3. M. Shino, Y. Choi, and K. Aizawa, Wavelet domain digital watermarking based on threshold-variable decision, Technical Report of IEICE, DSP2000-86, **100**(325) (2000) 29-34. (in Japanese)
4. S. Murata, Y. Yoshitomi, and H. Ishii, Audio watermarking using wavelet transform and gtenetic algorithm for realizing high tolerance to MP3 compression, *J. Inf. Sec.* **2**(3) (2011) 99-112.
5. M. Goto, H. Hashiguchi, T. Nishimura and R. Oka, RWC music database: database of copyright-cleared musical pieces and instrument sounds for research purposes, *Trans. IPSJ*, **45**(3) (2004) 728-738.