Audio Secret Sharing for 1-Bit Audio

Ryouichi Nishimura, Norihiro Fujita, and Yôiti Suzuki

Research Institute of Electrical Communication
Graduate School of Information Sciences, Tohoku University,
2-1-1 Katahira Aoba-ku Sendai 980-8577, Japan
ryou@ais.riec.tohoku.ac.jp
http://ais.riec.tohoku.ac.jp/

Abstract. In this paper, we propose a new secret sharing scheme (SSS) [1] for audio signals, called as "Binary audio secret sharing (BASS)." SSS is an encryption method and produces n shared data from an original data to hide useful information. Applying SSS to audio communications on the Internet can help to make it more robust against theft and the tapping of information. Thus, we focused on the 1-bit audio format and applied SSS to 1-bit audio signals to realize audio secret sharing. Moreover, we propose a method to make each shared data heard as its intended sound.

1 The BASS Algorithm

1.1 Visual Secret Sharing

The BASS algorithm is based on visual secret sharing (VSS). Of several proposals regarding VSS, "k out of k VSS [2]" is well known as a method for binary images and shares an original image into k random dot images. By stacking all of the shared images, the original information of the image can be easily obtained, but cannot be with any $k-1$ shared data. Figure 1 is a sample of 3 out of 3 VSS. The process of encryption of 3 out of 3 VSS is depicted schematically in Fig. 2, where sub-pixels are produced from each pixel of the original image. In our proposed method, each pixel of an image corresponds to each sample of a 1-bit audio signal.

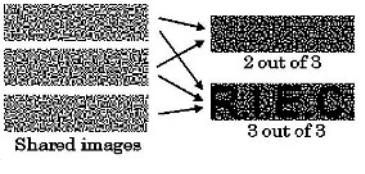

Fig. 1. A sample of 3 out of 3 VSS.

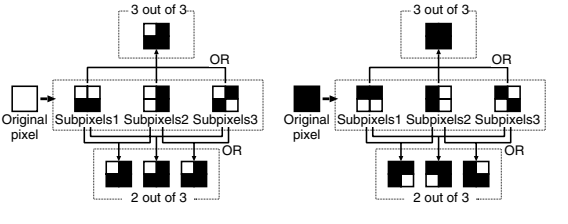

Fig. 2. A sample pattern of sub-pixels.

R. Khosla et al. (Eds.): KES 2005, LNAI 3682, pp. 1152–1158, 2005.

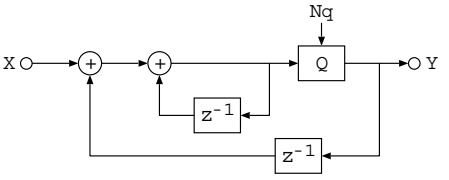

Fig. 3. Block diagram of a first-order $\Delta\Sigma$ converter.

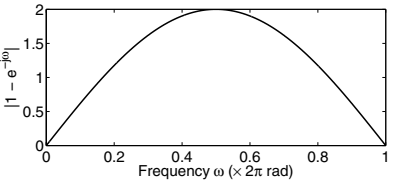

Fig. 4. Frequency response of the first-order $\Delta\Sigma$ converter.

1.2 1-Bit Audio

1-bit audio is a high quality digital audio format employed in the super audio CD (SACD). It has a sampling frequency of 2,822.4 kHz and a resolution of 1-bit. Quantization noise caused by 1-bit resolution can be reduced by the use of the $\Delta\Sigma$ converter. A first-order $\Delta\Sigma$ converter may be expressed by the block diagram depicted in Fig. 3, where Q denotes a quantizer of 1-bit resolution and N_q is the quantization noise. The input X is a discrete time signal sampled with a very high sampling rate. The output signal Y is expressed as

$$Y = X + (1 - z^{-1})N_q. \tag{1}$$

The second term on the right-hand side of this equation consists of the quantization noise N_q, which has a uniformly distributed frequency characteristic, multiplied by a filter having the frequency response depicted in Fig. 4. Consequently, 1-bit audio can realize its large dynamic range despite its low quantizing resolution because the quantization noise is concentrated on the region of high frequency.

1.3 Algorithm of Encryption and Decryption

BASS shares the original signal according to sharing tables $t_b (b = 0, 1)$ for each sample, which have a value of '0' or '1' because of the 1-bit audio. These sharing tables t_b are $k \times n$ boolean matrices, where k is the sharing number and $n = 2^{k-1}$. Defining $A(m, b)$ as the number of '1' in the logical sum of rows of $m (1 \leq m \leq k) \times n$ sub-matrices of t_b, t_b should be constructed so that $A(m, b)$ satisfies

$$A(m,b) = \begin{cases} \displaystyle\sum_{i=1}^{m} \frac{n}{2^i} & (m < k) \\ n - 1 & (m = k, b = 0) \\ n & (m = k, b = 1). \end{cases} \tag{2}$$

Defining T_b as collections of matrices obtained by exchanging any rows of t_b with other rows, shared signals $\boldsymbol{s}_i (1 \leq i \leq k)$ are obtained as

$$\begin{pmatrix} \boldsymbol{s}_1 \\ \boldsymbol{s}_2 \\ \vdots \\ \boldsymbol{s}_k \end{pmatrix} = \begin{pmatrix} \boldsymbol{t}_{o_1} \ \boldsymbol{t}_{o_2} \ \ldots \ \boldsymbol{t}_{o_L} \end{pmatrix} \quad \boldsymbol{t}_{o_j} \in \{\boldsymbol{T}_0, \boldsymbol{T}_1\} \tag{3}$$

from the original signal $\boldsymbol{o}$ represented by

$$\boldsymbol{o} = \begin{pmatrix} o_1 \ o_2 \ \ldots \ o_j \ \ldots \ o_L \end{pmatrix} \quad o_j \in \{0, 1\}. \tag{4}$$

When the sharing number k is $3(k = 3)$, for example, t_0 and t_1 may be 3×4 matrices like as

$$\begin{pmatrix} 1\,0\,0\,1 \\ 0\,1\,0\,1 \\ 1\,1\,0\,0 \end{pmatrix} \in \boldsymbol{T_0} \quad \begin{pmatrix} 1\,0\,0\,1 \\ 0\,1\,0\,1 \\ 0\,0\,1\,1 \end{pmatrix} \in \boldsymbol{T_1}. \tag{5}$$

Decryption of the original signal in BASS requires all the k shared signals. The $m \times (n \cdot L)$ matrix represented in Eq. (3) is separated at every n columns, resulting in L sets of $m \times n$ matrices. When $A(m, t)$ is calculated for each $m \times n$ matrices,

$$A(m, t, j) = \begin{cases} \displaystyle\sum_{i=1}^{m} \frac{n}{2^i} & (m < k) \\ n - 1 + o_j & (m = k) \end{cases} \tag{6}$$

is obtained for each $j(1 \leq j \leq L)$th $m \times n$ matrix. Consequently, we can decrypt the original signal $o_j(1 \leq j \leq L)$ from all the k shared signals using Eq. (6), since the information of the original signal o_j can be obtained only if $m = k$.

2 Shared Signals Heard as Intended Decoy Sound Signals

BASS based on the original k out of k VSS makes all shared signals heard as random noise. However, we can encrypt an original signal into $k - 1$ shared signals heard as distinct and intended decoy sounds and one random noise-like signal by properly constructing sharing tables. These $k - 1$ shared signals can be any audio signals. Note that making shared signals heard as intended sounds does not lessen any security of the BASS.

Defining $1 \times k$ matrix $\boldsymbol{p}_i(1 \leq i \leq k)$ as

$$\begin{pmatrix} \boldsymbol{p}_0 \\ \boldsymbol{p}_1 \\ \vdots \\ \boldsymbol{p}_k \end{pmatrix} = \boldsymbol{t}_0 \in \boldsymbol{T}_0 \tag{7}$$

using any sharing table $\boldsymbol{t}_0$. We also define $n \cdot L$ samples of arbitrarily selected audio signal as $\boldsymbol{d}_i$. Define $\boldsymbol{d}'_i(1 \leq i \leq k - 1)$ as $\boldsymbol{d_i}$ decimated by $1/n$, resulting $\boldsymbol{d}'_i$ in a signal of L samples. We further define shared signals $\boldsymbol{s}_i(1 \leq i \leq k - 1)$,

which should be heard as the intended sounds, and a random noise-like shared signal $\boldsymbol{s}_k$ as

$$\begin{pmatrix} \boldsymbol{s}_1 \\ \boldsymbol{s}_2 \\ \vdots \\ \boldsymbol{s}_i \\ \vdots \\ \boldsymbol{s}_k \end{pmatrix} = \begin{pmatrix} \boldsymbol{p}_1^{(d'_{11})} \cdots \boldsymbol{p}_1^{(d'_{1j})} \cdots \boldsymbol{p}_1^{(d'_{1L})} \\ \boldsymbol{p}_2^{(d'_{21})} \cdots \boldsymbol{p}_2^{(d'_{2j})} \cdots \boldsymbol{p}_2^{(d'_{2L})} \\ \vdots \quad \ddots \quad \vdots \quad \ddots \quad \vdots \\ \boldsymbol{p}_i^{(d'_{i1})} \cdots \boldsymbol{p}_i^{(d'_{ij})} \cdots \boldsymbol{p}_i^{(d'_{iL})} \\ \vdots \quad \ddots \quad \vdots \quad \ddots \quad \vdots \\ \boldsymbol{p}_k^{(d'_{k1})} \cdots \boldsymbol{p}_k^{(d'_{kj})} \cdots \boldsymbol{p}_k^{(d'_{kL})} \end{pmatrix}, \tag{8}$$

where

$$\boldsymbol{d}'_i = \left(d'_{i1}\ d'_{i2} \ldots d'_{iL}\right) \tag{9}$$

$$d'_{kj} = o_j + \sum_{i=1}^{k-1} d'_{ij} \tag{10}$$

$$\boldsymbol{p}_i^{(0)} = \boldsymbol{p}_i \tag{11}$$

$$\boldsymbol{p}_i^{(1)} = \overline{\boldsymbol{p}_i}. \tag{12}$$

In Eq. (12), $\overline{\boldsymbol{p}_i}$ denotes the logical negation of $\boldsymbol{p}_i$.

Finally, we verify that it is possible to decrypt the original signal from all of $\boldsymbol{s}_i (1 \leq i \leq k)$. We define any column of Eq. (8) as

$$\boldsymbol{t}'_j = \begin{pmatrix} \boldsymbol{p}_1^{(d'_{1j})} \\ \boldsymbol{p}_2^{(d'_{2j})} \\ \vdots \\ \boldsymbol{p}_i^{(d'_{kj})} \end{pmatrix}. \tag{13}$$

Here, $\boldsymbol{t}'_j$ is a matrix obtained by exchanging columns of any matrix included in T_0 and taking logical negation $\sum_{i=1}^{k} d'_{ij}$ times. Sharing tables T_0 and T_1 have the property that when we exchange any column of any sharing table included in T_0 with its logical negation, it becomes one included in T_1, and vice versa. This property is expressed as

$$\boldsymbol{t}_b^{(I)} \in \boldsymbol{T}_{((I+b) \bmod 2)} \quad (b = 0, 1). \tag{14}$$

Thus,

$$\begin{aligned} \boldsymbol{t}'_j \in T_{(\sum_{i=1}^{k} d'_{ij}) \bmod 2} &= T_{(o_j + 2 \cdot \sum_{i=1}^{k-1} d'_{ij}) \bmod 2} \\ &= T_{o_j}. \end{aligned} \tag{15}$$

is obtained.

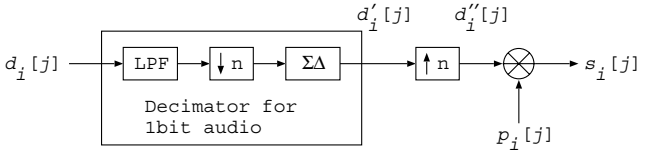

Fig. 5. Block diagram of the relationship between $\boldsymbol{d}_i$ and $\boldsymbol{s}_i$.

3 Verification of Decoy Sound

Figure 5 shows a block diagram to illustrate the method described in the previous section enabling $k-1$ shared signals to be heard as decoy sounds. In this figure, the input signal $d_i[j]$ and the output signal $s_i[j]$ are an intended decoy signal and an obtained shared signal, respectively. Here, we show that these two signals have close frequency characteristics. Notations of the variables used in this verification are summarized in Table 1.

Table 1. Notations of variables.

Variable	Description
$\boldsymbol{o}$	1-bit audio signal of the original signal {0, 1}
$\boldsymbol{s_i}$	1-bit audio signal of a shared signal {0, 1}
$s_i[j]$	Scalar representation of $\boldsymbol{s_i}$ ({0} is replaced with {−1})
$S_i[k]$	Frequency spectrum of $s_i[j]$
$S'_i[k]$	Frequency spectrum of $s_i[j]$ decimated by $1/n$
$\boldsymbol{p_i}$	Decomposition pattern {0, 1}
$p_i[j]$	Scalar representation of $\boldsymbol{p_i}$ ({0} is replaced with {−1})
$P_i[k]$	Frequency spectrum of $p_i[j]$ after zero-padding
$\boldsymbol{d_i}$	1-bit audio signal of a decoy signal {0, 1}
$d_i[j]$	Scalar representation of $\boldsymbol{d_i}$ ({0} is replaced with {−1})
$D_i[k]$	Frequency spectrum of $d_i[j]$
$\boldsymbol{d'_i}$	1-bit audio signal of d_i decimated by $1/n$ {0, 1}
$d'_i[j]$	Scalar representation of $\boldsymbol{d'_i}$ ({0} is replaced with {−1})
$D'_i[k]$	Frequency spectrum of $d'_i[j]$
$d''_i[j]$	$d'_i[j]$ up-sampled by n
$D''_i[k]$	Frequency spectrum of $d''_i[j]$

3.1 Analysis of the Time Domain

In Fig. 5, $d_i[j](j = 1, 2, \ldots, nL)$ is a signal obtained from a decoy signal $\boldsymbol{d_i} \in \{0, 1\}$ by replacing $\{0\}$ with $\{-1\}$. The output signal $s_i[j](j = 1, 2, \ldots, nL)$ is then represented using d'_i and $p_i[j]$ as follows:

$$\begin{aligned} s_i[j] &= (d'_i[1] \cdot \boldsymbol{p}'_i \quad d'_i[2] \cdot \boldsymbol{p}'_i \quad \ldots \quad d'_i[L] \cdot \boldsymbol{p}'_i) \\ \boldsymbol{p}'_i &= (p_i[1] \quad p_i[2] \quad \ldots \quad p_i[n]). \end{aligned} \tag{16}$$

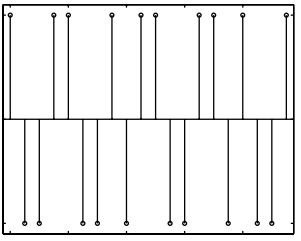

Fig. 6. An example of $s_i[j]$ actually obtained with the proposed method.

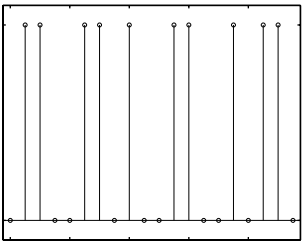

Fig. 7. Theoretically derived signal corresponding to Fig. 6.

For all components where $d'_i[j] = -1$, $\{1\}$ of $\boldsymbol{p}'_i$ is replaced with $\{-1\}$, and vice versa. Accordingly, $s_i[j]$ obtained by Eq.(16) is equivalent to $\boldsymbol{s}_i$ represented by Eq.(8). Note that the sign of the AC component of $s_i[j]$ is reciprocal of that of $\boldsymbol{s}_i$ because the sign of each component is inverted when $d_i[j]$ is -1, namely, d'_{ij} is 0.

Figure 6 shows an example of the waveform of $s_i[j]$ when $k = 3$, $L = 5$, $d'[j] = (1\ \ 1\ \ -1\ \ -1\ \ 1)$, and $p_i = (1\ \ 0\ \ 0\ \ 1)$. Accordingly, a $1 \times nL$ matrix $\boldsymbol{s}_i$ obtained with the parameters mentioned above is represented by

$$\begin{aligned} \boldsymbol{s}_i &= (\bar{\boldsymbol{p}}_i\ \ \bar{\boldsymbol{p}}_i\ \ \boldsymbol{p}_i\ \ \boldsymbol{p}_i\ \ \bar{\boldsymbol{p}}_i) \\ &= (0\ \ 1\ \ 1\ \ 0\ \ 0\ \ 1\ \ 1\ \ 0\ \ 1\ \ 0\ \ 0\ \ 1\ \ 1\ \ 0\ \ 0\ \ 1\ \ 0\ \ 1\ \ 1\ \ 0). \end{aligned} \tag{17}$$

Equation (17) can be depicted as in Fig. 7. Comparing these two figures, it is clear that the AC components of the signals are the same.

3.2 Analysis of the Frequency Domain

In Fig. 5, a decoy signal $d_i[j]$ with a length of $n \cdot L$ is fed to an LPF, a down-sampler by n and a $\Sigma\Delta$ converter. The role of the LPF is to remove high frequency components with the cut-off frequency of $f_c = f_s/2n$, not to introduce aliasing at the following down-sampler. In this operation, some audible sound components in a decoy signal might also be removed when the value of n is large. It is then down-sampled and requantized into a 1-bit sequence by the $\Delta\Sigma$ converter. Consequently, $d'[j]$ becomes a 1-bit audio signal of decoy sound $d_i[j]$ decimated by $1/n$.

Next, $d'[j]$ is up-sampled by n. When the signal obtained by this operation is defined as $d''[j]$, the frequency spectrum of $d''[j]$ becomes that of $d'[j]$ shrunk by $1/n$ on the frequency axis due to the up-sampling by n, and includes aliasing components.

Define the frequency spectrum of the signal obtained by zero-padding into $p_i[j]$ so as to have a length of $n \cdot L$ as in $P_i[k] (k = 0, 1, \ldots, nL-1)$, that of $d_i[j]$ as $D_i[k] (k = 1, 2, \ldots, nL-1)$, and that of $s_i[j]$ as $S_i[k]$. From Fig. 5, $S_i[k]$ can be derived by Eq. (18) as the product of $D''_i[k]$ and $P'_i[k]$.

$$S_i[k] = D''_i[k] \cdot P_i[k] \qquad (k = 0, 1, \ldots, L-1) \tag{18}$$

Since $D_i''[k](k = 0, 1, \ldots, nL - 1)$ can be expressed by $D_i'[k](k = 0, 1, \ldots, L - 1)$ with a cycle period of L, they are exactly the same within $(k = 0, 1, \ldots, L - 1)$. Since the audible sound components are retained after the sampling frequency becomes $f_s/2n$ because of the very high sampling frequency, $s_i'[j]$, which is obtained by decimating $s_i[j]$ by $1/n$, can be heard as the same as $s_i[j]$. Its frequency spectrum is represented as

$$\begin{aligned} S_i'[k] &= D_i''[k] \cdot P_i[k] \\ &= D'[k] \cdot P_i[k] \quad \text{for} \quad (k = 0, 1, \ldots, L - 1). \end{aligned} \tag{19}$$

Since $S_i'[k]$ and $D_i'[k]$ in Eq. (19) are the frequency spectra of $s_i[j]$ and $d_i[j]$ decimated by $1/n$, respectively, Eq. (19) can be represented using a LPF, $H_{LP}[k](k = 0, 1, \ldots, nL - 1)$, with a cut-off frequency of $f_s/2n$ as follows:

$$S_i[k] \cdot H_{LP}[k] = D_i[k] \cdot P_i[k] \cdot H_{LP}[k] \quad (k = 0, 1, \ldots, nL - 1). \tag{20}$$

Equation (20) states that a shared signal is heard as the decoy signal filtered by $P_i[k]$. Moreover, $p_i[k] = (1 \quad -1 \quad -1 \quad 1)$, which is employed in the previous section as an example, should provide sound without distortion because of its linear phase characteristics. However, a signal actually obtained by this method includes not only the decoy sound but also an additive noise. The reason for this may be the quantization noise brought about by the $\Delta\Sigma$ converter in Fig. 5 to requantize the multi-bit signal into a 1-bit audio signal. Namely, $D_i'[j]$ is $D_i[j]$ decimated by $1/n$ and added to a quantization noise. This effect is not considered in the previous discussion, and thus, further investigation is needed to verify it.

4 Conclusions

We propose a binary audio secret sharing method, BASS, which can share an original audio signal into k shared audio signals. A method to compose the $k-1$ shared signals heard as different and intended decoy sounds is also proposed. We have verified these sounds as the intended decoy sounds.

Acknowledgments

This study is partly supported by Strategic Information and Communications R&D Promotion Programme (SCOPE) by the Ministry of Internal Affairs and Communications of Japan.

References

1. Adi Shamir: How to share a secret. Communications of the ACM **22** (1979) 612–613
2. Moni Naor: Visual Cryptography. Advances in Cryptology - EUROCRYPT '94 LNCS 950 (1995) 1–12