Tally based Digital Audio Watermarking

Kotaro Sonoda and Shu Noguchi

Graduate School of Engineering, Nagasaki University.
14-1, Bunkyo-machi, Nagasaki-shi, Nagasaki 852-8521, Japan
sonoda-iihmsp17@cis.nagasaki-u.ac.jp

Abstract. In this paper, we propose a novel digital audio watermarking system inspired by the tally trade. In ordinary digital audio watermarking, a single stego signal is produced through the embedding process and the hidden message is extracted from the stego signal. In tally-based system we propose, multiple stego signals are tallied to produce and the hidden message which is extracted from the temporally mixed signal composed of the required number of stego tallies. When the extractor can't mix the required number of stego tallies, it misses the hidden message. The system's performances are evaluated in terms of the number of tallies (shares).

Keywords: Tally; Audio Secret Sharing; Time-spread Echo method; Digital Audio Watermarking;

1 Introduction

Due to the widespread presence of video sharing sites, anyone can easily share videos and music. Such illegal uploading, ignoring the copyright of contents, illegal copies without permission from the producer, illegal buying and selling correspond to infringement of copyright, etc. are regarded as problems. Copyright holders have used similar music search technology and embedding of digital watermarks of contents in order to detect such illegal usage and protect legitimate earnings from the distribution of the material.

Recently many algorithms for embedding watermarks in acoustic digital watermarks have been studied. However, even if the algorithm is complicated, once the algorithm becomes known, confidential information may be extracted. For that purpose, we have been taking countermeasures using secret parameters such as random numbers, but the random numbers are easily generated from predefined seed value. Once the seed value is leaked, the system becomes defenseless. In this paper, we aim to create a new audio tally which was inspired by tally-trade. A kind of tally is shown in figure 1, which was used for trade in medieval Japan. In tally trade, a watermark letter is written on a piece of wood, or paper, etc. and it is split/torn in two. The both parties held one piece each and then combined it to confirm it's validity. Here, a watermark letter is a kind of secret, but both of the corresponded connect pair and appropriate connect way are still required for evidence even if the watermark letter is known.

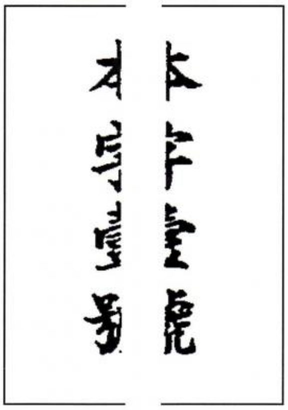

Fig. 1. Tally trade in medieval Japan

2 Tally-based watermarking system

In this section, we propose a tally-based watermarking system. The tally-based system aims to make the hidden message by decomposing it and embeddedding it to N pieces of stego-tally signals. The K pieces of stego-tally signals are mixed temporally and the hidden message is extracted by decoding the mixed stego signal.

Similar strategies are studied in a secret sharing system. The secret sharing is a distributed management method of hidden data and it was independently proposed by Blakley and Shamir in 1979. In $(K, N)-$ threshold secret sharing, secret datum s is divided into N shares of data. You can restore secret datum s when K shares are combined but cannot restore from $K-1$ shares or less.

With regard to an acoustic secret sharing method studied in order to make use of the characteristics of sound, Desmedt et al.[3] and Lin et al.[4] proposed an acoustic secret sharing of binary bit strings. Desmedt's method is $(2, 2)$-threshold secret sharing and they construct two shared signals such that any two shares have counter phases if the secret data is '0' or the same phases if the secret data is '1'. Lin extends Desmedt's method to $(2, N)$-threshold by representing a sharing bit with the main counter/same phase in consequent multiple sub-blocks. In Lin's study, this was implemented in Rock music.

These acoustic secret sharing methods are methods of acoustic information hiding. However these approaches do not consider the quality of the audio signals that are shared and results in quality degradiation. Therefore, we introduce this concept to an audio watermarking system in this report.

In some audio watermarking methods, the hidden message is spread by a pseudo-noise (PN) sequence (key), and the spread data is added or convolved with a host signal to produces a stego signal. The embedded hidden message is extracted by auto-correlation of time-sequences or the cepstrum of the stego data and the predefined pseudo-noise.

The PN sequence $p(n) \in \{-1, +1\}$ is used as a key here can be composed by the summation of N pieces of $p_k(n) \in \{-1, +1\}$ as $p(n) = \sum_{k=1}^{N} p_k(n)$. A signal group spread using $p_k(n)$ instead of $p(n)$ is the same as the signal diffused using $p(n)$ by addition synthesis. Therefore, by mixing and synthesizing all the stego signal groups $s_k(t)$ generated by using $p_k(n)$, the stego signal $s(t)$ generated by using $p(n)$ can be reproduced. Due to the noise immunity of the base watermarking technique, it is possible to detect the watermarks even when K or more pieces of $s_k(n)$ are used without using all N pieces. However, it should be noted that $\widehat{s(t)}_K$ obtained by mixing and synthesizing K items has an embedding strength of $1/K$ as compared with the base watermarking method.

2.1 Decomposion of pseudo-noise sequence

We decompose each element of the PN sequence $p(n)$ so that it is generated by addition of each element of N pieces of $p_k(n)$. Here, N is an odd number. This means that $(N + 1)/2$ pieces in $p_k(n)$ $(k = 1 \ldots N)$ are equal to $p(n)$ and the remaining $(N - 1)/2$ pieces are $-p(n)$. Therefore $(N - 1)/2$ pairs equaling $+p(n)$ and $-p(n)$ are canceled in summation and only one piece equaling $p(n)$ remains. You can choose any tally index k of $(N + 1)/2$ randomly for each element n, but you do not need to know how to choose this on the detection side.

For example, for an original PN sequence $p(n),\ n = 1 \ldots 5$ is decomposed to $N = 3$ sequences $p_k(n),\ k = 1 \ldots 3$ as following:

$$\begin{array}{c} p(n) \\ \downarrow \\ \begin{bmatrix} +1 \\ +1 \\ -1 \\ +1 \\ -1 \end{bmatrix} \end{array} = \begin{array}{c} \begin{array}{ccc} p_1(n) & p_2(n) & p_3(n) \\ \downarrow & \downarrow & \downarrow \end{array} \\ \begin{bmatrix} +1 & +1 & -1 \\ +1 & +1 & -1 \\ -1 & -1 & +1 \\ -1 & +1 & +1 \\ +1 & -1 & -1 \end{bmatrix} \end{array} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

3 Message detection performance in tally-system

Because the proposed tally-based system has not changed the base embedding scheme, the sound quality of stego-tally signals conform to the performance of the embedding scheme. On the other hand, the message detection performance is expected to depend on the proportion k/N. k is the number of mixing stego-tally signals, and N is the whole number of stego-tally signals.

In this section, we evaluate the message detection performance with several k values.

Remark 1. In this report, we use the Time-spread Echo method[5] as a base watermarking method. In the time-spread echo method, a PN sequence is used

Table 1. Parameters of Time-spread Echo method

Echo-gain β	0.08
Time-delay of watermark bit ”0” τ_0	4.5 ms
Time-delay of watermark bit ”1” τ_1	6.8 ms
Length of PN L_{PN}	1023 samples (23.2 ms)
Frame length L_{Fr}	8192 samples (185.7 ms)
embedding bit rates R_{emb}	5.38 bps

Table 2. SQAM tracks and genres

SQAM Track #	Genre
35	Glockenspiel
69	ABBA

as a component of the echo generating kernel with predefined time-delays representing the hidden bits. The component ”+1” yields same phase echo and the component ”-1” yields anti-phase echo. Therefore mixing the echo of a tally with the component ”+1” and one of another tally with component ”-1” cancels the echo. The parameters used in this method are listed in Table.1.

Remark 2. Payload data embedded in stego signal is 0/1 binary random bit sequence and it is published on the IHC website[6].

Remark 3. Tested host signals were monaural and selected from SQAM (Sound Quality Assessment Material)[7] published by EBU (European Broadcasting Union). Table.2shows the track numbers and genres.

3.1 K out of N tally System

We carried out the BER evaluation test with the proposed tally-based system with $N = 11$ tallies.

Figure.2 shows the result of BER in the case of k out of 11 tallies for SQAM Track #35 (a solo performance). The result shows that BER exponentially decreases and mixing whole tallies ($K = 11$) achieved under 10% and $K = 6$ ($K/N = 0.6$) or more tallies are required for a BER of under 20%.

While in the case of using another type of music (club music), similar results are achieved as shown in Figure.3. $K = 7$ ($K/N = 0.6$) or more tallies are required for under a BER of less than 20% for this case.

Regarding these results, mixing whole tallies ($K = N$ condition) resulted in missing a few hidden bits and BER is not zero. It is assumed that the embedding strength is insufficient to detect successfully after mixing tallies. The echo gain $\beta = 0.08$ set in the time-spread echo method was degraded to $\hat{\beta} = \beta/K$ in mixing K tallies. In condition of $K = N = 11$, that $\hat{\beta}$ is degraded to 0.08/11.

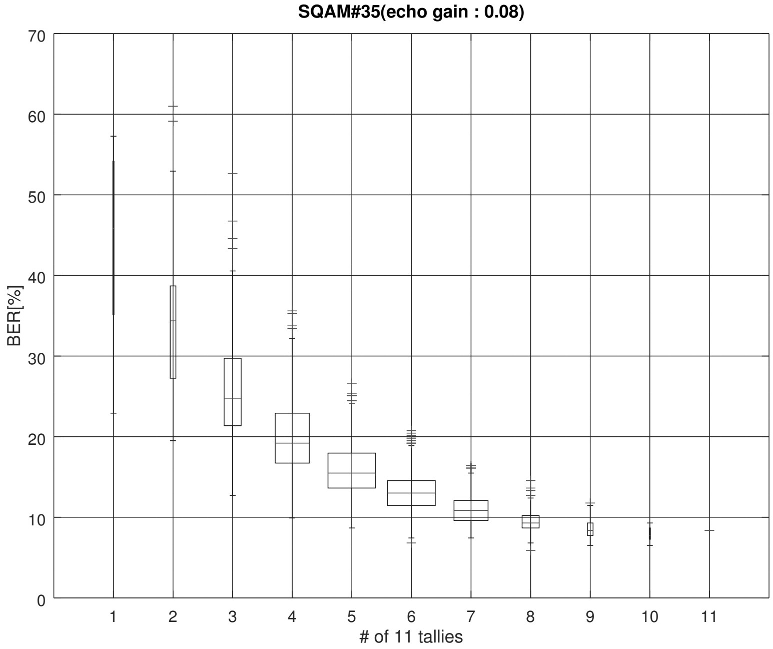

Fig. 2. BER in k out of 11 tallies; SQAM Track:35, echo gain:0.08

Larger echo-gain, longer PN sequences, or smaller sharing size are required for a lower BER.

We will expect the required relationships between the embedding strength β, threshold K and size of shares N from the experimental results in section 3.1.

The no-bit-error least strength of our base time-spread echo method is represented by $S_0 = \beta^2 L$ depending on a length of PN sequence (L) and a echo gain (β). In our N tally-based system, the echo can detect from N-mixed signal. N-mixed signal has N times bigger signal component and same level of echo component compared with normal stego signal. Thus, the tally signal must be embedded by the strength of greater than $N \cdot S_0 = N\beta^2 L$.

Here, echo gain β is usually bounded by required imperceptibility. Therefore, our system required to embed the echoes of N times longer PN sequence than base time-spread echo method.

Reason why the result of BER is not zero in $K = N = 11$ in figures of section 3.1 might because the experiment condition: $N = 11, \beta = 0.08, L = 1023$, has not suffered the relationship.

From the result of section.3.1, our tally-based system can detect many of bits from K-mixed signal ($K < N$). K-mixed signal has statistically correct $L_K = L \cdot K \cdot \frac{N+1}{2N}$ components of embedded PN sequence. Therefore the embedding

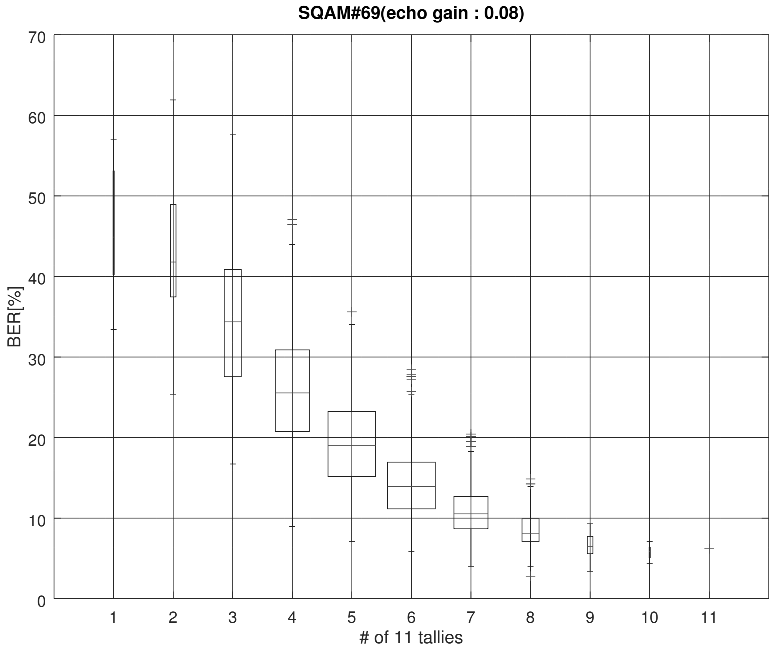

Fig. 3. BER in k out of 11 tallies; SQAM Track:69, echo gain:0.08

strength is $\beta^2 L_K$. Supposing all-bit-error threshold strength S_{100}, (K, N) tally system requires $\beta^2 L_{K-1} \leq S_{100}$ and $\beta^2 L_K \geq S_0$.

The practical setting of these parameters are feature works.

4 Discussion

Proposed our tally-based watermarking system is managing a hidden message by N tally audio signals in while conventional watermarking system managing by just one stego signal. Moreover, in the proposed method, larger echo-gain and longer PN sequence, those deteriorate the audio quality, are required for attaining the same correct detection rates with the conventional time-spread echo watermarking. However, the risk of leakage is expected to ramp down compared with the conventional one, because the proposed one requires the cooperative other tally signal physically for message detection.

5 Conclusion

In this report, we propose a novel tally-based audio watermarking system. Our system decomposes a PN sequence, which is used as a spread kernel for payload bits in a hybrid audio watermarking method, for N tally sequences. A stego signal watermarked using the tally sequence restores the stego signal to the original watermarking method by mixing. The results of BER for some music sources show that the proposed system supports 60% whole tallies to restore the hidden message.

References

1. G.-R. Blakley: Safeguarding cryptographic keys. Managing Requirements Knowledge, International Workshop on IEEE Computer Society, p.313, (1899)
2. A. Shamir: How to share a secret. Communications of the ACM, vol.22, no.11, pp.612–613, (1979)
3. Y. Desmedt, S. Hou, and J.-J. Quisquater: Audio and optical cryptography. Advances in Cryptography ASIACRYPT'98, Springer, pp.392–404, (1998)
4. C.-C. Lin, C.-S. Laih, and C.-N. Yang: New audio secret sharing schemes with time division technique. J. Inf. Sci. Eng., vol.19, no.4, pp.605–614, (2003)
5. B.S. Ko, R. Nishimura and Y. Suzuki : Time-spread echo method for digital audio watermarking using PN sequences. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP), pp. II-2001-II-2004 (2002)
6. IHC Evaluation Criteria and Competition, `http://www.ieice.org/iss/emm/ihc/IHC_criteriaVer5.pdf`, (2017)
7. European Broadcasting Union (EBU) : SQAM (Sound Quality Assessment Material). `http://tech.ebu.ch/publications/sqamcd` (2008)