This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Frequency Spectrum Modification Process-Based Anti-Collusion Mechanism for Audio Signals

Juan Zhao, *Graduate Student Member, IEEE*, Tianrui Zong, *Member, IEEE*, Yong Xiang, *Senior Member, IEEE*, Guang Hua, *Senior Member, IEEE*, Xinyu Lei, *Member, IEEE*, Longxiang Gao, *Senior Member, IEEE*, and Gleb Beliakov, *Senior Member, IEEE*

***Abstract*—The collusion attack combines multiple multimedia files into one new file to erase the user identity information. The traditional anti-collusion methods (which aim to trace the traitors) can defend the collusion attack, but they cannot well defend some hybrid collusion attacks (e.g., a collusion attack combined with desynchronization attacks). To address this issue, we propose a frequency spectrum modification process (FSMP) to defend the collusion attack by significantly downgrading the perceptual quality of the colluded file. The severe perceptual quality degradation can demotivate the attackers from launching the collusion attack. Because FSMP is orthogonal to the existing traitor-trace-based methods, it can be combined with the existing methods to provide a double-layer protection against different attacks. In FSMP, after several signal processing procedures (e.g., uneven framing and smoothing), multiple signals (called FSMP signals) can be generated from the host signal. Launching collusion attack using the generated FSMP signals would lead to the energy disturbance and attenuation effect (EDAE) over the colluded signals. Due to the EDAE, FSMP can significantly degrade the perceptual quality of the colluded audio file, thereby thwarting the collusion attack. In addition, FSMP can well defend different hybrid collusion attacks. Theoretical analysis and experimental results confirm the validity of the proposed method.**

***Index Terms*—Audio watermarking, collusion attack, desynchronization attacks, perceptual quality degradation.**

Manuscript received September 7, 2021; revised December 21, 2021; accepted February 26, 2022. This work was supported in part by the Australian Research Council under Grant LP170100458. This article was recommended by Associate Editor S. Ozawa. *(Corresponding authors: Yong Xiang; Longxiang Gao.)*

Juan Zhao, Yong Xiang, and Gleb Beliakov are with the School of Information Technology, Deakin University, Geelong, VIC 3125, Australia (e-mail: zhaojua@deakin.edu.au; yong.xiang@deakin.edu.au; gleb.beliakov@deakin.edu.au).

Tianrui Zong was with the School of Information Technology, Deakin University, Geelong, VIC 3125, Australia. He is now with the Department of Research, CNPIEC KEXIN Ltd., Beijing 100020, China (e-mail: zongziqin@hotmail.com).

Guang Hua is with the School of Electronic Information, Wuhan University, Wuhan 430072, China (e-mail: ghua@whu.edu.cn).

Xinyu Lei is with the Department of Computer Science, Michigan Technological University, Houghton, MI 49931 USA (e-mail: xinyulei@mtu.edu).

Longxiang Gao is with the Shandong Computer Science Center (National Supercomputer Center in Jinan), Qilu University of Technology (Shandong Academy of Sciences), Jinan 250000, China (e-mail: longx.gao@gmail.com).

Digital Object Identifier 10.1109/TCYB.2022.3156973

I. Introduction

In RECENT years, the developments of the multimedia technology and communication system have facilitated the proliferation of multimedia files. With the prevalence of electronic devices (e.g., smartphones, tablets, and computers), the multimedia files can be easily accessed and maliciously manipulated, which leads to serious security problems [1]–[3]. The copyright infringement problem [4] is one of the serious security problems. To tackle it, a commonly used solution is digital watermarking. Digital watermarking [5] enables to embed some ownership information in a multimedia file so that the embedded information can be extracted to prove the ownership of the file. The embedded ownership information typically has two types: 1) watermark and 2) fingerprint. In this article, the watermark is referred to as the general copyright information, while the fingerprint is referred to as the unique user identity information. Once the copyright is infringed, the embedded information can be extracted for copyright protection purposes (e.g., proof of ownership). Digital watermarking can be applied on different types of multimedia files, including image [6]–[8], video [9]–[11], and audio [12]–[14]. Among different types of files, audio files are our research focus.

The collusion attack toward audio combines multiple audio files with different fingerprints to produce a colluded audio file. From the colluded file, the fingerprint is hard to be detected and extracted. Thus, in the presence of a collusion attack, the fingerprint loses its ability to uniquely identify the audio file user. Due to the absence of effective methods to defend such an attack, there are strong financial incentives for some illegal users (i.e., traitors) to cooperatively launch the collusion attack without worrying about any punishment. Hence, it is highly demanded to design effective anti-collusion methods.

Most of the existing methods, such as the reversible data hiding methods [15]–[17] that aim to recover the original signal when extracting the embedded information, cannot survive collusion attacks. Some methods aim to tackle collusion attacks [18], [19], but they are not blind. The existing blind anti-collusion methods mainly work by offering the capability to trace the traitors from the colluded file. These methods can be divided into two categories: 1) the coded fingerprinting-based methods [20]–[23] and 2) the independent fingerprinting-based methods [24]–[26]. First, the coded fingerprinting-based methods focus on the design of the collusion-resistant fingerprint codes (e.g., collusion-secure

2168-2267 © 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: Michigan Technological University. Downloaded on March 24,2022 at 02:21:05 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

codes in [20], Tardos codes in [21], symmetric Tardos codes in [22], and Nuida codes in [23]). The collusion-resistant fingerprint codes support to trace the traitors from the colluded files. Second, the independent fingerprinting-based methods [24]–[26] focus on the design of the fingerprint codes embedding and extraction algorithms, which support traitors tracing.

The above-mentioned anti-collusion methods can only defend the collusion attack alone (called simple collusion attack), but they cannot survive if the collusion attack is combined with other common attacks (called hybrid collusion attacks). More specifically, the existing methods cannot well defend two types of hybrid collusion attacks: 1) collusion-signal-processing attacks and 2) collusion–desynchronization attacks. The former combines a collusion attack with common signal processing attacks (e.g., low-pass filtering (LPF) attack and MP3 compression attack), while the latter combines a collusion attack with common desynchronization attacks (e.g., time scaling attack and pitch scaling attack). Unfortunately, the existing methods fail to trace the traitors when encountering both types of hybrid collusion attacks, especially the collusion–desynchronization attacks. We note that researchers have developed robust watermarking methods [13], [27], [28] to cope with desynchronization attacks. However, these methods are vulnerable to the collusion attack, so they cannot resist the collusion–desynchronization attacks either.

The method in [29] tackles collusion attacks by significantly reducing the perceptual quality of the colluded audio file. This new perspective is promising, as the perceptual quality degradation removes the financial incentives and motivations of the traitors, thereby thwarting the collusion attack. However, there are three problems in [29]. First, the perceptual quality degradation mechanism is ineffective. Second, it has weak performance against the collusion attack when the number of traitors is odd. Third, it cannot generate enough copies for authorized distribution in real-world applications. Zong *et al.* [30] attempted to solve the problems in [29], but they also failed to generate enough copies for distribution. To improve the anti-collusion performance and generate enough copies for authorized distribution, we propose a frequency spectrum modification process (FSMP), which is an audio signal preprocessing process, to generate multiple signals[1] (called FSMP signals). The FSMP signals can resist both simple and hybrid collusion attacks, including the challenging collusion–desynchronization attacks. Besides, FSMP is orthogonal to the existing methods (i.e., tracing the traitor); thus, it can synthesize with the existing methods, providing a double-layer shield against attacks. More specifically, in FSMP, the host audio file is first divided into multiple segments and each segment is transformed into the discrete cosine transform (DCT) domain. Then, the selected low-frequency coefficients of each segment are modified to generate multiple FSMP signals. These modifications will lead to the energy disturbance and attenuation effect (EDAE) over the low-frequency components in the colluded signal. The EDAE not only helps to downgrade the perceptual quality but also helps to resist common signal processing attacks and common desynchronization attacks.

[1]The audio file can be treated as a segment of an audio signal, so we use the term "file" and "signal" interchangeably in this article.

TABLE I
COMPARISON BETWEEN THE EXISTING METHODS AND FSMP (●: DEFEND, ◐: PARTIAL DEFEND, ○: NOT DEFEND, NA: NOT APPLICABLE)

Attacks	[20]–[26]	[13], [27], [28]	[29], [30]	FSMP	FSMP + Watermarking
Simple collusion attack	●	○	◐	●	●
Desynchronization attacks	○ or Na	●	Na	Na	●
Collusion-signal-processing attacks	◐	○	◐	●	●
Collusion-desynchronization attacks	○	○	◐	●	●
Na: desynchronization attacks are not applicable to the methods without using watermarking.					

Our proposed FSMP has several key novelties compared with the existing methods. By using the EDAE and the novel strategies such as uneven framing and smoothing, the proposed FSMP has achieved three major improvements over [29]: 1) the use of EDAE significantly improves the perceptual quality degradation performance; 2) the uneven framing strategy guarantees that the EDAE will occur in the colluded signals, regardless of the number of traitors; and 3) the smoothing step facilitates the generation of sufficient copies for authorized distribution. Compared with the method in [30], apart from the significant increase in the number of copies, the proposed FSMP is also the first anti-collusion study that provides mathematical analysis to theoretically validate the effectiveness against various types of hybrid collusion attacks.

As aforementioned, the traitor-trace-based methods in [20]–[26] cannot defend desynchronization attacks. The robust watermarking methods in [13], [27], and [28] cannot defend collusion attacks. Unlike the existing methods, our proposed FSMP combining with watermarking can defend all types of attacks. For the sake of readability, we present the comparison results between the proposed FSMP and the existing methods in [13] and [20]–[30] in Table I. It can be seen from Table I that the proposed FSMP has superior features compared with the other methods.

In summary, this article makes the following four main contributions.

1) The proposed FSMP tackles collusion attacks by introducing the EDAE in the colluded signals to significantly degrade their perceptual quality, where the perceptual quality degradation mechanism is fundamentally different from the existing anti-collusion methods. By exploiting the HAS, the proposed FSMP is effective against both simple and hybrid collusion attacks, regardless of the number of traitors.
2) To further improve the performance, we propose the uneven framing strategy and the smoothing procedure that are well-designed and tailored for the proposed FSMP framework. These two mechanisms can greatly enhance the EDAE, largely preserve the perceptual quality of the FSMP signals, and dramatically increase the total number of the FSMP signals without compromising the perceptual quality.
3) We provide mathematical analysis to theoretically validate the effectiveness of the proposed FSMP against

Authorized licensed use limited to: Michigan Technological University. Downloaded on March 24,2022 at 02:21:05 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

various types of hybrid collusion attacks. To our best knowledge, this is the first anti-collusion study that provides such analysis.

4) Since the proposed FSMP can preserve the perceptual quality of the FSMP signals, the proposed FSMP can be combined with the existing traitor-trace-based methods to provide double-layer protection against attacks. We combine our proposed FSMP with one of the leading-edge robust watermarking techniques and compare the performance with both watermarking and fingerprinting methods. Experimental results validate the superiority of the proposed method.

The remainder of this article is organized as follows. The related work is illustrated in Section II. Section III elaborates the proposed FSMP. The effectiveness of FSMP is mathematically and experimentally demonstrated in Section IV. Section V presents that the proposed FSMP is superior to the traditional anti-collusion methods and robust watermarking technique. Finally, Section VI concludes this article.

II. Related Work

A. Coded Fingerprinting-Based Methods

Boneh and Shaw [20] proposed the collusion-secure codes using the randomization. However, they used the randomization in a restricted way as they only randomly picked a permutation for each column of the n-secure code matrix they constructed. Tardos [21] fully exploited the randomization to construct the binary fingerprint codes based on a bias distribution, which is referred to as Tardos codes in this article. The Tardos codes can trace the traitors through a scoring function with a universal threshold. Unfortunately, these codes require extremely long code lengths to effectively resist the collusion attack. Later, some researchers are dedicated to shortening the length of the fingerprint code. Škorić *et al.* [22] introduced a new construction of the fingerprint code, which is similar to the construction of the Tardos codes. It achieves a shorter fingerprint code length by expanding the alphabets from binary to q-ary. Nuida *et al.* [23] adopted a symmetric bias distribution called Gauss–Legendre distribution and introduced a new scoring function to further reduce the length of Tardos codes, which is referred as Nuida codes in this article.

Since the coded fingerprinting-based methods consider little about the watermark embedding and detection procedures, they have to be applied with the existing robust audio watermarking techniques to resist the collusion–desynchronization attacks. The existing robust watermarking methods exploit the statistical properties of the audio signal. They have to compromise the embedding capacity for the robustness against desynchronization attacks. Therefore, the fingerprint code length should not be too long. Although much effort has been done to reduce the lengths of the fingerprint codes, they are still too long for the robust watermarking methods. As a result, the coded fingerprint-based methods cannot be effectively applied with the robust watermarking methods, which makes the coded fingerprint-based methods vulnerable to the collusion–desynchronization attacks.

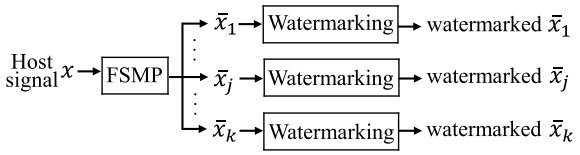


Fig. 1. Overview of the FSMP-based anti-collusion mechanism.

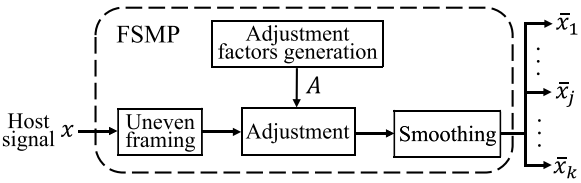


Fig. 2. Procedures of FSMP.

B. Independent Fingerprinting-Based Methods

The independent fingerprinting-based methods trace the traitors by using the orthogonal fingerprint codes, which are developed from the spread spectrum (SS)-based watermarking method in [24]. The traitors are identified based on the correlation between the colluded signal and the original fingerprint codes. Kirovski *et al.* designed an anti-collusion system in [25] where the watermark detection key is different from the secret key used for watermark embedding. This system can prevent the traitors from removing the watermark information. Wang *et al.* [26] adopted the independent and identically distributed (iid) Gaussian noise sequences as the fingerprint codes and analyzed the performance of the SS-based anti-collusion schemes under both of the maximum detector and the threshold detector.

The SS-based watermarking methods cannot resist desynchronization attacks, as desynchronization attacks can break the alignment between the encoder and the decoder. Since the independent fingerprinting-based methods adopt the SS-based watermark embedding and decoding mechanisms, they are also vulnerable to the collusion–desynchronization attacks.

III. Proposed FSMP

The overview of the FSMP-based anti-collusion mechanism is shown in Fig. 1. In the FSMP-based mechanism, the host signal x is first adjusted by FSMP to generate multiple FSMP signals in order to produce the EDAE in the colluded signal for anti-collusion purposes. Then, the robust watermarking techniques can be applied to the FSMP signals to further enhance the copyright protection ability of the proposed mechanism. Note the traditional anti-collusion methods do not have FSMP and the fingerprint codes are directly embedded in the host signal.

In the rest of this section, we introduce FSMP in detail, as shown in Fig. 2. FSMP consists of four main procedures, including: 1) uneven framing; 2) adjustment factors generation; 3) adjustment; and 4) smoothing.

A. Uneven Framing

The uneven framing procedure is shown in Fig. 3. This procedure is adopted to enhance the EDAE in the colluded

Authorized licensed use limited to: Michigan Technological University. Downloaded on March 24,2022 at 02:21:05 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

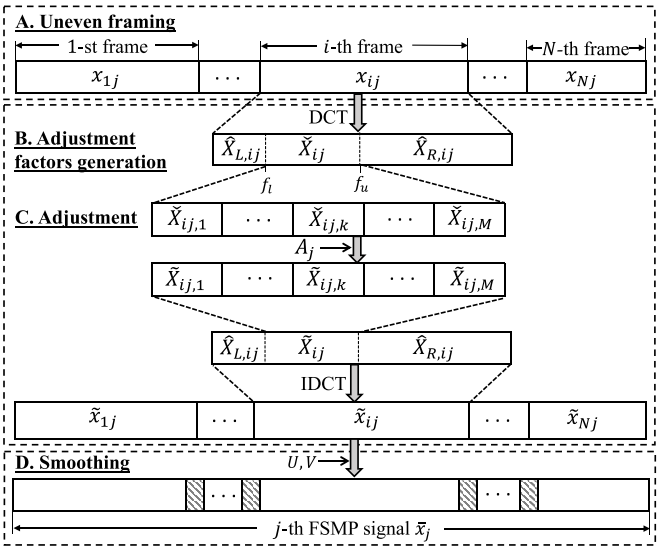


Fig. 3. Detailed procedures to generate the jth FSMP signal.

signals and its effectiveness will be analyzed later in Section IV-C in detail. Let x denote the host audio signal. The host signal x consists of L samples. To generate the jth FSMP signal, x is divided into N frames, which can be expressed as

$$x = \left[x_{1j}, x_{2j}, \ldots, x_{Nj}\right] \tag{1}$$

where x_{ij} is the ith frame in the jth FSMP signal, $i \in [1, N]$. For the first $N-1$ frames, the uneven framing strategy is achieved by randomly selecting the length of the ith frame in the jth FSMP signal $l_{i,j}$ from $[l_{lij}, l_{uij}]$, which satisfies

$$\begin{cases} l_{lij} \;\geq\; (1-\alpha)\cdot\frac{L}{N} \\ l_{uij} \;\leq\; (1+\alpha)\cdot\frac{L}{N} \end{cases} \tag{2}$$

where $\alpha \in (0, 1)$ is the parameter that controls the lengths of the frames and $i \in [1, N-1]$. Let $L_{i,j}$ denote as the total length of the first i frames in the jth FSMP signal, which can be expressed as

$$L_{i,j} = \sum_{k=1}^{i} l_{k,j}. \tag{3}$$

The length of the Nth frame $l_{N,j}$ can be calculated by

$$l_{N,j} = L - L_{N-1,j}. \tag{4}$$

To avoid a huge difference between $l_{N,j}$ and $l_{N-1,j}$, we constrain the value of $L_{i,j}$ for the first $N-1$ frames as

$$i\cdot\frac{L}{N} - T \leq L_{i,j} \leq i\cdot\frac{L}{N} + T \tag{5}$$

where T is the threshold to control the lengths of frames. Substituting (3) into (5), we have

$$i\cdot\frac{L}{N} - T - L_{i-1,j} \leq l_{i,j} \leq i\cdot\frac{L}{N} + T - L_{i-1,j}. \tag{6}$$

By combining (2) and (6), the lower boundary l_{lij} and upper boundary l_{uij} for the ith frame can be formulated as

$$\begin{cases} l_{lij} = \max\left\{(1-\alpha)\cdot\frac{L}{N},\; i\cdot\frac{L}{N} - T - L_{i-1,j}\right\} \\ l_{uij} = \min\left\{(1+\alpha)\cdot\frac{L}{N},\; i\cdot\frac{L}{N} + T - L_{i-1,j}\right\} \end{cases} \tag{7}$$

where $\max\{\cdot\}$ is the maximum value operator, $\min\{\cdot\}$ is the minimum value operator, and $i \in [1, N-1]$.

B. Adjustment Factors Generation

In order to plant the EDAE into the colluded signal to degrade the perceptual quality, the DCT coefficients of the host signal will be modified by a sequence of adjustment factors for each FSMP signal. To generate these adjustment factors, we introduce an adjustment matrix as A which takes values from $\{-1, 1\}$ with size $P \times N_c$, where P is the length of each adjustment sequence and N_c is the total number of copies required in the real-world application.

To generate A, we first design a linear codebook $C(P, d)$, where d is the minimum hamming distance between any two codewords. Then we randomly select N_c codewords from C to form the codebook matrix C_s with size $P \times N_c$. Finally, A is generated by mapping the values in C_s from $\{0, 1\}$ to $\{-1, 1\}$, which can be expressed as

$$A(i, j) = 2\cdot C_s(i, j) - 1 \tag{8}$$

where $i = 1, 2, \ldots, P$ and $j = 1, 2, \ldots, N_c$. The generated A can also be represented as

$$A = \left[A_1, A_2, \ldots, A_{N_c}\right] \tag{9}$$

where A_k is the adjustment sequence for the kth FSMP signal, $k = 1, 2, \ldots, N_c$. Then the adjustment factors in the kth adjustment sequence can be formulated as

$$A_k = \left[a_{k,1}, a_{k,2}, \ldots, a_{k,P}\right]^T \tag{10}$$

where $a_{k,m}$ is the mth adjustment factor in A_k, $m = 1, 2, \ldots, P$.

Note that this procedure only needs to run once to generate A, which can be used for multiple times. Once A is generated, it can be used to generate all the FSMP signals. It is also worth noting that the maximum number of codewords in C is determined by P and d. For example, when $d = 1$, the maximum value of N_c is 2^P; when $d = 2$, N_c can reach up to 2^{P-1}. Besides, when the value of P is fixed, a larger d value leads to the stronger EDAE and consequently a severer perceptual quality degradation in the colluded signal. To balance the total number of codewords and the strength of the EDAE, in this article, we choose the value of d as 2.

C. Adjustment

On input the frames (output from the uneven framing procedure) and A, the adjustment procedure intends to generate the FSMP signals for anti-collusion purposes. The detailed adjustment procedure is shown in Fig. 3. In this procedure, DCT [31] is first applied on each frame. Suppose the DCT is performed in the ith frame of the jth FSMP signal, then its frequency counterpart can be obtained as

$$X_{ij}(k) = h(k)\sum_{n=1}^{l_{i,j}} x_{ij}(n)\cos\frac{\pi(2n-1)(k-1)}{2l_{i,j}} \tag{11}$$

Authorized licensed use limited to: Michigan Technological University. Downloaded on March 24,2022 at 02:21:05 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

where $X_{ij}(k)$ represents the kth DCT coefficient of x_{ij}, $k = 1, 2, \ldots, l_{i,j}$, and

$$h(k) = \begin{cases} \frac{1}{\sqrt{l_{i,j}}}, & \text{if } k = 1 \\ \frac{2}{\sqrt{l_{i,j}}}, & \text{otherwise.} \end{cases} \tag{12}$$

According to [32], the HAS is sensitive to the low-frequency part of an audio signal. Therefore, we adjust the low-frequency coefficients of the host signal in order to produce a strong EDAE in the colluded signal. Let $\check{X}_{ij}$ denote as the low-frequency coefficients selected in the ith frame for the jth FSMP signal generation, which can be expressed as

$$\check{X}_{ij} = \left[X_{ij}(f_l), X_{ij}(f_l + 1), \ldots, X_{ij}(f_u)\right] \tag{13}$$

where f_l is the lower boundary of the selected DCT coefficients and f_u is the upper boundary of the selected DCT coefficients. The DCT coefficients which are not selected can be expressed as

$$\hat{X}_{L,ij} = \left[X_{ij}(1), X_{ij}(2), \ldots, X_{ij}(f_l - 1)\right] \tag{14}$$

and

$$\hat{X}_{R,ij} = \left[X_{ij}(f_u + 1), X_{ij}(f_u + 2), \ldots, X_{ij}(l_{ij})\right] \tag{15}$$

where $\hat{X}_{L,ij}$ and $\hat{X}_{R,ij}$ are the left and right unselected DCT frequency coefficients for the ith frame of the jth FSMP signal, respectively. We further divide $\check{X}_{ij}$ into M blocks with equal length, where $M = P/N$. Then, $\check{X}_{ij}$ can be expressed as

$$\check{X}_{ij} = \left[\check{X}_{ij,1}, \check{X}_{ij,2}, \ldots, \check{X}_{ij,M}\right] \tag{16}$$

where $\check{X}_{ij,k}$ is the selected DCT coefficients in the kth block of the ith frame in the jth FSMP signal, $k = 1, 2, \ldots, M$. Finally, we adjust $\check{X}_{ij,k}$ by the adjustment factors generated in Section III-B as

$$\tilde{X}_{ij,k} = \check{X}_{ij,k} \cdot a_{j,(i-1)\cdot N+k} \tag{17}$$

where $\tilde{X}_{ij,k}$ is the adjusted DCT coefficients in the kth block of the ith frame in the jth FSMP signal, $k = 1, 2, \ldots, M$. Apply (17) to all of the M blocks, then the adjusted selected DCT coefficients in the ith frame of the jth FSMP signal can be presented as

$$\tilde{X}_{ij} = \left[\tilde{X}_{ij,1}, \tilde{X}_{ij,2}, \ldots, \tilde{X}_{ij,M}\right]. \tag{18}$$

Combining (18) with (14) and (15), we can have

$$\bar{X}_{ij} = \left[\hat{X}_{L,ij}, \tilde{X}_{ij}, \hat{X}_{R,ij}\right] \tag{19}$$

where $\bar{X}_{ij}$ is the modified DCT coefficients in the ith frame of the jth FSMP signal. Apply the inverse DCT (IDCT) on $\bar{X}_{ij}$ to obtain the modified ith frame of the jth FSMP signal $\tilde{x}_{ij}$ in the time domain as

$$\tilde{x}_{ij} = \text{IDCT}\left(\bar{X}_{ij}\right) \tag{20}$$

where IDCT$(\cdot)$ is the IDCT operator. After performing IDCT on each frame, finally the jth FSMP signal can be generated by

$$\tilde{x}_j = \left[\tilde{x}_{1j}, \ \tilde{x}_{2j}, \ldots, \ \tilde{x}_{Nj}\right]. \tag{21}$$

TABLE II
ODG LEVELS AND ITS DESCRIPTION

ODG level	ODG value	Description
1	0.0	Imperceptible
2	-1.0	Perceptible, but not annoying
3	-2.0	Slightly annoying
4	-3.0	Annoying
5	-4.0	Very annoying

D. Smoothing

After the previous procedures, the junction between two consecutive frames may occur signal discontinuity [33]. In order to improve the perceptual quality of the FSMP signal, FSMP adopts the moving average filter [34] to smooth these discontinuities. The smoothing procedure for the jth FSMP signal is shown in Fig. 3. Assume a discontinuity occurs at $\tilde{x}_j(k)$, where $\tilde{x}_j(k)$ is the kth sample in $\tilde{x}_j$, the smoothed sample can be calculated by

$$\bar{x}_j(k) = \frac{1}{2V+1} \sum_{v=-V}^{V} \tilde{x}_j(k+v) \tag{22}$$

where $\bar{x}_j(k)$ is the smoothed kth sample in $\tilde{x}_j$ and $2V+1$ is the span. Apply (22) on $[\tilde{x}_j(k-U), \ \tilde{x}_j(k-U+1), \ldots, \ \tilde{x}_j(k+U)]$ to obtain the smoothed junction part between two consecutive frames, where U is the parameter controlling the length of the junction part. After the smoothing procedure, finally, we can obtain the smoothed jth FSMP signal $\bar{x}_j$. In Fig. 3, the shaded areas represent the smoothed junction parts between two consecutive frames.

IV. EFFECTIVENESS OF FSMP

A. Perceptual Quality Evaluation of the FSMP Signals

We adopt the widely used perceptual evaluation of audio quality (PEAQ) algorithm [35] to measure the perceptual quality of the FSMP signals. The PEAQ algorithm returns a mark called the objective difference grade (ODG) which ranges from -4 to 0. The ODG has 5 levels, which are illustrated in Table II. According to [36], if the ODG value is larger than -1, the perceptual quality degradation is imperceptible; If the ODG value is less than -3, the perceptual quality degradation is annoying. We randomly chose 160 audio clips as host signals to test the perceptual quality of the FSMP signals. All of the audio clips have a duration of 60 s, a sample rate of 44 100 Hz, and a resolution of 16 bits. The audio clips contain different genres, including jazz, classic, folk, funk, new age, pop, hip-hop, and rock. All of the experiments are performed using a Windows 10 laptop with an Intel Core-i7-8650U 2.11-GHz processor and 16.00-GB RAM.

To generate enough FSMP signals for authorized distribution, we set $N = 12$, $\alpha = 0.7$, $T = (L/2N)$, $M = 3$, $U = 50$, and $V = 15$. According to the digital single award requirements by the Recording Industry Association of America (RIAA) [37], the platinum award and the diamond award for a piece of music are 1×10^6 and 1×10^7 units, respectively. To guarantee the proposed FSMP can produce enough signals for real-world applications, we measure the average ODG values of the FSMP signals for each genre of the audio clips when

Authorized licensed use limited to: Michigan Technological University. Downloaded on March 24,2022 at 02:21:05 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE III
AVERAGED ODG VALUES OF THE FSMP SIGNALS UNDER DIFFERENT N_c

N_c	2×10^6	4×10^6	6×10^6	8×10^6	1×10^7
ODG values	-0.144	-0.154	-0.160	-0.165	-0.177

N_c is up to 1×10^7, which is standard for the diamond award. Table III shows the average ODG values of the FSMP signals under different N_c. From Table III, it is easy to observe that all of the averaged ODG values are around -0.2 even though the total number of the FSMP signals reaches up to $N_c = 1 \times 10^7$, which implies the perceptual quality of the FSMP signals is largely preserved. Therefore, our proposed method can generate enough signals for authorized distribution without compromising the perceptual quality of the FSMP signals.

B. EDAE Removal Attack

The most straightforward attack for the proposed FSMP is the EDAE removal attack. In the attack, the traitor attempt to remove the EDAE from the colluded signal. The removal of EDAE can recover the perceptual quality of the colluded signal. If the EDAE is removed, our proposed method will fail to protect the audio files from collusion attacks. Our proposed EDAE has strong robustness against the collusion attack because the traitors can never remove the EDAE in the colluded signals. There are two reasons.

1) Since the proposed FSMP does not require a detection phase at the receiver end, there is no need to transmit or even store the parameters used in FSMP. Therefore, all of the parameters, such as N, f_l, f_u, and $C(P, d)$, can be considered as secret keys in the proposed FSMP. Without the access to these secret keys, the traitors cannot remove the EDAE from the colluded signals.
2) Even if the traitors obtain the knowledge of the secret keys, due to the use of the uneven framing strategy, the length of each frame $l_{i,j}$ is still not accessible to the traitors, as $l_{i,j}$ is a random value chosen from $[l_{lij}, l_{uij}]$. Without the knowledge of $l_{i,j}$, the traitors cannot remove the EDAE from the colluded signals.

C. FSMP Against Averaging Attack

Simple collusion attacks include averaging attack, minimum attack, maximum attack, median attack, and interleaving attack. Among those, averaging attack is the most common collusion attack [38] as no one wants to share more risks than others [39]. Therefore, we mainly focus on averaging attack in this article. The mathematical model of averaging attack can be formulated as

$$\ddot{x}_{\text{avg}} = \frac{1}{N_t} \cdot \sum_{i=1}^{N_t} \bar{x}_i, \tag{23}$$

where $\ddot{x}_{\text{avg}}$ represents the colluded signal after averaging attack, $\bar{x}_i$ is the ith smoothed FSMP signal included in averaging attack, and N_t is the total number of the traitors. Fig. 4 illustrates the hybrid collusion attack model for averaging attack.

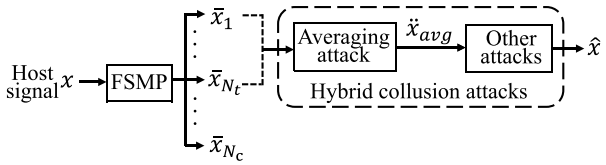


Fig. 4. Hybrid collusion attack model for averaging attack.

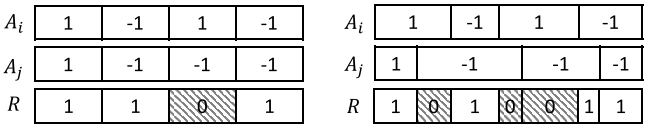


Fig. 5. Examples to illustrate two framing strategies. (a) Even framing. (b) Uneven framing.

1) FSMP Against Simple Averaging Attack: Our proposed method tackles the collusion attack by introducing the EDAE into the colluded signal to significantly reduce the perceptual quality of the colluded signal. In this section, we explain the mechanism of the EDAE in detail. For the illustration purpose, we start from $N_t = 2$, $N = 4$, and $M = 1$ using the even framing strategy with all of the DCT coefficients selected for FSMP. Denote the first FSMP signal involved in the collusion attack as $\bar{x}_i$, the second FSMP signal as $\bar{x}_j$, the adjustment sequence for $\bar{x}_i$ as A_i, and the adjustment sequence for $\bar{x}_j$ as A_j, $(i, j) \in [1, N_c]$ and $i \neq j$. Since the frames are evenly segmented, denote the frame length as L_s. As shown in Fig. 5(a), assume that A_i is $[1, -1, 1, -1]$ and A_j is $[1, -1, -1, -1]$. R stands for the ratio between the energy of the selected DCT coefficients in the colluded signal and the energy of the selected DCT coefficients in the host signal for each frame. The energy ratio for the kth frame $R(k)$ can be calculated as

$$\begin{aligned} R(k) &= \left(\sum_{l=1}^{L_s} \ddot{X}_k(l)^2\right) \Big/ \left(\sum_{l=1}^{L_s} X_k(l)^2\right) \\ &= \left(\sum_{l=1}^{L_s} \left[\frac{X_{k,i}(l) + X_{k,j}(l)}{2}\right]^2\right) \Big/ \left(\sum_{l=1}^{L_s} X_k(l)^2\right) \\ &= \left(\sum_{l=1}^{L_s} \left[\frac{(a_{i,k} + a_{j,k})X_k(l)}{2}\right]^2\right) \Big/ \left(\sum_{l=1}^{L_s} X_k(l)^2\right) \\ &= \frac{(a_{i,k} + a_{j,k})^2}{4} \end{aligned} \tag{24}$$

where $\ddot{X}_k(l)$ is the lth selected DCT coefficient in the kth frame of the colluded signal, $X_k(l)$ is the lth selected DCT coefficient in the kth frame of the host signal, and $k = 1, 2, 3, 4$. From (24), it is obvious that

$$R(k) = \begin{cases} 0, & \text{if } a_{i,k} \neq a_{j,k} \\ 1, & \text{otherwise.} \end{cases} \tag{25}$$

From Fig. 5(a), we can see that the values of $R(1)$, $R(2)$, and $R(4)$ are 1, while the value of $R(3)$ is 0, which means after FSMP, the energy of the frequency coefficients is nonuniformly reduced from frame to frame in the colluded signal. This effect is called as the EDAE in this article. Because

Authorized licensed use limited to: Michigan Technological University. Downloaded on March 24,2022 at 02:21:05 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

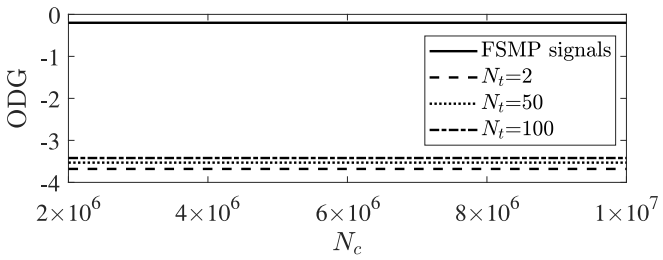


Fig. 6. ODG values of the FSMP signals and the colluded signals generated by averaging attack when $N_t = 2, 50, 100$ under different N_c.

the HAS is sensitive to the low-frequency components of the audio signal, the EDAE can significantly degrade the perceptual quality of the colluded signal. Since the adjustment sequences are generated from the codebook $C(P, d)$, when $d > 0$, $\exists k$ satisfies $a_{i,k} \neq a_{j,k}$ when $i \neq j$. Therefore, no matter which FSMP signals are involved in the collusion attack, the EDAE will always be planted into the colluded signal. As the EDAE can be generated using any two FSMP signals, it is obvious that the colluded signal will also suffer from the EDAE when $N_t > 2$.

Now, we discuss the case when the uneven framing strategy is adopted. As shown in Fig. 5(b), although both $\bar{x}_i$ and $\bar{x}_j$ are divided into four frames, due to the uneven framing strategy, R has seven elements instead of 4. More elements in R means possibly more disturbance attenuation effects are planted into the colluded signal, which can enhance the EDAE in the colluded signal and can consequently cause severer perceptual quality degradation.

In addition, from Fig. 5(b), we can see that the 2nd element in R is 0. This means by using the uneven framing strategy, even when $a_{i,k} = a_{j,k}$ (e.g., $a_{i,1} = a_{j,1}$ and $a_{i,2} = a_{j,2}$), the EDAE can still be generated in the colluded signal, which cannot be achieved using the even framing strategy. Therefore, compared to the even framing strategy, our proposed uneven framing strategy is more effective in causing the perceptual quality degradation.

Fig. 6 illustrates the experimental results of the ODG values of the colluded signals generated by averaging attack under different N_t and the averaged ODG values of the FSMP signals. It is easy to observe that all of the ODG values of the FSMP signals are close to 0, which implies that the FSMP signals have high perceptual quality. Quite the contrary, all of the ODG values of the colluded signal are below -3.4, which shows that the EDAE generated by our proposed FSMP using the uneven framing strategy can effectively degrade the perceptual quality of the colluded signals.

2) Averaging-Signal-Processing Attacks: The averaging-signal-processing attack is averaging attack combined with common signal processing attacks, including LPF, MP3 compression, AAC compression, additive Gaussian noise addition (AWGN), quantization, and amplitude scaling attacks. As the proposed FSMP resists collusion attacks by introducing the EDAE in the colluded signal, we theoretically and experimentally validate that the averaging-signal-processing attacks have little impact on the EDAE and thus cannot remove it from the colluded signals. Denote $\ddot{E}(i)$ as the energy of the selected DCT coefficients in the ith frame in the colluded signal. Then after the averaging-signal-processing attack, the energy of the selected DCT coefficients in the ith frame in the colluded signal $\ddot{E}_s(i)$ can be formulated as

$$\ddot{E}_s(i) = \ddot{E}(i) + \ddot{O} \tag{26}$$

where $\ddot{O}$ is a distortion sequence introduced by signal processing attacks. For LPF, MP3 compression, and AAC compression attacks, they only suppress the high-frequency component without significantly modifying the low-frequency component. Since FSMP is performed on the low-frequency band, we can have $\ddot{O} \approx 0$ under these attacks, which implies FSMP is robust against the simple averaging attack combined with those low-frequency preserving attacks. As for AWGN and quantization attacks, the distortions they introduce can be neglected compared to the energy of the signal, which means $\ddot{O} \ll \ddot{E}(i)$ and thus $\ddot{E}_s(i) \approx \ddot{E}(i)$. Therefore, the proposed FSMP is resilient to the simple averaging attack combined with AWGN and quantization attacks. Regarding the amplitude scaling attack, due to the property of DCT, $\ddot{O}$ is linear to $\ddot{E}(i)$ and consequently $\ddot{E}_s(i)$ is also linear to $\ddot{E}(i)$. As a result, we can easily have $\ddot{E}_s(i)/\ddot{E}(i) = \ddot{E}_s(j)/\ddot{E}(j)$, which means the EDAE will be preserved under the simple averaging attack combined the amplitude scaling attack. In summary, the averaging-signal-processing attacks cannot remove the EDAE from the colluded signals. Table IV shows the ODG values of the colluded signals generated by averaging-signal-processing attacks when $N_t = 10$ under different N_c. As shown in Table IV, when the simple averaging attack is applied with LPF (cut-off frequency of 8 kHz), MP3 compression (128 kb/s), AAC compression (128 kb/s), AWGN (SNR = 30 dB), quantization (from 16 to 8 bits), and amplitude scaling (scaling factors: 80%, 90%, 110%, 120%) attacks, all ODG values are less than -3.5, which validates the robustness of the proposed FSMP against averaging-signal-processing attacks.

TABLE IV
ODG VALUES OF THE COLLUDED SIGNALS GENERATED BY THE AVERAGING-SIGNAL-PROCESSING ATTACKS WHEN $N_t = 10$ UNDER DIFFERENT N_c

Attacks	N_c				
	2×10^6	4×10^6	6×10^6	8×10^6	1×10^7
LPF 8 kHz	-3.672	-3.821	-3.716	-3.612	-3.774
MP3 128 kbps	-3.658	-3.689	-3.641	-3.701	-3.645
AAC 128 kbps	-3.625	-3.689	-3.568	-3.649	-3.675
AWGN SNR 30 dB	-3.614	-3.683	-3.614	-3.646	-3.659
Quantization 8 bits	-3.592	-3.654	-3.652	-3.656	-3.569
Amplitude scaling 80%	-3.561	-3.569	-3.625	-3.615	-3.659
Amplitude scaling 90%	-3.658	-3.618	-3.665	-3.618	-3.648
Amplitude scaling 110%	-3.645	-3.599	-3.596	-3.715	-3.615
Amplitude scaling 120%	-3.655	-3.615	-3.598	-3.654	-3.685

3) Averaging–Desynchronization Attacks: The averaging–desynchronization attack is another type of hybrid collusion attack, where the averaging attack is combined with common desynchronization attacks, including cropping, jittering, time scaling, and pitch scaling attacks. Cropping and jittering attacks continuously or randomly remove some samples from

Authorized licensed use limited to: Michigan Technological University. Downloaded on March 24,2022 at 02:21:05 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE V
ODG VALUES OF THE COLLUDED SIGNAL GENERATED BY THE AVERAGING–DESYNCHRONIZATION ATTACKS WHEN $N_t = 10$ UNDER DIFFERENT N_c

Attacks	N_c				
	2×10^6	4×10^6	6×10^6	8×10^6	1×10^7
Cropping 10%	-3.625	-3.648	-3.659	-3.698	-3.653
Cropping 20%	-3.715	-3.759	-3.789	-3.798	-3.746
Jittering 1/1000	-3.618	-3.645	-3.648	-3.659	-3.648
Jittering 1/100	-3.615	-3.701	-3.706	-3.711	-3.748
Jittering 1/10	-3.689	-3.684	-3.759	-3.755	-3.759
Time scaling 80%	-3.555	-3.648	-3.628	-3.659	-3.648
Time scaling 90%	-3.648	-3.656	-3.654	-3.696	-3.695
Time scaling 110%	-3.622	-3.657	-3.645	-3.649	-3.655
Time scaling 120%	-3.625	-3.519	-3.537	-3.637	-3.685
Pitch scaling 80%	-3.595	-3.666	-3.658	-3.649	-3.685
Pitch scaling 90%	-3.618	-3.658	-3.648	-3.649	-3.648
Pitch scaling 110%	-3.694	-3.701	-3.644	-3.695	-3.648
Pitch scaling 120%	-3.705	-3.699	-3.658	-3.645	-3.706

the signal. Time scaling attack modifies the duration of the signal without changing the pitch while pitch scaling attack modifies the pitch without changing the duration. Similar to Section IV-C2, we demonstrate its impact on the EDAE via theoretical analysis and experimental results. Denote $\ddot{X}_i(k)$ as the kth selected DCT coefficients of the ith frame in the colluded signal generated by the simple averaging attack. According to [40], desynchronization attacks can be modeled as stretching operations in the frequency domain [40] as

$$\ddot{Y}_i(k) = \ddot{X}_i(\beta k) \tag{27}$$

where $\ddot{Y}_i(k)$ is the kth selected DCT coefficient of the ith frame in the colluded signal after the averaging–desynchronization attack and β is the scaling factor introduced by the desynchronization attack. Using (27), we can have the energy relationship between the selected DCT coefficients of the ith frame and the jth frame as

$$\frac{\sum \ddot{Y}_i^2(k)}{\sum \ddot{Y}_j^2(k)} = \frac{\sum \ddot{X}_i^2(\beta k)}{\sum \ddot{X}_j^2(\beta k)} \approx \frac{\sum \ddot{X}_i^2(k)}{\sum \ddot{X}_j^2(k)}. \tag{28}$$

From (28), we can see that the energy relationship is preserved, which means the averaging–desynchronization attack cannot remove the EDAE from the colluded signal. Table V shows the ODG values of the colluded signal generated by the averaging–desynchronization attacks when $N_t = 10$ under different N_c. From Table V, we can conclude that proposed FSMP has the robustness against the averaging–desynchronization attacks when the simple averaging attack is applied with cropping (rates: 10%, 20%), jittering (rates: 1/1000, 1/100, 1/10), time scaling (scaling factors: 80%, 90%, 110%, 120%), and pitch scaling (scaling factors: 80%, 90%, 110%, 120%) attacks.

D. Improvements of the Proposed FSMP Over [29]

In this section, we will demonstrate the improvements of the proposed FSMP over the method in [29] from two aspects, which are the perceptual quality of the generated signals and the effectiveness under hybrid collusion attacks.

TABLE VI
ODG VALUES OF THE PAP SIGNALS IN [29] AND THE FSMP SIGNALS UNDER DIFFERENT N_c

N_c	1×10^5	5×10^5	1×10^6	5×10^6	1×10^7
ODG of PAP	-0.609	-0.826	-1.081	-1.204	-1.375
ODG of FSMP	-0.129	-0.137	-0.141	-0.158	-0.177

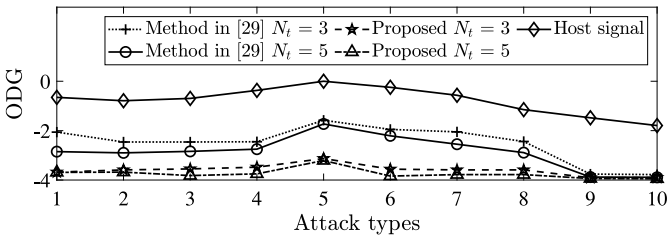


Fig. 7. ODG values of host signals, the PAP signals in [29], and the FSMP signals under various hybrid collusion attacks when $N_t = 3$ and $N_t = 5$ (Attack types: 1-LPF 8 kHz; 2-MP3 128 kb/s; 3-AAC 128 kb/s; 4-AWGN SNR 40 dB; 5-Amplitude scaling 90%; 6-Quantization 8 bits; 7-Cropping 1%; 8-Jittering 1/1000; 9-Time scaling 99.5%; 10-Pitch scaling 99.5%).

1) Perceptual Quality of the Generated Signals: Table VI shows the ODG values of the preadjustment process (PAP) signals [29] and the FSMP signals under different N_c. To be effectively combined with robust watermarking methods, the ODG values of the generated signals should be larger than -0.6. However, when N_c is only 1×10^5, the ODG value of the PAP signals is still less than -0.6. Due to the platinum and diamond award for a piece of music are 1×10^6 and 1×10^7 units, respectively, which are much larger than 1×10^5. Hence, the method in [29] can be hardly applied to real-world applications. Quite the contrary, due to the uneven framing strategy and the smoothing procedure, the ODG value of the FSMP signals is greater than -0.2 even if N_c reaches 1×10^7, which is a significant improvement compared to [29] and sufficient for real-world applications.

2) Effectiveness Under Hybrid Collusion Attacks: Fig. 7 illustrates the ODG values of the host signals, the PAP signals generated using [29], and the FSMP signals under various hybrid collusion attacks when $N_t = 3$ and $N_t = 5$. For a fair comparison, the ODG values of the PAP signals and the FSMP signals are adjusted to -0.5 before attacks. From Fig. 7, we can clearly see that the ODG values for all of the colluded host signals are higher than -2, which means the hybrid collusion attacks will not remove the commercial value from the colluded host audio signals. Therefore, it is necessary to further degrade the perceptual quality. However, since [29] transforms the host signal prior to the segmentation, it cannot effectively tackle hybrid collusion attacks when the number of traitors is odd. From Fig. 7, it is obvious that for most of the hybrid collusion attacks, [29] cannot significantly degrade the perceptual quality of the colluded signals. Especially, when $N_t = 3$, the ODG values for most of the colluded PAP signals are greater than -2.5, which shows that [29] is incapable of removing the commercial value from the colluded signals. Fundamentally, different from [29], the proposed FSMP resists collusion attacks by introducing the EDAE in the colluded signals, which can significantly degrade the perceptual quality of

Authorized licensed use limited to: Michigan Technological University. Downloaded on March 24,2022 at 02:21:05 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE VII
ODG VALUES OF THE COLLUDED SIGNALS GENERATED BY MINIMUM ATTACK, MAXIMUM ATTACK, AND MEDIAN ATTACK UNDER DIFFERENT N_t

		Minimum attack	Maximum attack	Median attack
N_t	2	-3.842	-3.889	-3.865
	20	-3.812	-3.854	-3.825
	40	-3.921	-3.856	-3.855
	60	-3.894	-3.819	-3.854
	80	-3.899	-3.816	-3.891
	100	-3.867	-3.901	-3.912

TABLE VIII
ODG VALUES OF THE COLLUDED SIGNALS GENERATED BY INTERLEAVING ATTACK UNDER DIFFERENT SEGMENT LENGTHS AND N_t

		Segment length (Samples)				
		200	400	600	800	1000
N_t	5	-3.794	-3.678	-3.548	-3.389	-3.234
	10	-3.813	-3.720	-3.626	-3.528	-3.381
	20	-3.824	-3.741	-3.649	-3.555	-3.411

the colluded signal, regardless of the number of the traitors. As a result, even when $N_t = 3$, the ODG values of the colluded FSMP signals under all kinds of hybrid collusion attacks are still below -3.5, which not only outperforms the method in [29] by a large margin but also validates the effectiveness of the proposed FSMP under hybrid collusion attacks.

E. FSMP Against Other Simple Collusion Attacks

Other simple collusion attacks include maximum, minimum, median, and interleaving [41], [42] attacks. These attacks splice the samples from different FSMP signals into one colluded signal. Since each FSMP signal is modified by a unique adjustment sequence in the frequency domain, one FSMP signal will be quite different from the other FSMP signals in the time domain. The uneven framing step further increases the difference between different FSMP signals. Therefore, splicing the samples from different FSMP signals will lead to the discontinuity between samples in the colluded signal, which can severely degrade the perceptual quality of the colluded signal. Table VII shows the ODG values of the colluded signals generated by maximum attack, minimum attack, and median attack under different N_t. Table VIII shows the ODG values of the colluded signals generated by interleaving attack under different N_t and segment lengths. From the experimental results, it is clear that our proposed method is effective against the other simple collusion attacks.

V. FSMP WITH WATERMARKING

The robust audio watermarking methods embed the copyright information by slightly modifying the host audio signal. As the EDAE is robust to the minor changes in the host signal, the proposed FSMP can be effectively combined with the robust watermarking methods to enhance its copyright protection capability. In this section, we combine the proposed FSMP with one of the leading-edge robust watermarking techniques in [27] and compare the experimental results with the existing methods. Since watermarking and fingerprinting are two methods to embed the ownership information, the existing methods we used for comparison include the fingerprinting methods in [23] and [26] and the watermarking method in [27], where [23] is a coded fingerprinting-based method, [26] is an independent fingerprinting-based method, and [27] is a leading-edge robust watermarking method. Because the Nuida codes in [23] have to be implemented with the existing watermarking techniques, we combine the Nuida codes with the traditional SS-based watermarking method in [24] and the latest robust watermarking technique in [27], respectively. For illustration purpose, we refer the combination of the Nuida codes with the method in [24] as SS-Nuida, the combination of the Nuida codes with the method in [27] as Robust-Nuida, and the combination of the proposed FSMP with the method in [27] as Robust-FSMP.

The performance of all of the methods is evaluated by the robustness against common signal processing attacks, common desynchronization attacks, the collusion-signal-processing attacks, and the collusion–desynchronization attacks. The colluded signals are generated using the averaging strategy. For a fair comparison, the embedding rate is set to 10 bps and the ODG values for the watermarked signals are adjusted to -0.9 for all of the methods. The average running time for FSMP, Robust-FSMP, SS-Nuida, Robust-Nuida, the method in [26], and the method in [27] are 0.114, 0.732, 0.716, 0.759, 0.609, and 0.618 s, respectively. It is clear that the running time of FSMP is the smallest. Besides, the running time of Robust-FSMP is in a small proximity to that of the other existing methods, which indicates that it is feasible to combine the proposed FSMP with the existing watermarking methods.

A. Robustness Against Common Signal Processing Attacks and Common Desynchronization Attacks

In this section, we evaluate the robustness of each method against common signal processing attacks and common desynchronization attacks without considering collusion attacks. The robustness is measured by the bit error rate (BER) and the detection rate (DR). The BER can be calculated by

$$\text{BER} = \frac{\text{Number of error bits}}{\text{Total number of embedded bits}} \times 100\%. \quad (29)$$

A smaller value of BER indicates better robustness against attacks. The DR represents the probability of identifying at least one of the traitors. A larger value of DR implies a higher probability of identifying the traitor, while a smaller value of DR implies the innocent user is more likely to be arrested. For the method in [27] and Robust-FSMP, the traitor is considered as successfully identified when the BER is less than 15%. Note that the method in [26] uses the iid Gaussian noise sequences instead of the binary sequences as the fingerprint codes, and the BER is not applicable to the method in [26]. Tables IX–XII illustrate the BER and DR of each method under common signal processing attacks and common desynchronization attacks.

1) Common Signal Processing Attacks: From Table IX, it is easy to observe that SS-Nuida has strong robustness against AWGN attack (SNR = 30 dB) as well as quantization attack (from 16 bits to 8 bits). However, it is not

Authorized licensed use limited to: Michigan Technological University. Downloaded on March 24,2022 at 02:21:05 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE IX
BER (%) OF SS-NUIDA, ROBUST-NUIDA, THE METHOD IN [27], AND ROBUST-FSMP UNDER COMMON SIGNAL PROCESS ATTACKS

Attacks	SS-Nuida	Robust-Nuida	Method in [27]	Robust-FSMP
LPF 8 kHz	13	1	1	3
MP3 128 kbps	20	0	0	2
AAC 128 kbps	9	3	3	3
AWGN SNR 30 dB	2	2	2	3
Amplitude scaling 80%	31	0	0	0
Amplitude scaling 90%	10	0	0	0
Amplitude scaling 110%	12	0	0	0
Amplitude scaling 120%	29	0	0	0
Quantization 8 bits	0	2	2	5

TABLE X
DR (%) OF SS-NUIDA, ROBUST-NUIDA, THE METHOD IN [26], THE METHOD IN [27], AND ROBUST-FSMP UNDER COMMON SIGNAL PROCESS ATTACKS

Attacks	SS-Nuida	Robust-Nuida	Method in [26]	Method in [27]	Robust-FSMP
LPF 8 kHz	100	100	66	100	100
MP3 128 kbps	100	100	59	100	100
AAC 128 kbps	100	100	60	100	100
AWGN SNR 30 dB	100	100	100	100	100
Amplitude scaling 80%	100	100	70	100	100
Amplitude scaling 90%	100	100	93	100	100
Amplitude scaling 110%	100	100	90	100	100
Amplitude scaling 120%	100	100	65	100	100
Quantization 8 bits	100	100	100	100	100

robust against amplitude scaling attacks (scaling factors: 80%, 90%, 110%, 120%). This is because the watermark extraction algorithm of the method in [24] relies on comparing the difference between the received signal and the host signal, which makes SS-Nuida have little resistance to amplitude scaling attacks. Besides, SS-Nuida is vulnerable to the LPF attack, MP3 compression attack, and AAC compression attack because the method in [24] adopts the entire frequency band for embedding. Although the BER values of SS-Nuida are not satisfactory, it still achieves 100% DR for all kinds of common signal processing attacks, which is shown in Table X. According to Table X, the method in [26] is also vulnerable to LPF (cut-off frequency of 8 kHz), MP3 compression (128 kb/s), AAC compression attack (128 kb/s), and amplitude scaling (scaling factors: 80%, 90%, 110%, 120%) attacks, which is similar to SS-Nuida.

As the method in Section I, the method in [27] is designed to resist attacks. Hence, Robust-Nuida, the method in [27], and Robust-FSMP are resilient to common signal processing attacks, which is illustrated in Tables IX and X. The BER values of the method in [27] are slightly less than the BER values of Robust-FSMP, because we compromise the embedding strength of Robust-FSMP to maintain the perceptual quality of the watermarked signal. Although the embedding strength is reduced, Robust-FSMP still achieves 100% DR for all of common signal processing attacks, which is shown in Table X.

2) Desynchronization Attacks: The SS-based watermarking methods cannot resist desynchronization attacks because desynchronization attacks break the alignment between the encoder and the decoder. According to Table XII, SS-Nuida and the method in [26] lose the ability to identify the traitors (DR = 0%) under common desynchronization attacks. As the robust watermarking method in [27] embeds the watermark into the statistical properties of the host signal repeatedly, it can achieve the robustness against common desynchronization attacks. Therefore, Robust-Nuida, the method in [27], and Robust-FSMP are robust against common desynchronization attacks. The experimental results are illustrated in Tables XI and XII. Similar to common signal processing attacks, the BER values of Robust-FSMP are slightly larger than the method in [27]. However, the ability of Robust-FSMP to identify the traitor is not affected as the DR is 100% for all kinds of common desynchronization attacks.

TABLE XI
BER (%) OF SS-NUIDA, ROBUST-NUIDA, THE METHOD IN [27], AND ROBUST-FSMP UNDER COMMON DESYNCHRONIZATION ATTACKS

Attacks	SS-Nuida	Robust-Nuida	Method in [27]	Robust-FSMP
Cropping 10%	43	0	0	0
Cropping 20%	49	0	0	0
Jittering 1/1000	48	1	1	1
Jittering 1/100	50	1	1	2
Jittering 1/10	55	2	2	4
Time scaling 80%	51	3	3	5
Time scaling 90%	47	2	3	6
Time scaling 110%	46	2	2	4
Time scaling 120%	48	4	3	4
Pitch scaling 80%	56	7	9	10
Pitch scaling 90%	47	5	5	7
Pitch scaling 110%	49	6	7	9
Pitch scaling 120%	59	9	8	9

TABLE XII
DR (%) OF SS-NUIDA, ROBUST-NUIDA, THE METHOD IN [26], THE METHOD IN [27], AND ROBUST-FSMP UNDER COMMON DESYNCHRONIZATION ATTACKS

Attacks	SS-Nuida	Robust-Nuida	Method in [26]	Method in [27]	Robust-FSMP
Cropping 10%	0	100	0	100	100
Cropping 20%	0	100	0	100	100
Jittering 1/1000	0	100	0	100	100
Jittering 1/100	0	100	0	100	100
Jittering 1/10	0	100	0	100	100
Time scaling 80%	0	100	0	100	100
Time scaling 90%	0	100	0	100	100
Time scaling 110%	0	100	0	100	100
Time scaling 120%	0	100	0	100	100
Pitch scaling 80%	0	100	0	100	100
Pitch scaling 90%	0	100	0	100	100
Pitch scaling 110%	0	100	0	100	100
Pitch scaling 120%	0	100	0	100	100

B. Robustness Against the Collusion-Signal-Processing Attacks and the Collusion–Desynchronization Attacks

In this section, the performance against hybrid collusion attacks is evaluated using DR and the ODG values of the colluded signal. A larger ODG value indicates a higher monetary value of the colluded signal and consequently, the traitors will be more motivated to launch the collusion attack. On the contrary, a smaller ODG value indicates a lower commercial value of the colluded signal, which will remove the motivation from the traitors to launch the collusion attack. Note that since the proposed FSMP resists the collusion attack by significantly degrading the perceptual quality of the colluded signal, in this section the performance of Robust-FSMP will be evaluated using the ODG values of the colluded signal only.

Authorized licensed use limited to: Michigan Technological University. Downloaded on March 24,2022 at 02:21:05 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE XIII
DR (%) OF SS-NUIDA, ROBUST-NUIDA, METHOD IN [26], THE METHOD IN [27], AND ROBUST-FSMP UNDER HYBRID COLLUSION ATTACKS WHEN $N_t = 10$ AND $N_t = 20$

Attacks	$N_t = 10$				$N_t = 20$			
	SS-Nuida	Robust-Nuida	Method in [26]	Method in [27]	SS-Nuida	Robust-Nuida	Method in [26]	Method in [27]
LPF 8 kHz	100	86	0	0	100	77	0	0
MP3 128 kbps	100	45	0	0	100	30	0	0
AAC 128 kbps	100	96	12	0	100	91	6	0
AWGN SNR 30 dB	100	92	27	0	100	78	8	0
Amplitude scaling 80%	65	96	0	0	42	93	0	0
Amplitude scaling 90%	73	97	0	0	45	95	0	0
Amplitude scaling 110%	70	97	0	0	48	97	0	0
Amplitude scaling 120%	62	95	0	0	40	93	0	0
Quantization 8 bits	100	80	97	0	100	72	22	0
Cropping 10%	0	95	0	0	0	92	0	0
Cropping 20%	0	90	0	0	0	88	0	0
Jittering 1/1000	0	87	0	0	0	80	0	0
Jittering 1/100	0	84	0	0	0	75	0	0
Jittering 1/10	0	81	0	0	0	73	0	0
Time scaling 80%	0	67	0	0	0	60	0	0
Time scaling 90%	0	80	0	0	0	69	0	0
Time scaling 110%	0	79	0	0	0	73	0	0
Time scaling 120%	0	70	0	0	0	65	0	0
Pitch scaling 80%	0	52	0	0	0	49	0	0
Pitch scaling 90%	0	59	0	0	0	56	0	0
Pitch scaling 110%	0	60	0	0	0	57	0	0
Pitch scaling 120%	0	49	0	0	0	46	0	0

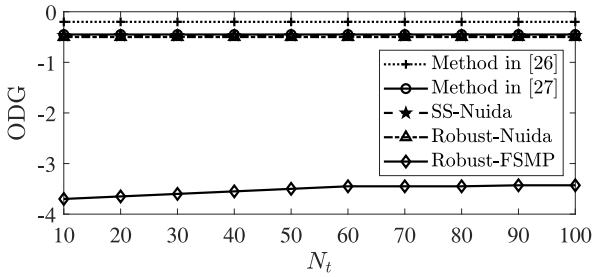


Fig. 8. ODG values of the colluded signals for SS-Nuida, Robust-Nuida, the method in [26], the method in [27], and Robust-FSMP under different N_t.

Table XIII illustrates the DR of each method under hybrid collusion attacks when $N_t = 10$ and $N_t = 20$. Fig. 8 shows the ODG values of the colluded signals for each method under different N_t.

1) Detection Performance: Table XIII demonstrates the DR of each method against hybrid collusion attacks. Since SS-Nuida and the method in [26] cannot resist common desynchronization attacks even when the collusion attack is not presented, it is obvious that they have no resistance against the collusion–desynchronization attacks. Although the method in [27] is resilient to common attacks, it has no mechanisms to tackle the collusion attack. As a result, the method in [27] cannot resist any kind of hybrid collusion attacks (DR = 0%).

As mentioned in Section II, the Nuida codes need to be extremely long to tackle the collusion attack. As the coalition size increases, the required code length grows tremendously. However, the robust watermarking methods achieve the robustness against common desynchronization attacks by exploiting the statistical properties of the audio signal, which limits their embedding capacity. In order to enhance the robustness, the copyright information is embedded by a repeated manner in the robust watermarking methods, which further lowers the embedding capacity. As a result, when the Nuida codes are applied with the robust watermarking methods, they become ineffective against the collusion–desynchronization attacks due to the limitation on the code length, especially under a large coalition size. From Table XIII, we can see that the DR of Robust-Nuida is awful against the collusion–desynchronization attacks when $N_t = 20$. The false-positive probability is too high to be accepted by real-world applications.

In conclusion, none of SS-Nuida, Robust-Nuida, the method in [26], and the method in [27] can effectively identify the traitors under hybrid collusion attacks, especially the collusion–desynchronization attacks.

2) Perceptual Quality of the Colluded Signal: Fig. 8 illustrates the ODG values of the colluded signals for each method. Since the method in [26] adopts the iid Gaussian noise sequences as the fingerprint codes, which will offset each other after averaging attack, the ODG values of the colluded signals for the method in [26] are very close to 0. For SS-Nuida, Robust-Nuida, and the method in [27], the ODG values of the colluded signals are around -0.5, which are better than the watermarked signals whose ODG values are around -0.9. The reason is that some of the modifications caused by the watermark embedding procedures will be attenuated by averaging attack. On the contrary, from Fig. 8, it is easy to observe that the ODG values of the colluded signals for Robust-FSMP are below -3.4. This is because the EDAE generated by our proposed FSMP can significantly degrade the perceptual quality of the colluded signal and is resistant toward hybrid collusion attacks and the modifications caused by the watermark embedding procedure.

In summary, none of SS-Nuida, Robust-Nuida, the method in [26], and the method in [27] can tackle hybrid collusion attacks, especially the collusion–desynchronization attacks. However, our proposed FSMP can significantly degrade the perceptual quality of the colluded signal after the collusion-signal-processing attacks and the

Authorized licensed use limited to: Michigan Technological University. Downloaded on March 24,2022 at 02:21:05 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

collusion–desynchronization attacks. As a result, our proposed FSMP can effectively resist hybrid collusion attacks by removing the commercial value from the colluded signal and consequently demotivate the traitors from launching collusion attacks.

VI. Conclusion

In this article, we proposed the FSMP-based mechanism to tackle the hybrid collusion attacks, especially the collusion–desynchronization attacks. The proposed mechanism resists the collusion attack by planting the EDAE into the colluded signal to significantly degrade the perceptual quality of the colluded signal. To maximize the perceptual quality degradation in the colluded signal while maintaining the perceptual quality of the FSMP signals, we proposed the uneven framing strategy to enhance the EDAE and a customized smoothing procedure to improve the perceptual quality of the FSMP signals. Furthermore, FSMP can be effectively combined with the existing trace-traitor-based methods to reinforce the protection against different attacks. Experimental results demonstrated that the proposed FSMP outperforms the conventional anti-collusion methods. We hope that our initial study can attract more subsequent researches on collusion attack resistance.

References

[1] L. Y. Zhang *et al.*, "On the security of a class of diffusion mechanisms for image encryption," *IEEE Trans. Cybern.*, vol. 48, no. 4, pp. 1163–1175, Apr. 2018.

[2] L. Shanmugam, P. Mani, R. Rajan, and Y. H. Joo, "Adaptive synchronization of reaction–diffusion neural networks and its application to secure communication," *IEEE Trans. Cybern.*, vol. 50, no. 3, pp. 911–922, Mar. 2020.

[3] W. Lu, J. Chen, J. Zhang, J. Huang, J. Weng, and Y. Zhou, "Secure halftone image steganography based on feature space and layer embedding," *IEEE Trans. Cybern.*, early access, Oct. 23, 2020, doi: 10.1109/TCYB.2020.3026047.

[4] J. Jia *et al.*, "RIHOOP: Robust invisible hyperlinks in offline and online photographs," *IEEE Trans. Cybern.*, early access, Dec. 14, 2020, doi: 10.1109/TCYB.2020.3037208.

[5] Y.-G. Wang, G. Zhu, S. Kwong, and Y.-Q. Shi, "A study on the security levels of spread-spectrum embedding schemes in the WOA framework," *IEEE Trans. Cybern.*, vol. 48, no. 8, pp. 2307–2320, Aug. 2018.

[6] H. Tian, Y. Zhao, R. Ni, L. Qin, and X. Li, "LDFT-based watermarking resilient to local desynchronization attacks," *IEEE Trans. Cybern.*, vol. 43, no. 6, pp. 2190–2201, Dec. 2013.

[7] X. Cao, L. Du, X. Wei, D. Meng, and X. Guo, "High capacity reversible data hiding in encrypted images by patch-level sparse representation," *IEEE Trans. Cybern.*, vol. 46, no. 5, pp. 1132–1143, May 2016.

[8] J. Wang, J. Ni, X. Zhang, and Y.-Q. Shi, "Rate and distortion optimization for reversible data hiding using multiple histogram shifting," *IEEE Trans. Cybern.*, vol. 47, no. 2, pp. 315–326, Feb. 2017.

[9] H.-Y. Huang, C.-H. Yang, and W.-H. Hsu, "A video watermarking technique based on pseudo-3-D DCT and quantization index modulation," *IEEE Trans. Inf. Forensics Security*, vol. 5, pp. 625–637, 2010.

[10] M. Asikuzzaman, M. J. Alam, A. J. Lambert, and M. R. Pickering, "Imperceptible and robust blind video watermarking using chrominance embedding: A set of approaches in the DT CWT domain," *IEEE Trans. Inf. Forensics Security*, vol. 9, pp. 1502–1517, 2014.

[11] M. Asikuzzaman, M. J. Alam, A. J. Lambert, and M. R. Pickering, "Robust DT CWT-based DIBR 3D video watermarking using chrominance embedding," *IEEE Trans. Multimedia*, vol. 18, no. 9, pp. 1733–1748, Sep. 2016.

[12] Y. Xiang, I. Natgunanathan, D. Peng, W. Zhou, and S. Yu, "A dual-channel time-spread echo method for audio watermarking," *IEEE Trans. Inf. Forensics Security*, vol. 7, pp. 383–392, 2012.

[13] Y. Xiang, I. Natgunanathan, S. Guo, W. Zhou, and S. Nahavandi, "Patchwork-based audio watermarking method robust to de-synchronization attacks," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 22, no. 9, pp. 1413–1423, Sep. 2014.

[14] Z. Su, G. Zhang, F. Yue, L. Chang, J. Jiang, and X. Yao, "SNR-constrained heuristics for optimizing the scaling parameter of robust audio watermarking," *IEEE Trans. Multimedia*, vol. 20, no. 10, pp. 2631–2644, Oct. 2018.

[15] S. Xiang, L. Yang, and Y. Wang, "Robust and reversible audio watermarking by modifying statistical features in time domain," *Adv. Multimedia*, vol. 2017, Apr. 2017, Art. no. 8492672.

[16] A. Bobeica, I. C. Dragoi, I. Caciula, D. Coltuc, F. Albu, and F. Yang, "Capacity control for prediction error expansion based audio reversible data hiding," in *Proc. Int. Conf. Syst. Theory Control Comput.*, Sinaia, Romania, Oct. 2018, pp. 810–815.

[17] X. Liang and S. Xiang, "Robust reversible audio watermarking based on high-order difference statistics," *Signal Process.*, vol. 173, Aug. 2020, Art. no. 107584.

[18] M. Li, H. Chang, Y. Xiang, and D. An, "A novel anti-collusion audio fingerprinting scheme based on Fourier coefficients reversing," *IEEE Signal Process. Lett.*, vol. 27, pp. 1794–1798, Sep. 2020.

[19] J. Zhao, T. Zong, Y. Xiang, L. Gao, and G. Hua, "Segmental DCT coefficient reversal based anti-collusion audio fingerprinting mechanism," *IEEE Signal Process. Lett.*, vol. 28, pp. 1833–1837, Sep. 2021, doi: 10.1109/LSP.2021.3108903.

[20] D. Boneh and J. Shaw, "Collusion-secure fingerprinting for digital data," *IEEE Trans. Inf. Theory*, vol. 44, no. 5, pp. 1897–1905, Sep. 1998.

[21] G. Tardos, "Optimal probabilistic fingerprint codes," *J. ACM*, vol. 55, no. 2, pp. 1–24, May 2008.

[22] B. Škorić, S. Katzenbeisser, and M. U. Celik, "Symmetric Tardos fingerprinting codes for arbitrary alphabet sizes," *Des. Codes Cryptogr.*, vol. 46, no. 2, pp. 137–166, Oct. 2008.

[23] K. Nuida *et al.*, "An improvement of discrete Tardos fingerprinting codes," *Des. Codes Cryptogr.*, vol. 52, no. 3, pp. 339–362, Mar. 2009.

[24] I. J. Cox, J. Kilian, F. T. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia," *IEEE Trans. Image Process.*, vol. 6, pp. 1673–1687, 1997.

[25] D. Kirovski, H. Malvar, and Y. Yacobi, "Multimedia content screening using a dual watermarking and fingerprinting system," in *Proc. ACM Int. Conf. Multimedia*, Dec. 2002, pp. 372–381.

[26] Z. J. Wang, M. Wu, H. V. Zhao, W. Trappe, and K. J. R. Liu, "Anti-collusion forensics of multimedia fingerprinting using orthogonal modulation," *IEEE Trans. Image Process.*, vol. 14, pp. 804–821, 2005.

[27] Z. Liu, Y. Huang, and J. Huang, "Patchwork-based audio watermarking robust against de-synchronization and recapturing attacks," *IEEE Trans. Inf. Forensics Security*, vol. 14, pp. 1171–1180, 2019.

[28] W. Jiang, X. Huang, and Y. Quan, "Audio watermarking algorithm against synchronization attacks using global characteristics and adaptive frame division," *Signal Process.*, vol. 162, pp. 153–160, Sep. 2019.

[29] J. Zhao, T. Zong, Y. Xiang, L. Gao, and G. Beliakov, "Pre-adjustment based anti-collusion mechanism for audio signals," in *Proc. Int. Netw. Syst. Security*, Dec. 2019, pp. 305–319.

[30] T. Zong, Y. Xiang, I. Natgunanathan, L. Gao, G. Hua, and W. Zhou, "Non-linear-echo based anti-collusion mechanism for audio signals," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 29, pp. 969–984, Feb. 2021.

[31] J.-L. Wu and J. Shin, "Discrete cosine transform in error control coding," *IEEE Trans. Commun.*, vol. 43, no. 5, pp. 1857–1861, May 1995.

[32] M. Fallahpour and D. Megías, "Secure logarithmic audio watermarking scheme based on the human auditory system," *Multimedia Syst.*, vol. 20, no. 2, pp. 155–164, Jun. 2014.

[33] W.-N. Lie and L.-C. Chang, "Robust and high-quality time-domain audio watermarking based on low-frequency amplitude modification," *IEEE Trans. Multimedia*, vol. 8, no. 1, pp. 46–59, Feb. 2006.

[34] K. Mivule and C. Turner, "Applying moving average filtering for non-interactive differential privacy settings," *Procedia Comput. Sci.*, vol. 36, pp. 409–415, Sep. 2014.

[35] T. Thiede *et al.*, "PEAQ-the ITU standard for objective measurement of perceived audio quality," *J. Audio Eng. Soc.*, vol. 48, nos. 1–2, pp. 3–29, Feb. 2000.

[36] V. Bhat, I. Sengupta, and A. Das, "An audio watermarking scheme using singular value decomposition and dither-modulation quantization," *Multimedia Tools Appl.*, vol. 52, nos. 2–3, pp. 369–383, Apr. 2011.

[37] "RIAA Digital Single Award." Feb. 2016. [Online]. Available: https://www.riaa.com/wp-content/uploads/2016/02/DIGITAL-SINGLE-AWARD-RIAA-AND-GRF-CERTIFICATION-AUDIT-REQUIREMENTS.pdf (Accessed: Jun. 2, 2020).

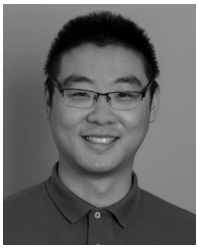
Authorized licensed use limited to: Michigan Technological University. Downloaded on March 24,2022 at 02:21:05 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

[38] S. Behseta, C. Lam, J. E. Sutton, and R. L. Webb, "A time series intra-video collusion attack on frame-by-frame video watermarking," in *Proc. Int. Workshop Digit. Watermarking*, Nov. 2008, pp. 31–44.
[39] M. Cheng and Y. Miao, "On anti-collusion codes and detection algorithms for multimedia fingerprinting," *IEEE Trans. Inf. Theory*, vol. 57, no. 7, pp. 4843–4851, Jul. 2011.
[40] X. Kang, R. Yang, and J. Huang, "Geometric invariant audio watermarking based on an LCM feature," *IEEE Trans. Multimedia*, vol. 13, no. 2, pp. 181–190, Apr. 2011.
[41] B.-H. Cha and C.-C. J. Kuo, "Robust MC-CDMA-based fingerprinting against time-varying collusion attacks," *IEEE Trans. Inf. Forensics Security*, vol. 4, pp. 302–317, 2009.
[42] B. Mathon, P. Bas, F. Cayre, and B. Macq, "Impacts of watermarking security on Tardos-based fingerprinting," *IEEE Trans. Inf. Forensics Security*, vol. 8, pp. 1038–1050, 2013.

Juan Zhao (Graduate Student Member, IEEE) received the B.S. degree in soft engineering from the Chongqing University of Posts and Telecommunications, Chongqing, China, in 2015, and the M.S. degree in computer science from Chongqing University, Chongqing, in 2018. She is currently pursuing the Ph.D. degree with the School of Information Technology, Deakin University, Geelong, VIC, Australia.

Her research interests include information security, signal processing, and machine learning.



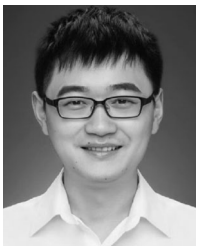
Tianrui Zong (Member, IEEE) received the B.Eng. degree in automation science and electrical engineering from Beihang University, Beijing, China, in 2009, the M.Sc. degree in signal processing and communications from the University of Edinburgh, Edinburgh, U.K., in 2010, and the Ph.D. degree in signal processing and communications from Deakin University, Geelong, VIC, Australia, in 2015.

He is currently a Senior Researcher with CNPIEC KEXIN Ltd., Beijing, China. His research interests include multimedia forensics and security, signal processing, and machine learning.

Yong Xiang (Senior Member, IEEE) received the Ph.D. degree in electrical and electronic engineering from The University of Melbourne, Melbourne, VIC, Australia, in 2003.

He is currently a Professor with the School of Information Technology, Deakin University, Geelong, VIC, Australia. His research interests include information security and privacy, data analytics and machine learning, Internet of Things, and blockchain. He has published six monographs, over 190 refereed journal articles, and numerous conference papers in these areas.

Prof. Xiang has served as the honorary chair, general chair, program chair, technical program committee chair, symposium chair, and track chair for many conferences, and was invited to give keynotes at several international conferences. He is the Senior Area Editor of IEEE SIGNAL PROCESSING LETTERS and the Associate Editor of IEEE COMMUNICATIONS SURVEYS AND TUTORIALS. He was the Associate Editor of IEEE SIGNAL PROCESSING LETTERS and IEEE ACCESS, and the Guest Editor of IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS and IEEE MULTIMEDIA.



Guang Hua (Senior Member, IEEE) received the B.Eng. degree in communication engineering from Wuhan University, Wuhan, China, in 2009, and the Ph.D. degree in information engineering from Nanyang Technological University (NTU), Singapore, in 2014.

From July 2013 to November 2015, he was a Scientist I with the Department of Cyber Security and Intelligence, Institute for Infocomm Research, A*Star, Singapore. After that, he was with the School of Electrical and Electronic Engineering, NTU, as a Research Fellow, until 2017. He is currently with the School of Electronic Information, Wuhan University. His research interests include multimedia forensics and security, security issues of deep learning, applied machine learning, and general signal processing topics.

Dr. Hua serves as an Associate Editor for the IEEE SIGNAL PROCESSING LETTERS.

Xinyu Lei (Member, IEEE) received the Ph.D. degree with the Department of Computer Science and Engineering, Michigan State University, East Lansing, MI, USA, in 2021.

He is currently an Assistant Professor with the Department of Computer Science, Michigan Technological University, Houghton, MI, USA. He worked as a Research Assistant with the Texas A&M University at Qatar, Doha, Qatar, in 2013. In 2017, he worked as a Research Intern with Ford Motor Company, Dearborn, MI, USA. His current research focuses on machine learning and cybersecurity.

Longxiang Gao (Senior Member, IEEE) received the Ph.D. degree in computer science from Deakin University, Geelong, VIC, Australia, in 2014.

He is currently a Professor with the Qilu University of Technology (Shandong Academy of Sciences), Jinan, China, and Shandong Computer Science Center (National Supercomputer Center in Jinan), Jinan. He was a Senior Lecturer with the School of Information Technology, Deakin University, and a Postdoctoral Research Fellow with IBM Research and Development, Melbourne, VIC, Australia. He has been the Chief Investigator for more than 20 research projects (the total awarded amount is over U.S. $5 million), from pure research project to contracted industry research. He has over 90 publications, including patent, monograph, book chapter, journals, and conference papers. Some of his publications have been published in the top venue, such as IEEE TRANSACTIONS ON MOBILE COMPUTING, IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, IEEE INTERNET OF THINGS JOURNAL, IEEE TRANSACTIONS ON DEPENDABLE AND SECURE COMPUTING, IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, IEEE TRANSACTIONS ON COMPUTATIONAL SOCIAL SYSTEMS, IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, and IEEE TRANSACTIONS ON NETWORK SCIENCE AND ENGINEERING. His research interests include fog/edge computing, blockchain, data analysis, and privacy protection.

Gleb Beliakov (Senior Member, IEEE) received the Ph.D. degree in physics and mathematics from the Peoples' Friendship University of Russia, Moscow, Russia, in 1992.

He is currently a Professor with the School of Information Technology, Deakin University, Geelong, VIC, Australia, where he is the Director of the Data to Intelligence Research Centre. His research interests are in the areas of aggregation operators, fuzzy measures, multivariate approximation, global optimization, and decision support systems. He is the author of over 200 research articles and three monographs in the mentioned areas, and a number of software packages.

Prof. Beliakov serves as an Area Editor for *Fuzzy Sets and Systems* and an Associate Editor for the IEEE TRANSACTIONS ON FUZZY SYSTEMS and *International Journal of Approximate Reasoning*.

Authorized licensed use limited to: Michigan Technological University. Downloaded on March 24,2022 at 02:21:05 UTC from IEEE Xplore. Restrictions apply.