Collusion Resistant Fingerprinting of Digital Audio

Andrew Tirkel
Scientific Technology
8 Cecil St
East Brighton, Vic 3187, Australia
+61395922206
atirkel@bigpond.net.au

Thomas Hall
Monash University
School of Mathematical Sciences
Victoria 3800, Australia
+61399054412
tom.hall@monash.edu

Charles Osborne
Monash University
School of Physics
Victoria 3800, Australia
+61412713690
charlesoz1@optusnet.com.au

Nicholas Meinhold
79 Old Healesville Rd
Yarra Glen, Vic 3775
Australia
+61405128047
nick.meinhold@gmail.com

Oscar Moreno
Gauss Research Laboratory
4 Calle Georgetti
San Juan, PR, 009253606
+17875023452
moreno@nic.pr

ABSTRACT
Digital fingerprinting is a technique for tracing the distribution of multimedia content, and protecting it from unauthorized manipulation. Unique identification information is embedded into each distributed copy of the signal. In a collusion attack, fingerprints are combined to remove or distort the fingerprints. Audio signals are good candidates for fingerprinting, because of the forgiving nature of the human auditory system to cross-talk between channels. We use principal components of the audio signal to construct an abstract vector space. The fingerprints are ordered rotations in that space. The rotations are determined by arrays with good correlation properties. These arrays are embedded in real audio, and are imperceptible, according to a panel of experts. These fingerprints are resistant to an averaging collusion attack by hundreds or thousands of colluders, and can withstand a worst case RandNeg attack by up to 30 colluders.

Categories and Subject Descriptors
K.4.4 Electronic Commerce (J.1), Distributed commercial transactions, Electronic data interchange (EDI), Intellectual property, Payment schemes, Security.

General Terms

Keywords
fingerprint, watermark, collusion, traitor tracing, audio, array.

1. INTRODUCTION
A digital watermark is any signal that is added to a document in order to embed some information. This is a broad definition. The original digital watermark [21], contained information about the owner of the document, or about the document itself. Since then, watermarks have been embedded and recovered from still images, video, audio and many other information carrying formats, and research into watermarking and the related field of information hiding (steganography) has been prolific, and the nature and application of watermarks has diversified dramatically. One such application is digital fingerprinting [1], where the specific objective of the watermark is the identification of a recipient of a document, providing the ability to trace the document from its originator, resulting in an audit trail. One reason for the proliferation of watermarks/fingerprints is that the fingerprint should be matched to the nature of the document, and the response characteristic of the recipient. There are many different types of documents and many ways of processing and interpreting them, including compression, cropping and innocent or deliberate distortions. Another is that as watermarking has gained acceptance, countermeasures have developed. Resistance to processing distortions and countermeasures has been a major focus of research into this area. There are fundamental differences between still image and audio watermarking, in that the image is presented in parallel fashion (all pixels are available to the eye), whilst audio is presented serially to the ears. This paper focuses on digital fingerprints for audio applications. A digital audio fingerprint must be unobtrusive, so that, first, it does not distract the listener from the audio content, and second, it is difficult for a potential attacker to detect in the electronic version. Kirovski [8] describes key requirements of fingerprints, and the consequences of collusion attacks: "Imperceptiveness, robustness, and reliability are the key requirements for fingerprints. One major difference with respect to content screening is that the robustness requirement is significantly easier to satisfy - fingerprint detection is done in the presence of the original clip, not "blindly". A major problem for fingerprinting systems is the collusion attack. To launch such an attack, an adversarial clique of malicious users colludes their copies in order to create a copy which is statistically clean of any fingerprint traces (e.g., the original) or a copy that incriminates another innocent user. Collusion resistance for multimedia content is typically low [1]. Because of this deficiency, fingerprinting systems are commonly restricted to small distribution lists. Finally, one of the most devastating problems for fingerprinting systems is surprisingly, successful identity theft. An adversary with a stolen identity can purchase a multimedia clip and then illegally distribute it, leaving multimedia studios without a target for legal action. Collusion is usually the most effective effort to defeat fingerprinting schemes as opposed to other signal processing attacks that target removal or
obfuscation of the embedded secret. For example, while the estimation attack typically produces a pirated copy of inferior quality, the result of collusion is of equal or even better quality than the distributed content. The adversary can have two types of goals: (i) removal of their fingerprints from the pirated copy and (ii) framing an innocent user. The latter attack is of particular importance because it governs the number of copies the copyright owner can distribute. Once innocent users can be framed, the entire system is rendered dysfunctional."
The paper is organized as follows: Section 2 defines the concepts of digital audio fingerprinting used throughout the rest of the paper, and reviews the conventional fingerprinting techniques found in the literature. Section 3 describes the new method of fingerprinting proposed in this paper. Section 4 describes implementation of the new scheme. Section 5 describes the performance of the new scheme when subjected to collusion attacks, which are likely to be encountered in the distribution of audio files, and compares that with traditional schemes. Section 6 describes the construction of arrays used for audio fingerprinting, and the construction of new arrays, which are designed specifically for fingerprinting video and multimedia. These arrays are multi-dimensional and well matched to such data.

2. Definitions & Explanation of Terms

2.1 Definitions

Fingerprint scheme: A fingerprinting scheme for a document \( D \) and a set of users \( \{1,...,n\} \), is a pair of algorithms, mark and detect, combined with a code book \( \Gamma^c \) where:

- Given a user \( i \in \{1,...,n\} \) and a document \( D \), the distributor uses mark to add a fingerprint \( W \) to the document. For user \( i \) mark generates the document \( D^w_i \).
- Given a marked document \( D^* \) detect returns a set of users corresponding to potential owners of the document.

The code book stores any information created in the marking process that is required by the detection process [4].

Collusion: Any form of attack where multiple marked documents \( \{D_1,D_2,...,D_n\} \) are processed to create another document \( D^* \) is called a collusion attack. The set of users \( \{u_1,...,u_k\} \) corresponding to the owners of the documents is called the colluding coalition, or more simply a coalition.

c-secure codes: A code \( \Gamma \) is totally c-secure if there exists a tracing algorithm \( A \) satisfying the following condition: if a coalition \( C \) of at most \( c \) users generates a modified fingerprint \( x \) then \( A(x) \in C \) [2].

1.1.2 Explanation of Terms

Human Auditory System (HAS): Human perception of sounds (or a model to mimic the processing of sound in the ear/brain) is complex, but must be taken into account in designing a fingerprint/watermark. Warby, [7], states "For example, because of the physical nature of the basilar membrane in the ear, a psycho-acoustic masking effect exists whereby small signals close by in frequency to large signals, cannot be perceived. The compression obtained by MP3 compression is the result of utilizing this effect to ignore redundant parts of sounds that will anyway not be perceived by listeners. Such frequency components, definitely present but inaudible to the human being, can be used to make an audio watermark."

Binaural Sensitivity: McAlpine et al [12] explain: "For many terrestrial mammals, and particularly for humans, localization of sound sources in the horizontal plane is achieved by an exquisite sensitivity to differences in the fine-time structure of low-frequency (< 1500 Hz) spectral components between the two ears. The accepted model to account for this remarkable binaural sensitivity is the coincidence detection model. In this model, an array of neurons receives inputs from the two ears such that a given neuron fires maximally when the difference in arrival time at the two ears resulting from the location of a sound source offsets the difference in neural transmission time."

Intel High Definition Audio: (IHD or "Azalia") refers to the specification released by Intel in 2004 for delivering high-definition audio that is capable of playing back more channels at higher quality than previous integrated audio codecs like AC97. Hardware based on Intel HD Audio specifications is capable of delivering 192 kHz/32 bit quality for two and 96 kHz/32 bit for up to eight channels [6].

2.2 Review of Audio Fingerprinting Methods

Audio fingerprinting methods have involved orthogonal codes, BIBD (Balanced Incomplete Block Design), error correcting codes [24,10,11,22,9] or subtle but deliberate distortion (e.g. time warping) [2,3].

2.2.1 Additive Fingerprinting

Traditional audio fingerprinting methods involve the use of spread spectrum type sequences added to the signal or its transform. Detection is achieved through correlation. Such techniques are prone to additive attacks, such as averaging. Various schemes have been devised to make this technique more robust. However, the method has drawbacks when it comes to coalition identification. This is because, for \( n \) recipients, there are \( 2^n \) potential coalitions. It is impractical to construct and embed \( 2^n \) different additive fingerprints, to defeat such an attack. It is also difficult to find a sufficient number of fingerprints with the required good autocorrelation and cross-correlation. A theoretical limit on the family size is the Welch Bound. The authors have spent the last 15 years constructing families of sequences and arrays which meet or approximately meet the Welch Bound, and have concluded that such constructions are unlikely to lead to fingerprints secure from large coalition attacks.

2.2.2 Time Warping

T. Parker [6] has studied a distortion-based audio fingerprint, which uses time warping. It appears to work, but has 2 fundamental limitations:

1. It is computationally horrendous
2. It has not yet passed the audio test (time-warp may be objectionable)

A set of colluders can attempt to get rid of the fingerprints by averaging, which causes echoes to appear. Echo cancellation techniques may get rid of these echoes and defeat the method.

2.3 Fingerprint Detection

Traditionally, the presence and content of a fingerprint is detected by correlation or measure of distance, when a copy under test is compared with marked or unmarked copies of the document using a maximum likelihood or least error comparison.

2.4 Collusion Attack

2.4.1 Averaging Attack

Consider a coalition of \( K \) users \( C = \{1,2,..,K\} \) with corresponding watermarks \( \{w',w'',...,w\} \). The document that the coalition will produce by an averaging type attack, \( D^* \) is given by

\[
D^* = \frac{1}{2^K} \sum_{j=1}^{2^K} D^{w_j} \quad (j \in C)
\]

(1)

More typically, there exists a transform \( T \), in which the watermark appears as an additive entity. In such a domain
\[ D'_T = \frac{1}{2} \sum_{j=1}^{K} (D_T + w_j) = D_T + \frac{1}{2} \sum_{j=1}^{K} w_j \rightarrow D_T \]  

(2)

Clearly, as \( K \) increases, the result of the averaging attack approaches the original (unmarked) document, especially when LSB type embedding is used to make the watermark unobtrusive. Of course, the coalition is free to produce a weighted average, with each document being given a different weight.

### 2.4.2 Min-Max & RandNeg Attacks

Instead of averaging, the colluders can perform a max-min attack, by producing followed by scaling.

\[ D^* = \frac{1}{2} \left[ \max(D^{\text{w}k}) + \min(D^{\text{w}k}) \right], j, k \in \mathcal{C}. \]  

(3)

Kirovski [5] proves that the max-min attack creates a better expected estimate of the original audio (and hence a weaker fingerprint), than the averaging attack for large classes of fingerprints and \( K > 2 \), but with a higher variance than the averaging attack. For the randomized negative attack each component of the attacked signal takes the value of the maximum of the corresponding components of the fingerprinted signals with probability \( p \), and takes the minimum with probability \( 1 - p \). [10]

### 2.4.3 Collusion Resistance

A fingerprinting scheme where the existence of a fingerprint(s) is not compromised by a collusion attack is known as a collusion resistant fingerprint. Here, we consider collusion resistance against averaging attack only.

### 2.4.4 Traitor Tracing

The identification of traitors is a difficult problem, because of the potentially enormous number of possible coalitions of colluders and possibility that natural distortions can cloud the issue, or that colluders may conspire to confuse or deliberately mislead the distributor. The effectiveness of traitor tracing can be measured in statistical terms: probability of correct identification, probability of false identification and probability of missed detection. In addition, a coalition of colluders may involve passive colluders i.e. recipients whose documents have been used in the collusion process by allowing inadvertent or deliberate access of a copy of the document to active colluders (for example, where a supervisor, or by contrast, an IT assistant has access to the files of department members). It is not possible to determine the roles of individual colluders, without resorting to sequential "trials of trust", where deliberate distribution of multiple documents to carefully selected recipients is calculated to reveal a traitor or traitors. Of course, the traitors may be aware of such schemes, and have their own countermeasures against them.

### 3. NEW METHOD

In this section, we describe a new method of audio fingerprinting, with resistance to many collusion attacks and potential identification of colluders. The signal flow for the algorithms is presented in graphical form in Section 3.

#### 3.1 Audio Signal Analysis

In this section we present a method of audio signal analysis, which prepares audio signals for marking. Consider a typical audio signal. The method applies to any audio signal, but here we consider the popular IHD Audio standard. This has at least 2 channels, each sampled at 96 or 192 KHz. Currently, each channel is represented by a time series of integers, with each integer having 20-24 bit resolution. The channels represent signals controlling Left and Right speakers, and therefore what is intended to reach the left and right ear of the listener, respectively.

Averaged over sufficient periods, the magnitude (norm) of the left and right channel are approximately equal (balanced).

\[ S(i) = [L(i), R(i)], i \in 1, \text{for time interval } 1, \text{and } \mathcal{T} = \mathbb{R} \]  

(4)

The signals in both channels can be transformed by projecting them onto a discrete orthonormal basis. Examples of such bases are: sines and cosines which are harmonically related (Fourier), wavelets, fractal functions. Audio signals range from 20 Hz to 20 KHz, with the majority of the information and hence energy content being concentrated between 300 Hz and 3 KHz. The sampling must be at least twice per period of the highest frequency (Nyquist) which is higher than required for the fingerprinting process. Hence, it is a good idea to down-sample the audio by a factor of 10 or 20, before the spectral analysis. Consider a time segment of say 1021 samples (a prime number, we use for cryptographic resistance) although any other value would do. This represents 100 ms (milliseconds) of audio data.

After applying appropriate windowing (such as Tukey), perform the discrete Fourier transform (DFT) on the above signal

\[ S(f) = \{L(f), R(f)\} f = 0,1,\ldots,1024. \]  

(5)

Consider the two entries as components of a vector \( S \), with magnitude \( S \) and direction \( \theta \).

\[ S(f) = \sqrt{[L(f)]^2 + [R(f)]^2}, \theta = \tan^{-1} \frac{L(f)}{R(f)} \]  

(6)

Rank all the 1024 entries representing that 100 ms of audio signal, according to magnitude, and select the largest \( n \) of them (\( n \) must be at least as large as the number of recipients of the audio, any of whom can collude with each other). Should there be components with equal norm, rank these according to frequency. Check that these \( n \) principal components for that time interval are all above some suitable threshold, and that \( 35^\circ < \theta < 55^\circ \) (balance criterion: the angles correspond to ±3dB balance between L and R channels) for all \( n \) chosen components (if a component with suitable norm does not meet the balance criterion, skip it and go to the next ranked component). If \( n \) suitable principal components cannot be found, do not mark that time interval. This is to prevent marking silent or quiet intervals, in which the changes could be obtrusive, and difficult to decode, or marking highly unbalanced passages. There are about 10 intervals per second of audio, so the loss of a few marking opportunities should not be a major issue. It should also be noted that the start of each analysis period (time interval) need not be regular. Random or pseudo-random gaps in the analysis periods may serve to confuse potential attackers and the pattern of gaps can act as a secret key. For time intervals where suitable principal components are found, there are \( n \) components which are to be modified (marked). Denote the set of components, in descending order as:

\[ \{S_j\} \text{ with } 0 = 1, \ldots, n - 1. \]  

(7)

Perform a permutation of this set, according to a row of a Latin Square, or any other suitable permutation. Basically \( j \) gets mapped to \( j' \). The re-ordered set is:

\[ \{S_{j'}\} \text{ with } 0 = 1, \ldots, n - 1. \]  

(8)

#### 3.2 Principal Component Marking

In this section, we present an algorithm that marks the principal components of chosen time segments of the audio signals.

Consider recipient \( k \). The set \( \{S_j\} \) will be marked by altering it to

\[ S_{j'}^M = \sqrt{[R(f)]^2 + [L(f)]^2}, \theta^M = \tan^{-1} \frac{L(f)}{R(f)} + \phi^M \]  

for \( j' = 0,1,\ldots, n - 1 \)  

(9)

The angles \( \phi^M \) are functions of the recipient, the principal component \( j' \) and the time interval. Note, that the magnitude (norm) is not affected by the marking, and therefore neither is the
There are many possible methods of controlling the frequency order, and hence the ranking of these is unaffected by the marking process. The phase angle of each component is recovered for each time slot. The components are assigned their order according to the permutation, and a time series of phases is obtained for each component. The unmarked phases are subtracted from these. The remaining phase sequence is scaled by the reciprocal scaling factor (inverse of the embedding process) and the phase sequence for each component is correlated with that for each recipient in the embedding process.

Presence of unambiguous peaks in this correlation identifies the existence of marks from members of any colluding coalition. Note, that the nature of the marking is multiplicative. The RMS value of an audio signal during the marking period is of the order of 1/32d of full scale. This allows for volume adjustments, sound peaks and troughs, otherwise described as dynamic range headroom. Assuming uniform distribution of 20 principal components, each component has an RMS magnitude of $L(f) \approx 1 \times \sqrt{\frac{M}{50}}$. This means that, for a 24 bit signal, the RMS value of a principal component is of the order of 100,000 units. Therefore, the marking introduces modifications of the order of 1000 units (10 LSB’s). This is a worst case scenario, since the largest principal component values of R and L are likely to be significantly greater. Anyway, the modification to the audio is not in just the LSB, but to the 10 lowest bits! Therefore, a coalition would need about 2000 members in order to average such marks to below LSB status (after quantization). Current fingerprinting schemes are not designed to accommodate such large distribution, so the system, proposed here, is collusion resistant to a vastly larger number of colluders, and should cater for a coalition of the order of 50. By comparison, the brute force technique would need to construct a fingerprint set with 250 marks in order to achieve the same objective, which is impossible.

### 3.4 Synchronization

In order to perform the recovery of the mark, or to prove the existence of a mark, or marks, in a data stream, the transform must be performed in synchronicity, and using the same pitch scale, as that used in the marking. Tachibana et al [18] have shown that a pseudonoise array (in time and frequency) can be used to provide this synchronization, and carry some secret data to identify the originator (distributor) of the audio. Once synchronization is achieved, this mark can be removed before performing the analysis, although this is not essential.

### 3.5 Mono and Multi-Channel Audio

#### 3.5.1 Mono

In some circumstances, the audio signal can be monophonic. In such cases, it is still possible to perform the marking and detection as outlined in Section 2. As an example, consider a slightly different definition of the vector S

$$S(f) = \sqrt{||DCT(f)||^2 + ||DST(f)||^2}, \theta = \tan^{-1}\frac{||DCT(f)||}{||DST(f)||}$$

where $||DCT(f)||$ and $||DST(f)||$ are the magnitudes of the Discrete Cosine Transform (DCT) and the Discrete Sine Transform (DST) coefficients calculated at frequency $f$. Other methods of constructing the vector $S$ are also possible. For example, sampling the monophonic data stream, to partition it into suitable pairs of numbers can produce a set of vectors in an abstract space. Such a technique was used to embed and recover watermarks in greyscale and colour images [15,16].
3.5.2 Multi-Channel
IHD Audio format supports 7 audio channels, corresponding to anything between regular stereophonic reproduction, and full home theatre surround sound. The presence of multiple channels offers more opportunities of natural contructions, involving higher dimensional vector spaces, which are discussed next. Multi-channel sound reproduction is used to simulate a real acoustic environment, where binaural signal processing in the human cortex is used to localize the source(s) of sound. At this stage, it is not known what psycho-acoustic effects are produced by slight balance alterations between the channels, nor what are the thresholds of detection of such changes. Also, sometimes the other 5 channels can be "phantoms": synthesized from the 2 principal stereo channels by a combination of filtering, delay and phasing, and may therefore not be truly independent.

3.6 Marking Strength and Fingerprint Perceptibility
We use measured data from Hater et al [5], to determine the sensitivity of Human Auditory System (HAS) to variations in the balance angle $\phi = \tan^{-1}(\frac{1}{2})$. A threshold of perceptibility is obtained. The scaling factor in Section 2.2 is adjusted to ensure that $\phi^H$ (or marking angle) is below that threshold. Hafter et al [5] explore human sensitivity to differences in intensity of signals applied differentially to the two ears (dichotic). These differences were applied as sequences of clicks at 40dB SPL each. Inter-click intervals (ICI) and number of clicks in a sequence were varied. The subjects were trained to recognize the clicks, by a calibration sequence of diotic clicks (equal intensity to both ears). Fig.1 [5] shows the results for one of 5 subjects (one with the best acuity). This scenario corresponds directly with the proposed modulation of balance between the L and R stereo channels by using the marking angle method of fingerprinting described in this paper. These results indicate that fluctuations of the order of 0.25dB should be imperceptible under most conditions. This takes into account that the marking does not affect the whole signal power, but just the principal components, the fact that the number of clicks is likely to be larger, and that the ICI is likely to be longer than 10ms. Fluctuations of 0.25dB translate to a ±1.45% variation, and therefore $|\phi| < 1^\circ$. It should be noted that the tests in [5] were performed using headphones, in a controlled acoustic environment, optimized for detection of such effects.

4. IMPLEMENTATION

4.1 Fingerprint Embedding
Fig.2. below shows a conceptual signal flow diagram for the combined embedding of the proposed fingerprint and a synchronization watermark. The synchronization watermark can be a pseudonoise array in frequency-time domain, which guards against cropping, sync attack, and pitch shifting [22].

![Fig. 2. Fingerprint and Watermark Embedding Scheme](image)

**Fig. 2. Fingerprint and Watermark Embedding Scheme**

Such a synchronization watermark can carry a limited amount of data, which can be used to establish proof of ownership of the audio clip by the originator. This is shown in Fig.3. All recipients are issued with the same synchronization and data watermarks, so that collusion attacks are ineffective against these. Since these marks are common, they are embedded on an un-fingerprinted audio clip, to serve as a reference signal in the fingerprint recovery process (shown in Fig.5). The fingerprinted copies are distributed, whilst the reference and the raw audio is kept by the distributor.

![Fig. 3. Reference (Watermarked) Audio Generator](image)

**Fig. 3. Reference (Watermarked) Audio Generator**

4.2 Fingerprint Recovery

4.2.1 Synchronization Recovery
In order to recover the fingerprint, the distributed audio is processed, firstly to recover the synchronization mark. This is performed by transforming the audio and looking for a peak in the cross-correlation of a template with the embedded pseudorandom array, as shown in Fig.4. Absence of a suitable peak indicates likely tampering. The location of the peak determines timing and frequency corrections, which must be applied before attempting to extract the fingerprint.
4.2.2 Data Recovery

In order to verify that the audio clip under test originated from the distributor, the audio is re-sampled (resynchronized) and corrected for pitch errors, before an attempt to decode the data is performed. The (secret) data is embedded in a pattern of polarity reversals of the synchronization array. This pattern may be augmented by a checksum, or a more sophisticated error detection or correction code. If the correct data is decoded, as shown in Fig. 5, the corrected, transformed audio is suitable for fingerprint recovery in Fig. 6.

4.2.2 Fingerprint Recovery

The corrected, transformed audio clip under test is analyzed and compared with the stored (watermarked) reference clip. Both signals are converted into polar co-ordinates, principal components are found and ranked. The ranking in both signals is reordered according to the permutation generator. Any minor discrepancies in the reordered ranking are corrected. The polar angles in the reference signal are then subtracted from those in the fingerprinted signal, leaving mainly the original fingerprint (possibly scaled and distorted by collusion and/or other attacks) and possibly fingerprints of other colluders, similarly distorted by the attack. All these fingerprints are different cyclic shifts of a pseudonoise array, and are thus recoverable by a complete cross-correlation of the pattern under test, with the reference array. The cyclic shift of each significant peak in this cross-correlation indicates the identity of each participant in a collusion attack.

5. RESULTS

"Linear collusion by averaging is a simple and effective way for a coalition of users to attenuate embedded fingerprints. Averaging, however, is not the only form of collusion attacks available to a coalition of adversaries. In fact, for each component of the multimedia signal, the colluders can output any value between the minimum and maximum corresponding values, and have high confidence that the spurious value they get will be within the range of the just-noticeable-different since each fingerprinted copy is expected to have high perceptual quality. An important class of nonlinear collusion attacks is based upon such operations as taking the maximum, minimum, and median of corresponding components of the colluders’ fingerprinted copies."[10] Our fingerprinting scheme was implemented in Matlab® and tested against collusion attacks: averaging, median, MinMax, ModNeg, Min, Max, and RandNeg attacks. By using sufficiently large arrays, embedding in a suitable transform domain (where the data components are significant and easy to hide) and hiding that domain, we can detect and identify hundreds of colluders using a linear attack. An example of our scheme applied to audio fingerprinting is shown in Fig. 7. The graph shows the results of a linear collusion attack on the Signal to Noise Ratio (SNR). More than 500 colluders can be detected in less than 5 seconds of a popular song. An SNR of 3 is sufficient for a detection probability exceeding 90%. The SNR is well above 10 for 500 colluders. Our fingerprinted audio files were tested by independent audio experts and compared with the original files. The fingerprints were deemed to be imperceptible.

[10] shows the effects of various nonlinear attacks on the probability of detection of a single colluder from a coalition of
Fig. 9. Our Probability of Colluder Detection RandNeg Attack

colluders. In order to maintain a 90% detection probability for Min, Max or RandNeg attacks, the number of colluders must be below 24, as seen from Fig. 8. [12]. Fig. 9 shows the probability of detection of a single colluder from colluders applying the RandNeg attack to our fingerprints. Here, at most, 37 colluders can be accommodated for a detection probability exceeding 90% (i.e. 13 more than in [12]). This is difficult to achieve, since these probabilities follow an exponential behavior. (e^{13/442,412}).

6. ARRAY CONSTRUCTIONS

6.1 Frank Array

Arrays of large size are preferable for fingerprinting, so that the matched autocorrelation of the array is larger than the cross-correlation with the audio. The principal components are restricted in number, so a long, non-square array is preferable. In order to evaluate the performance of the fingerprinting scheme, a custom two dimensional almost perfect autocorrelation array was designed. This array is based on a Frank sequence of length $p^2$ over $p$ roots of unity [4] applied as a shift sequence [20]. The rotations around the unit circle in the complex plane were replaced by cyclic shifts of a binary, or almost binary Legendre sequence of length $p$. Such an array has an autocorrelation of 0 for all shifts with a non-zero horizontal cyclic shift component. Each used was assigned one of these $p(p-1)$ shifts. We chose $p=23$, yielding an array which was applied to 529 DFT time apertures with 23 principal components in each. For cryptographic robustness, the marking was not applied directly, but using a scrambling method based on the discrete logarithm. Primes and prime powers have primitive roots, so this was appropriate for this array geometry. The choice of primitive root and exact details of the permutation mapping can optionally be used as private keys.

6.2 Three Dimensional Constructions

Log Map on a Multidimensional Grid

Fingerprinting video, or multimedia, multi-dimensional arrays with good correlation are required. These arrays should be long in the “time” dimension compared to the other dimensions. Recently, Moreno et al [13] developed a new method of constructing $n$ dimensional arrays of size $(p \times p \ldots \times p) \times (p^{n-1} - 1)$. The constructions are based on substituting the columns of new multi-dimensional Welch-Costas array constructions also discovered by the authors [14]. These constructions use logarithmic (Fig. 10) and exponential map (Fig. 11).
7. CONCLUSIONS

This paper presents a new method of embedding imperceptible fingerprints in audio signals. Using a mix of signal processing techniques, steganography and cryptography, the fingerprints are highly resistant to linear and non-linear collusion attacks. The fingerprints make use of families of arrays with good correlation properties, unlike other methods which use random number generators. New, multi-dimensional arrays have been constructed, so that the fingerprints can be extended to video and multi-media.

8. ACKNOWLEDGMENTS

The authors acknowledge the Australian Research Council for their financial support under Discovery Grant DP0770857, and Monash University for their infrastructure support for this project.

9. REFERENCES