COLLUSION-RESISTANT FINGERPRINTING FOR MULTIMEDIA

Wade Trappe, Min Wu, K.J. Ray Liu

Department of Electrical and Computer Engineering,
University of Maryland, College Park, MD 20742

ABSTRACT

Digital fingerprinting is an effective method to identify users who might try to redistribute multimedia content, such as images and video. These fingerprints are typically embedded into the content using watermarking techniques that are designed to be robust to a variety of attacks. A cheap and effective attack against such digital fingerprints is collusion, where several differently marked copies of the same content are averaged or combined to disrupt the underlying fingerprint. In this paper, we study the problem of designing fingerprints that can withstand collusion, yet trace colluders. Since, in antipodal CDMA-type watermarking, the correlation contributions only decrease where watermarks differ, by constructing binary code vectors where any subset of k or fewer of these vectors have unique overlap, we may identify groups of k or less colluders. Our construction of such anti-collusion codes (ACC) uses the theory of combinatorial designs, and for n users requires $\mathcal{O}(\sqrt{n})$ bits. Further, we explore a block matrix structure for the ACC that reduces the computational complexity for identifying colluders and improves the detection capability when colluders belong to the same subgroup.

1. INTRODUCTION

The advancement of communication technologies, coupled with recent investments in building an infrastructure of ubiquitous communication networks, promises to facilitate the development of a digital marketplace where a broad range of multimedia content, such as images and video, will be available. However, these communication networks not only improve the ability to distribute content, but also make more difficult the task of insuring that content is appropriately used.

In order to control the redistribution of content, digital fingerprinting is used to trace the consumers who use their content for unintended purposes. These fingerprints can be embedded in multimedia content through a variety of watermarking techniques[1, 2]. Conventional watermarking techniques are concerned with robustness against a variety of attacks such as filtering, but do not emphasize robustness against a coalition of users with the same content that contains different marks. These attacks, known as *collusion* attacks, can provide a cost-effective approach to removing an identifying watermark. One of the simplest approaches to perform a collusion attack is to average multiple copies of the content together[3]. Other collusion attacks might involve forming a new content by selecting different pixels or blocks from the different colluders' content.

The authors may be contacted at wxt, minwu, and kjrliu @eng.umd.edu.

In this paper, we study the problem of making fingerprints resilient to a collusion attack by averaging, although our approach holds for collusion attacks that interleave pixels from different images. In Section 2 we review several important related issues involved in spread spectrum watermarking. In Section 3, we introduce the concept of an anti-collusion code (ACC), which is the data embedded into the content to identify users. ACC are designed to be resistant to averaging, and able to exactly identify groups of colluders. Next, we consider grouping the users into subgroups that are likely to collude, and present a code design that improves the detection probability when colluders are within the same subgroup, and provides an efficient method to determine which subgroups are involved in a collusion. Finally, we present simulations in Section 5, and draw conclusions in Section 6.

2. WATERMARKING AND COLLUSION

In this section we review the basics of digital watermarking. Digital watermarking can be considered from a communication perspective where the watermark, or data, is embedded in a host signal. If the host signal is unknown to the detector, then the host signal serves as a noise that hinders the ability to detect the watermark. While it is possible that the host signal is available in centralized fingerprint verification, in this paper we shall consider the worst-case scenario where the host signal is unavailable. Blind detection may allow for more flexibility in fingerprint verification.

We shall consider two types of embedding in this section that are both based upon the basic principles of modulation and detection. The first method is orthogonal signaling. Here M orthogonal signals s_j are used to convey $B = \log_2 M$ bits by inserting one of the M signals into the host signal. The classical method for estimating which signal was embedded in the host signal is done via M correlators, and identifies the B bit message.

The second approach is a TDMA/CDMA-type modulation, where B bits are encoded into a watermark signal w via

$$w = \sum_{j=1}^{B} b_j \mathbf{s}_j, \tag{1}$$

where $b_j \in \{\pm 1\}$, and the signal vectors $\mathbf{s}_j$ are orthogonal. If $b_j = 0$, this is equivalent to having no contribution in the $\mathbf{s}_j$ direction. Determining each bit b_j is done by correlating with the $\mathbf{s}_j$, and comparing against a decision threshold. Orthogonal signaling has better detection capabilities than CDMA-type modulation, but does so at an expense in computational efficiency[1].

[1]Suppose $\mathcal{E}$ energy is used to convey B bits. Orthogonal signaling requires correlating with 2^B waveforms and produces constellation points

A different bit sequence $\{b_j\}$ is employed for each user. We may view the assignment of the bits b_j for different watermarks in a matrix C, where each column contains the bit sequence for a different user. This viewpoint allows us to capture both the orthogonal and CDMA-type watermarking. The identity matrix describes the orthogonal signaling case since each user u_j is only associated with one signal vector $\mathbf{s}_j$.

Consider the case where we are averaging two watermarked signals that have the same host signal, but different watermarks. If orthogonal signaling is used, then each watermark has energy $\mathcal{E}$, but the average of the watermarks has total energy $\mathcal{E}/2$. When correlating with each of the M orthogonal signals, each of the two correlations will be scaled by $1/2$, and the remainder will have 0 correlation. In general, if we average K different watermarks, then the energy of the watermarks will be $\mathcal{E}/K$, and the correlation with a corresponding orthogonal waveform will be scaled by $1/K$. Therefore, as K increases, the identifying watermarks will increasingly attenuate, and become harder to detect.

On the other hand, in CDMA-type modulation, if we average two watermarks, w^1 and w^2 corresponding to bit sequences b^1 and b^2, then where $b^1_j \neq b^2_j$ the contributions cancel, while where $b^1_j = b^2_j$ the contributions do not attenuate. The result is that many of the components will still have $\sqrt{\mathcal{E}/B}$ amplitude, while some will have 0 amplitude. When we average K watermarks, those components in the bit sequences that are all the same will not experience any cancellation, and their amplitude will remain $\sqrt{\mathcal{E}/B}$, while others will experience diminishing (though not necessarily complete cancellation).

3. ANTI-COLLUSION CODES

In this section we focus on the CDMA-type modulation since when we average different CDMA-watermarked signals, the components that agree between the different watermarks do not experience any loss in amplitude, while those bits that differ do. We use this observation to design a family of bit sequences $\{c^j\}$ whose overlap with each other can identify groups of colluding users. A similar idea was proposed in [5], where projective geometry was used to construct such code sequences. As we will explain in this section, the codes constructed in [5] were less efficient than the proposed.

For this section, we shall describe codes using the binary symbols $\{0, 1\}$, which are mapped to $\{-1, 1\}$ via $f(x) = 2x - 1$ for use in CDMA-based watermarking. We assume that when a sequence of watermarks is averaged, the effect it has is that the resulting binary message is the logical AND of the codewords c^j. For example, when the codes (1110) and (1101) are combined, the result is (1100). When we perform 3 or more averages, this assumption does not necessarily hold since the average of many 1's and a few 0's may produce a decision statistic large enough to pass through the detector as a 1. However, soft-decision detection may be used to achieve our assumption.

Definition 1. *A binary code $\mathcal{C} = \{c^1, \cdots, c^n\}$ such that the logical AND of any subset of k or fewer codevectors is non-zero and distinct from the logical AND of any other subset of k or fewer codevectors is a k-resilient anti-collusion code, or an ACC code.*

We first present a $(n-1)$-resilient ACC code. Let $\mathcal{C}$ consist of all n-bit binary vectors that have only a single 0 bit. For example, when $n = 4$, $\mathcal{C} = \{1110, 1101, 1011, 0111\}$. It is easy to see when $k \leq n - 1$ of these vectors are combined under AND, that this combination is unique. This code has cardinality n, and hence can produce at most n differently watermarked media. We shall refer to this code as the trivial ACC code for n users.

It is desirable to squeeze more users into fewer bits. To do this, we need to give up some resiliency. We now present a construction of a k-resilient ACC code that, for n users, requires $\mathcal{O}(\sqrt{n})$ bits. This construction uses balanced incomplete block designs[6]:

Definition 2. *A (v, k, λ) balanced incomplete block design (BIBD) is a pair $(\mathcal{X}, \mathcal{A})$, where $\mathcal{A}$ is a collection of k-element subsets (blocks) of a v-element set $\mathcal{X}$, such that each pair of elements of $\mathcal{X}$ occur together in exactly λ blocks.*

A (v, k, λ)-BIBD has $b = \lambda(v^2 - v)/(k^2 - k)$ blocks. Corresponding to a block design is the $v \times b$ incidence matrix $M = (m_{ij})$ defined by

$$m_{ij} = \begin{cases} 1 & \text{if the } i\text{th element belongs to the } j\text{th block,} \\ 0 & \text{otherwise.} \end{cases}$$

If we assign the codewords as the bit-complement of the column vectors of M then we have a $(k-1)$-resilient ACC.

Theorem 1. *Let $(\mathcal{X}, \mathcal{A})$ be a $(v, k, 1)$-BIBD, and M the corresponding incidence matrix. If the codevectors are assigned as the bit complement of the columns of M, then the resulting scheme is a $(k-1)$-collusion resistant code.*

The proof is provided in the appendix. We now present an example. The following is the bit-complement of the incidence matrix for a $(7, 3, 1)$-BIBD:

$$M = \begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{pmatrix}. \qquad (2)$$

This code requires 7 bits for 7 users and provides 2-resiliency since any two column vectors share a unique pair of 1 bits. Each column vector c of M is mapped to $\{\pm 1\}$ by $f(x) = 2x - 1$. The CDMA watermark is then $\mathbf{w} = \sum_{j=1}^{v} f(c_j)\mathbf{s}_j$. When two watermarks are averaged, the locations where the corresponding ACC codes agree and have a value of 1 identify the colluding users. For example, let

$$\mathbf{w}^1 = -\mathbf{s}_1 - \mathbf{s}_2 + \mathbf{s}_3 - \mathbf{s}_4 + \mathbf{s}_5 + \mathbf{s}_6 + \mathbf{s}_7 \qquad (3)$$

$$\mathbf{w}^2 = -\mathbf{s}_1 + \mathbf{s}_2 - \mathbf{s}_3 + \mathbf{s}_4 + \mathbf{s}_5 - \mathbf{s}_6 + \mathbf{s}_7 \qquad (4)$$

be the watermarks for the first two columns of the above $(7, 3, 1)$ code, then $(\mathbf{w}^1 + \mathbf{w}^2)/2$ has coefficient vector $(-1, 0, 0, 0, 1, 0, 1)$. The fact that a 1 occurs in the 5th and 7th location uniquely identifies user 1 and user 2 as the colluders.

The example that we presented had no improvement in bit efficiency over the trivial ACC code for 7 users, and it had less collusion resilience. A useful metric for evaluating the efficiency of an ACC code for a given resiliency is its rate $R = v/b$, which describes amount of spreading sequences needed per user. For codes (v, k, λ)-BIBD ACC, their rate is $R = (k^2 - k)/(\lambda(v-1))$. ACC codes with lower rates are better. By Fisher's Inequality[6], we

separated by $\sqrt{2\mathcal{E}}$, whereas CDMA requires correlating with B waveforms and has constellation points separated by $2\sqrt{\mathcal{E}/B}$. See [4].

know that $b \geq v$ for a (v, k, λ)-BIBD, and thus $R \leq 1$ using the BIBD construction. In contrast, the k-resilient construction in [5] has rate much larger than 1, and thus requires more spreading sequences (or marking locations) to accommodate the same amount of users than our scheme. Additionally, for the same amount of users, the use of CDMA watermarking with an ACC code constructed using a $(v, k, 1)$-BIBD requires less spreading sequences than orthogonal signaling. A CDMA-ACC scheme would need v orthogonal sequences for $b = (v^2 - v)/(k^2 - k)$ users, while orthogonal signaling would require b sequences.

In general, (v, k, λ)-BIBDs do not necessarily exist for an arbitrary choice of v, and k. The condition that b must be an integer restricts some possibilities for v and k, and for a given (v, k, λ) triple there may not exist a (v, k, λ)-BIBD. We may, however, construct infinite families of BIBDs. For example, $(v, 3, 1)$ systems (also known as Steiner triple systems) are known to exist if and only if $v \equiv 1$ or $3 \pmod 6$. The Bose construction builds Steiner triple systems when $v \equiv 3 \pmod 6$, and the Skolem construction builds Steiner triple systems when $v \equiv 1 \pmod 6$[7]. Additionally, $(p^d, p, 1)$-BIBD can be constructed when p is of prime power[6].

4. SUBGROUP-BASED CONSTRUCTIONS

In this section, we present an approach that decreases the computation requirements needed to identify an individual's or group of individuals' fingerprints. This approach has the added advantage of allowing us to increase the detection statistics when colluders come from the same subgroup.

Suppose the code matrix C is constructed as a block diagonal matrix $C = diag(C_1, C_2, \cdots C_s)$, where the matrices C_j on the diagonal correspond to matrices for smaller codes. If the bits b_j are assigned to the users according to the columns of C, then by correlating with the ensemble of the spreading sequences corresponding to each block, we only need s correlations to determine to which block, or subgroup of users, the fingerprint belongs. Once a subgroup has been identified, the specific user's identity can be determined by correlating with each of the spreading sequences for that block matrix. If each matrix C_j is $t \times t$, then we need $\mathcal{O}(s) + \mathcal{O}(t)$ correlations, as opposed to $\mathcal{O}(st)$ if we had correlated with each of the spreading sequences. This technique also provides a much more efficient method for decoding watermarks that are embedded using orthogonal signaling.

In many applications, a group of users will be suspected of likely colluding, but not with others. This could be due to geographical or social variables, or based upon a previous precedent. In such scenarios, it is desirable to improve the ability to detect collusion within subgroups of users. By using the block matrix construction, we can achieve this by assigning users who might collude to code vectors from the same block matrix.

As an example, let $Q_n = \mathbf{1}_n - 2I_n$ be the $n \times n$ matrix that is all ones except for -1 on the diagonal. This matrix corresponds to the trivial ACC code for n users after applying the $f(x)$ map. Now let us assign the bits b_j according to the columns of

$$\begin{pmatrix} Q_8 & \mathbf{0}_8 \\ \mathbf{0}_8 & Q_8 \end{pmatrix}. \tag{5}$$

Let the subgroups of users likely to collude be assigned to the same Q_8 matrix. Suppose we allocate $\mathcal{E}$ energy for each watermark, and have two colluders from the first eight users. In this case, when they average their watermarks, there will be six components with amplitude $\sqrt{\mathcal{E}/8}$, and the other components will have 0 magnitude. However, if one user is from the first 8 and the other is from the second 8, then their average watermark will consist of all 16 components at magnitude $\sqrt{\mathcal{E}/32}$. Since the first case results in two components with larger magnitude, the detection probability is higher than in the second case. In comparison, when we assign the bits b_j according to the columns of Q_{16}, the collusion of any two members will result in a watermark that has magnitude $\sqrt{\mathcal{E}/16}$ in 14 of the 16 components, and 0 magnitude in the other two.

Therefore, in the block matrix approach, when colluders belong to the same subgroup, the energy of the averaged watermark in some of the components is higher than when colluders belong to different subgroups, or when all users are treated as being in the same subgroup. However, if we have users from different subgroups, the energy in each component is less than if we had used a full code for the entire group. Thus, since there tends to be few colluders compared to the total amount of users, allocating codes according to a block matrix approach improves the ability to detect colluders from the same subgroup at the expense of decreasing the ability to detect colluders from different subgroups.

5. SIMULATION

In order to demonstrate the capabilities of using an ACC code with CDMA watermarking to fingerprint users and detect colluders, we used an additive spread spectrum watermarking scheme similar to that in[2], where the perceptually weighted watermark was added to block DCT coefficients. The detection of the watermark is performed without the knowledge of the host image[8] via the Z detection statistic[3]. We used the 512×512 Lenna as the host image for the fingerprints. The fingerprinted images had no visible artifacts with average PSNR of 39.6dB.

We assigned the code vectors as the column vectors of the bit complement of the incidence matrix for a $(15, 3, 1)$-BIBD that was constructed using the Bose method with a symmetric idempotent quasigroup structure on Z_5 given by the binary operation $x \cdot y = (3x + 3y) \pmod 5$. Two example code vectors that were assigned to user 1 and 6 are

$$\begin{array}{ll} \text{User 1:} & (-1,-1,-1,\ 1,1,1,1,1,1,1,\ 1,1,1,1,1) \\ \text{User 6:} & (-1,\ 1,\ 1,-1,1,1,1,1,1,1,-1,1,1,1,1) \\ \text{Average:} & (-1,\ 0,\ 0,\ 0,1,1,1,1,1,1,\ 0,1,1,1,1)\,. \end{array}$$

This code is designed for 2 colluders, and was used to simulate the collusion of two users who average their fingerprinted versions of the Lenna image. An example of the behavior of the Z statistic when two users collude is depicted in Figure 1, where we depict the values of the Z statistic for the 15 different spreading sequences used when user 1 and user 6 average their differently marked images and then compress using JPEG with quality factor 50%. If we average the code vectors above, the resulting vector will consist of 1's where the two vectors both have a value of 1, a -1 where they have a -1, and 0 where they differ. Large positive statistics are mapped to 1, and large negative statistics are mapped to -1 by the detector. Thus, we expect to observe that the first sequence will have large negative correlation. The second, third, and fourth sequences will have small magnitude since cancellation is taking place, while sequences 5 through 10 will have large magnitude.

We present a histogram containing the Z statistics from roughly 200 pairs of users in Figure 2 when the watermarked images are

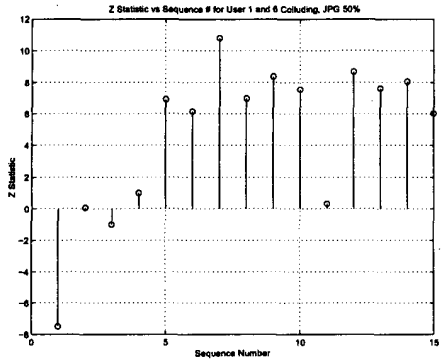


Fig. 1. Example detection statistic values for when user 1 and user 6 average their watermarked images with a $(15, 3, 1)$-BIBD ACC fingerprint. The output of the detector would be $(-1, 0, 0, 0, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1)$.

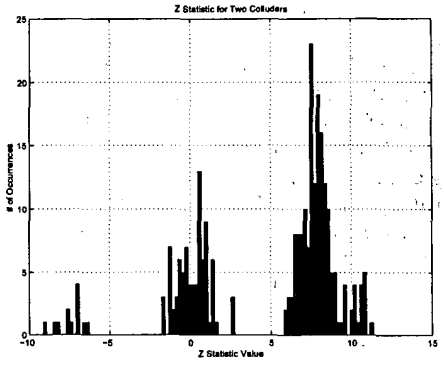
compressed using JPEG with a quality factor of 50%. In the table, we present the mean and variance for the magnitude of the Z statistics for 2 colluders. The Match columns refers to cases where the bits b_j for two different users agreed, while the Mismatch refers to cases where the bits b_j differed for two different users. We see that the average magnitude of the Z statistic when the bit values agree is much larger than the average magnitude for where the bit values disagree. This is also seen in the histogram, where there is a clear distinction between the three detector decision regions.

6. CONCLUSION

In this paper we have proposed anti-collusion codes, which provide collusion resistance for fingerprinting multimedia. The primary attack that we focused on was the averaging of differently marked versions of the same content. We observed that when orthogonal signaling is used to convey the fingerprint, the energy in the watermark decreases with the amount of colluders, which will cause the fingerprints to disappear with enough colluders. On the other hand, when CDMA-type signaling is used, the correlation contributions only decrease where the bits from different watermarks differ. We designed codes so that the logical AND of the codes for k colluders would uniquely identify the colluders. We also presented an approach to decrease the computational requirements needed to identify an individual fingerprint or group of colluders. Finally, we presented simulations showing that our code construction is able to identify 2 colluders when used in CDMA signaling.

7. APPENDIX

We prove the theorem by working with the blocks A_j of the BIBD. The bitwise complementation of the column vectors corresponds to complementation of the sets $\{A_j\}$. We would like for $\cap_{j\in J} A_j^C$ to be distinct over all sets J with cardinality less than or equal to $k-1$. By De Morgan's Law, this corresponds to uniqueness of $\cup_{j\in J} A_j$ for all sets J with cardinality less than or equal to $k-1$. Suppose we have a set of $k-1$ blocks $A_1, A_2, \cdots, A_{k-1}$, we must show that there does not exist another set of blocks whose union produces the same set. There are two cases to consider. First, assume there is another set of blocks $\{A_i\}_{i\in I}$ with $\cup_{j\in J} A_j = \cup_{i\in I} A_i$ such that $I \cap J = \emptyset$ and $|I| \leq k-1$. Suppose we take a block A_{i_0} for $i_0 \in I$. Then A_{i_0} must share at most one element with each A_j, otherwise it would violate the $\lambda = 1$ assumption of the BIBD. Therefore, the cardinality of A_i is at most $k-1$, which contradicts the requirement that each block have k elements. Thus, there does not exist another set of blocks $\{A_i\}_{i\in I}$ with $\cup_{j\in J} A_j = \cup_{i\in I} A_i$ and $I \cap J = \emptyset$. Next, consider $I \cap J \neq \emptyset$. If we choose $i_0 \in I \backslash (I \cap J)$ and look at A_{i_0}, then again we have that A_{i_0} can share at most 1 element with each A_j for $j \in J$, and thus A_{i_0} would have fewer than k elements, contradicting the fact that A_{i_0} belongs to a $(v, k, 1)$-BIBD. Thus, $\cup_{j\in J} A_j$ is unique.



	Match		Mismatch	
	$\overline{\lvert Z\rvert}$	$\sigma^2_{\lvert Z\rvert}$	$\overline{\lvert Z\rvert}$	$\sigma^2_{\lvert Z\rvert}$
JPEG 50%	7.96	1.25	0.82	0.32

Fig. 2. Histogram of Z statistic values for many different pairs of colluders using the $(15, 3, 1)$-BIBD construction.

8. REFERENCES

[1] I. Cox, J. Kilian, F. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia," *IEEE Tran. on Image Proc.*, vol. 6(12), pp. 1673–1687, December 1997.

[2] C. Podilchuk and W. Zeng, "Image adaptive watermarking using visual models," *IEEE Journal on Selected Areas in Communications*, vol. 16(4), pp. 525–540, May 1998.

[3] H. S. Stone, "Analysis of attacks on image watermarks with randomized coefficients," *NEC Technical Report*, 1996.

[4] M. Wu and B. Liu, "Modulation and multiplexing techniques for multimedia data hiding," in *Proc. of SPIE ITcom'01, SPIE vol 4518*, Aug. 2001.

[5] J. Dittmann, P. Schmitt, E. Saar, J. Schwenk, and J. Ueberberg, "Combining digital watermarks and collusion secure fingerprints for digital images," *SPIE Journal of Electronic Imaging*, vol. 9, pp. 456–467, 2000.

[6] J. H. Dinitz and D. R. Stinson, *Contemporary Design Theory: A Collection of Surveys*, John Wiley and Sons, 1992.

[7] C.C. Lindner and C.A. Rodger, *Design Theory*, CRC Press, 1997.

[8] W. Zeng and B. Liu, "A statistical watermark detection technique without using original images for resolving rightful ownerships of digital images," *IEEE Tran. on Image Proc.*, vol. 8(11), pp. 1534–1548, November 1999.