

Construction of Anti-Collusion Codes Based on Cover-Free Families

Qiaoliang Li

Department of Computer Science,
Hunan University, Changsha,
40082, P. R. of China
liqiaoliang@eyou.com

Yingshu Li, Yi Pan

Department of Computer Science,
Georgia State University, Atlanta,
GA, 30303, USA
yli@cs.gsu.edu, pan@cs.gsu.edu

Xiaoming Wang

Department of Computer Science,
Shaanxi Normal University, Xian,
710062, P. R. of China
wangxmsnnu@hotmail.com

Pingzhi Fan

Institute of Mobile Communications,
Southwest Jiaotong University,
Chendu 610031, China
P.fan@ieee.org *

Abstract

Digital fingerprinting is a technique for identifying users who use multimedia content for unintended purpose, such as redistribution. In order to characterize the fingerprints that can withstand collusion attack, Dittmann et. al proposed the definition of Anti-Collusion Codes (ACCs). In this paper, we present a new approach to construct anti-collusion codes based on cover-free families. All the previous constructions of ACCs are special cases of our construction. Especially, we can construct ACCs by using error-correcting codes. Collusion resistance comparison of different schemes show that our scheme is more efficient than previous ACCs. Some related problems on this subject are proposed.

1 Introduction

With the advancement of multimedia technologies, coupled with the development of an infrastructure of ubiquitous broadband communication networks, a large amount of multimedia contents, such as image, video, audio and speech, will be available in digital marketplace. However, such an advantage also poses the challenge of insuring that content is appropriately used. Devising techniques for copyright protection of digital contents have been an important issue.

As one of the prominent solutions, 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. Conventional watermarking techniques are concerned with the robustness against a variety of attacks, such as filtering but do not always address robustness to attacks mounted by a coalition of users with the same content that contains different marks. These attacks, which are known as collusion attacks, can provide a cost-effective approach to remove an identifying watermark. One of the simplest approaches to perform a collusion attack on multimedia is to average multiple copies of the content together. By gathering enough coalition of colluders, it is possible to sufficiently attenuate each of the colluders' identifying fingerprints.

W. Trappe et al.[1] proposed AND Anti-Collusion Codes (AND-ACC) and devised collusion secure codes against the averaging attack by using bit complement of the incident matrices of Balanced Incomplete Block Designs (BIBD). A similar idea was proposed in [2], by J. Dittmann et al. [2] and by Yagi et al. [3] in using finite projective geometries to construct AND-ACC codes. Recently, I. Kang et al. [4] gave a construction of AND-ACC using group-divisible design.

In this paper, we propose a new class of AND-ACC based on cover free families. All the previous constructions of AND-ACCs are special cases of our construction. Especially, we can construct AND-ACC by using error-correcting codes.

The paper is organized as follows: in section 2, we introduce the fingerprint model we adopted. Section 3 is our main result. In this section, we introduce the AND-ACC

*Supported by National Natural Science Foundation of China (NSFC) under Grant No.60773224, 10571052, the Key Research Project of Ministry of Education of China under Grant No.107106, and the 111 Project of China under Grant No.111-2-14.)

based on cover free families. In section 4, we show some applications of our main results. Finally, we present conclusions in Section 5.

2 Fingerprinting Model

Let us consider additive embedding, where a watermark signal is added to a host signal. Suppose that the host signal is a vector denoted as \mathbf{x} and we have a family of watermarks $\{\mathbf{w}_j\}$ that are fingerprints associated with different users who purchase the rights to access \mathbf{x} . Before the watermarks are added to the host signal, every component of each \mathbf{w}_j is scaled by an appropriate factor, i.e., $s_j(l) = \alpha(l)w_j(l)$, where we refer to the l -th component of a vector \mathbf{w}_j by $w_j(l)$. One possibility for $\alpha(l)$ is to use the Just-Noticeable-Difference (JND) from human visual system models. Corresponding to each user is a marked version of the content $\mathbf{y}_j = \mathbf{x} + \mathbf{s}_j$. The content may experience additional distortion before it is tested for the presence of the watermark \mathbf{s}_j . We represent this additional distortion by \mathbf{z} . There are therefore two possible sources of interference hindering the detection of the watermark: the underlying host signal \mathbf{x} and the distortion \mathbf{z} . For simplicity of notation, we gather both of these possible distortions into a single term denoted by \mathbf{d} . A test content \mathbf{y} that originates from user j can thus be modeled by

$$\mathbf{y} = \mathbf{s}_j + \mathbf{d}. \quad (1)$$

The fingerprints \mathbf{w}_j are often chosen to be orthogonal noise-like signals directly or are represented using a basis of v orthogonal noise-like signals \mathbf{u}_i via

$$\mathbf{w}_j = \sum_{i=1}^v b_{ij} \mathbf{u}_i \quad (2)$$

where $b_{ij} \in \{0, 1\}$ (On-Off Keying (OOK)) or $b_{ij} \in \{\pm 1\}$ (antipodal form).

We may regard the assignment of the bit b_{ij} for different fingerprints as a matrix $\mathbf{B} = (b_{ij})$, which is called the derived code matrix. Each column of \mathbf{B} contains a derived codevector for a different user. In the following section, we will design a code matrix C whose elements are either 0 or 1. By applying a suitable mapping that depends on whether the OOK or antipodal form of the code modulation is used, the code matrix C is used to derive the matrix B that is used in forming fingerprinting signals.

The purpose is to design codes using the binary symbols $\{0, 1\}$. According to the description of the above paragraph, the designed codes should satisfy that when k or fewer users collude, we can identify the colluders.

Definition 1. let $G = \{0, 1\}$. A code $C = \{c_1, c_2, \dots, c_n\}$ of vectors belonging to G^v is called a k -resilient AND anti-collusion code (AND-ACC) when any

subset of k or fewer code-vectors combined element-wise under AND is distinct from the element-wise AND of any other subset of k or fewer code-vectors.

We prefer shorter codes since embedding longer fingerprints would distribute energy over more basis vectors.

W. Trappe et. al. have also constructed a class of AND-ACC based on the bitwise complement of adjacent matrix of balanced incomplete block design (BIBD) The definition of BIBD is as follows.

Definition 2. A (v, k, λ) Balanced Incomplete Block Design (BIBD) is a pair (X, \mathcal{A}) , where \mathcal{A} is a collection of k -element subsets (block) of a v -element of X , such that each pair of elements of X occur together in exactly λ blocks.

For a complete comprehensive survey of BIBD, readers is referred to [5]. A (v, k, λ) -BIBD has a total of $n = \lambda(v^2 - v)/(k^2 - k)$ blocks. Corresponding to a block design is a $v \times n$ 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 define the codematrix C as the bit-complement of M and assign the codevectors \mathbf{c}_j as the columns of C . W. Trappe et al. proved the following

Theorem 1. Let (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 $(k - 1)$ -resilient AND-ACC.

3 ACCs based on cover free families

A set system (X, F) with n elements and N blocks is called a r -cover free family (or r -CFF(n, N)) provided that, for any r blocks $A_1, A_2, \dots, A_r \in F$ and any other block $B_0 \in F$, we have

$$B_0 \not\subseteq \cup_{j=1}^r A_j.$$

Cover-free families were considered from different subjects, such as information theory, combinatorics and group testing. In 1964, cover-free families were first introduced by Kautz and Singleton. These codes may be used for retrieval files, data communication and magnetic memories. After that, some researchers conducted research in this area. We refer to [6] for a complete survey on this subject. In the following, we propose a new class of AND-ACC based on cover free families.

Theorem 2. (X, F) is a r -cover free family and M is the corresponding incidence matrix. If the code-vectors are assigned as the bit complement of the columns of M , then the resulting scheme is a r -resilient AND-ACC.

Proof. Let A_1, A_2, \dots, A_r be the blocks of a r -cover free family and M be its incidence matrix. C be the bit complement of M . We want to prove that any subset of k

or fewer code-vectors combined element-wise under AND is distinct from the element wise AND of any other subset of k or fewer code-vectors. Let I, J be any two subset of $\{1, 2, \dots, r\}$ with $|I| = t$ and $|J| = k$ $1 \leq t, k \leq r$, $I \cap J = \phi$. We want to prove that $\cap_{i \in I} A_i^C$ and $\cap_{j \in J} A_j^C$ are distinct. By De Morgan's law, this corresponds to prove that $\cup_{i \in I} A_i$ and $\cup_{j \in J} A_j$ are distinct. Assume to the contrary that $\cup_{i \in I} A_i = \cup_{j \in J} A_j$. It follows that for any $i \in I$ we have $A_i \subseteq \cup_{j \in J} A_j$, which contradicts to the concept of cover-free families.

For applications, researchers tried to find efficient constructions of cover-free families. Three basic methods were used to construct CFF: packing design, separating hash families and the methods from coding theory. As an application of our main result, we have the following construction of AND-ACC based on packing designs:

Definition 3. A $t - (v, k, 1)$ packing design is an ordered pair (V, \mathcal{B}) , where V is a v -element set and \mathcal{B} is a collection of k -subsets of V (called blocks) such that every t -subset of V occurs in at most one block of \mathcal{B} .

The maximum number of blocks $D(v, k, t)$ in any $t - (v, k, 1)$ packing designs is called packing number. The packing problem is mainly to determine the packing number $D(v, k, t)$. The results of $D(v, k, 2)$, $3 \leq k \leq 5$, $D(v, 4, 3)$ are stated in [5].

In [6], the authors proved the following:

Theorem 3. If there exists a $t - (v, k, 1)$ packing design having b blocks, then there exists a r -CFF, where $r = \lfloor (k-1)/(t-1) \rfloor$.

By Theorem 1 and Theorem 2, we have the following corollary 1, which is the main result of [3].

Corollary 1. If there exists a $t - (v, k, 1)$ packing design with b blocks, then there exists a r resilient AND-ACC with code length v and b codewords, where $r = \lfloor (k-1)/(t-1) \rfloor$.

In [7], the authors used orthogonal arrays to construct packing designs. An orthogonal array $OA(t, k, s)$ is $k \times s^t$ array, with entries from a set of $s \geq 2$ symbols, such that in any t rows, every $t \times 1$ column vector appears exactly once.

Suppose a column in an $OA(t, k, s)$ is $\{s_1, s_2, \dots, s_k\}$. Define a block as $\{(s_1, 1), (s_2, 2), \dots, (s_k, k)\}$ accordingly. In this way, a $t - (ks, k, 1)$ packing design can be obtained from an $OA(t, k, s)$

It is well known [5] that if q is a prime power and $t < q$, then there exists an $OA(t, q+1, q)$, and hence a $t - (q^2 + q, q+1, 1)$ packing design with q^t blocks exists. Therefore we have the following:

Corollary 2. For any prime power q and integer $t < q$, there exists a $\lfloor \frac{q}{t-1} \rfloor$ -resilient AND-ACC with code length $q^2 + q$ and q^t codewords.

Since BIBD and the construction in [3] are special class of packing design, so our construction contains their construction.

Using separating hash families, the author proved [6]

that for any positive integer r , there exists an explicit construction for an infinite class of r -CFF with n elements and $(O(n)^{\log(r+1)})$ blocks. So we have

Corollary 3. For any positive integers r , there exists a r -resilient AND-ACC with code length n and $(O(n)^{\log(r+1)})$ codewords.

Suppose \mathcal{C} is an (n, N, q) code on an alphabet set Q with q elements. Define $X = \{1, 2, \dots, n\} \times Q$, and for each codeword $c = (c_1, c_2, \dots, c_n)$, define an n -subset of X as $B_c = \{(i, c_i) : 1 \leq i \leq n\}$. Finally, define $\mathcal{B} = \{B_c : c \in \mathcal{C}\}$. In [8], the authors proved the following:

Theorem 5. Suppose that \mathcal{C} is an (n, N, q) -code having minimum distance d . Then there is a r -CFF($nq; N$), where $r = \lfloor \frac{n-1}{n-d} \rfloor$.

Using shorten Reed-Solomon codes in [8], we have:

Corollary 4. Suppose q, r and d are given, with q a prime power. Then there exists an r -resilient AND-ACC with code length $(q+1-s)q$ and the number of codewords is $q^{\lfloor (q+r-d-s)/r \rfloor}$.

We give a concrete construction of AND-ACC based on coding theory.

Example 1. Let $q = 4, n = 3$ and $k = 2$. The polynomial $\varphi = x^2 + x + 1$ is primitive for F_4 over F_2 . If w is a root of this polynomial, we have $F_4 = \{0, 1, w, w^2\}$. A 4-ary $[3, 2, 2]$ Reed-Solomon code φ has generator polynomial of degree 1. Let us take $g(x) = x - w$. The code is then expressed as $\varphi = \{p(x)(x - w) | \deg(p(x)) \leq 1\}$.

It can be verified that the 16 codewords in φ are

$$\begin{array}{cccccc} 000, & w10, & w^2w0, & 1w^20, & 0w1, & 0w^2w, \\ 01w^2, & ww^21 & w^201, & 111, & ww^w, & w^21w, \\ 10w, & ww^w2, & w^2w^2w^2, & 1ww^2 & & \end{array}$$

Let

$$\begin{array}{cccc} (1, 0) \mapsto 1 & (2, 0) \mapsto 5 & (3, 0) \mapsto 9 & (1, 1) \mapsto 2 \\ (2, 1) \mapsto 6 & (3, 1) \mapsto 10 & (1, \omega) \mapsto 3 & (2, \omega) \mapsto 7 \\ (3, \omega) \mapsto 11 & (1, \omega^2) \mapsto 4 & (2, \omega^2) \mapsto 8 & (3, \omega^2) \mapsto 12 \end{array}$$

Then

$$\begin{array}{lll} B_1 = \{1, 5, 9\} & B_2 = \{3, 6, 9\} & B_3 = \{4, 7, 9\} \\ B_4 = \{2, 8, 9\} & B_5 = \{1, 7, 10\} & B_6 = \{1, 8, 11\} \\ B_7 = \{1, 6, 12\} & B_8 = \{3, 8, 10\} & B_9 = \{4, 5, 10\} \\ B_{10} = \{2, 6, 10\} & B_{11} = \{3, 7, 11\} & B_{12} = \{4, 6, 11\} \\ B_{13} = \{2, 5, 11\} & B_{14} = \{3, 7, 12\} & \\ B_{15} = \{4, 8, 12\} & B_{16} = \{2, 7, 12\} & \end{array}$$

The above is a $\lfloor (3-1)/(3-2) \rfloor = 2$ -CFF. Corresponding to a block design is the 12×16 incident matrix $M = (m_{ij})$. Define the codematrix C as the bit-complement of M , then we obtain the required codematrix.

4 Analysis and Comparison

From section 4, we can construct some new classes of ACCs based on packing designs, orthogonal arrays,

separating hash families and codings. Let $f_{EG}(\mu) = p^{(m-\mu)s} \prod_{i=1}^{\mu} \frac{p^{(m-i+1)s}-1}{p^{(\mu-i+1)s}-1}$, $f_{PG}(\mu) = \prod_{i=0}^{\mu} \frac{p^{(m-i+1)s}-1}{p^{(\mu-i+1)s}-1}$ and $P(k, v) = \lfloor v \lfloor (v-1)/(k-1) \rfloor / k \rfloor = J(k, v)$. In the following table, we list some of the parameters of the known ACCs.

Type	Elements	Blocks	Resilience
BIBD	v	$\frac{v^2-v}{k^2-k}$	$(k-1)$
$EG(m, p^s)$	p^{ms}	$f_{EG}(\mu)$	$p^s - 1$
$PG(m, p^s)$	$\frac{p^{(m+1)s}-1}{p^s-1}$	$f_{PG}(\mu)$	p^s
Packing	v	$P(k, v, 1)$	$(k-1)$
SHF	n	$O(n^{\log(r+1)})$	r
codes	$(q+1-s)q$	$q^{\lfloor (q+r-d-s)/r \rfloor}$	r

Table 1. parameter of the known AND-ACC.

A useful metric for evaluating the efficiency of an ACCs for a given collusion resistance r is $\beta = n/v$, where n denoted the number of users, and v is the code length. The parameter $\beta = n/v$ describes the amount of users that can be accomodated per basis vector. ACCs with a higher β are better. The following is a efficiency comparison between the ACCs based on codes and the ACCs based on BIBD. For BIBD, we have $k-1 = r$, and $\beta = n/v = (v-1)/(k^2-k)$. For ACCs based on Codes, we choose $s=0$, then the length of the codes is $v = q(q+1)$. Solving this equation, we have $q = (-1 + \sqrt{1+4v})/2$. So $n = ((-1 + \sqrt{1+4v})/2)^{\lfloor ((-1 + \sqrt{1+4v})/2 + r)/r \rfloor}$ and $\beta = n/v = ((-1 + \sqrt{1+4v})/2)^{\lfloor ((-1 + \sqrt{1+4v})/2 + r)/r \rfloor} / v$. In figure 1, we give the difference of the β value of ACCs based on codes and the β value of ACCs based on BIBD. From the figure, we can see that the difference is larger than 0. With the increasing of v and r , the difference becomes larger.

5 Conclusion

In this paper, we present a new approach to construct ACC based on cover free families. As the applications of our main results, we give three ways to construct ACC, that is, using packing designs, separating hash families and coding theory. We also give a parameter comparison of all the methods. From our results, we can get low bounds on the maximum number of codewords of ACC with a fixed code length. It is an interesting problem to characterize ACC using combinatorial theory. Realizing that secure frameproof codes and traceability schemes also have tight connection with cover-free families, and it is interesting to study the connection of ACC and all these codes.

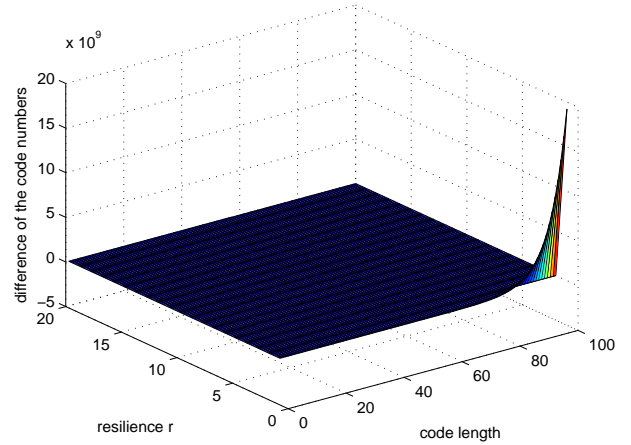


Figure 1. Difference of the number of ACCs based on codes and packing designs

References

- [1] W. Trappe, M. Wu, Z. Wang and K. J. R. Liu, "Anti-collision Fingerprinting for Multimedia," *IEEE Trans on Signal Processing*, vol.51, pp.1069-1087, Apr. 2003.
- [2] J. Dittmann, P. Schmitt, E. Saar, J. Schwenk, and J. Ueberberg, "Combining digital watermarks and collusion secure fingerprints for digital images," *SPIE J. Electron. Imag.*, vol. 9, pp. 456-467, 2000.
- [3] H. Yagi, T. Matsushima and S. Hirawsawa, "Improved Collusion Secure Codes for Digital Based on Finite Geometries," *IEEE Int. Conf. on System, Man and Cybernetics*, 2007, 948-953.
- [4] I. Kang, K. Sinha, and H. Lee, "New digital fingerprint code construction scheme using group-divisible design," *IEICE Trans. Fundamentals*, vol. E89-A, pp. 3732-3735, Dec.2006.
- [5] C. J. Colbourn and J. H. Dinitz, *The CRC Handbook of Combinatorial Designs*. Boca Taton: CRC, 1996.
- [6] D. R. Stinson, Tran van Trung and R. Wei, Secure frameproof codes, key distribution patterns, group testing algorithms and related structures, *J. Statist. Planning Inference*, vol. 86, pp. 595-617, 2000
- [7] D. R. Stinson and R. Wei, Combinatorial properties and constructions of traceability schemes and frameproof codes, *SIAM J. Discr. Math.*, vol. 11, pp. 41-53, 1998.
- [8] A. G. Dyachkov, A. J. Macula and V. V. Rykov, New constructions of supercomposed codes, *IEEE Trans. Inf Theory*, vol.46, pp.248-290, Jan. 2000