Improved Collusion-Secure Codes for Digital Fingerprinting Based on Finite Geometries

Hideki Yagi, Toshiyasu Matsushima and Shigeichi Hirasawa

***Abstract*— Digital fingerprinting, a copyright protection technique for digital contents, is considered. Digital fingerprinting should deter collusion attacks, where several fingerprinted copies of the same content are mixed to disturb their fingerprints. In this paper, we consider the averaging attack, which has effect for multimedia fingerprinting. We propose new collusion-secure fingerprinting codes based on finite geometries (FGs) which increase the rate of conventional collusion-secure codes, while they guarantee to identify the same number of colluders. Due to the new FG-based fingerprinting codes, the system can deal with a larger number of users to distribute a digital content.**

I. Introduction

With the high advances of information technologies, a large amount of digital contents can be processed by computers and devising techniques for copyright protection of digital contents has been an important problem to be solved. As one of the most prominent solutions, *digital fingerprinting* has attracted a great deal of attention. The digital fingerprinting embeds a user's ID called a *fingerprint* into an original content with a watermarking technique and the fingerprinted contents are distributed to users.

Digital fingerprinting requires robustness against *collusion attacks*, where more than one illicit user collude to take illegal actions to the distributed contents. One of well-known collusion attacks are the *interleaving attack* [1], [4], [8], [9] and the *averaging attack* [4], [11], [12], [13]. W. Trappe et al. have devised collusion-secure codes against the averaging attack based on balanced incomplete block design (BIBD) [11]. The collusion-secure codes devised by Trappe et al. is called BIBD-based *anti-collusion* (AC) codes and it has been reported that these codes have robustness against the averaging attack [12].

In this study we propose, based on finite geometries, methods for improving BIBD-based AC codes devised by Trappe et al. or Yang et al. [15] by increasing its coding rate while their resilience is maintained. Consequently, we can realize content distribution system which provides services for greater number of users.

II. Fingerprinting Model

A. Digital Fingerprinting

When distributing a digital content to users, a codeword corresponding to each user is embedded into an original content by a watermarking technique. The codeword allocated for each user is called the user's *fingerprint*. Some illicit users may collude and attempt to disturb their fingerprints so that their fingerprints are not revealed from an illegally utilized content. This action is called a *collusion attack*. The detector of colluders estimates colluders' fingerprints from the disturbed fingerprint.

Let $\Gamma = \{1, 2, \ldots, |\Gamma|\}$ be a set of users of a digital content and we denote a codeword to the user $j \in \Gamma$ by $\boldsymbol{b}_j = (b_{j1}, b_{j2}, \ldots, b_{jN})^T \in \{0,1\}^N$, where T denotes the transposition. The fingerprint watermark $\boldsymbol{w}_j$ is created by using N orthogonal bases $\{\boldsymbol{u}_i \in \mathcal{R}^N | \, i = 1, 2, \ldots, N\}$ with equal energy and a codeword $\boldsymbol{b}_j$ as

$$\boldsymbol{w}_j = \sum_{i=1}^{N} (2b_{ij} - 1)\boldsymbol{u}_i. \tag{1}$$

Next, regarding the distributed content to users as the host signal, the created watermark signal is embedded into it. Denoting the host signal by a vector $\boldsymbol{x} \in \mathcal{R}^N$, the distributed content to the user $j \in \Gamma$ is[1] $\boldsymbol{y}_j = \boldsymbol{x} + \boldsymbol{w}_j$.

Since the fingerprint is embedded by using a watermarking technique, any users cannot detect their own fingerprint $\boldsymbol{w}_j$ from the watermarked content $\boldsymbol{y}_j$. Therefore illicit users may collude to disturb their fingerprints by creating an illegal content from their distributed contents.

B. Assumed Collusion Attack

We consider a set of colluders with the size $h \geq 1$, denoted by $\mathcal{S}_c \subseteq \Gamma$, and without loss of generality, we assume $\mathcal{S}_c = \{1, 2, \ldots, h\}$. The attacked host signal by a set of colluders $\mathcal{S}_c$ is expressed as

$$\boldsymbol{y} = \frac{1}{h}\sum_{j=1}^{h} \boldsymbol{y}_j = \boldsymbol{x} + \frac{1}{h}\sum_{j=1}^{h}\sum_{i=1}^{N} (2b_{ij} - 1)\boldsymbol{u}_i. \tag{2}$$

The detector of the colluders estimates the set of colluders $\mathcal{S}_c$ from the attacked host signal $\boldsymbol{y} \in \mathcal{R}^N$. This attack is called the *averaging attack*, which is one of well-known collusion attacks[2] [4], [11], [12], [13].

This work is supported by Waseda University and the Telecommunications Advancement Foundation (TAF).

H. Yagi is with Media Network Center, Waseda University, Tokyo 169-8050, Japan yagi@hirasa.mgmt.waseda.ac.jp

T. Matsushima and S. Hirasawa are with the School of Science and Engineering, Waseda University, Tokyo 169-8555, Japan

[1] In this paper, for simplicity, the fingerprinted content is defined in this manner. More precisely, each $\boldsymbol{w}_j$ is multiplied by some value called Just-Difference Noticeable (JDN) coefficient [7], before it is added to the host signal.

[2] For simplicity, although we only state the case of the averaging attack, the argument here can hold for the *logical OR attack* [15]. The AC codes devised by J. Yang et al. [15] are also based on BIBD, and the proposed method in this paper can also improve their performance.

III. Anti-Collusion Codes against Averaging Attack

A. BIBD-based AC Codes

Trappe et al. have proposed BIBD-based anti-collusion (AC) codes [11]. First, we introduce the definition of the AC codes.

Definition 1: Assume that the host signal $\boldsymbol{x}$ is known to the detector. If the size of a set of colluders $\mathcal{S}_c$ satisfies $|\mathcal{S}_c| \leq \ell$ for some positive constant ℓ, the code which can reveal all the colluders in $\mathcal{S}_c$ is referred to as an *ℓ-resilient AC code*. The parameter ℓ is called the *resilience* of the AC codes. □

Trappe et al. have constructed AC codes whose codewords have a constant Hamming weight k and whose two distinct codewords have at most one "1-entry" in common based on BIBD. It has been shown that this code becomes a $(k-1)$-resilient AC code.

Consider a set $\mathcal{X}$ of v elements and call a set of k elements *block*. If any pairs of two distinct elements are contained in exactly λ blocks, the system of the elements and the blocks is called a (v, k, λ) *balanced incomplete block design (BIBD)*. For a j-th block, let $\boldsymbol{b}_j = (b_{1j}, b_{2j}, \ldots, b_{vj})^T$ of the length v take 1-entry in the i-th position if this block contains the i-th element in $\mathcal{X}$, and 0-entry otherwise. We call a matrix $B_1 = [b_{ij}]$ whose columns are composed of these vectors of all blocks the *incident matrix* of the BIBD. Trappe et al. have proposed an AC code whose codewords are columns of the incident matrix of an $(N, k, 1)$ BIBD [11]. The resilience of this AC code can be guaranteed by the following lemma.

Lemma 1 ([11]): Letting $B_1 = [b_{ij}]$ be the incident matrix of an $(N, k, 1)$ BIBD, we denote the j-th column of B_1 by $\boldsymbol{b}_j$ and its Hamming weight by $w_H(\boldsymbol{b}_j) = k$ for some $k > 2$. If $\boldsymbol{b}_j$ is the j-th user's fingerprint, a set of column vectors, denoted by $\mathcal{B}_1 = \{\boldsymbol{b}_j\}$, becomes a $(k-1)$-resilient AC code. i.e., if $|\mathcal{S}_c| \leq k-1$, any set of colluders $\mathcal{S}_c$ can be uniquely detected. □

We here consider a mechanism for detecting a set of colluders by using a $(k-1)$-resilient AC code.

Remark 1: Let $\mathcal{Q}(\mathcal{S}_c)$ be a set of symbol positions where any fingerprints in $\mathcal{S}_c$ equally take 0-entry. A $(k-1)$-resilient AC code uniquely determines the set $\mathcal{Q}(\mathcal{S}_c)$ for any $\mathcal{S}_c$ with $|\mathcal{S}_c| \leq k-1$. □

In order to explain the principle of the detecting method, we use an example for simplicity. Let $k = 3$ and $\mathcal{S}_c = \{i, j\}$ whose fingerprints are given by $\boldsymbol{b}_i = (1, 1, 0, 1, 0, 0, 0), \boldsymbol{b}_j = (1, 0, 1, 0, 0, 1, 0)$. Then from eq. (1),

$$\boldsymbol{w}_i = \boldsymbol{u}_1 + \boldsymbol{u}_2 - \boldsymbol{u}_3 + \boldsymbol{u}_4 - \boldsymbol{u}_5 - \boldsymbol{u}_6 - \boldsymbol{u}_7 \tag{3}$$

$$\boldsymbol{w}_j = \boldsymbol{u}_1 - \boldsymbol{u}_2 + \boldsymbol{u}_3 - \boldsymbol{u}_4 - \boldsymbol{u}_5 + \boldsymbol{u}_6 - \boldsymbol{u}_7. \tag{4}$$

As a result of collusion, an illegal content $\boldsymbol{y}$ is produced. The detected sequence $\boldsymbol{y} - \boldsymbol{x} = (\boldsymbol{w}_i + \boldsymbol{w}_j)/2$ has a coefficient vector[3] $(1, 0, 0, 0, -1, 0, -1)$, whose position set of the (-1)-entry, $\{5, 7\}$, corresponds to the position set $\mathcal{Q}(\mathcal{S}_c)$, which expresses positions of 0-entry in the both $\boldsymbol{b}_i$ and $\boldsymbol{b}_j$. From Remark 1, since a $(k-1)$-resilient AC code uniquely identifies $\mathcal{Q}(\mathcal{S}_c)$ for any $\mathcal{S}_c$ of the size less than k, the set of positions of (-1)-entry reveals that the user i and j take participate in the collusion. Even for a general case, any $(k-1)$-resilient AC code can identify colluders [11], [12].

B. Class of AC Codes Based on Finite Geometries

A subclass of BIBD-based AC codes by Trappe et al. can be algebraically constructed by using finite geometries. In this paper, we focus on this subclass of the Trappe's ℓ-resilient AC codes based on finite geometries. We briefly describe two kinds of finite geometries. Refer to [5], [10] for detail.

For a prime p and two positive integers m and s ($m \geq 2, s \geq 1$), a m-dimensional *Euclidean geometry* $\mathrm{EG}(m, p^s)$ over a Galois field $\mathrm{GF}(p^s)$ consists of *points*, *lines*, and *hyperplanes*. Any points in $\mathrm{EG}(m, p^s)$ are p^{ms} m-dimensional vectors over $\mathrm{GF}(p^s)$, and they form an m-dimensional vector space V over $\mathrm{GF}(p^s)$. For μ such that $0 \leq \mu \leq m$, since a μ-dimensional hyperplane (generally, called a *μ-flat*) is a μ-dimensional subspace of V and its cosets, any μ-flat contains $p^{\mu s}$ points. Points and lines correspond to 0-flats and 1-flats, respectively.

For a given $\mu < m$, let $\boldsymbol{a}_0, \boldsymbol{a}_1, \ldots, \boldsymbol{a}_\mu$ be $\mu + 1$ linear independent points in $\mathrm{EG}(m, p^s)$. Then using μ elements $\beta_1, \beta_2, \ldots, \beta_\mu$ of $\mathrm{GF}(p^s)$, $p^{\mu s}$ points given by

$$\boldsymbol{a}_0 + \beta_1 \boldsymbol{a}_1 + \beta_2 \boldsymbol{a}_2 + \cdots + \beta_\mu \boldsymbol{a}_\mu \tag{5}$$

forms a μ-flat.

Any pair of two μ-flats, (F_1, F_2), has at most one $(\mu-1)$-flat in common, which implies F_1 and F_2 have at most $p^{(\mu-1)s}$ points in common. In a Euclidean geometry $\mathrm{EG}(m, p^s)$, there are

$$f_{\mathrm{EG}}(\mu) = p^{(m-\mu)s} \prod_{i=1}^{\mu} \frac{p^{(m-i+1)s} - 1}{p^{(\mu-i+1)s} - 1} \tag{6}$$

μ-flats in total.

Denoting a m-dimensional *projective geometry* over a Galois field $\mathrm{GF}(p^s)$ by $\mathrm{PG}(m, p^s)$, $\mathrm{PG}(m, p^s)$ contains $(p^{(m+1)s} - 1)/(p^s - 1)$ points. In a projective geometry $\mathrm{PG}(m, p^s)$, there are

$$f_{\mathrm{PG}}(\mu) = \prod_{i=0}^{\mu} \frac{p^{(m-i+1)s} - 1}{p^{(\mu-i+1)s} - 1} \tag{7}$$

μ-flats in total. Any pair of two μ-flats, (F_1, F_2), has at most one $(\mu-1)$-flat in common, which implies F_1 and F_2 have at most $(p^{\mu s} - 1)/(p^s - 1)$ points in common.

For simplicity, we use the notation $\mathrm{FG}(m, p^s)$ to express either the Euclidean geometry $\mathrm{EG}(m, p^s)$ or the projective geometry $\mathrm{PG}(m, p^s)$. In a similar manner, $f_{\mathrm{FG}}(\mu)$ expresses either $f_{\mathrm{EG}}(\mu)$ or $f_{\mathrm{PG}}(\mu)$.

Letting $N_0 = f_{\mathrm{FG}}(0)$, suppose an $N_0 \times f_{\mathrm{FG}}(\mu)$ matrix $B_\mu = [b_{ij}]$. We allocate the rows and columns of B_μ to points and μ-flats in $\mathrm{FG}(m, p^s)$, respectively. An entry b_{ij}

[3]The coefficient vector can be obtained by calculating the inner product with orthogonal bases $\{\boldsymbol{u}_i\}$.

in a matrix B_μ takes $b_{ij} = 1$ if the points i is contained in the μ-flat j, or takes $b_{ij} = 0$ otherwise. This matrix B_μ is referred to as the *incident matrix of μ-flats over points* in $\mathrm{FG}(m, p^s)$.

As mentioned in Sect. III-B, Trappe et al. arrange each column vector of the incident matrix B_1 of 1-flats (lines) over points in $\mathrm{FG}(m, p^s)$ as a codeword of an AC code. They utilize the two properties: (i) any 1-flat in $\mathrm{FG}(m, p^s)$ has a constant number of points, (ii) any pair of two 1-flats has at most one point in common. Any AC code constructed from $\mathrm{EG}(m, p^s)$ becomes a $(p^s - 1)$-resilient AC code. By using $\mathrm{PG}(m, p^s)$ to construct an AC code, this code becomes a p^s-resilient AC code.

We mention parameters of the AC code $\mathcal{B}_1$. The code length of $\mathcal{B}_1$ is N_0, which equals to the number of points in $\mathrm{FG}(m, p^s)$. The number of codewords (the number of accommodated users) is $f_{\mathrm{FG}}(1)$, which expresses the number of 1-flats in $\mathrm{FG}(m, p^s)$. Thus the coding rate r_1 is given by $r_1 = \log_2 f_{\mathrm{FG}}(1)/N_0$. With the constant resilience and code length, as the number of codewords increases, the system can provide services to more users.

IV. Relaxation of Conditions on AC Codes

In this section, we relax the conditions of the BIBD-based AC codes [11], [12], which provides flexible construction of the AC codes[4].

A. Relaxation for General AC Codes Based on BIBD

For a real number v, we express $\lceil v \rceil$ as the minimum integer not less than v.

Lemma 2: Assume a binary matrix satisfies: (i) the Hamming weight of each column is at least k, and (ii) any pair of distinct two column vectors has at most t 1-entries in common. Then, the AC code obtained from this matrix is a $(\lceil k/t \rceil - 1)$-resilient AC code.

(*Proof*) The proof is an extension of that in [11] for the case $t \geq 1$.

We denote the position sets in which the j-th column vector has 1-entries by $\mathcal{A}_j$. Suppose a set of colluders $\mathcal{S}_c$ whose size satisfies $|\mathcal{S}_c| \leq \lceil k/t \rceil - 1$. As in Remark 1, if $\bigcap_{j \in \mathcal{S}_c} \overline{\mathcal{A}_j} \neq \bigcap_{i \in \mathcal{I}} \overline{\mathcal{A}_i}$ for arbitrary subset $\mathcal{I} \subseteq \Gamma$ whose size is less than or equal to $\lceil k/t \rceil - 1$, $\mathcal{S}_c$ is uniquely identified. Furthermore, from De Morgan's low, this is equivalent to

$$\bigcup_{j \in \mathcal{S}_c} \mathcal{A}_j \neq \bigcup_{i \in \mathcal{I}} \mathcal{A}_i, \qquad \forall \mathcal{I},\ s.t.\ |\mathcal{I}| \leq \lceil k/t \rceil - 1. \quad (8)$$

Thus it suffices to show eq. (8).

We suppose temporarily that a set $\mathcal{I} \neq \mathcal{S}_c$ with size $|\mathcal{I}| \leq \lceil k/t \rceil - 1$ satisfies $\bigcup_{j \in \mathcal{S}_c} \mathcal{A}_j = \bigcup_{i \in \mathcal{I}} \mathcal{A}_i$. From the assumption of the lemma, if $\mathcal{S}_c \cap \mathcal{I} = \emptyset$ for some $\mathcal{A}_{j^o}, j^o \in \mathcal{S}_c$ (if $\mathcal{S}_c \cap \mathcal{I} \neq \emptyset$ for some $\mathcal{A}_{j^o}, j^o \in \mathcal{S}_c \backslash (\mathcal{S}_c \cap \mathcal{I})$), any $\mathcal{A}_i, i \in \mathcal{I}$, has at most t elements in common with $\mathcal{A}_{j^o}$. Therefore it requires $|\mathcal{I}| \geq \lceil k/t \rceil$ to satisfy $\mathcal{A}_{j^o} \subseteq \bigcup_{i \in \mathcal{I}} \mathcal{A}_i$ from the assumption $|\mathcal{A}_{j^o}| \geq k$. Thus it contradicts the assumption $|\mathcal{I}| \leq \lceil k/t \rceil - 1$, and eq. (8) holds. □

[4]This extension is not limited to the AC codes using finite geometries but can apply to any AC codes proposed by Trappe et al.

The AC codes assumed in Lemma 2 are reduced to the AC codes of Trappe et al. if their codewords have a constant Hamming weight k and $t = 1$. Therefore this extension provides flexibility to construct ℓ-resilient AC codes.

B. Distortion Given by the AC Codes with Relaxed Condition

In this subsection, we mention distortion to the original content given by the AC codes with the relaxation of the conditions.

The distortion of the digital content for the users can be measured with $E[||\boldsymbol{y}_j - \boldsymbol{x}||^2]$, where $E[\cdot]$ and $||\cdot||^2$ denote the expectation by $\{\boldsymbol{b}_j\}$ and the square of norm, respectively. It follows from $\boldsymbol{y}_j = \boldsymbol{x} + \boldsymbol{w}_j$ and eq. (1) that the distortion to the contents can be calculated as

$$E\Big[||\boldsymbol{y}_j - \boldsymbol{x}||^2\Big] = \sum_{i=1}^{N} E\Big[||(2b_{ij} - 1)\boldsymbol{u}_i||^2\Big] = \sum_{i=1}^{N} ||\boldsymbol{u}_i||^2,$$

where the first equality is obtained by the linearity of the expectation. It can be seen that the distortion takes a constant value $\sum_{i=1}^{N} ||\boldsymbol{u}_i||^2$ regardless of the probability of symbols of $\{\boldsymbol{b}_j\}$.

Therefore the distortion to the original content given by the AC codes of Lemma 2 is equal to that of Trappe et al. even if the Hamming weight of the codewords becomes greater.

V. Improvement of AC Codes Using Finite Geometries

We propose code construction for increasing the coding rate of the conventional AC codes with keeping resilience.

A. AC Codes Based on Finite Geometries

In this subsection, we describe an explicit code construction in Lemma 2 by using finite geometries.

When we construct the ℓ-resilient AC codes of Trappe et al. by using finite geometries $\mathrm{FG}(m, p^s)$, relationship between points and lines (1-flats) in $\mathrm{FG}(m, p^s)$ is considered. In the new code construction, relationship between points and μ-flats ($\mu \geq 1$) in $\mathrm{FG}(m, p^s)$ will be utilized.

Definition 2: For $\mu \geq 1$, let B_μ be the incident matrix of μ-flats over points in a finite geometry $\mathrm{FG}(m, p^s)$, and we denote its j-th column vector by $\boldsymbol{b}_j$. Allocating $\boldsymbol{b}_j$ to the j-th user's fingerprint, the obtained code $\mathcal{B}_\mu = \{\boldsymbol{b}_j\}$ is called the μ-th order *FG-AC code*. In particular, the AC code $\mathcal{B}_\mu$ constructed from the Euclid geometry and the projective geometry are called the *μ-th order EG-AC code* and the *μ-th order PG-AC code*, respectively. □

We then have the following theorem.

Theorem 1: For some $\mathrm{EG}(m, p^s)$, the μ-th order EG-AC code $\mathcal{B}_\mu$ is a $(p^s - 1)$-resilient AC code. For some $\mathrm{PG}(m, p^s)$, the μ-th order PG-AC code $\mathcal{B}_\mu$ is a p^s-resilient AC code.

(*Proof*) As explained in Sect. III-B, any pair of two μ-flats, F_1 and F_2, in $\mathrm{FG}(m, p^s)$ has at most one $(\mu - 1)$-flat in common. Therefore it can be found that $k = p^{\mu s}, t = p^{(\mu-1)s}$ for EG-AC codes and $k = (p^{(\mu+1)s} - 1)/(p^s - 1), t = (p^{\mu s} - 1)/(p^s - 1)$ for PG-AC codes. From Lemma 2, the claim can be proven. □

From Theorem 1, it can be found that the resilience of a μ-th order FG-AC code $\mathcal{B}_\mu$ is independent of the order μ of the flats.

We here mention parameters of the μ-th order FG-AC codes. For a given $\mathrm{FG}(m, p^s)$, we can have $m-1$ μ-th order FG-AC codes $\mathcal{B}_\mu$ for $\mu = 1, 2, \ldots, m-1$ with the same resilience. From the properties of the incident matrix of μ-flats over points, the code lengths of these FG-AC codes $\mathcal{B}_\mu$ are equally N_0 and the numbers of codewords are $f_{\mathrm{FG}}(\mu)$. The coding rate, denoted by r_μ, is given by $r_\mu = \log_2 f_{\mathrm{FG}}(\mu)/N_0$. i.e., parameter of $\mathcal{B}_\mu$ depending on the order μ is only the number of codewords, which determines the best FG-AC code $\mathcal{B}_{\mu^*}$ with the maximal size for a given $\mathrm{FG}(m, p^s)$. Therefore we call such order μ^* the *maximal order* of the FG-AC codes for a given $\mathrm{FG}(m, p^s)$.

Thus we have the following theorem.

Theorem 2: The maximal order μ^* of the FG-AC codes for a given $\mathrm{FG}(m, p^s)$ satisfies:

- the case $m \leq 3$: $\mu^* = 1$;
- the case $m = 4$: $\mu^* = 2$ or $\mu^* = 1, 2$;
- the case $m > 4$: $\mu^* \geq 2$.

(*Proof*) It can be easily verified from the definition since the number of codewords of a μ-th order FG-AC code $\mathcal{B}_\mu$ is given by eq. (6) or eq. (7). If $m = 3$, the maximal order of the EG-AC codes satisfies $\mu^* = 2$ while that of the PG-AC codes is $\mu^* = 1, 2$. □

It follows from Theorem 2 that there always exists a better FG-AC code than AC codes $\mathcal{B}_1$ of Trappe et al. when $m > 3$.

B. Examples of FG-AC Codes with the Maximal Order

For a given $\mathrm{EG}(m, p^s)$, we show examples of EG-AC codes $\mathcal{B}_{\mu^*}$ with the maximal order in Table I. For $m > 3$, we display codes with the resilience $(p^s - 1)$ greater than one in the increasing order of their length. In the table, the columns of "$\log_2 f_{\mathrm{EG}}(1)$" and "$\log_2 f_{\mathrm{EG}}(\mu^*)$" express the logarithm of the number of codewords of $\mathcal{B}_1$ (the Trappe's AC code) and that of the EG-AC code $\mathcal{B}_{\mu^*}$ with the maximal order.

It follows from the property of the function $f_{\mathrm{EG}}(\mu)$ that the number of codewords of $\mathcal{B}_{\mu^*}$ becomes larger than that of $\mathcal{B}_1$ as the dimension m of $\mathrm{EG}(m, p^s)$ increases. In particular, the number of codewords of $\mathcal{B}_{\mu^*}$ is 2^2 times larger than that of $\mathcal{B}_{\mu^*}$ when $m \geq 5$, $2^{6.5}$ times larger for $\mathrm{EG}(7, 3)$, and 2^{10} times larger for $\mathrm{EG}(8, 3)$.

We also show examples of PG-AC codes $\mathcal{B}_{\mu^*}$ with the maximal order in Table II for a given $\mathrm{PG}(m, p^s)$. Although the PG-AC codes with the maximal order behave similar to the EG-AC codes, there exist two maximal orders when m is even.

TABLE I
EXAMPLES OF THE EG-AC CODES WITH THE MAXIMAL ORDER

(m, p^s)	N_0	μ^*	$\log_2 f_{\mathrm{EG}}(1)$	$\log_2 f_{\mathrm{EG}}(\mu^*)$
$(4, 3^1)$	81	2	10.08	10.19
$(5, 3^1)$	243	2	13.26	15.00
$(4, 2^2)$	256	2	12.41	12.48
$(4, 5^1)$	625	2	14.25	14.30
$(6, 3^1)$	729	3	16.43	19.80
$(5, 2^2)$	1024	2	16.41	18.50
$(7, 3^1)$	2187	3	19.60	26.16
$(5, 5^1)$	3125	2	18.90	21.28
$(4, 2^3)$	4096	2	18.19	18.21
$(6, 2^2)$	4096	3	20.41	24.52
$(8, 3^1)$	6561	4	22.77	32.52

TABLE II
EXAMPLES OF THE PG-AC CODES WITH THE MAXIMAL ORDER

(m, p^s)	N_0	μ^*	$\log_2 f_{\mathrm{PG}}(1)$	$\log_2 f_{\mathrm{PG}}(\mu^*)$
$(4, 3^1)$	121	1, 2	10.24	10.24
$(4, 2^2)$	341	1, 2	12.50	12.50
$(5, 3^1)$	364	2	13.43	15.05
$(4, 5^1)$	781	1, 2	14.31	14.31
$(6, 3^1)$	1093	2, 3	16.60	19.82
$(5, 2^2)$	1365	2	16.51	18.52
$(7, 3^1)$	3280	3	19.77	26.18
$(5, 5^1)$	3906	2	18.96	21.29
$(4, 2^3)$	4681	1, 2	18.21	18.21
$(6, 2^2)$	5461	2, 3	20.51	24.53
$(8, 3^1)$	9841	3, 4	22.94	32.52

VI. AC CODE BASED ON QUASI-CYCLIC LDPC MATRIX

Conventional $(k-1)$-resilient AC codes by Trappe et al. utilize the following property: (i) the Hamming weight of each column weight is k, and (ii) any pair of two codewords has at most one 1-entry in common. In other words, any code with this property is a $(k-1)$-resilient AC code. A regular low-density parity check (LDPC) matrix without cycles of length four [5] can be used for constructing a $(k-1)$-resilient AC code. In this section, we show how to improve such AC codes when we use LDPC matrices with quasi-cyclic (QC) structure.

A. Quasi-Cyclic LDPC Matrix over Galois Field

Let α be a primitive element over a Galois field $\mathrm{GF}(p^{ms})$ and we denote the zero element over this field by $0 = \alpha^{-\infty}$. Then any non-zero element can be expressed as α^i for $i = 0, 1, \ldots, p^{ms} - 2$. For any element α^i, let $\boldsymbol{z}_i = (z_{i,-\infty}, z_{i,0}, z_{i,1}, \ldots, z_{i,p^{ms}-2})$ be a p^{ms}-tuple over $\mathrm{GF}(2)$ such that it takes $z_{i,j} = 1$ if $i = j$, and $z_{i,j} = 0$ otherwise. The vector $\boldsymbol{z}_i$ is called the *location vector* of α^i. Arrange p^{ms} cyclic-shifted versions of the location vector $\boldsymbol{z}_i$ to form a $p^{ms} \times p^{ms}$ circulant matrix, where the first row is $\boldsymbol{z}_i$ itself and the j-th row is the right-shifted version of $\boldsymbol{z}_i$ by $j-1$ times. We denote this matrix of α^i by $\pi^i(I)$, where I corresponds to the $p^{ms} \times p^{ms}$ circulant matrix of $0 = \alpha^{-\infty}$ (i.e., the identity matrix) and π expresses the cyclic permutation.

For two integers $\gamma \geq 1$ and $\rho \geq 1$, A regular QC-LDPC matrix defined over a Galois field $\mathrm{GF}(p^{ms})$ is given by

$$M_0 = \begin{bmatrix} \pi^{a_{1,1}}(I) & \pi^{a_{1,2}}(I) & \cdots & \pi^{a_{1,\rho}}(I) \\ \pi^{a_{2,1}}(I) & \pi^{a_{2,2}}(I) & \cdots & \pi^{a_{2,\rho}}(I) \\ \vdots & \vdots & \ddots & \vdots \\ \pi^{a_{\gamma,1}}(I) & \pi^{a_{\gamma,2}}(I) & \cdots & \pi^{a_{\gamma,\rho}}(I) \end{bmatrix}, \quad (9)$$

where $\pi^{a_{i,j}}(I)$ for $i = 1, 2, \ldots, \gamma$ and $j = 1, 2, \ldots, \rho$ is a $p^{ms} \times p^{ms}$ circulant matrix of $\alpha^{a_{i,j}}$. Thus the size of M_0 is

$\gamma p^{ms} \times \rho p^{ms}$. If any pair of two columns of the matrix M_0 has at most one 1-entry in common, M_0 is called a *regular* (γ, ρ) *QC-LDPC matrix*.

The QC-LDPC matrices is used for constructing error-correcting codes [5] and there have been proposed many types of QC-LDPC matrices. QC-LDPC matrices based on the structure of the Reed-Solomon code [2], [6] or based on the structure of the Array codes [3], [14] are one of the examples assumed in this paper.

Another QC-LDPC matrix defined over a Galois field $\mathrm{GF}(p^{ms})$ can be considered. This matrix consists of two types of columns, namely the first type is of the Hamming weight γ and the other is of the Hamming weight $\gamma - 1$. Let O and I' be the $(p^{ms}-1) \times (p^{ms}-1)$ all zero matrix and the $(p^{ms}-1) \times (p^{ms}-1)$ identity matrix, respectively. We substitute each submatrix $\pi^{a_{i,j}}(I)$ with O if $a_{i,j} = -\infty$ and with $\pi^{a_{i,j}}(I')$ otherwise for $i = 1, 2, \ldots, \gamma, j = 1, 2, \ldots, \rho$. The resultant matrix is of the size $\gamma(p^{ms}-1) \times \rho(p^{ms}-1)$. We call this type of LDPC matrices *partially regular* $(\{\gamma, \gamma - 1\}, \rho)$ QC-LDPC matrices.

B. Improvement of AC Codes Based on Regular Quasi-Cyclic LDPC Matrix

Since a regular (γ, ρ) QC-LDPC matrix satisfies (i) each column weight is γ, (ii) any pair of two columns has one 1-entry in common, any AC code whose codewords are arranged from column vectors of M_0 is a $(\gamma-1)$-resilient AC code. We denote this AC code by $\mathcal{M}_0$ and call conventional AC codes based on QC-LDPC matrices.

We here propose a method for increasing the number of codewords while maintaining the code length, the resilience by using a similar technique in Sect. V. For a Euclidian geometry $\mathrm{EG}(m, p^s)$, let B_μ be a $p^{ms} \times f_{\mathrm{EG}}(\mu)$ incident matrix of the μ-flats over points. We substitute the (i, j)-th circulant matrix $\pi^{a_{i,j}}(I)$ $(i = 1, 2, \ldots, \gamma, j = 1, 2, \ldots, \rho)$ of M_0 with a matrix $\pi^{a_{i,j}}(B_\mu)$, which can be obtained to right-shift the matrix B_μ $a_{i,j}$ times. We denote the resultant $\gamma p^{ms} \times \rho f_{\mathrm{EG}}(\mu)$ matrix by M_μ. Let $\mathcal{M}_\mu$ be an AC code whose codewords are column vectors of M_μ, and we have the following theorem.

Theorem 3: The AC code $\mathcal{M}_\mu$ for a given $\mathrm{EG}(m, p^s)$ has (i) the code length γp^{ms}, (ii) the number of codewords $\rho f_{\mathrm{EG}}(\mu)$ and (iii) the resilience $\ell = \min\{\gamma - 1, p^s - 1\}$.

(*Proof*) Since both (i) the code length and (ii) the number of codewords are easily verified, we here mention (iii), the resilience.

We partition γp^{ms} rows of the matrix M_μ by p^{ms} rows into γ groups, and we call the ν-th group (the $(\nu p^{ms}+1)$-th row to $(\nu+1)p^{ms}$-th row) *the ν-th row section*. By expressing each column vector $\boldsymbol{m}_j \in \{0,1\}^{\gamma p^{ms}}$ of M_μ with γ vectors $\boldsymbol{m}_{j,\nu} \in \{0,1\}^{p^{ms}}$ as $\boldsymbol{m}_j^T = (\boldsymbol{m}_{j,1}^T, \boldsymbol{m}_{j,2}^T, \ldots, \boldsymbol{m}_{j,\gamma}^T)$, we have $w_H(\boldsymbol{m}_{j,\nu}) = p^{\mu s}$ for $\nu = 1, 2, \ldots, \gamma$.

(i) the case of $\gamma \le p^s$:

Consider a set of colluders, $\mathcal{S}_c$ of size $|\mathcal{S}_c| \le \gamma - 1$. Denoting the support of $\boldsymbol{m}_{j,\nu}$ $(\nu = 1, 2, \ldots, \gamma)$ by $\mathcal{A}_{j,\nu}$, we suppose there exists a set of users, $\mathcal{I}$ of size $|\mathcal{I}| \le \gamma - 1$ satisfying

$$\bigcup_{j \in \mathcal{S}_c} \bigcup_{\nu \in [1,\gamma]} \mathcal{A}_{j,\nu} = \bigcup_{i \in \mathcal{I}} \bigcup_{\nu \in [1,\gamma]} \mathcal{A}_{i,\nu}. \tag{10}$$

In this case, it requires $\mathcal{A}_{j^*,\nu} \subseteq \bigcup_{i \in \mathcal{I}} \bigcup_{\nu \in [1,\gamma]} \mathcal{A}_{i,\nu}$ for column vectors $\boldsymbol{m}_{j^*}$, $\forall j^* \in \mathcal{S}_c \backslash (\mathcal{S}_c \cap \mathcal{I})$, of M_μ.

The set $\mathcal{A}_{j^*}$ and each $\mathcal{A}_i, i \in \mathcal{I}$, have at most $p^{\mu s} + (\gamma - 1)p^{(\mu-1)s}$ elements in common. If $\mathcal{A}_{j^*}$ and $\mathcal{A}_i$ have one μ-flat in common at some ν-th row section, they have at most one $(\mu - 1)$-flats in common at other $(\gamma - 1)$ row sections. Since $|\mathcal{I}| \le \gamma - 1$, there exists at least one row section (say, ν^*-th row section) in which only $(\mu - 1)$-flats are shared in common. In this row section, we have $\mathcal{A}_{j,\nu^*} \nsubseteq \bigcup_{i \in \mathcal{I}} \bigcup_{\nu \in [1,\gamma]} \mathcal{A}_{i,\nu}$. Thus it contradicts to eq. (10), and the code is a $(\gamma - 1)$-resilient AC code when $\gamma \le p^s$.

(ii) The case of $\gamma > p^s$:

Taking a similar steps to the case (i), it can be shown that the code should be a $(p^s - 1)$-resilient AC code. Thus the theorem holds. □

It follows from Theorems 3 that we can improve AC codes which are constructed based on a QC-LDPC matrix over a Galois field by using Euclidean geometry $\mathrm{EG}(m, p^s)$. In particular, if $\gamma - 1 < p^s$ and we utilize the relationship between the μ-flats and points over $\mathrm{EG}(m, p^s)$, the AC codes $\mathcal{M}_\mu$ are always more efficient than the conventional code $\mathcal{M}_0$. About the obtained AC codes in this section, we can assert a similar effectiveness mentioned in Sect. V-B.

We state some relationship between the conventional AC code $\mathcal{M}_0$ and the AC code $\mathcal{M}_\mu$ for a given $\mathrm{EG}(m, p^s)$. If we denote the incident matrix of 0-flats (namely, points) over points in $\mathrm{EG}(m, p^s)$ by B_0, there is a relationship $I = B_0$. Therefore, it follows from substituting $\mu = 0$ in the matrix M_μ that we can obtain the matrix M_0. This fact implies that the conventional AC code $\mathcal{M}_0$ based on QC-LDPC matrix is an instance of the AC codes $\mathcal{M}_\mu$ when $\mu = 0$. Thus a class of the AC codes $\mathcal{M}_\mu$ includes the AC code $\mathcal{M}_0$ as a special case.

As illustration of the obtained EG-AC codes based on QC-LDPC matrices, we show some examples by assuming QC-LDPC matrix based on Reed-Solomon code [2] in Table III. For $\mathrm{GF}(p^{ms})$, (γ, ρ) QC-LDPC matrices with $1 \le \gamma \le p^{ms} - 1$ and $1 \le \rho \le p^{ms}$ can be constructed. We choose some γ and fix the value of ρ as $\rho = p^{ms}$ to construct AC codes as large as possible. We show the logarithm of the code size for M_0 in the column of "$\log_2 \rho f_{\mathrm{EG}}(0)$" and for M_{μ^*} in that of "$\log_2 \rho f_{\mathrm{EG}}(\mu^*)$".

Using the projective geometry $\mathrm{PG}(m, p^s)$, we can obtain a similar result. We only show the result without the proof.

Theorem 4: Assume that we allocate each column vector of the incident matrix of μ-flats over points in $\mathrm{PG}(m, p^s)$ to a user's codeword. This AC code has (i) the code length $\gamma(p^{(m+1)s} - 1)/(p^s - 1)$, (ii) the number of codewords $\rho f_{\mathrm{PG}}(\mu)$, and (iii) the resilience

$$\ell = \begin{cases} \gamma - 1, & \text{if } \gamma - 1 < (p^{(\mu+1)s} - 1)/(p^{\mu s} - 1), \\ p^s, & \text{otherwies.} \end{cases}$$

TABLE III
EXAMPLES OF EG-AC CODES BASED ON QC-LDPC MATRICES

γ	ρ	(m, p^s)	N	$\log_2 \rho f_{\mathrm{EG}}(0)$	$\log_2 \rho f_{\mathrm{EG}}(\mu^*)$
3	26	$(3, 3^1)$	81	9.46	11.57
3	80	$(4, 3^1)$	243	12.66	18.10
3	242	$(5, 3^1)$	729	15.84	24.50
4	63	$(3, 4^1)$	256	11.98	14.37
4	255	$(4, 4^1)$	1024	15.99	22.47
4	1023	$(5, 4^1)$	4096	20.00	30.50
5	124	$(3, 5^1)$	625	13.92	16.55
5	624	$(4, 5^1)$	3125	18.57	23.58

If $\gamma - 1 < (p^{(\mu+1)s} - 1)/(p^{\mu s} - 1)$, this code is also a $(\gamma - 1)$-resilient AC code. □

C. AC Codes Based on Partially Regular Quasi-Cyclic LDPC Matrix

Since a partially regular $(\{\gamma, \gamma - 1\}, \rho)$ QC-LDPC matrix satisfies (i) each column weight is at least $\gamma - 1$, (ii) any pair of two columns has at most one 1-entry in common, any AC code obtained from this matrix is a $(\gamma - 2)$-resilient AC code from Lemma 2. However we can show its resilience is greater than $\gamma - 2$.

Lemma 3: AC codes constructed from a partially regular $(\{\gamma, \gamma - 1\}, \rho)$ QC-LDPC matrix are $(\gamma - 1)$-resilient AC codes.

(*Proof*) Noting that any pair of two columns with the Hamming weight $\gamma - 1$ has no 1-entries in common, we can prove the lemma. □

As in Sect. VI-B, we can use an improved method for increasing the number of codewords, while keeping the code length and the resilience. For a Euclidean geometry $\mathrm{EG}(m, p^s)$, consider extracting the zero point from this geometry and the all lines including the zero point are reduced to lines without this point. Let B'_μ be a $(p^{ms} - 1) \times f_{\mathrm{EG}}(\mu)$ matrix whose rows corresponds to non-zero points in $\mathrm{EG}(m, p^s)$ and columns corresponds to the obtained lines.

Theorem 5: Assume that we substitute any submatrix $\pi^{a_{i,j}}(I')$ with $\pi^{a_{i,j}}(B'_\mu)$, then the AC code from this matrix satisfies: (i) the code length $\gamma(p^{ms} - 1)$, (ii) the number of codewords $\rho f_{\mathrm{EG}}(\mu)$ and (iii) the resilience $\ell = \{\gamma - 1, p^s - 1\}$. □

If $\gamma - 1 < p^s$, the AC code is also a $(\gamma - 1)$-resilient AC code.

VII. CONCLUSION AND FUTURE IMPROVEMENTS

In this paper, for a class of AC codes proposed by Trappe et al., novel methods for increasing their coding rate were proposed based on finite geometries, while their resilience is maintained. We showed examples of the AC codes with the maximal number of codewords for a given finite geometry. The obtained AC code can have the greater number of codewords than the conventional AC code by Trappe et al. as the dimension m of the finite geometry $\mathrm{FG}(m, p^2)$ increases. Taking a similar approach to this construction method, other methods for constructing efficient AC codes based on quasi-cyclic LDPC matrices were also proposed. In this method, although all the case does not necessarily guarantee the same resilience, conditions on parameters which provides the same resilience were derived. Consequently, we can construct a fingerprinting system which can provide service of distributing a digital content for more users, while keeping both the resilience and the distortion to the original digital contents.

Unfortunately, the codes obtained by the proposed method have comparatively large code lengths, which implies the distortion to the original content by these codes may be large. An effective shortening method of the code while the resilience is maintained should be devised.

In this paper, the resilience is guaranteed by assuming no noise sequence. The performance of the AC codes should be analyzed when there occurs a noise sequence.

REFERENCES

[1] D. Boneh and J. Shaw, "Collusion-secure fingerprinting for digital data," *IEEE Trans. Inform. Theory*, vol. 44, pp. 1897–1905, Sep. 1998.

[2] I. Djurdjevic, J. Xu, K. Abdel-Ghaffar, and S. Lin, "A class of low-density parity check codes constructed based on Reed-Solomon codes with two information symbols," *IEEE Commun. Letters*, vol. 7, no. 7, pp. 317–319, June 2003.

[3] H. Fujita and K. Sakaniwa, "An efficient encoding method for LDPC codes based on cyclic shift," *Proc. of 2004 IEEE Int. Symp. on Inform. Theory (ISIT2004)*, p. 275, Chicago, USA, June–July 2004.

[4] S. He and M. Wu, "Joint coding and embedding techniques for multimedia fingerprinting," *IEEE Trans. on Information Forensics and Security*, vol. 1, pp. 231–247, June 2006.

[5] S. Lin and D.J. Costello Jr., *Error Control Coding: Fundamentals and Applications, 2nd ed.*, Upper Saddle River, NJ: Prentice-Hall, 2004.

[6] T. Mittelholzer, "Efficient encoding and minimum distance bounds of Reed-Solomon-type array codes," *Proc. of 2002 IEEE Int. Symp. on Inform. Theory (ISIT2002)*, p. 282, Lausanne, Switzerland, June–July 2003.

[7] C. Podilchuk and W. Zeng, "Image adaptive watermarking using visual models," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 525–540, May 1998.

[8] R. Safavi-Naini and Y. Wang, "New results on frame-proof codes and traceability schemes. ," *IEEE Trans. Inform. Theory*, vol. 47, no. 7, pp. 3029–3033, Nov. 2001.

[9] J.N. Staddon, D.R. Stinson, and R. Wei, " Combinatorial properties of frameproof and traceability codes," *IEEE Trans. Inform. Theory*, vol. 47, no. 3, pp. 1042–1049, Mar. 2001.

[10] H. Tang, J. Xu, S. Lin, and K.A.S. Abdel-Ghaffar, "Codes on finite geometries," *IEEE Trans. Inform. Theory*, vol. 51, pp. 572–596, Feb. 2005.

[11] W. Trappe, M. Wu, Z.J. Wang, and K.J.R. Liu, "Anti-collusion fingerprinting for multimedia," *IEEE Trans. Signal Processing*, vol. 51, pp. 1069–1087, Apr. 2003.

[12] M. Wu, W. Trappe, Z.J. Wang, and K.J.R. Liu "Collusion-resistant fingerprinting for multimedia," *IEEE Signal Processing Magazine*, vol. 21, pp. 15–27, Mar. 2004.

[13] H. Yagi, T. Matsushima, and S. Hirasawa, "New traceability codes against a generalized collusion attack for digital fingerprinting," *Proc. of 2006 Int. Workshop on Information Security Applications (WISA2006)*, pp.569–584, Jeju Island, Korea, Aug. 2006.

[14] K. Yang and T. Helleseth, "On the minimum distance of array codes as LDPC codes," *IEEE Trans. on Inf. Theory*, vol. 49, no. 12, pp. 3268–3271, Dec. 2003.

[15] J. Yang, P. Liu, and G.Z. Tan, "The digital fingerprint coding based on LDPC," *Proc. of 2004 7th Int. Conf. on Signal Processing (ICSP2004)*, pp. 2600–2603, Beijing, China, Aug.–Sept. 2004.