2007 Hawaii and SITA Joint Conference on Information Theory, HISC2007
Hawaii, USA, May 29–31, 2007

Shortening Methods of Collusion-Secure Codes for Digital Fingerprinting

Hideki YAGI*, Toshiyasu MATSUSHIMA† and Shigeichi HIRASAWA†

* Media Network Center
Waseda University
1-104 Totsuka-Machi, Shinjuku-ku,
Tokyo, 169-8050 Japan
E-mail: yagi@hirasa.mgmt.waseda.ac.jp

† School of Science and Engineering
Waseda University
3-4-1 Ohkubo, Shinjuku-ku,
Tokyo, 169-8555 Japan

Abstract

In this paper, code construction for digital fingerprinting is considered. Digital fingerprinting is a copyright protection technique of digital contents. For digital fingerprinting, measures against collusion attacks, where several fingerprinted copies of the same content are mixed to disturb their fingerprints, should be taken. In this paper, we propose shortening methods of collusion-secure fingerprinting codes based on finite geometries (FGs). These methods reduce the code lengths but increase the coding rates of conventional collusion-secure codes with keeping the codes' resilience. Due to the new fingerprinting codes, the system can deal with a larger number of users to distribute a digital content.

1. Introduction

In recent years, with the high advances of digital technologies, a large amount of digital contents can be processed by computers. Protecting the copyrights 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**, in which more than one illicit user colludes to take illegal actions to the distributed contents. One of the well-known collusion attacks are the **interleaving attack** [1, 3, 6] and the **averaging attack** [3, 8, 9, 10, 11, 12]. W. Trappe et al. have devised collusion-secure codes against the averaging attack by using incident matrices of block designs [8], which is equivalent to regular low-density parity check (LDPC) matrices without cycles of length four [4]. The collusion-secure codes devised by Trappe et al. are called **anti-collusion** (AC) codes [8, 9]. Subsequently, Yagi et al. have proposed a method for increasing coding rates of Trappe's AC codes based on finite geometries [10, 12]. Although these codes can guarantee to capture colluders if the number of colluders is smaller than a designed value, the code length becomes larger, which causes degradation of the original content.

In this study we propose shortening methods of AC codes devised in [8] and [10, 12] while their number of codewords and security are maintained. Consequently, we can realize a fingerprinting system which gives smaller distortion to distributed contents.

2. Fingerprinting Model

2.1. Digital Fingerprinting

When distributing a digital content to users, a codeword corresponding to each user is embedded into the original content by a watermarking technique. The codeword arranged 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 $\boldsymbol{b}_j = (b_{j1}, b_{j2}, \ldots, b_{jN})^{\mathrm{T}} \in \{0,1\}^N$, where T denotes the transposition. The fingerprint watermark $\boldsymbol{w}_j$ is created by using N orthogonal bases $\{\boldsymbol{u}_i \in \mathcal{R}^M | \, i = 1, 2, \ldots, N\}$ with the unit 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 $\boldsymbol{x} \in \mathcal{R}^M$, the distributed content to a user $j \in \Gamma$ is[1] $\boldsymbol{y}_j = \boldsymbol{x} + \boldsymbol{w}_j \in \mathcal{R}^M$.

This work is supported by Waseda University Grant for Special Research Project No. 2006B-293 and the Telecommunications Advancement Foundation (TAF).

[1] More precisely, each $\boldsymbol{w}_j$ is multiplied by some value called Just-Difference Noticeable (JDN) coefficient [5], before it is added to the host signal.

Since the fingerprint is embedded with a watermark technique, any user 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.

2.2. 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

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

The detector of the colluders estimates the set of colluders $\mathcal{S}_c$ from the attacked host signal $\boldsymbol{y} \in \mathcal{R}^M$. This attack is called the **averaging attack**, which is one of well-known collusion attacks[2] [3, 8, 9, 11, 12].

3. Anti-Collusion Codes against Averaging Attack

3.1. General Class of AC Codes

Trappe et al. [8] and Yagi et al. [10, 12] have devised anti-collusion (AC) codes. 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 integer ℓ, the code which can reveal all the colluders of $\mathcal{S}_c$ is referred to as an ℓ-**resilient AC code**. The parameter ℓ is called the **resilience** of the AC codes. □

Consider a binary matrix B whose codewords have a Hamming weight greater than or equal to k and whose two distinct codewords have at most t "1-components" for some integer $t \geq 1$. Let $\lceil v \rceil$ for a real number v express the minimum integer not less than v.

Lemma 1 ([10, 12]) Assume a binary matrix to satisfy (i) the Hamming weight of each column is at least k, and (ii) any pair of distinct two columns has at most t 1-components in common. Then, the AC code whose codewords are all column vectors of this matrix is a $(\lceil k/t \rceil - 1)$-resilient AC code. i.e., if $|\mathcal{S}_c| \leq \lceil k/t \rceil - 1$, any set of colluders $\mathcal{S}_c$ can be uniquely detected. □

If $t = 1$, the matrix B is called an LDPC matrix without cycles of length four [4], and the ℓ-resilient AC codes in [8] are contained in this class.

We here consider a mechanism for detecting a set of colluders by using an ℓ-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 the 0-component. Then an ℓ-resilient AC code in [8, 10, 12] uniquely identifies the set $\mathcal{Q}(\mathcal{S}_c)$ for any $\mathcal{S}_c$ with $|\mathcal{S}_c| \leq \ell$. □

From Remark 1, since an ℓ-resilient AC code uniquely identifies $\mathcal{Q}(\mathcal{S}_c)$ for any $\mathcal{S}_c$ of the size less than or equal to ℓ, the code reveals the set of colluders $\mathcal{S}_c$. For a detailed procedure of decoding algorithm, refer to [8].

3.2. Special Class of AC Codes Based on Finite Geometries

We review a subclass of ℓ-resilient AC codes based-on finite geometry [10, 12].

For a prime p and two positive integers m and s ($m \geq 2, s \geq 1$), the 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 vector 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 over $\mathrm{EG}(m, p^s)$. Then using μ elements $\beta_1, \beta_2, \ldots, \beta_\mu$ of $\mathrm{GF}(p^s)$, $p^{\mu s}$ points given by

$$\boldsymbol{a}_0 + \beta_1\boldsymbol{a}_1 + \beta_2\boldsymbol{a}_2 + \cdots + \beta_\mu\boldsymbol{a}_\mu \quad (3)$$

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

$$f_{\mathrm{EG}}(\mu) = p^{(m-\mu)s}\prod_{i=1}^{\mu}\frac{p^{(m-i+1)s} - 1}{p^{(\mu-i+1)s} - 1} \quad (4)$$

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

$$f_{\mathrm{PG}}(\mu) = \prod_{i=0}^{\mu}\frac{p^{(m-i+1)s} - 1}{p^{(\mu-i+1)s} - 1} \quad (5)$$

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 $N_0 = f_{\rm FG}(0)$, suppose an $N_0 \times f_{\rm FG}(\mu)$ matrix $B_\mu = (b_{ij})$. An element b_{ij} 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 FG(m, p^s).

For $\mu \geq 1$, let B_μ be the incident matrix of μ-flats over points in a finite geometry 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.

Lemma 2 For some EG(m, p^s), the μ-th order EG-AC code $\mathcal{B}_\mu$ is a $(p^s - 1)$-resilient AC code. For some PG(m, p^s), the μ-th order PG-AC code $\mathcal{B}_\mu$ is a p^s-resilient AC code. □

3.3. Special Class of AC Codes Based on Quasi-Cyclic LDPC Matrices

A quasi-cyclic low-density parity check (QC-LDPC) matrix without cycles of length four [4] can be used for constructing other types of ℓ-resilient AC codes [12].

Let α be a primitive element over a Galois field 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 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 GF(p^{ms}) is given by

$$M_0 = \begin{bmatrix} \pi^{a_{1,1}}(I) & \pi^{a_{1,2}}(I) & \cdots & \pi^{a_{1,\rho}}(I) \\ \pi^{a_{2,1}}(I) & \pi^{a_{2,2}}(I) & \cdots & \pi^{a_{2,\rho}}(I) \\ \vdots & \vdots & \ddots & \vdots \\ \pi^{a_{\gamma,1}}(I) & \pi^{a_{\gamma,2}}(I) & \cdots & \pi^{a_{\gamma,\rho}}(I) \end{bmatrix}, \quad (6)$$

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-component in common, M_0 is called a **regular (γ, ρ) QC-LDPC matrix**. The QC-LDPC matrices are used for constructing error-correcting codes [4] and there have been proposed many types of QC-LDPC matrices.

For a Euclidean geometry EG(m, p^s), let B_μ be a $p^{ms} \times f_{\rm EG}(\mu)$ 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 $\gamma p^{ms} \times \rho f_{\rm EG}(\mu)$ 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 lemma.

Lemma 3 The AC code $\mathcal{M}_\mu$ for a given EG(m, p^s) has (i) the code length γp^{ms}, (ii) the number of codewords $\rho f_{\rm EG}(\mu)$ and (iii) the resilience $\ell = \min\{\gamma - 1, p^s - 1\}$ for $\gamma \geq 2$ and $\ell = p^s - 1$ for $\gamma = 1$. □

When we use a projective geometry PG(m, p^s) to construct an ℓ-resilient AC code, we have a similar result.

Lemma 4 Assume that we allocate each column vector of the incident matrix of μ-flats over points in 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 $\rho f_{\rm PG}(\mu)$, and (iii) the resilience

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

for $\gamma \geq 2$ and $\ell = p^s$ for $\gamma = 1$. □

When $\gamma = 1$ and $\rho = 1$, the code $\mathcal{M}_\mu$ is equivalent to the μ-th order FG-AC code $\mathcal{B}_\mu$ for a given FG(m, p^s). In general, as the dimension m of a geometry increases, the effectiveness of an AC code becomes higher, however in this case the code length grows exponentially. Since a large code length leads to a significant distortion to the original digital content, the code length should be remained short.

4. Shortening Methods for AC Codes

4.1. Basic Idea

The length of ℓ-resilient AC codes in [8], [12] might become larger. In this section, we propose shortening methods for these AC codes, while their resilience and the number of codewords are maintained.

We will first illustrate the basic idea of proposed shortening methods. Consider a code matrix given by

$$B = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix}. \quad (7)$$

From Lemma 1, since the each column has Hamming weight three and $t = 1$, this matrix gives a 2-resilient AC code.

We can choose any one row, say the bottom row. Even though the chosen row is removed from the matrix, it can be easily verified that the remaining matrix still gives a 2-resilient AC code. This matrix has two types of columns; the columns with the Hamming weight three and those with the Hamming weight two. Moreover, any pair of two column vectors with the Hamming weight two has no 1-components in common. If we further remove one additional row, the remaining matrix is unable to give a 2-resilient AC code. Thus removing one bit of any ℓ-resilient code AC with a constant Hamming weight and $t = 1$ keeps its resilience while reducing the code length. As for cases with general $t \geq 1$, we have the following result:

Proposition 1 Assume that a binary matrix $\tilde{B}$ has column vectors with the Hamming weight at most k and for each column vector with the Hamming weight $k - \delta$ with $\delta = 0, 1, \ldots, t$, other column vectors with the Hamming weight at most $k-\delta$ have at most $(t-\delta)$ 1-components in common. Then the matrix $\tilde{B}$ gives a $(\lceil k/t \rceil - 1)$-resilient AC code.

(**Proof**) We denote the position sets in which the j-th column vector has 1-components 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 the De Morgan's low, this is equivalent to

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

Thus it suffices to show that eq. (8) holds.

We suppose temporarily that a set $\mathcal{I} \neq \mathcal{S}_c$ with size $\lceil k/t \rceil - 1$ satisfies $\bigcup_{j\in\mathcal{S}_c} \mathcal{A}_j = \bigcup_{i\in\mathcal{I}} \mathcal{A}_i$.

(i) **The case there exists some $j^o \in \mathcal{S}_c \setminus (\mathcal{S}_c \cap \mathcal{I})$ with $|\mathcal{A}_j| = k$:**
From the assumption of the proposition, for some $\mathcal{A}_{j^o}, j^o \in \mathcal{S}_c \setminus (\mathcal{S}_c \cap \mathcal{I})$, any $\mathcal{A}_i, i \in \mathcal{I}$, has at most t elements in common with $\mathcal{A}_{j^o}$. Therefore it requires $|\mathcal{I}| > \lceil k/t \rceil - 1$ to satisfy $\bigcup_{j\in\mathcal{S}_c} \mathcal{A}_j = \bigcup_{i\in\mathcal{I}} \mathcal{A}_i$ from the assumption $|\mathcal{A}_{j^o}| \geq k$. Otherwise, $\mathcal{A}_{j^o} \not\subseteq \bigcup_{i\in\mathcal{I}} \mathcal{A}_i$. Thus it contradicts the assumption $|\mathcal{I}| \leq \lceil k/t \rceil - 1$, and eq. (8) holds.

(ii) **The case there exists some $i^o \in \mathcal{I} \setminus (\mathcal{S}_c \cap \mathcal{I})$ with $|\mathcal{A}_j| = k$:**
A similar argument to the case (i) holds for this case and eq. (8) holds.

(iii) **The case there does not exist $j \in \mathcal{S}_c \cup \mathcal{I} \setminus (\mathcal{S}_c \cap \mathcal{I})$ with $|\mathcal{A}_j| = k$:**
We denote the maximum Hamming weight of codewords in $\mathcal{S}_c \setminus (\mathcal{S}_c \cap \mathcal{I})$ by $k - \delta^*$ for some $\delta^* = 1, 2, \ldots, t-1$. From the assumption of the proposition, any $\mathcal{A}_{j^*}$ with $|\mathcal{A}_{j^*}| = k - \delta^*$ and $\mathcal{A}_i, i \in \mathcal{I} \setminus (\mathcal{S}_c \cap \mathcal{I})$, have at most $t - \delta^*$ elements in common. Therefore it requires $|\mathcal{I}| > \lceil (k-\delta^*)/(t-\delta^*) \rceil - 1$ to satisfy $\bigcup_{j\in\mathcal{S}_c} \mathcal{A}_j = \bigcup_{i\in\mathcal{I}} \mathcal{A}_i$ from the assumption $|\mathcal{A}_{j^*}| = k - \delta^*$. Otherwise, $\mathcal{A}_{j^*} \not\subseteq \bigcup_{i\in\mathcal{I}} \mathcal{A}_i$. Note that $\lceil (k-\delta)/(t-\delta) \rceil > \lceil k/t \rceil$ for $\delta = 1, 2, \ldots, t-1$ since

$$\frac{k-\delta}{t-\delta} - \frac{k}{t} = \frac{\delta(k-t)}{t(t-\delta)} > 0. \quad (9)$$

Thus it contradicts the assumption $|\mathcal{I}| \leq \lceil k/t \rceil - 1$, and eq. (8) holds. □

We may obtain a code matrix $\tilde{B}$ which satisfies the assumption of Proposition 1 by removing t rows from a matrix B with the Hamming weight k. Therefore we require a structured method for constructing such shortened AC codes.

4.2. Structured Shortening Method of AC Codes Based on Finite Geometries

We can shorten the μ-th order FG-AC codes explained in Sect. 3.2.

Consider an incident matrix B_μ of μ-flats over points in a finite geometry $\mathrm{FG}(m, p^s)$. Choose a $(\mu-1)$-flat from $\mathrm{FG}(m, p^s)$ and eliminate rows which correspond to all points contained in this $(\mu-1)$-flat from B_μ. We denote the resultant matrix by $\tilde{B}_{\mu,\mu-1}$. We denote an AC code whose codewords are column vectors of the matrix $\tilde{B}_{\mu,\mu-1}$ by $\tilde{\mathcal{B}}_{\mu,\mu-1}$.

Theorem 1 The code $\tilde{\mathcal{B}}_{\mu,\mu-1}$ for a given $\mathrm{EG}(m, p^s)$ is a (p^s-1)-resilient AC code with (i) the code length $\tilde{N} = p^{ms} - p^{(\mu-1)s}$, (ii) the number of codewords $f_{\mathrm{EG}}(\mu)$, and (iii) the resilience $(p^s - 1)$.

(**Proof**) As explained in Sect. 3.2, any pair of two μ-flats (say $\mathcal{F}_1$ and $\mathcal{F}_2$) in an Euclidean geometry $\mathrm{EG}(m, p^s)$ has a hyperplane with a smaller dimension in common. We denote a flat of the maximum dimension contained in $\mathcal{F}_1$ and $\mathcal{F}_2$ by $\mathcal{F}'$, then no other points are contained in both of them.

After a $(\mu-1)$-flat $\mathcal{F}'$ is eliminated from this geometry, points contained in a μ-flat are changed depending on how many points this μ-flat and $\mathcal{F}'$ have in common. If this μ-flat and $\mathcal{F}'$ have no points in common, the number of points in this μ-flat is unchanged. If a μ-flat and $\mathcal{F}'$ have a ν-flat in common where $\nu = 0, 1, \ldots, \mu-1$, the number of points in this μ-flat is changed from $p^{\mu s}$ to $p^{\mu s} - p^{\nu s}$. We call this flat the (μ, ν)-flat over an Euclidean geometry.

From the properties of Euclidean geometries, a (μ, ν)-flat and a (μ, ν')-flat for $\nu' = \nu, \nu+1, \ldots, \mu-1$ have at most $p^{(\mu-1)s} - p^{\nu s}$ points in common. Therefore the code matrix $\tilde{B}_{\mu,\mu-1}$ satisfies the condition of Proposition 1 with $k = p^{\mu s}$, $t = p^{(\mu-1)s}$, and $\delta = 0, 1, p^s, \ldots, p^{(\mu-1)s}$. Thus the theorem is proven. □

Theorem 1 indicates that we can shorten any μ-th order EG-AC code by $p^{(\mu-1)s}$ bits while keeping the number of codewords and the resilience.

Similar to the case using Euclidean geometries $\mathrm{EG}(m, p^s)$, we can shorten the μ-th order PG-AC codes.

Theorem 2 The code $\tilde{\mathcal{B}}_{\mu,\mu-1}$ for a given $\mathrm{PG}(m, p^s)$ is an AC code with (i) the code length

$$\tilde{N} = \frac{p^{(m+1)s} - p^{\mu s}}{p^s - 1}, \tag{10}$$

(ii) the number of codewords $f_{\mathrm{PG}}(\mu)$, and (iii) the resilience p^s.

(**Proof**) Similar to an Euclidean geometry, any pair of two μ-flats (say $\mathcal{F}_1$ and $\mathcal{F}_2$) in a projective geometry $\mathrm{PG}(m, p^s)$ has a hyperplane with a smaller dimension in common. We denote a flat of the maximum dimension contained in $\mathcal{F}_1$ and $\mathcal{F}_2$ by $\mathcal{F}'$.

After a $(\mu-1)$-flat $\mathcal{F}'$ is eliminated from this geometry, points contained in a μ-flat are changed depending on how many points this μ-flat and $\mathcal{F}'$ have in common. If a μ-flat and $\mathcal{F}'$ have no points in common, the number of points in this μ-flat is unchanged. If a μ-flat and $\mathcal{F}'$ have a ν-flat in common where $\nu = 0, 1, \ldots, \mu-1$, the number of points in this μ-flat is changed from $(p^{(\mu+1)s} - 1)/(p^s - 1)$ to

$$\frac{p^{(\mu+1)s} - 1}{p^s - 1} - \frac{p^{(\nu+1)s} - 1}{p^s - 1} = \frac{p^{(\mu+1)s} - p^{(\nu+1)s}}{p^s - 1}. \tag{11}$$

We call this flat the (μ, ν)-flat over a projective geometry.

From the properties of projective geometries, a (μ, ν)-flat and a (μ, ν')-flat for $\nu' = \nu, \nu+1, \ldots, \mu-1$ have at most $(p^{\mu s} - p^{\nu s})/(p^s - 1)$ points in common. Therefore the code matrix $\tilde{B}_{\mu,\mu-1}$ satisfies the condition of Proposition 1 with $k = (p^{(\mu+1)s} - 1)/(p^s - 1)$, $t = (p^{\mu s} - 1)/(p^s - 1)$, and $\delta = 0, 1, (p^{2s} - 1)/(p^s - 1), \ldots, (p^{\mu s} - 1)/(p^s - 1)$. Thus the theorem is proven. □

4.3. Structured Shortening Method of AC Codes Based on QC-LDPC Matrices

We can shorten AC codes based on QC-LDPC matrices in Sect. 3.3.

Consider a code matrix M_μ which is obtained from a (γ, ρ) QC-LDPC matrix and the incident matrix of μ-flats over points in a finite geometry $\mathrm{FG}(m, p^s)$. For $i = 1, 2, \ldots, \gamma$ and $j = 1, 2, \ldots, \rho$, the (i, j)-th sub-block of the matrix M_μ is expressed as $\pi^{a_{i,j}}(B_\mu)$. We denote a matrix obtained by eliminating rows which correspond to points contained in a $(\mu-1)$-flat from $\pi^{a_{i,j}}(B_\mu)$ by $\pi^{a_{i,j}}(\tilde{B}_{\mu,\mu-1})$ for $i = 1, 2, \ldots, \gamma$ and $j = 1, 2, \ldots, \rho$. We denote the over-all code matrix by $\tilde{M}_{\mu,\mu-1}$ and the AC code whose codewords are all column vectors of the matrix $\tilde{M}_{\mu,\mu-1}$ by $\tilde{\mathcal{M}}_{\mu,\mu-1}$. We then obtain the following result.

Table 1: Examples of Original and Shortened EG-AC Codes Based on (γ, ρ) QC-LDPC Matrices

γ	ρ	(m, p^s)	μ	N	$\tilde{N}$	$\log_2 \rho f_{\mathrm{EG}}(\mu)$
3	26	$(3, 3^1)$	1	81	78	11.57
3	80	$(4, 3^1)$	2	243	234	16.51
3	242	$(5, 3^1)$	2	729	720	22.92
4	63	$(3, 2^2)$	1	256	252	14.37
4	255	$(4, 2^2)$	2	1024	1008	20.47
4	1023	$(5, 2^2)$	2	4096	4080	28.49
5	124	$(3, 5^1)$	1	625	620	16.55
5	624	$(4, 5^1)$	2	3125	3100	23.58

Theorem 3 The code $\tilde{\mathcal{M}}_{\mu,\mu-1}$ for a given (γ, ρ) QC-LDPC matrix and $\mathrm{EG}(m, p^s)$ is an AC code with (i) the code length $\tilde{N} = \gamma(p^{ms} - p^{(\mu-1)s})$, (ii) the number of codewords $\rho f_{\mathrm{EG}}(\mu)$, and (iii) the resilience $\ell = \min\{\gamma - 1, p^s - 1\}$ for $\gamma \geq 2$ and $\ell = p^s - 1$ for $\gamma = 1$.

(**Proof**) Since $\pi^{a_{i,j}}(\tilde{B}_{\mu,\mu-1})$ for $i = 1, 2, \ldots, \gamma$ and $j = 1, 2, \ldots, \rho$ consists of $p^{ms} - p^{(\mu-1)s}$ rows, the code length is easily verified. Since $\tilde{\mathcal{B}}_{\mu,\mu-1}$ is a $(p^s - 1)$-resilient AC code from Theorem 1, the resilience of $\tilde{\mathcal{M}}_{\mu,\mu-1}$ is not altered from $\mathcal{M}_\mu$. □

Theorem 4 The code $\tilde{\mathcal{M}}_{\mu,\mu-1}$ for a given (γ, ρ) QC-LDPC matrix and $\mathrm{PG}(m, p^s)$ is an AC code with (i) the code length

$$\tilde{N} = \frac{\gamma(p^{(m+1)s} - p^{\mu s})}{p^s - 1}, \tag{12}$$

(ii) the number of codewords $\rho f_{\mathrm{PG}}(\mu)$, and (iii) the resilience

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

for $\gamma \geq 2$ and $\ell = p^s$ for $\gamma = 1$.

(**Proof**) The theorem can be proven in a similar way to the proof of Theorem 3. □

Example 1 We show some example of original and shortened AC codes. Tables 1 and 2 show the EG-AC codes and PG-AC codes, respectively, obtained from (γ, ρ) QC-LDPC matrices. We assume the QC-LDPC matrices are constructed by the method of [2]. In Tables, N and $\tilde{N}$ indicate the code lengths of original codes and shortened codes, respectively. The columns "$\log_2 \rho f_{\mathrm{EG}}(\mu)$" and "$\log_2 \rho f_{\mathrm{PG}}(\mu)$" express the logarithms of the number of codewords, i.e., the number of information symbols. The effectiveness of the shortened codes becomes high as the dimension m of the finite geometry $\mathrm{FG}(m, p^2)$ increases. □

Table 2: Examples of Original and Shortened PG-AC Codes Based on (γ, ρ) QC-LDPC Matrices

γ	ρ	(m, p^s)	μ	N	$\tilde{N}$	$\log_2 \rho f_{\mathrm{PG}}(\mu)$
3	26	$(3, 3^1)$	1	120	117	11.72
3	80	$(4, 3^1)$	2	363	351	16.56
3	242	$(5, 3^1)$	2	1092	1080	22.97
4	63	$(3, 2^2)$	1	340	336	14.46
4	255	$(4, 2^2)$	2	1364	1344	20.50
4	1023	$(5, 2^2)$	2	5460	5440	28.52
5	124	$(3, 5^1)$	1	780	775	16.61
5	624	$(4, 5^1)$	2	3905	3875	23.85

4.4. Comparison with Distortions between Conventional and Shortened AC Codes

In this subsection, we will compare distortion to the original content given by the conventional AC codes and the shortened AC codes.

We denote the code lengths of original AC codes and shortened AC codes by N and $\tilde{N}$, respectively. 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 [12].

It follows from $\boldsymbol{y}_j = \boldsymbol{x} + \boldsymbol{w}_j$ and eq. (1) that the distortion of an original AC code to a content can be calculated as

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

where the first equality is obtained by the linearity of the expectation. Similarly, the distortion of a shortened AC code to a content can be calculated as

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

Therefore the distortion to an original content given by the shortened AC codes of Proposition 1 (and hence, Theorems 1–4) is reduced compared with that of original AC codes. As the number of shortened symbols becomes large, the effectiveness of the shortening becomes greater.

5. Conclusion and Future Improvements

In this paper, for a class of AC codes devised by Trappe et al. and Yagi et al., novel methods for shortening the code length were proposed, while their number of codewords and the resilience are maintained. We proposed structured methods for shortening AC codes by using properties of finite geometries. The shortened AC codes can have the shorter codewords than original AC codes. Due to a smaller code length, the distortion provided by AC codes to an original content can be reduced and the coding rate becomes high. The effectiveness of the shortened codes becomes high as the dimension m of a finite geometry $\mathrm{FG}(m, p^2)$ increases.

In this paper, the resilience is guaranteed by assuming that there occur no additive noises. Therefore, the performance of the AC codes should be analyzed by assuming additive noises.

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 commn. letters*, vol. 7, no. 7, pp.317–319, June 2003.

[3] S. He and M. Wu, "Improving collusion resistance of error correcting code based multimedia fingerprinting," *Proc of Int.. Conf. on Acoustics, Speech, and Signal Processing 2005 (ICASSP2005)*, vol. 2, pp. 1029–1032, USA, Mar. 2005.

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

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

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] H. Yagi, T. Matsushima, and S. Hirasawa, "Collusion-secure codes for fingerprinting based on finite geometries," (in Japanese) *Proc. of 28th Symp. on Information Theory and its Applications (SITA2005)*, pp.189-192, Okinawa, Japan, Nov. 2005.

[11] 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.

[12] H. Yagi, T. Matsushima, and S. Hirasawa, "Improved collusion-secure codes for digital fingerprinting based on finite geometries," submitted to *2007 IEEE Int. Conf. on Systems, Man, and Cybernetics (SMC 2007)*, Montreal, Canada, Oct. 2007.