Best Security Index for Digital Fingerprinting
(Extended Abstract)

Kozo Banno[1], Shingo Orihara[2], Takaaki Mizuki[3], and Takao Nishizeki[1]

[1] Graduate School of Information Sciences, Tohoku University,
Aramaki-Aza-Aoba 6-6-05, Aoba-ku, Sendai 980-8579, Japan
[2] NTT Information Sharing Platform Laboratories,
Midori-Cho 3-9-11, Musashino-Shi, Tokyo 180-8585, Japan
[3] Information Synergy Center, Tohoku University,
Aramaki-Aza-Aoba 6-3, Aoba-ku, Sendai 980-8578, Japan
tm-paper@rd.isc.tohoku.ac.jp

Abstract. Digital watermarking used for fingerprinting may receive a collusion attack; two or more users collude, compare their data, find a part of embedded watermarks, and make an unauthorized copy by masking their identities. In this paper, assuming that at most c users collude, we give a characterization of the fingerprinting codes that have the best security index in a sense of "$(c, p/q)$-secureness" proposed by Orihara *et al.* The characterization is expressed in terms of intersecting families of sets. Using a block design, we also show that a distributor of data can only find asymptotically a set of c users including at least one culprit, no matter how good fingerprinting code is used.

1 Introduction

Various kinds of data such as documents, music, movies, etc. are digitized, and are processed as digital contents. The digital data can be sent to millions of people instantly through the Internet, and copyright violation is now a serious social problem. One of the key techniques for the problem is digital watermarking. It embeds a secret mark in the digital contents so that the secret mark cannot be detected when the resulting contents are conventionally replayed. The digital watermarking usually embeds either "author's ID" or "user's ID" as a secret mark. In the former case, the author of the contents can insist that the contents are produced by himself/herself. In the latter, a distributor of the contents can identify a user from his/her contents. The latter is called fingerprinting.

Digital watermarking used for fingerprinting may receive a collusion-attack; two or more users collude, compare their data, find a part of embedded watermarks, and make an unauthorized copy by masking their identities. In this paper we assume that at most c users collude for some number c. The "$(c, p/q)$-secureness" has been proposed as an index to measure the resilience of fingerprinting codes for such a collusion attack; a code for fingerprinting is $(c, p/q)$-*secure* for integers $p \geq 0$ and $q \geq 1$ if a distributor can find a set of q users such that at least p of them are surely collusive [5]. The largest fraction p/q among all

M. Barni et al. (Eds.): IH 2005, LNCS 3727, pp. 398–412, 2005.

fingerprinting codes is called the *best security index* and denoted by $s(c)$. Some upper and lower bounds on $s(c)$ are given, and it is known that $s(1) = 1/1$, $s(2) = 2/3$ and $s(3) = 3/7$ [5]. However, it has been remained open to determine accurately the value of $s(c)$ for $c \geq 4$.

In this paper, we first characterize the fingerprinting codes that have the best security index $s(c)$, and then show that $s(c)$ is determined by the intersecting families of sets. Using a block design, we furthermore show that $s(c) \leq c/(c^2 - c + 1)$ for an infinite number of c and hence $s(c) = 1/c$ holds asymptotically. Thus a distributor can only find a set of c users including at least one culprit, no matter how good fingerprinting code is used.

The remainder of the paper is organized as follows. In Sect. 2, we formally describe a model of watermarking and define the "$(c, p/q)$-secureness" and the best security index $s(c)$. In Sect. 3, we present a characterization of fingerprinting codes that have the best security index $s(c)$. In Sect. 4, we show that $s(c) = 1/c$ holds asymptotically. Finally, in Sect. 5, we conclude with discussions.

2 Preliminaries

In this section, we first present a model of watermarking used in the paper, and then define some terms.

2.1 The Model of the Watermark

Assume that there are a number n of (legal) *users*, $u_1, u_2, \cdots, u_n$, and a *distributor* of contents. A *watermark* w is a binary sequence of length $l \geq 1$: $w \in W = \{0, 1\}^l$. The distributor chooses a watermark $w_i \in W$ for each user u_i, $1 \leq i \leq n$. The watermarks $w_1, w_2, \cdots, w_n$ are distinct with each other, and are called the *legal watermarks*. The set $\Gamma = \{w_1, w_2, \cdots, w_n\}$ is called an (l, n)-*code* or simply a *code*. The distributor embeds a watermark w_i in the contents, and distributes the resulting contents to each user u_i. The i-th bit of a watermark $w \in W$ is denoted by $\langle w \rangle_i$.

We make the following assumption throughout the paper.

Assumption 1 (Marking Assumption [3]). *Any single user cannot find out where his/her watermark is embedded in the contents. However, if two or more users collude, then, since their watermarks are different from each other, they can realize some of the bit positions of their contents in which their watermarks are embedded by comparing their data and finding some differences in their data. These discovered bits cannot be deleted, but can be arbitrarily changed to either 0 or 1.*

We call a nonempty subset $C \subseteq \Gamma$ a *coalition* of a code Γ. Let $r = |C|$, and let $C = \{w_{c_1}, w_{c_2}, \cdots, w_{c_r}\}$. Thus the r users $u_{c_1}, u_{c_2}, \cdots, u_{c_r}$ are collusive. If all the i-th bits of their watermarks are same, i.e. $\langle w_{c_1} \rangle_i = \langle w_{c_2} \rangle_i = \cdots = \langle w_{c_r} \rangle_i$, then the users in coalition C cannot change the i-th bits of their watermarks because they cannot know where their i-th bits are embedded in the contents.

Otherwise, the users in C can change the i-th bits of their watermarks to either 0 or 1 arbitrarily because they can know where the i-th bits of their watermarks are embedded. The set of all watermarks that are obtained in this way is called the set of *falsified watermarks* by coalition C, and is denoted by $F(C)$. Thus, each falsified watermark $w \in F(C)$ satisfies

$$\langle w \rangle_i = \begin{cases} 0 & \text{if } \langle w_{c_1} \rangle_i = \langle w_{c_2} \rangle_i = \cdots = \langle w_{c_r} \rangle_i = 0; \\ 1 & \text{if } \langle w_{c_1} \rangle_i = \langle w_{c_2} \rangle_i = \cdots = \langle w_{c_r} \rangle_i = 1; \\ 0 \text{ or } 1 & \text{otherwise} \end{cases}$$

for each bit position i, $1 \leq i \leq l$. We hence have

$$F(C) = \{w \in W \mid \text{for each } i,\ 1 \leq i \leq l, \\ \text{there is } w' \in C \text{ with } \langle w \rangle_i = \langle w' \rangle_i\}. \tag{1}$$

One can observe from Eq. (1) that the set $F(C)$ of bit sequences can be represented by a sequence of characters $0, 1$ and $*$ of length l; the i-th character $\langle F(C) \rangle_i$ of $F(C)$, $1 \leq i \leq l$, satisfies

$$\langle F(C) \rangle_i = \begin{cases} 0 \text{ if } \langle w_{c_1} \rangle_i = \langle w_{c_2} \rangle_i = \cdots = \langle w_{c_r} \rangle_i = 0; \\ 1 \text{ if } \langle w_{c_1} \rangle_i = \langle w_{c_2} \rangle_i = \cdots = \langle w_{c_r} \rangle_i = 1; \\ * \text{ otherwise,} \end{cases} \tag{2}$$

where $*$ means DON'T CARE. It should be noted that $C \subseteq F(C)$ for any coalition $C \subseteq \Gamma$.

When a distributor finds an unauthorized copy, he/she detects an illegal watermark $w \in (W - \Gamma)$ embedded in the copy and finds a coalition C such that $w \in F(C)$. We assume that a bounded number of users, say at most c users, take part in the coalition C.

An illegal watermark $w \in (W - \Gamma)$ may be contained in $F(C)$ for several coalitions C of at most c users. So we define a set $\mathcal{S}(c, w; \Gamma)$ of coalitions as follows.

Definition 1. *For a code Γ, a watermark $w \in W$ and an integer $c \geq 1$, we define a* suspected family *for w as*

$$\mathcal{S}(c, w; \Gamma) = \{C \subseteq \Gamma \mid 1 \leq |C| \leq c,\ w \in F(C)\}.$$

We often denote $\mathcal{S}(c, w; \Gamma)$ simply by $\mathcal{S}(c, w)$.

Thus $\mathcal{S}(c, w; \Gamma) \subseteq 2^{\Gamma}$. If $\mathcal{S}(c, w; \Gamma) = \emptyset$, then there is no coalition of at most c users that can make the watermark w. From Definition 1 and Eq. (2) we immediately have the following lemma.

Lemma 1. *Let $w \in W$, $C \subseteq \Gamma$, $1 \leq |C| = r \leq c$ and $C = \{w_{c_1}, w_{c_2}, \cdots, w_{c_r}\}$. Then $C \notin \mathcal{S}(c, w; \Gamma)$ if and only if there exists a bit position i, $1 \leq i \leq l$, such that*

$$\langle w_{c_1} \rangle_i = \langle w_{c_2} \rangle_i = \cdots = \langle w_{c_r} \rangle_i \neq \langle w \rangle_i. \tag{3}$$

A distributor can find $\mathcal{S}(c, w; \Gamma)$ as follows.

- First, the distributor considers a family $\mathcal{E}_0^w = \{C \mid C \subseteq \Gamma, 1 \leq |C| \leq c\}$.
- Then, for each bit-position i, $1 \leq i \leq l$, the distributor removes from $\mathcal{E}_0^w$ all sets $C \in \mathcal{E}_0^w$ such that

$$\langle w_{c_1} \rangle_i = \langle w_{c_2} \rangle_i = \cdots = \langle w_{c_r} \rangle_i \neq \langle w \rangle_i.$$

By Lemma 1, the resulting family is the suspected family $\mathcal{S}(c, w; \Gamma)$.

The *c-coalition detection problem* is to detect a coalition that made the unauthorized copy, assuming that at most c users collude.

2.2 Secureness of Codes

Various research has been done on the secureness of codes (e.g. [3,4,5,6,7,8]). Boneh and Shaw defined "c-secureness" as an index to measure the resilience of watermarks for collusion attacks [3]; a code is *c-secure* if a distributor can detect at least one of the collusive users when at most c users collude. However, they showed that there is indeed no c-secure code [3]. They also defined "ϵ-error c-secureness"; a code is *ϵ-error c-secure* if a distributor can detect at least one of the collusive users with probability at least $1 - \epsilon$ when at most c users collude. They constructed an example of an ϵ-error c-secure code [3]. If a code is ϵ-error c-secure, then a distributor can detect at least one of the collusive users with small error, but cannot surely detect a definitely collusive user. Orihara *et al.* introduced the "$(c, p/q)$-secureness" as an index to measure the quality of a code; if a code is $(c, p/q)$-secure, then the distributor may not detect all the collusive users, but can detect a group of q users including at least p collusive users [5]. Yoshioka *et al.* [7,8] investigated the relationships among c-secureness, ϵ-error c-secureness, $(c, p/q)$-secureness, c-frameproofness [3], c-secure frameproofness [6], and so on. Note that the more basic collusion problem was discussed first by Blakley, Meadows and Purdy [2].

In the remainder of this section, we explain $(c, p/q)$-secureness.

We first define some terms.

Definition 2. *For integers $p \geq 0$ and $q \geq 1$, we call $[p/q]$ an* index. *For a set V, we say that a family $\mathcal{S} \subseteq 2^V$ is* $[p/q]$-detectable *if there exists a set $X \subseteq V$ such that $|X| = q$ and $|C \cap X| \geq p$ for any set $C \in \mathcal{S}$.*

If a suspected family $\mathcal{S}(c, w)$ is $[p/q]$-detectable, then there is a set X of q suspicious users and a distributor can insist that at least p of them are surely culprits.

For a family $\mathcal{S} \subseteq 2^V$, there are many pairs of integers p and q for which $\mathcal{S}$ is $[p/q]$-detectable. For example, if $V = \{w_1, w_2, w_3, w_4\}$ and

$$\mathcal{S} = \big\{\{w_1, w_2\}, \{w_2, w_3\}, \{w_3, w_1\}\big\},$$

then $\mathcal{S}$ is $[1/2]$-detectable and $[2/3]$-detectable. So we wish to specify a pair of integers p and q best to describe the feature of $\mathcal{S}$. We thus define a total order "$\preceq$" on the set of indices as follows.

Definition 3. *Let $p \geq 0$ and $q \geq 1$. If either $\frac{p}{q} < \frac{r}{s}$, or $\frac{p}{q} = \frac{r}{s}$ and $q < s$, then $[p/q] \prec [r/s]$. If $p = r$ and $q = s$, then $[p/q] = [r/s]$. If $[p/q] \prec [r/s]$ or $[p/q] = [r/s]$, then $[p/q] \preceq [r/s]$.*

For example, we have $[0/4] \prec [1/4] \prec [1/3] \prec [1/2] \prec [2/4] \prec [2/3] \prec [3/4] \prec [1/1] \prec [2/2] \prec [3/3] \prec [4/4]$.

We then define the "detectable index" of a family $\mathcal{S} \subseteq 2^V$ as follows.

Definition 4. *For a set V and a nonempty family $\mathcal{S} \subseteq 2^V$, we define a* detectable index $d(\mathcal{S})$ *of $\mathcal{S}$ to be the "best" index $[p/q]$ such that the $\mathcal{S}$ is $[p/q]$-detectable, that is*

$$d(\mathcal{S}) = \max \{[p/q] \mid \mathcal{S} \text{ is } [p/q]\text{-detectable}\}$$

where max $(\preceq)$ *is taken over all indices. We define $d(\mathcal{S}) = [\infty/\infty]$ if $\mathcal{S} = \emptyset$, and define $[p/q] \preceq [\infty/\infty]$ for any indices $[p/q]$.*

We now define a $(c, p/q)$-secureness as follows.

Definition 5. *Let Γ be a code and let c is a natural number. We say that Γ is* $(c, p/q)$-secure *if $d(\mathcal{S}(c, w; \Gamma)) \succeq [p/q]$ for any watermark $w \in W$.*

If a code Γ is $(c, p/q)$-secure, then for any (illegal) watermark $w \in W$ there is a set X of q suspicious users such that at least p of them are surely culprits, under an assumption that at most c users collude.

We now define a "security index" $s(\Gamma, c)$ of a code Γ as follows.

Definition 6. *For a natural number c, a* security index $s(\Gamma, c)$ *of a code Γ is*

$$s(\Gamma, c) = \min \{d\left(\mathcal{S}(c, w; \Gamma)\right) \mid w \in W\}$$

where min $(\preceq)$ is taken over all watermarks $w \in W$.

The security index $s(\Gamma, c)$ is the minimum detectable index $d(\mathcal{S}(c, w; \Gamma))$ for all watermarks $w \in W$. Clearly, $s(\Gamma, c)$ is also the maximum one for all indices $[p/q]$ such that a code Γ is $(c, p/q)$-secure.

We now define the best security index $s(c)$ as follows.

Definition 7. *The* best security index $s(c)$ *for collusions of at most c users is*

$$s(c) = \max\{s(\Gamma, c) \mid \Gamma \text{ is a code}\}$$

where max *$(\preceq)$ is taken over all codes Γ.*

3 c-Intersecting Code

In this section, we present a characterization of fingerprinting codes that have the best security index $s(c)$.

We first define some terms.

Definition 8. *A family $\mathcal{S}$ of sets is* intersecting *if $C \cap C' \neq \emptyset$ for any sets $C, C' \in \mathcal{S}$. An intersecting family $\mathcal{S}$ is* c-intersecting *if $|C| \leq c$ for every set $C \in \mathcal{S}$.*

Definition 9. *A code Γ is* c-intersecting *if the suspected family $\mathcal{S}(c, w; \Gamma)$ is intersecting for every watermark $w \in W$.*

The code Γ_c in Sect. 4, the c-secure frameproof code in [6], and the (c,c)-separating code in [4] are examples of c-intersecting codes.

For a set V, we denote by $\mathcal{F}(V, c)$ the set of all c-intersecting families $\mathcal{E} \subseteq 2^V$:

$$\mathcal{F}(V, c) = \{\mathcal{E} \subseteq 2^V \mid \mathcal{E} \text{ is } c\text{-intersecting}\}.$$

We define an index $d(n, c)$ as follows:

$$d(n, c) = \min\{d(\mathcal{E}) \mid \mathcal{E} \in \mathcal{F}(V, c)\}$$

where V is a set of n elements, i.e., $|V| = n$. The index $d(n, c)$ is determined only by n and c, and does not depend on the set V. For example, $d(3, 2) = [2/3]$, because $d(\mathcal{E}) = [2/3]$ for a 2-intersecting family

$$\mathcal{E} = \big\{\{w_1, w_2\}, \{w_2, w_3\}, \{w_3, w_1\}\big\} \in \mathcal{F}(V, 2)$$

and $d(\mathcal{E}') = [1/1] \succ [2/3]$ for any other 2-intersecting family $\mathcal{E}' \in \mathcal{F}(V, 2)$ where $V = \{w_1, w_2, w_3\}$. Note that $d(\mathcal{E}') = [1/1]$ for $\mathcal{E}' = \big\{\{w_1, w_2\}, \{w_1, w_3\}\big\} \in \mathcal{F}(V, 2)$.

A main result of this section is the following theorem.

Theorem 1. *If a code Γ is c-intersecting, then $s(\Gamma, c) = s(c) = d(n, c)$ and hence the $s(\Gamma, c)$ is the maximum among all codes.*

We give a proof of Theorem 1 in the remainder of this section. For a Boolean value $x \in \{0, 1\}$, we define $\overline{x}$ as follows:

$$\overline{x} = \begin{cases} 1 & \text{if} \quad x = 0; \\ 0 & \text{if} \quad x = 1. \end{cases}$$

We then have the following lemma, the proof of which is omitted in this extended abstract due to the page limitation.

Lemma 2. *A code Γ is c-intersecting if and only if, for any coalitions $C_1, C_2 \subseteq \Gamma$ such that $C_1 \cap C_2 = \emptyset$ and $|C_1| = |C_2| = c$, there exists a bit position i, $1 \leq i \leq l$, such that $\langle F(C_1)\rangle_i = x$ and $\langle F(C_2)\rangle_i = \overline{x}$, $x \in \{0, 1\}$.*

If $\mathcal{S}(c, w; \Gamma)$ is intersecting, then $C \cap C' \neq \emptyset$ for any coalitions $C, C' \in \mathcal{S}(c, w; \Gamma)$. For a legal watermark $w_i \in \Gamma$,

$$\{w_i\} \in \mathcal{S}(c, w_i; \Gamma)$$

and hence $\bigcap\{C \mid C \in \mathcal{S}(c, w_i; \Gamma)\} = \{w_i\}$. Thus every coalition that can make a legal watermark w_i includes w_i. On the other hand, if a coalition $C \in \Gamma$

could make a legal watermark $w_i \in (\Gamma - C)$, then an innocent user u_i would be suspected. However, if Γ is c-intersecting, then such a false charge would not occur.

The fewer coalitions included in a suspected family are, the more accurate information a distributor obtains about the collusive users. However, we have the following lemma.

Lemma 3. *For any code Γ and any c-intersecting family $\mathcal{E} \subseteq 2^{\Gamma}$, there is a watermark $w \in W$ such that $\mathcal{E} \subseteq \mathcal{S}(c, w; \Gamma)$.*

Proof. If $\mathcal{E} = \emptyset$, then clearly $\emptyset = \mathcal{E} \subseteq \mathcal{S}(c, w; \Gamma)$ for any watermark $w \in W$. One may thus assume that $|\mathcal{E}| = m \geq 1$ and $\mathcal{E} = \{C_1, C_2, \cdots, C_m\}$. Since $\mathcal{E}$ is c-intersecting,

$$C_i \cap C_j \neq \emptyset \tag{4}$$

for any indices i and j, $1 \leq i < j \leq m$. If

$$\bigcap_{i=1}^{m} F(C_i) \neq \emptyset \tag{5}$$

then there is a watermark $w' \in W$ such that $w' \in \bigcap_{i=1}^{m} F(C_i)$, and $\mathcal{E} = \{C_1, C_2, \cdots, C_m\} \subseteq \mathcal{S}(c, w'; \Gamma)$. It thus suffices to verify Eq. (5).

Suppose for a contradiction that $\bigcap_{i=1}^{m} F(C_i) = \emptyset$. Since $C_1 \subseteq F(C_1) \neq \emptyset$, there is an integer r, $1 < r \leq m$, such that $\bigcap_{i=1}^{r-1} F(C_i) \neq \emptyset$ and $\bigcap_{i=1}^{r} F(C_i) = \emptyset$. Thus there exists a bit position k, $1 \leq k \leq l$, such that $\langle \bigcap_{i=1}^{r-1} F(C_i) \rangle_k = x$ and $\langle F(C_r) \rangle_k = \bar{x}$, where $x \in \{0, 1\}$. Since $\langle \bigcap_{i=1}^{r-1} F(C_i) \rangle_k = x$, we have $\langle F(C_j) \rangle_k = x$ for some index j, $1 \leq j \leq r-1$. Therefore by Eq. (2) we have $\langle w \rangle_k = x$ for every watermark $w \in C_j$. On the other hand, since $\langle F(C_r) \rangle_k = \bar{x}$, we have $\langle w \rangle_k = \bar{x}$ for every watermark $w \in C_r$. We thus have $C_j \cap C_r = \emptyset$, contrary to Eq. (4). □

If, for any watermark $w \in W$, $\mathcal{S}(c, w; \Gamma)$ is intersecting and is of a *star type* in particular, that is, there is a legal watermark $w_i \in \Gamma$ which is included in every coalition $C \in \mathcal{S}(c, w; \Gamma)$, then a distributor can surely detect the user u_i as one of the collusive users. However, when $n \geq 3$ and $c \geq 2$, there is no code Γ such that $\mathcal{S}(c, w; \Gamma)$ is of a star type for every watermark $w \in W$, because $\mathcal{E} = \big\{\{w_1, w_2\}, \{w_2, w_3\}, \{w_3, w_1\}\big\}$ is intersecting but is not of a star type, and by Lemma 3 there is a watermark $w \in W$ such that $\mathcal{E} \subseteq \mathcal{S}(c, w; \Gamma)$.

The following lemma is known [5].

Lemma 4 ([5]). *If $\mathcal{S}_1 \subseteq \mathcal{S}_2 \subseteq 2^{\Gamma}$, then $d(\mathcal{S}_1) \succeq d(\mathcal{S}_2)$.*

Using Lemmas 2, 3 and 4, we now prove the following Lemma 5 on the secureness of an intersecting code.

Lemma 5. *If Γ is a c-intersecting code and $|\Gamma| = n$, then $s(\Gamma, c) = d(n, c)$.*

Proof. Let $\Gamma = \{w_1, w_2, \cdots, w_n\}$. One may assume that

$$d(n, c) = d(\mathcal{E}_{\min}) \tag{6}$$

for a c-intersecting set $\mathcal{E}_{\min} \in \mathcal{F}(\Gamma, c)$. Then,

$$d(\mathcal{E}_{\min}) \preceq d(\mathcal{E}) \tag{7}$$

for every $\mathcal{E} \in \mathcal{F}(\Gamma, c)$. It should be noted that the value $d(\mathcal{E}_{\min})$ is determined only by n and c and does not depend on what bit-sequence each watermark $w_i \in \Gamma$ is.

We first verify $d(n, c) \preceq s(\Gamma, c)$. By Definition 6,

$$s(\Gamma, c) = \min_{w \in W} \big\{ d\big(\mathcal{S}(c, w; \Gamma)\big) \big\}.$$

One may assume that a watermark $w_{\min} \in W$ attains the minimum above. Then

$$s(\Gamma, c) = d\big(\mathcal{S}(c, w_{\min}; \Gamma)\big) \preceq d\big(\mathcal{S}(c, w; \Gamma)\big) \tag{8}$$

for every watermark $w \in W$. Since the code Γ is c-intersecting, $\mathcal{S}(c, w_{\min}; \Gamma)$ is c-intersecting and hence

$$\mathcal{S}(c, w_{\min}; \Gamma) \in \mathcal{F}(\Gamma, c). \tag{9}$$

By Eqs. (6) – (9), we have

$$d(n, c) = d(\mathcal{E}_{\min}) \preceq d\big(\mathcal{S}(c, w_{\min}; \Gamma)\big) = s(\Gamma, c).$$

We then verify $d(n, c) \succeq s(\Gamma, c)$. Since $\mathcal{E}_{\min}$ is c-intersecting, by Lemma 3 there is a watermark $w' \in W$ such that

$$\mathcal{E}_{\min} \subseteq \mathcal{S}(c, w'; \Gamma).$$

Therefore, by Lemma 4, we have

$$d(\mathcal{E}_{\min}) \succeq d\big(\mathcal{S}(c, w'; \Gamma)\big),$$

and hence

$$\begin{aligned} d(n, c) = d(\mathcal{E}_{\min}) &\succeq d\big(\mathcal{S}(c, w'; \Gamma)\big) \\ &\succeq \min_{w \in W} \big\{ d\big(\mathcal{S}(c, w; \Gamma)\big) \big\} = s(\Gamma, c), \end{aligned}$$

as desired. □

We are now ready to prove Theorem 1.

Proof of Theorem 1. Let $\Gamma = \{w_1, w_2, \cdots, w_n\}$. By Lemma 5 $s(\Gamma, c) = d(n, c)$. Therefore it suffices to verify $s(\Gamma, c) = s(c)$.

Suppose for a contradiction that there is a code $\Gamma_a = \{w_1^a, w_2^a, \cdots, w_n^a\}$ such that $s(\Gamma_a, c) \succ s(\Gamma, c)$. Let l_a be the length of the code Γ_a, and let l be the length of the code Γ. From Γ_a and Γ we construct a new code $\Gamma_b = \{w_1^b, w_2^b, \cdots, w_n^b\}$ of length $l_b = l_a + l$ where $w_i^b = w_i^a \parallel w_i$, $1 \le i \le n$, that is, the bit-sequence w_i^b is a concatenation of w_i^a and w_i.

We claim that Γ_b is c-intersecting. Let $w \in \{0,1\}^{l_b}$ be an arbitrary watermark of length l_b. Let $\mathcal{E}_0^w = \{C \mid C \subseteq \Gamma_b, 1 \leq |C| \leq c\}$. Compare the i-th bits of w and $w_i^b \in \Gamma_b$ for each i, $l_a + 1 \leq i \leq l_b$, and remove C from $\mathcal{E}_0^w$ if Eq. (3) holds for C. (See Lemma 1.) The resulting family is intersecting, because the bit-subsequences of Γ_b from the (l_a+1)-th position to the l_b-th correspond to the bit-sequences of a c-intersecting code Γ. $\mathcal{S}(c, w; \Gamma_b)$ can be obtained from the resulting intersecting family by repeating the operation above from the first position to the l_a-th, and hence $\mathcal{S}(c, w; \Gamma_b)$ is a subset of the resulting intersecting family. Thus $\mathcal{S}(c, w; \Gamma_b)$ is intersecting, and hence Γ_b is c-intersecting.

Let $\mathcal{E}_{l_a}^w$ be a family obtained from $\mathcal{E}_0^w$ by repeating the operation for the watermark $w \in \{0,1\}^{l_b}$ and each watermark $w_i^b \in \Gamma_b$ from the first position to the l_a-th, and let $\mathcal{E}_{l_b}^w$ be the family obtained from $\mathcal{E}_{l_a}^w$ by repeating the operation from the (l_a+1)-th position to the l_b-th. Then $\mathcal{E}_{l_b}^w = \mathcal{S}(c, w; \Gamma_b)$, and hence $d(\mathcal{E}_{l_b}^w) = d(\mathcal{S}(c, w; \Gamma_b))$. Since $\mathcal{E}_{l_b}^w \subseteq \mathcal{E}_{l_a}^w$, by Lemma 4 we have

$$d\big(\mathcal{S}(c, w; \Gamma_b)\big) = d(\mathcal{E}_{l_b}^w) \succeq d(\mathcal{E}_{l_a}^w). \tag{10}$$

Let w' be the first l_a bits sequence of w. Let $\mathcal{E}_{l_a}^{w'}$ be the family obtained from $\mathcal{E}_0^{w'} = \{C \mid C \subseteq \Gamma_a, 1 \leq |C| \leq c\}$ by repeating the operation for w' and $w_i^a \in \Gamma_a$ from the first position to the l_a-th. Then $\mathcal{E}_{l_a}^{w'} = \mathcal{S}(c, w'; \Gamma_a)$. Although $\mathcal{E}_{l_a}^{w'} \subseteq 2^{\Gamma_a}$ and $\mathcal{E}_{l_a}^w \subseteq 2^{\Gamma_b}$, the families $\mathcal{E}_{l_a}^{w'}$ and $\mathcal{E}_{l_a}^w$ are isomorphic. We therefore have

$$d(\mathcal{E}_{l_a}^w) = d(\mathcal{E}_{l_a}^{w'}) = d\big(\mathcal{S}(c, w'; \Gamma_a)\big). \tag{11}$$

By Eqs. (10) and (11), for an arbitrary watermark $w \in \{0,1\}^{l_b}$ and the watermark w' that is the first l_a bits sequence of w, we have

$$d\big(\mathcal{S}(c, w; \Gamma_b)\big) \succeq d(\mathcal{E}_{l_a}^w) = d\big(\mathcal{S}(c, w'; \Gamma_a)\big). \tag{12}$$

Let $\hat{w}$ be a watermark $w \in \{0,1\}^{l_b}$ that minimizes the index $d\big(\mathcal{S}(c, w; \Gamma_b)\big)$. Then

$$\begin{aligned} d\big(\mathcal{S}(c, \hat{w}; \Gamma_b)\big) &= \min_{w \in \{0,1\}^{l_b}} \big\{d\big(\mathcal{S}(c, w; \Gamma_b)\big)\big\} \\ &= s(\Gamma_b, c). \end{aligned} \tag{13}$$

Let $\hat{w}'$ be the watermark that is the first l_a bits sequence of $\hat{w}$, then by Eq. (12) we have

$$d\big(\mathcal{S}(c, \hat{w}; \Gamma_b)\big) \succeq d\big(\mathcal{S}(c, \hat{w}'; \Gamma_a)\big). \tag{14}$$

By Eqs. (13) and (14) we have

$$\begin{aligned} s(\Gamma_b, c) &\succeq d\big(\mathcal{S}(c, \hat{w}'; \Gamma_a)\big) \\ &\succeq \min_{w' \in \{0,1\}^{l_a}} \big\{d\big(\mathcal{S}(c, w'; \Gamma_a)\big)\big\} \\ &= s(\Gamma_a, c). \end{aligned} \tag{15}$$

Since both Γ and Γ_b are c-intersecting, by Lemma 5

$$s(\Gamma, c) = s(\Gamma_b, c) = d(n, c). \tag{16}$$

By Eqs. (15) and (16) we have

$$s(\Gamma, c) \succeq s(\Gamma_a, c).$$

However, this contradicts to the assumption $s(\Gamma_a, c) \succ s(\Gamma, c)$. □

4 The Best Security Index

In this section, using theory of block designs, we show that $s(c) = [1/c]$ holds asymptotically.

Orihara *et al.* obtained the following Theorems 2 and 3 for the upper bound on the security index $s(\Gamma, c)$ [5].

Theorem 2 ([5]). *If $c \leq (n+1)/2$, then $s(\Gamma, c) \preceq [c/(2c-1)]$ for every code Γ. If $n \geq 7$ and $c \geq 3$, then $s(\Gamma, c) \preceq [3/7]$ for every code Γ.*

A code $\Gamma_c = \{w_1, w_2, \cdots, w_n\}$ is defined by the following $n \times l$ binary matrix

$$\Gamma_c = \begin{array}{c} \\ w_1 \\ w_2 \\ \vdots \\ w_n \end{array} \begin{array}{c} \begin{array}{cccc} \underbrace{}_{\binom{n}{1}} & \underbrace{}_{\binom{n}{2}} & \cdots & \underbrace{}_{\binom{n}{c}} \end{array} \\ \begin{bmatrix} 1000\cdots & 0110\cdots & & 0 \\ 0100\cdots & 0100\cdots & & 0 \\ & \vdots & \cdots\cdots & \vdots \\ 0000\cdots & 1000\cdots & & 1 \end{bmatrix} \end{array},$$

where $l = \binom{n}{1} + \binom{n}{2} + \cdots + \binom{n}{c}$. The i-th row represents w_i for each i, $1 \leq i \leq n$. Each column corresponds to a set $C \subseteq \Gamma$ such that $1 \leq |C| \leq c$. The first $\binom{n}{1} = n$ columns list all bit patterns of length n, each having exactly one 1. The succeeding $\binom{n}{2}$ columns list all bit patterns, each having exactly two 1's, and so on. The last $\binom{n}{c}$ columns list all bit patterns, each having exactly c 1's. Thus each watermark has length l. The following theorem is known for $s(c)$ and $s(\Gamma_c, c)$ [5].

Theorem 3 ([5]). *For any natural number c, $s(c) \succeq s(\Gamma_c, c) \succeq [1/c]$. If $n \geq 3$ then $s(2) = s(\Gamma_2, 2) = [2/3]$, and if $n \geq 8$ then $s(3) = s(\Gamma_3, 3) = [3/7]$.*

Thus a lower bound $[1/c]$ on $s(c)$ is known, but the exact value of $s(c)$ has not been known for $c \geq 4$. We now have the following theorem on $s(c)$.

Theorem 4. *If $1 \leq c \leq n/2$, then $s(c) = s(\Gamma_c, c) = d(n, c)$.*

Proof. Let C_1, $C_2 \subseteq \Gamma_c$ be any coalitions such that $C_1 \cap C_2 = \emptyset$ and $|C_1| = |C_2| = c$. Then there exists a bit position i such that every watermark $w \in \Gamma_c$ satisfies

$$\langle w \rangle_i = \begin{cases} 1 \text{ if } \ w \in C_1 \\ 0 \text{ otherwise} \end{cases}$$

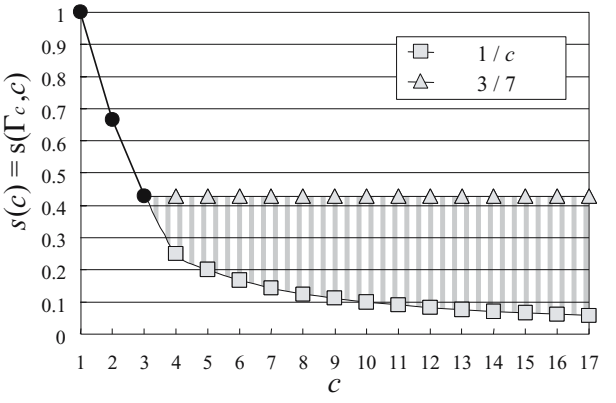

Fig. 1. The best security index $s(c) = s(\Gamma_c, c)$

and

$$\binom{n}{1}+\binom{n}{2}+\cdots+\binom{n}{c-1}<i\leq\binom{n}{1}+\binom{n}{2}+\cdots+\binom{n}{c}.$$

Since $\langle w\rangle_i = 1$ for every watermark $w \in C_1$, we have $\langle F(C_1)\rangle_i = 1$. Since $\langle w\rangle_i = 0$ for every watermark $w \in C_2$, we have $\langle F(C_2)\rangle_i = 0$. Thus Γ_c is c-intersecting by Lemma 2, and hence $s(c) = s(\Gamma_c, c) = d(n, c)$ by Theorem 1. □

The results in Theorems 2, 3 and 4 are illustrated in Figure 1. Note that $s(c) = s(\Gamma_c, c)$, $c \geq 4$, takes some value in the shaded region in Fig. 1.

We then give a new upper bound on $s(c)$. Remember that $s(c) = d(n, c)$, and that

$$d(n, c) = \min\{d(\mathcal{E}) \mid \mathcal{E} \in \mathcal{F}(\Gamma, c)\} \tag{17}$$

for an arbitrary set Γ with $|\Gamma| = n$. Therefore a detectable index $d(\mathcal{E})$ for any family $\mathcal{E} \in \mathcal{F}(\Gamma, c)$ is an upper bound on $s(c)$. We thus wish to find a family $\mathcal{E} \in \mathcal{F}(\Gamma, c)$ for which $d(\mathcal{E})$ is as smaller as possible in order to obtain a good upper bound on $s(c)$. For the purpose, we use "block designs."

Definition 10. *Let $V = \{x_1, x_2, \cdots, x_v\}$ be a set of v elements $x_1, x_2, \cdots, x_v$. We call a family $\mathcal{E} = \{B_1, B_2, \cdots, B_b\}$ of b subsets $B_1, B_2, \cdots, B_b$ of V a* block design on V with parameters (b, v, r, k, λ) *or a* (b, v, r, k, λ)-block design *if*

(1) each block B_i, $1 \leq i \leq b$, contains exactly k elements;
(2) each element x_i, $1 \leq i \leq v$, belongs to exactly r blocks; and
(3) any two distinct elements x_i and x_j, $i \neq j$, belong to exactly λ blocks.

The parameters b, v, r, k and λ must satisfy the following two equations [1]:

$$vr = kb \tag{18}$$

and

$$(k-1)r = (v-1)\lambda. \tag{19}$$

If $b = v$, $r = k$ and $\lambda \geq 1$, then any two distinct blocks B_i and B_j, $i \neq j$, in a (v, v, k, k, λ)-block design $\mathcal{E} = \{B_1, B_2, \cdots, B_v\}$ have exactly λ common elements [1], and hence a family $\mathcal{E} = \{B_1, B_2, \cdots, B_v\}$ is intersecting.

We have the following lemma.

Lemma 6. *Let $\mathcal{E} = \{B_1, B_2, \cdots, B_v\}$ be a (v, v, k, k, λ)-block design. If $d(\mathcal{E}) = [p/q]$ for integers $p \geq 0$ and $q \geq 1$, then*

$$p \leq \begin{cases} \left\lfloor \dfrac{kq}{v} \right\rfloor & \text{if} \quad 1 \leq q \leq v \\ k & \text{if} \quad v < q. \end{cases}$$

Proof. Let $V = \bigcup_{i=1}^{v} B_i$. We shall consider only the case of $1 \leq q \leq v$, because the case of $v < q$ is similar (and easier).

Let $1 \leq q \leq v$. Suppose for a contradiction that $p > \left\lfloor \dfrac{kq}{v} \right\rfloor$. Then

$$p \geq \left\lfloor \frac{kq}{v} \right\rfloor + 1.$$

Since $d(\mathcal{E}) = [p/q]$, there exists a set $X \subseteq V$ such that $|X| = q$ and

$$|B_i \cap X| \geq p \geq \left\lfloor \frac{kq}{v} \right\rfloor + 1$$

for every block $B_i \in \mathcal{E}$. We thus have

$$\sum_{i=1}^{v} |B_i \cap X| \geq pv \geq \Big(\left\lfloor \frac{kq}{v} \right\rfloor + 1 \Big) v > kq. \tag{20}$$

Since $|X| = q$, one may assume that $X = \{x_1, x_2, \cdots, x_q\}$. Then we have

$$\begin{aligned} \sum_{i=1}^{v} |B_i \cap X| &= \sum_{i=1}^{v} \sum_{j=1}^{q} |B_i \cap \{x_j\}| \\ &= \sum_{j=1}^{q} \sum_{i=1}^{v} |B_i \cap \{x_j\}|. \end{aligned} \tag{21}$$

Since each element in V belongs to exactly k blocks, we have

$$\sum_{i=1}^{v} |B_i \cap \{x_j\}| = k. \tag{22}$$

Thus, by Eqs. (21) and (22), we have

$$\sum_{i=1}^{v} |B_i \cap X| = \sum_{j=1}^{q} k = kq,$$

contrary to Eq. (20). □

We can prove the following Lemma 7 by Lemma 6.

Lemma 7. *Let Γ be a set, and let $\mathcal{E} = \{B_1, B_2, \cdots, B_v\}$ be a (v, v, k, k, λ)-block design on $V \subseteq \Gamma$. Then $d(\mathcal{E}) = [k/v]$.*

Proof. $V = \bigcup_{i=1}^{v} B_i$, $|V| = v$, and $|B_i \cap V| = k$ for each block $B_i \in \mathcal{E}$. Therefore $\mathcal{E}$ is $[k/v]$-detectable, and hence $d(\mathcal{E}) \succeq [k/v]$. Thus it suffices to verify that $d(\mathcal{E}) \preceq [k/v]$, that is, $[p/q] \preceq [k/v]$ for any index $[p/q]$ such that $\mathcal{E}$ is $[p/q]$-detectable.

We first consider the case where $1 \leq q \leq v$. In this case $p \leq \left\lfloor \frac{kq}{v} \right\rfloor$ by Lemma 6, and hence

$$\frac{p}{q} \leq \frac{\left\lfloor \frac{kq}{v} \right\rfloor}{q} \leq \frac{\frac{kq}{v}}{q} = \frac{k}{v}.$$

We thus have $[p/q] \preceq [k/v]$.

We then consider the case where $q > v$. In this case $p \leq k$ by Lemma 6, and hence

$$\frac{p}{q} \leq \frac{k}{q} < \frac{k}{v}.$$

We thus have $[p/q] \prec [k/v]$. □

If there exists a (v, v, c, c, λ)-block design $\mathcal{E} = \{B_1, B_2, \cdots, B_v\}$ on a set V such that $V \subseteq \Gamma$ and $|\Gamma| = n$, then $\mathcal{E} \in \mathcal{F}(\Gamma, c)$ and hence $d(\mathcal{E})$ is an upper bound on $s(c)$ and $d(\mathcal{E}) = [c/v] \succeq s(c)$ by Lemma 7. By Eq. (19), the parameter v of a (v, v, c, c, λ)-block design satisfies

$$v = \frac{c^2 - c}{\lambda} + 1. \tag{23}$$

We thus have

$$s(c) \preceq d(\mathcal{E}) = [c/v] = \left[c \Big/ \left(\frac{c^2 - c}{\lambda} + 1 \right) \right].$$

We wish to make the index $d(\mathcal{E}) = [c/v]$ as smaller as possible in order to obtain a good upper bound on $s(c)$. We thus wish to make v bigger and hence λ smaller by Eq. (23). Hence we let $\lambda = 1$, because by Eq. (19) $\lambda \geq 1$ when $c \geq 2$. Then

$$v = c^2 - c + 1,$$

and we have

$$s(c) \preceq d(\mathcal{E}) = [c/(c^2 - c + 1)].$$

There does not always exist a $(c^2 - c + 1, c^2 - c + 1, c, c, 1)$-block design for every natural number c. However, there exists a $(c^2 - c + 1, c^2 - c + 1, c, c, 1)$-block design if $c - 1$ is a prime power: $c - 1 = p^q$ for some prime p and natural number q [1]. We thus have the following theorem.

Theorem 5. *Let n be any natural number. If c is a natural number such that $c^2 - c + 1 \leq n$ and $c - 1$ is a prime power, then $s(c) \preceq [c/(c^2 - c + 1)]$.*

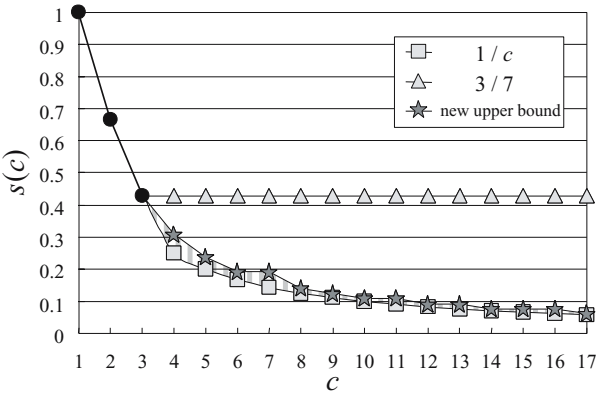

Fig. 2. The best security index $s(c)$

Proof. Since $c-1$ is a prime power, there exists a $(c^2-c+1, c^2-c+1, c, c, 1)$-block design $\mathcal{E} = \{B_1, B_2, \cdots, B_{c^2-c+1}\}$ [1]. By Lemma 7, $d(\mathcal{E}) = [c/(c^2-c+1)]$. Since $\mathcal{E} \in \mathcal{F}(\Gamma, c)$ for a set Γ with $|\Gamma| = n$, we have $s(c) = d(n, c) \preceq d(\mathcal{E}) = [c/(c^2-c+1)]$ by Eq. (17). □

We immediately have the following corollary on $s(c)$ for every natural number c.

Corollary 1. *Let n be any natural number, and let c be any natural number such that $3 \le c \le n$. If c' is a natural number such that*

$$c' = \max\{c'' \mid c'' \le c,\ c''-1 \text{ is a prime power},\ c''^2 - c'' + 1 \le n\},$$

then $s(c) \preceq [c'/(c'^2 - c' + 1)]$.

The results of Theorem 5 and Corollary 1 are illustrated in Figure 2. If $c \ge 4$, then $s(c)$ takes some value in the shaded region in Fig. 2. Theorem 5 and Corollary 1 imply that $s(c) = [1/c]$ holds asymptotically when c becomes large. Hence, a distributor can only find a set of c users including at least one culprit, no matter how good fingerprinting code is used.

5 Conclusions

This paper deals with the problem of fingerprinting codes for collusion attacks. We presented a characterization of fingerprinting codes that have the best security index $s(c)$, that is, we showed that every c-intersecting code has the best security index $s(c)$. We also showed that the value $s(c)$ depends only on the number c of collusive users and the number n of users, and that $s(c) = [1/c]$ holds asymptotically. Thus a distributor can find only a set of c users such that at least one of them is surely collusive, regardless of how good code is used.

Stinson *et al.* introduced a "c-secure frameproof code" [6], and Cohen *et al.* studied a "(t, u)-separating code" [4]. One can easily know that the following (a), (b) and (c) are equivalent with each other:

(a) Γ is a c-intersecting code;
(b) Γ is a c-secure frameproof code; and
(c) Γ is a (c, c)-separating code.

Acknowledgment

We thank the anonymous referees whose comments and suggestions helped us to improve the presentation of the paper.

References

1. T. Beth, D. Jungnickel and H. Lenz, "Design Theory, Second edition," Cambridge University Press, 1999.
2. G. R. Blakley, C. Meadows, and G. B. Purdy, "Fingerprinting long forgiving messages," Proc. CRYPTO '85, Lecture Notes in Computer Science, vol. 218, pp. 180–189, Springer-Verlag, 1986.
3. D. Boneh and J. Shaw, "Collusion secure fingerprinting for digital data," IEEE Trans. on Information Theory, vol. 44, no. 5, pp. 1897–1905, 1998.
4. G. D. Cohen and H. G. Schaathun, "Upper bounds on separating codes," IEEE Trans. on Information Theory, vol. 50, no. 6, pp. 1291–1294, 2004.
5. S. Orihara, T. Mizuki, and T. Nishizeki, "New security index for digital fingerprinting and its bounds," IEICE Trans. Fundamentals, vol. E86-A, no. 5, pp. 1156–1163, 2003.
6. D. R. Stinson, T. van Trung, and R. Wei, "Secure frameproof codes, key distribution patterns, group testing algorithms and related structures," Journal of Statistical Planning and Inference, vol. 86, no. 2, pp. 595–617, 2000.
7. K. Yoshioka, J. Shikata, and T. Matsumoto, "Collusion secure codes: Systematic security definitions and their relations," IEICE Trans. Fundamentals, vol. E87-A, no. 5, pp. 1162–1171, 2004.
8. K. Yoshioka, J. Shikata, and T. Matsumoto, "On collusion security of random codes," IEICE Trans. Fundamentals, vol. E88-A, no. 1, pp. 296–304, 2005.