デジタルコンテンツの著作権保護のための符号化手法に関する研究

八　木　秀　樹　　電気通信大学 先端領域教育研究センター 特任助教

1 はじめに

情報技術の発展に伴い，大量のデジタルコンテンツがコンピュータによって処理されるようになった．そのなか，デジタルコンテンツの著作権を保護することは重要な課題となっている．ひとつの解決策として，**デジタル指紋**（digital fingerprinting）が多くの注目を集めている．デジタル指紋の技術では，ユーザ固有の ID 情報 (fingerprint) がオリジナルのコンテンツに電子透かしの情報を用いて埋め込まれる．その透かし情報を含んだコンテンツが各ユーザに配布される．

デジタル指紋技術では，複数人が結託してそれぞれの配布コンテンツに対して不正行為を行う結託攻撃（collusion attacks）に対する耐性が求められる．結託攻撃としては，シンボル選択攻撃 (interleaving attack) [1],[4],[8],[9]や平均化攻撃 (averaging attack) [4][11][12][13]などが著名である．特に平均化攻撃はマルチメディアに対するデジタル指紋に対する攻撃として有用なことが知られている．W. Trappe らは balanced incomplete ブロックデザイン (BIBD) を用いて，結託攻撃に対して耐性を持つ**結託耐性符号**を提案した[11]. Trappe らによって提案された符号は，BIBD に基づく結託耐性符号 (BIBD-based AC 符号) と呼ばれる．この符号は平均化攻撃に対してロバストであることが示されている．特に，結託者数がある定数以下であれば，全ての結託者を検出できる特徴を持つ[12]．この定数は特に**結託耐性**と呼ばれる．

本論文では，Trappe らや Yang らによって提案された BIBD-based AC 符号に対して，有限幾何の概念を用いて符号長，結託耐性を同一に保ったまま，その符号化レートを増加させる手法を提案する．結果として，デジタル指紋システムの安全性やコンテンツに与える劣みを同一に保ったまま，多くのユーザに対してサービスを提供できるシステムを実現できる．

2 システムモデル

2-1 デジタル指紋

利用者にデジタルコンテンツを配布する際，各利用者に固有の符号語を電子透かしの技術により，オリジナルのコンテンツに埋め込む．各利用者に割り当てられた符号語をユーザの **fingerprint** (**デジタル指紋**)と呼ぶ．悪意をもったユーザは不正に使用したコンテンツから自分の符号語 (fingerprint) が容易にあばかれないよう，結託して攻撃をすることが予想される．この行為を**結託攻撃** (collusion attack) と呼ぶ．結託攻撃によって不正に作成されたコンテンツが発見された場合，結託者の検出器はそのコンテンツから結託者達の符号語を推定する．この推定が成功すると，結託者を捕らえることが可能となる．

集合 $D=\{1,2,\ldots,|D|\}$をコンテンツが配信される利用者の集合とする．利用者 $j\in D$ に対する符号語を $\mathbf{b}_j = (b_{j1},b_{j2},\ldots,b_{jN})\in\{0,1\}^N$ と表す．利用者の fingerprint 透かし情報 $\mathbf{w}_j\in R^N$ は定数エネルギーをもつ N 本の直交基底$\{\mathbf{u}_i\in R^N \mid i=1,2,\ldots,N\}$と符号語 b_j から，次のように生成される．

$$w_j = \sum_{i=1}^{N}(2b_{ij}-1)\mathbf{u}_i \qquad (1)$$

次に，各利用者に配布されたコンテンツをホスト信号と見なし，得られた透かし情報をそこに埋め込む．ホスト信号（の一部）のベクトルを $\mathbf{x}\in R^N$ と表す時，利用者 $j\in D$ へ配信されるコンテンツは $\mathbf{y}_j=\mathbf{x}+\mathbf{w}_j$ のように表される．

各 fingerprint は電子透かし技術を用いて埋め込まれるため，全ての利用者は自分に配布された透かし済みコンテンツ $\mathbf{y}_j$ から，自分の fingerprint 透かし情報 $\mathbf{w}_j$ を読み取ることはできない．したがって，悪意をもった利用者達は自分達に配布された複数のコンテンツから不正なコンテンツ（すなわち不正な fingerprint）を作成して不正利用に用いる．

2-2 結託攻撃

サイズ h の結託者集合を考え，この集合を $S \subseteq D$ と表す．結託者集合 S から攻撃を受けたホスト信号は次式のように表わされる．

$$\mathbf{y} = \frac{1}{h}\sum_{j\in S}\mathbf{y}_j = \mathbf{x} + \frac{1}{h}\sum_{j\in S}\sum_{i=1}^{N}(2b_{ij}-1)\mathbf{u}_i \qquad (2)$$

この結託攻撃法は**平均化攻撃**と呼ばれ，マルチディアのデジタル指紋に対して有効な攻撃法であることが知られている[4][11][12][13]．結託者の検出器は攻撃によって不正に作成されたコンテンツ $\mathbf{y}\in R^N$ から結託者集合 S を推定する．

3　平均化攻撃に耐性を持つ AC 符号

3-1 BIBD-based AC 符号

Trappe らは BIBD-based AC 符号と呼ばれる平均化攻撃に耐性を持つ結託耐性符号を提案した[11]. 以下に，AC 符号に関する定義を与える．ここで，集合 Q(S)を S に属する全ての符号語が等しく 0-成分を持つシンボル位置の集合と定義する．

[定義 1]

ホスト信号 **x** と N 本の直交基底{$\mathbf{u}_i$}は検出器に既知であると仮定する．このとき，与えられた正定数 c>0 に対し，結託者集合 S のサイズが$|S|\le c$ のとき，Q(S)が唯一に定まる符号を **c-resilient** AC **符号**と呼ぶ．また，パラメータ c を AC 符号の**結託耐性**と呼ぶ．

□

c−resilient AC **符号**を用いれば，Q(S)を不正なコンテンツ **y** から計算することにより S に参加している全ての結託者を誤りなく検出できる．

Trappe らは各符号語の Hamming 重みが k で，異なる 2 つの符号語が高々1 つの“1-成分"を共通位置に持つ AC 符号を BIBD に基づいて構成する方法を提案した[11]．この符号は(k-1)-resilient AC 符号となることが示されている（すなわち$|S|\le$k-1 のとき，それぞれの S に対して Q(S)が唯一に定まる）．

3-2　有限幾何に基づく AC 符号

Trappe らによる BIBD-based AC 符号のサブクラスは有限幾何を用いて代数的に構成することができる．本論文では，このクラスの AC 符号に限定して議論を進める．ここで，2 種類の有限幾何について簡単に説明する．詳細については[5][10]などを参照されたい．

ある素数 p と 2 つの正整数 $m\ge 2$, $s\ge 1$ に対し，有限体 GF(p^s)上で定義される m 次元**ユークリッド幾何** EG(m, p^s)は**点・線・超平面**から構成される．EG(m, p^s)内の任意の点は GF(p^s)上の p^{ms} 個の m 次元ベクトルで表現される． $0\le r\le m$ となる r に対し，r 次元超平面 (一般に，r-**flat** と呼ばれる) は r 次元部分ベクトル空間 V とそのコセットを表し，ひとつの r-flat はちょうど p^{rs} 個の点を含む．点と線はそれぞれ 0-flat，1-flat に対応する．

与えられた次元 r<m に対し，$\mathbf{a}_0, \mathbf{a}_1, \ldots, \mathbf{a}_r$ を EG(m,p^s)内の r+1 個の線形独立な点とする．このとき，GF(p^s)上の r 個の点 $b_1, b_2, \ldots, b_r$ を動かすと，$\mathbf{a}_0 + b_1\mathbf{a}_1 + b_2\mathbf{a}_2 + \ldots + b_r\mathbf{a}_r$ で表現される p^{rs} 個の点はある r-flat を成す．2 つの r-flat のペア(F_1, F_2)は，高々1 つの(r-1)-flat を共通に持つ．このことは，F_1 と F_2 が高々$p^{(r-1)s}$ 個の点を共通に持つことを意味する．ユークリッド幾何 EG(m,p^s)内には全部で

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

個の r-flat が含まれる．

有限体 GF(p^s)上の m **次元射影幾何**を PG(m,p^s)で表すとき，PG(m,p^s)には$(p^{(m+1)s}-1)/(p^s-1)$個の点が含まれる．射影幾何 PG(m,p^s)内には全部で

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

個の r-flat が含まれる．2 つの r-flat (F_1, F_2)は高々1 つの(r-1)-flat を共通に持つ．このことは，F_1 と F_2 が高々$(p^{rs}-1)/(p^s-1)$個の点を共通に持つことを意味する．

以降特に区別する必要がない場合は，"FG(m,p^s)"という表記により，ユークリッド幾何 EG(m,p^s)もしくは，射影幾何 PG(m,p^s)を表すものとする．同様に，"$f_{FG}(r)$"という表記によって，$f_{EG}(r)$もしくは $f_{PG}(r)$を表す．

ここで， $N_0 = f_{FG}(0)$と定義し，2 元 $N_0 \times f_{FG}(r)$行列 $B_r = [b_{ij}]$を考える．この行列の各列に対し FG(m,p^s)内の各点を，各行に対しては各 r-flat を対応させる．成分 b_{ij}は点 i が r-flat j に含まれていれば $b_{ij}=1$，そうでなければ $b_{ij}=0$ となるものとする．この行列 B_rは FG(m,p^s)における r-flat の点に対する**接続行列** (incident matrix) と呼ばれる．

先に述べた Trappe らの AC 符号は 1-flat (線)の点に対する接続行列 B_1の列ベクトルを AC 符号の符号語に割り当てることで実現できる．この AC 符号を $\mathbf{B}_1$と表す．ここで，以下の 2 つの性質を利用する：

(1) FG(m, p^s)における任意の 1-flat は定数個の点を持つ

(2) 2 つの 1-flat は高々1 つの点を共通に持つ

これらの性質から，EG(m,p^s)から構成される任意の AC 符号は (p^s-1)-resilient AC 符号となることが分かる．また，PG(m,p^s)を用いると p^s-resilient AC 符号が得られることも確認できる．

ここで， AC 符号 $\mathbf{B}_1$のパラメータについて確認する．符号長は N_0となり，これは FG(m,p^s)内の点の総数に対応する．また，符号語数 (サービスを提供できる利用者数)は $f_{FG}(1)$となり，これは FG(m,p^s)内の 1-flat の総数と一致する．したがって，符号の効率性を表す**符号化レート** R_1は $R_1=(\log_2 f_{FG}(1))/N_0$ となる．結託耐性と符号長を固定した元では，符号語数を増やせば，安全性とコンテンツに対する歪みを同一に保ったままサービスを利用できる利用者を増やすことができる．

4 AC 符号に対する条件の緩和

本節では, Trappe らの BIBD-based AC 符号の条件を緩和して, より柔軟性のある AC 符号の構成を与える．本節で述べる内容は，有限幾何から構成される AC 符号に限定せず，任意の AC 符号に対して適用することができる．

いま，ある実数 v に対し，$\lceil v \rceil$によって v を下回らない最小の整数を表す．

[補題 1]

ある 2 元行列が以下の 2 つの条件を満足すると仮定する：

(1) 各列ベクトルの Hamming 重みは少なくとも k となる

(2) 任意の 2 つの列ベクトルは高々t 個の 1-成分を共通に持つ

このとき，この 2 元行列の列ベクトルから構成される AC 符号は$\lceil k/t \rceil$-resilient AC 符号となる．

□

各符号語の Hamming 重みが k でかつ t=1 のとき, 補題 1 で仮定する AC 符号は Trappe らの AC 符号と一致する．すなわち，補題 1 はより広いクラスの l-resilient AC 符号を与える．

5 有限幾何を用いた AC 符号 の改良

5-1 有限幾何に基づく AC 符号

本節では，先に説明した有限幾何を利用し，補題 1 の条件を満たす代数的な符号の構成法を与える．

有限幾何 FG(m,p^s)を用いて Trappe らの c-resilient AC 符号を構成する際, FG(m,p^s)における各点と各線(1-flat)の関係を利用する．これに対し，新しい構成では FG(m,p^s)における各点と各 r-flat ($r \geq 1$) の関係を利用する．

[補題 1]

ある整数 r ($m-1 \geq r \geq 1$)に対し，B_rを有限幾何 FG(m,p^s)内の点に対する r-flat の接続行列とし，その j 列目をベクトル $\mathbf{b}_j$によって表す. 列ベクトル $\mathbf{b}_j$を j 番目の利用者の符号語に割り当てることにより得られる符号 $\mathbf{B}_r=\{\mathbf{b}_j\}$を r **次の FG-AC 符号**と呼ぶ．特に区別する必要がある場合は， ユークリッド幾何もしくは射影幾何から構成される AC 符号 $\mathbf{B}_r$を r **次の EG-AC 符号**もしくは r **次の PG-AC 符号**と呼ぶ.

□

ここで r 次の FG-AC 符号について以下の定理を示す.

[定理 1]

任意の EG(m,p^s)に対し，r 次の EG-AC 符号 $\mathbf{B}_r$は(p^s-1)-resilient AC 符号となる．また，任意の PG(m,p^s)に対し，r 次の PG-AC 符号 $\mathbf{B}_r$は p^s-resilient AC 符号となる．

(証明)

前節で述べた通り, FG(m,p^s)内の 2 つの r-flat F_1と F_2は高々1 つの (r-1)-flat を共通に持つ. したがって, EG-AC 符号に対しては，$k=p^{rs}$, $t=p^{(r-1)s}$ となり，PG-AC 符号に対しては $k=(p^{(r+1)s}-1)/(p^s-1)$, $t=(p^{rs}-1)/(p^s-1)$となる．補題 1 を用いると定理 1 の内容が示される．

□

定理 1 から r 次の FG-AC 符号 $\mathbf{B}_r$ の結託耐性は超平面の次数 r には依らない定数となることが分かる．

ここで，r 次の FG-AC 符号のパラメータについて述べる．与えられた FG(m,p^s)に対し，m-1 通りの r 次の FG-AC 符号 $\mathbf{B}_r$ (r=1,2,..., m-1) は同じ結託耐性を持つ．また，r-flat の点に対する接続行列の性質から，FG-AC 符号 $\mathbf{B}_r$ の符号長はどれも等しく N_0 となり，符号語数は $f_{FG}(r)$となる．AC 符号 $\mathbf{B}_r$ の符号化レートを R_r で表すとき，$R_r=(\log_2 f_{FG}(r))/N_0$ となる．すなわち，$\mathbf{B}_r$ のパラメータのうち，次数 r に依存するのは符号語数のみである．この事実により，与えられた FG(m,p^s)に対し，符号化レートを最大にするという意味において最良の FG-AC 符号 $\mathbf{B}_{r*}$が得られる．以上の議論から，この次数 r*を FG(m, p^s)に対する FG-AC 符号の**最良次数** (best order) と呼ぶことにする．

ここで最良次数に関して，以下の定理を示す．

[定理 2]

与えられた EG(m,p^s)に対し，FG-AC 符号の最良次数は以下の通り．

(1) m≤3 のとき，r* = 1 となる．

(2) m = 4 のとき，EG に対して r* = 2，PG に対して r* = 1,2 となる．

(3) m > 4 のとき，r*≥2 となる．

（証明）

r 次の FG-AC 符号 $\mathbf{B}_r$ の符号語数は $f_{EG}(r)$もしくは $f_{PG}(r)$で与えられるため，これらの関数の性質による．□

定理 2 より，m>3 の場合には，Trappe らの AC 符号 $\mathbf{B}_1$ に比べて大きい符号化レートを持つ FG-AC 符号が必ず存在することが分かる．

5-2 最良次数を持つ FG-AC 符号の例

ある EG(m,p^s)に対し，最良次数の EG-AC 符号 $\mathbf{B}_{r*}$の例を表 1 に示す．表 1 において，m>3 のユークリッド幾何 EG(m,p^s)に対し，結託耐性(p^s-1)が 2 以上になる符号を符号長の小さい順に掲載する．この表において，"$\log_2 f_{EG}(1)$"の列と"$\log_2 f_{EG}(r^*)$"の列はそれぞれ，$\mathbf{B}_1$ (Trappe らの AC 符号)と最良次数を持つ EG-AC 符号 $\mathbf{B}_{r*}$ の符号語数を底 2 の対数で表す．

図 1.　最良次数を持つ EG-AC 符号の例

Examples of EG-AC Codes with Maximal Order

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

符号数を表す関数 $f_{EG}(r)$の性質から，EG(m,p^s)の次元数 m が大きくなると符号語数は $\mathbf{B}_1$ のそれに比べてはるかに大きくなることが確かめられる．特に，m≥5 のとき $\mathbf{B}_{r*}$の符号語数は $\mathbf{B}_1$ の符号語数の 2^2 倍以上となり，EG(7,3)に対しては $2^{6.5}$ 倍以上，EG(8,3)に対しては 2^{10} 倍以上となることが分かる．

また，与えられた PG(m,p^s)に対して最良次数を持つ PG-AC 符号 $\mathbf{B}_{r*}$の例を表 2 に示す．最良次数を持つ PG-AC 符号の振る舞いは EG-AC 符号とほぼ同様なことが分かるが，m が偶数のときには最良次数が 2 つ存在する点に注意が必要である．

図 2.　最良次数を持つ PG-AC 符号の例

Examples of PG-AC Codes with Maximal Order

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

5 まとめと今後との課題

本論文では，Trappe らによって提案された AC 符号の 1 クラスに対して，有限幾何に基づいて結託耐性を同一に保ったまま，その符号化レートを増加させる新しい手法を提案した．与えられた有限幾何において，最大の符号語数を持つ AC 符号の例を紹介した．得られた AC 符号は有限幾何 FG(m,p^s)の次元 m が大きくなるほど，従来の AC 符号に対してその符号語数は増大する．この手法と同様なアプローチから，効率的な AC 符号を疑似巡回疎行列（LD）行列[2][3][6]から構成する手法も併せて提案している．この手法では，先の手法と異なり必ずしも結託耐性を同一に保つことができるとは限らないが，そのためのパラメータに関する条件を導出している．結果として，結託耐性とオリジナルのデジタルコンテンツに対する歪みを同一に保ったまま，より多くの利用者に対してコンテンツ配信のサービスを提供するシステムが実現できる．

本論文の提案手法によって得られた AC 符号は比較的大きい符号長になる．一方，符号長が大きくなると fingerprint の埋め込みによるコンテンツの歪みの影響も大きくなるため，符号長は短く抑えることが求められる．本研究では，有限幾何や疑似巡回 LD 行列の性質を利用して，もとの AC 符号の符号長を短くできる条件，それを実現する一つの手法も併せて提案している．

本論文では，不正利用されたコンテンツには雑音がないと仮定して議論を進めた．実際の応用を想定した場合，多くは画像処理や加法的雑音の影響で，必ずしも不正利用されたコンテンツが誤り無く検出できるとは限らない. そのような状況下で, 提案した AC 符号の性能を評価することも今後の課題として挙げられる．

【参考文献】

[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] 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.

〈発 表 資 料〉

題　　名	掲載誌・学会名等	発表年月
Improved collusion-secure codes for digital fingerprinting based on finite geometries	Proc. of 2007 IEEE Int. Conf. on System, Man, Cybernetics	2007 年 10 月
Short concatenated fingerprinting codes for multimedia data	Proc. of 45th Annual Allerton Conf. on Commun., Control, and Computing	2007 年 9 月
Shortening methods of collusion-secure codes for digital fingerprinting	Proc. of 2007 Hawaii and SITA Joint Conference on Information Theory	2007 年 5 月