様　式　Ｃ－１９、Ｆ－１９、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

平成 ２６ 年　 ６ 月　 ６ 日現在

機関番号：１２１０２
研究種目：基盤研究(C)
研究期間：2011～2013
課題番号：２３５０００４

研究課題名（和文）不正者全員を特定可能な電子指紋符号の構成法とその性能解析

研究課題名（英文）Construction and Performance Analysis of Digital Fingerprinting Code with Identifiability of All Malicious Users

研究代表者
古賀　弘樹（Koga, Hiroki）

筑波大学・システム情報系・准教授

研究者番号：２０２７２３８８

交付決定額（研究期間全体）：（直接経費）　3,100,000円、（間接経費）　930,000円

研究成果の概要（和文）：電子指紋符号は，ライセンスのあるディジタルコンテンツの不正流出を防ぐためにコンテンツに埋め込まれる符号である．電子指紋符号では，通常，悪意のある複数のユーザが結託して海賊版のコンテンツを生成しても，その海賊版の生成に関わったユーザの一部または全部を特定できることが求められる．
本研究では，情報理論的な電子指紋符号のモデルにおいて，ある仮定のもとで，漸近的に1に近い確率で不正者全員を特定できるための最大のユーザのレート(電子指紋符号の容量)を導出する．また，特別な場合に対しては確率1で不正者善意を特定するための容量（ゼロエラー容量）の上界と下界が導出できることを示す．

研究成果の概要（英文）：Digital fingerprinting codes are embedded in licensed digital contents for preventing illegal distribution by malicious users. Digital fingerprinting codes are usually required to have the ability to specify a part (or all) of a collusion of malicious users who generate the pirated copy.
In this study we consider an information-theoretic model of the digital fingerprinting code and characterize the maximum rate of the users (the capacity of the digital fingerprinting code) such that all the malicious users in a collusion can be specified with probality close to one under a certain assumption. We also discuss the zero-error capacity of the digital fingerprinting code for the case where all the malicious users in a collusion are specified without error under certain attack models.

研究分野：情報理論

科研費の分科・細目：情報学基礎・数理情報学

キーワード：電子指紋符号　コンテンツ保護　電子透かし　不正者特定　符号化定理　結託耐性符号

１．研究開始当初の背景

電子指紋符号(digital fingerprinting code)は，ライセンスつきのディジタルコンテンツの不正配信を抑止するための技術であり，情報理論的な解析の枠組みはBonehとShawにより1998年に提案された．電子指紋符号では，ライセンスつきの情報を配信する配信者とその情報を有償で購入する多数のユーザからなる次の状況を想定する．

想定１：配信者は，ユーザ毎に異なる電子指紋符号が埋め込まれたディジタルコンテンツを，各ユーザに配信する．各ユーザは，配信されたコンテンツを受信する．

想定２：悪意のあるユーザの集合(以下Ｋと書く)は結託して，各メンバーがもつディジタルコンテンツから新たなコンテンツを偽造し，不正に配信する．この際，Ｋは，マーキング仮定という制約のもとで，Ｋの各メンバーができるだけ特定されないように，コンテンツを偽造する．

想定３：配信者は，不正に配信されたコンテンツを発見した場合には，そのコンテンツから上書きされた符号語（不正符号語）を抽出し，コンテンツの偽造に関わったユーザの一部または全員を特定する．

従来，情報理論的な立場または離散数学的な立場から，電子指紋符号の構成に関する多くの研究がなされてきた．しかしながら，それらの研究のほとんどは，配信者が不正なコンテンツを発見したときに，コンテンツの偽造に関わったユーザの少なくとも１人を特定する，という規範を用いているものがほとんどであった．例えばBonehとShawは，提案した電子指紋符号により，不正に関わったユーザの少なくとも１人が，１に近い確率で特定できることを示している．

最近になって，研究代表者は，有限射影平面として知られる離散構造に基づく電子指紋符号を用いると，多くの場合に不正に関わったユーザ全員が特定できることを示した．特に，結託グループＢが２名のユーザから構成されるときは，不正なコンテンツから抽出される上書きされた符号語のパターンにより，どのような特定誤りが生じるか完全に記述できる．この事実は，電子指紋符号の種類によっては，電子指紋符号性能のより詳細な解析が可能であることを示唆している．

２．研究の目的

本研究では，上記の問題設定のもとで，電子指紋符号の容量の情報理論的な解析を行うことを目的とする．例えばAnthapadmanabhamらの論文では，電子指紋符号の容量は，結託グループのメンバーの少なくとも一人が１に任意に近い確率で特定される場合のユーザ数の上限に基づき定義されているが，この定義は，結託グループのメンバー全員が１に任意に近い確率で特定できる場合にも拡張できる．この場合の電子指紋符号の容量に関する研究はこれまでになく，新しい知見が得られることが期待される．

また，研究代表者は，結託グループがAND攻撃として知られる攻撃を行う場合を考察し，結託グループのメンバー全員を誤りなく確率で特定できるとした場合の電子指紋符号の容量（ゼロエラー容量）の下界を，情報理論における同定符号の問題と関連付けて導出できることをすでに示している．この結果を精密化して，よりタイトな容量の上界と下界を評価することは，電子指紋符号の新しい情報理論的な側面を明らかするという意味をもつ．

３．研究の方法

本研究では，次の設定のもとで，電子指紋符号の容量に対する理論的考察を行った．

(1) 符号器と復号器

$n \geqq 1$を任意の整数とし，ユーザ集合を$U_n = \{1,2,\cdots,M_n\}$と表す．M_nはユーザの総数である．各ユーザは，コンテンツ供給者から，あるディジタルコンテンツのライセンスを購入する．コンテンツ供給者からユーザｊに配布されるディジタルコンテンツには，長さｎの符号語x_jが埋め込まれているとする．以下では，ユーザｊに対して符号語x_jを対応させる写像を符号器と呼ぶ．

ユーザ集合U_nの中には，自分たちに配布されたディジタルコンテンツを持ち寄り，結託して不正な海賊版のディジタルコンテンツを生成する不正者グループ$K=\{k_1,k_2,\cdots,k_t\}$が存在する．海賊版の中から抽出される，不正に上書きされた長さｎの符号語をｙとする．不正者グループは，後述の３つの制約のもとで，自分たちに配信されたコンテンツを加工して，不正符号語ｙが埋め込まれたディジタルコンテンツを生成する．

コンテンツ供給者は，海賊版ディジタルコンテンツを発見したとき，海賊版の生成に寄与した不正者グループＫのすべてのメンバーを特定することを試みる．我々は，コンテンツ供給者が，正規版と海賊版を問わず，ディジタルコンテンツに埋め込まれた(不正)符号語を誤りなく抽出できると仮定する．また，コンテンツ供給者は不正者グループの人数の上界ｕは知っているものと仮定する．この状況のもとで，復号器は不正符号語ｙを入力とし，不正者グループの推定値Ｋ’を出力する写像として定義される．

(2) 不正者に対する仮定

我々は，不正者グループ $K=\{k_1,k_2,\cdots,k_t\}$が，符号語 x_{k_1}，x_{k_2}，…，x_{k_t} から不正符号語 y を生成すると考える．不正者グループが生成する符号語が，次の形に制限される状況を想定する．

仮定１(定常独立性)：不正者グループは，不正符号語 y の第 i 成分を，符号語 x_{k_1}，x_{k_2}，…，x_{k_t}それぞれの第 i 成分から，他の成分とは独立に，同一の条件つき確率分布 $P_{Y|X_1\cdots X_t}$に従って生成する．

仮定２(マーキング仮定)：不正者グループは，検出不能な記号に対して書き換えを行わない．すなわち，不正者グループが選ぶことができるすべての条件つき確率分布 $P_{Y|X_1\cdots X_t}$は，すべての記号 x に対して $P_{Y|X_1\cdots X_t}(x|x,x,\cdots,x)=1$ を満たす．

仮定３(不正者グループの対称性)：不正者グループが選ぶことができる条件つき確率分布 $P_{Y|X_1\cdots X_t}$は，条件部分を任意に置換しても値は変わらない．

仮定１～３はいずれも既存の論文でも使われている仮定である．特に，仮定３のもとでは，不正者グループの特定のメンバーだけが特定されやすいということはなく，どのメンバーも同等になる．

(3) 電子指紋符号の容量

本稿で考えるのは，M_n 人のユーザに配信されるディジタルコンテンツに埋め込まれたすべての符号語の成分が，ある一定の確率分布 P_Xに従って定常独立に生成される場合である．t 人の不正者グループ K が条件つき確率分布 $P_{Y|X_1\cdots X_t}$を用いて不正符号語を生成するとき，復号器が不正者グループの特定に失敗する確率を $P_{n,t}$と書く．符号語の構成法から，この確率は K の要素数 t だけに依存する．

不正者グループ K は，与えられた符号語生成の確率分布 P_Xのもとで，最もグループ全体が特定されにくい $P_{Y|X_1\cdots X_t}$を選び，不正符号語 y を生成する．他方，コンテンツ供給者は，不正符号語 y からできるだけ高い確率で不正者グループを検出できるように P_X を選ぶ．したがって，不正者グループのメンバーが t 人のときに，コンテンツ供給者が不正者グループの特定に失敗する確率は $P_{n,t}$ をまず $P_{Y|X_1\cdots X_t}$で最大化し，さらにそれを P_Xで最小化したものとなる．本稿では不正者グループ K の人数 t は未知であるが，t の上界 u は与えられている状況を考えるので，不正者グループ K の特定失敗確率を，上記の確率をさらに $1\leqq t\leqq u$ に関して最大化した P_nとして定義する．

我々は，$n\to\infty$のもとで $P_n \to 0$ を満たすという制約のもとで，どのくらい多くのユーザに対して電子指紋符号を適用できるのかに興味がある．具体的には次の定義で与えられる電子指紋符号の容量を求めることが目標になる．

定義１（電子指紋符号の容量）
u を不正者グループの人数の上界とする．レート R が達成可能であるとは，ある確率分布 P_X と符号器・復号器の組の列が存在して，$M_n = 2^{nR}$かつ $P_n \to 0(n \to\infty)$を満たすことをいう．ここに 2^{nR}は 2 の nR 乗を表す．達成可能なレート R の上限を電子指紋符号の容量と呼び，C_u と表す．

この電子指紋符号の容量を特徴づけるためには，相互情報量の min-max 問題を考える必要がある．具体的には，確率分布 P_Xに従う独立な確率変数 $X_1,X_2,\cdots,X_t$ と，条件つき確率 $P_{Y|X_1\cdots X_t}$により定義される Y との間の相互情報量 $I(X_1\cdots X_t;Y)$を考え，$1/t\ I(X_1\cdots X_t;Y)$を $P_{Y|X_1\cdots X_t}$に関して最小化し，P_X に関して最大化した値を D_t と書く．$X_1,X_2,\cdots,X_t,Y$ がすべて２値のときには，$D_t = 2^{-(t-1)}/t$ と具体的に求まる．

(4) 電子指紋符号のゼロエラー容量

３(3)項で述べた電子指紋符号の容量は，不正者グループ全員の特定を目標とするが，漸近的にゼロに収束する微小な特定誤り確率を許した．本節では，電子指紋符号の問題において特定誤り確率を一切許さないゼロエラー容量を定義する．我々が特に興味をもつのは，不正符号語 y が不正者グループのメンバーの符号語 x_{k_1}，x_{k_2}，…，x_{k_t}の成分ごとの AND または OR を用いて生成される場合である．

定義２(u-resilient AND Anti-Collusion Code)
$u\geqq 2$ を不正者グループの人数の上界とする．符号$\{x_1, x_2, \cdots, x_{M_n}\}$は，$\{1,2,\cdots,M_n\}$以下の相異なる要素数 u 以下の任意の部分集合 I,J に対して，添字が I に属する符号語の AND と，J に属する符号語の AND が異なるとき，u-resilient (n,M_n) AND Anti-Collusion Code (u-resilient (n,M_n) AND-ACC)と呼ぶ．符号語の OR に対して同じ性質が成立する場合は u-resilient (n,M_n) OR-ACC と呼ぶ．

符号が u-resilient (n,M_n) AND-ACC であれば，u 人以下の任意の結託に対して不正符号語 y が異なるので，総当り等の手法を用いることで不正者全員が誤りなく特定できることになる．u-resilient AND-ACC のゼロエラー容量を次で定義する．

定義３（u-resilient AND-ACC のゼロエラー容量）
u≧2 を不正者グループの人数の上界とする．レート R は，u-resilient (n,M_n) AND-ACC の列が存在して 1/n log M_n ≧ R －γを満たすとき達成可能であるという．ここにγ>0 は任意に小さい定数である．達成可能なレート R の上限を u-resilient AND-ACC のゼロエラー容量と呼び，C_u*と表す．

ド・モルガンの法則より容易にわかることであるが，符号が u-resilient (n,M_n) AND-ACC であれば，各符号語のビットを反転することで u-resilient (n,M_n) OR-ACC が得られる．ゆえに u-resilient AND-ACC と u-resilient OR-ACC のゼロエラー容量は等しい．

４．研究成果

(1) 電子指紋符号の容量

まず最も簡単な u=2 の場合について，電子指紋符号の容量を考察する．我々は次の結果を得た(文献[4])．

定理１：D_2 の定義における最大値を達成する P_X が一様分布であれば，C_2 = D_2 が成り立つ．

定理１を示すためには順定理と逆定理を示す必要がある．順定理（D_2 を漸近的に達成する復号器の具体的な構成）の証明では，情報理論でよく知られたタイプ理論を用いる．逆定理（どんな復号器を用いても C_2 が D_2 を超えないこと）の証明では，通信路符号化定理の逆定理の証明で用いる手法と同様の手法を用いる．
u ≧2 の一般の場合については次の定理が成り立つ．

定理２：u ≧2 を任意定数とする．もし D_t の最大値を達成する P_X が一様分布であり，かつ D_t が t に関して単調非増加であれば，C_t = D_t が成り立つ．

定理２より，符号語が２値の場合には，定理の仮定が成り立ち C_t = 2^{-t+1}/t であることがわかる．

(2) AND-ACC のゼロエラー容量

C_u*については上界と下界を得た．下界は次の定理で与えられる．

定理３：u ≧2 を任意定数とすると，C_u*≧ max α/(2^{αu}+1)が成り立つ．ここに max はα≧1/u log_2 5 なるαに関して取る．

定理３は，情報理論において同定符号として知られる符号の存在証明で用いる補題と，AND-ACC を初めて定義した Trappe らの結果を用いて証明される．定理３において特にα＝２とすると C_u*≧2/(4^k+1)が得られ，この定理３は C_u*について著者が従来得ていた結果 C_u*≧1/(4^k+1)よりも大きな下界を与えることに成功している．
C_u*の上界は次の定理で与えられる．

定理４：u ≧2 を任意定数とすると，C_u*≦ 1/u が成り立つ．

定理４は組合せ論的な議論で得られる．すなわち，M_n 個の符号語から u 個以下の要素を選び成分ごとの AND を得られる符号語は，u-resilient (n,M_n) AND-ACC の定義よりすべて異ならなければならない．この符号語の組み合わせは M_n 個の中から u 個を選ぶ組み合わせ未満である．他方，不正符号語は長さ n であるから，相異なるものは高々2^n 個にすぎない．この大小関係を評価していくと，C_u*≦1/u が得られる．
定理３の C_u*の下界は u とともに指数関数オーダで急速に減少し，定理４の C_u*の上界は 1/u でゆっくり減少するので，特に大きい u に対しては定理３と定理４の上界と下界のギャップは大きくなる．現在のところ，このゼロエラー容量に関するギャップを埋めるだけの考察はできていない．
しかしながら，例えば特定誤りを許す場合の電子指紋符号の容量において，P_{Y|X_1…X_t}に関する最小化を行わず AND または OR に限定したときには，別の手法を用いて電子指紋符号の容量を求めることができる．具体的には 1/u ≦ C_u ≦ 2/u が導かれる．すなわち，微小な特定誤りを許せば 1/u は上界としての意味をもつ．これらの結果は現在国際会議 2014 IEEE Information Theory Workshop に投稿中である．

(3) その他の成果

本研究ではまた，電子指紋符号の問題を含む情報セキュリティ分野でよく用いられる smooth Renyi entropy の評価も行い，smooth Renyi entropy の定義に含まれる最小値を達成する分布が Majorization の理論を用いて導出できることを示した（文献[1]）．
一般に，α次の Renyi entropy は確率分布 P=(P_1,P_2,…,P_m)に対して
H_α(P) = 1/(α-1)log (∑ P_i^α)
と定義される．α=1,0,∞に対してはそれぞれ極限値を用いて定義される．これに対して smooth Renyi entropy は
K_α(P) = min 1/(α-1)log (∑ Q_i^α)
として定義される．ここに min は劣確率分布 Q=(Q_1,Q_2,…,Q_m)に関して取り，Q は与えられた正数εに対して Q は確率が 1-ε以上ですべての i に対して 0 ≦Q_i ≦P_i を満たす範囲を動く．

いま，P が $P_1 \geqq P_2 \geqq \cdots \geqq P_m$ を満たすとし，$1 \leqq J \leqq m$ を $\sum P_i \geqq 1-\varepsilon$ を満たす最小の数として選ぶ．そして劣確率分布 Q^* を

$$Q^*_i = P_i \quad (i=1,2,\cdots,J-1)$$
$$Q^*_J = 1-\varepsilon-(P_1+P_2+\cdots+P_{J-1})$$
$$Q^*_i = 0 \quad (i=J+1,\cdots,m)$$

と定義する．また，J を後で定めるβを用いて $J = \max\{1 \leqq j \leqq m: P_i \geqq \beta\}$ と定め，Q^+ を

$$Q^+_i = \beta \quad (i=1,2,\cdots,J)$$
$$Q^+_i = P_i \quad (i=J+1,\cdots,m)$$

と定める．ここにβは $\sum Q^+_i = 1-\varepsilon$ を満たすようにとる．Q^* も Q^+ も，P とεが与えられれば一意的に定まることに注意せよ．

このとき次の定理が成り立つ．

定理５：任意定数ε$(0 \leqq \varepsilon < 1)$に対して，次数αの smooth Renyi Entropy は次のように書ける．

(i) $0 < \alpha < 1$ のとき

$$K_\alpha(P) = 1/(\alpha-1)\log\left(\sum Q^*_i{}^\alpha\right)$$

(ii) $1 < \alpha < \infty$ のとき

$$K_\alpha(P) = 1/(\alpha-1)\log\left(\sum Q^+_i{}^\alpha\right)$$

$0 < \alpha < 1$ のときと $1 < \alpha < \infty$ のときで，x のα乗の凹凸が変化し，これに伴って，和 $\sum Q_i{}^\alpha$ は Schur concave，Schur convex になる．これらの性質を用いて，smooth Renyi Entropy の定義に含まれる最小値を達成する劣確率分布がそれぞれ Q^*，Q^+ であることが示される．

定理５はα=0，∞のときに知られている既存の結果を特殊ケースとして含む．実際，定理５においてα→0，α→∞の極限を考えると既存の結果が得られる．

また，定理５を用いると，定常無記憶情報源に対する smooth Renyi entropy の公式を導くことができる．実際，任意の $0 \leqq \alpha \leqq \infty$ と $0 \leqq \varepsilon < 1$ に対して，1/n 倍した smooth Renyi Entropy は $n \to \infty$ の極限で情報源のエントロピーに一致することを容易に示すことができる．

５．主な発表論文等

（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計３件）

[1] H. Koga，“Characterization of the Smooth Renyi Entropy Using Majorization，” Proc. 2013 IEEE Information Theory Workshop，pp.604–608，セビーリャ（スペイン），査読有，2013 年 9 月 13 日．

[2] H. Koga and K. Koyano，“On the Role of Mutual Information between the Shares in a Robust Secret Sharing Scheme，” Proc. 2012 International Symposium on Information Theory，pp. 260–265，ホノルル(アメリカ合衆国)，査読有，2012 年 10 月 29 日．

〔学会発表〕（計７件）

[1] 金井紘平，古賀弘樹，“定常性を仮定しない確率的な電子指紋符号の性能に関する一考察，”電子情報通信学会情報理論研究会，信学技報 IT2013-93，名古屋大学，愛知県，2014 年 3 月 11 日．

[2] 古賀弘樹，児矢野和也，“不正者の対称性のもとでの確率的な電子指紋符号に対する符号化定理，”電子情報通信学会情報理論研究会，信学技報 IT2013-4，あわら温泉まつや千千，福井県，2013 年 5 月 24 日．

[3] 金井紘平，古賀弘樹，“確率的攻撃モデルにおける Boneh-Shaw 符号の不正者追跡アルゴリズム，” 電子情報通信学会情報理論研究会，信学技報 IT2012-10，豊田工業大学，愛知県，2012 年 7 月 19 日．

[4] 児矢野和也，古賀弘樹，“確率的電子指紋符号の性能に関する情報理論的考察，”第３４回情報理論とその応用シンポジウム，pp. 560–565，鶯宿温泉ホテル森の風，岩手県，2011 年 12 月 2 日．

６．研究組織

(1)研究代表者

古賀　弘樹（Koga，Hiroki）

筑波大学・システム情報系・准教授

研究者番号：20272388