照合タグを用いた秘匿共通集合計算プロトコルとその応用

千田 浩司 †　　五十嵐 大 †　　高橋 克巳 †

†NTT 情報流通プラットフォーム研究所
180-8585 東京都武蔵野市緑町 3-9-11

あらまし　集合 X, Y を互いに開示する事無く共通集合 $X \cap Y$ を得る秘匿共通集合計算プロトコル，及びその応用について考察する．既存研究として，集合の各要素を推定困難な「照合タグ」に効率良く変換し，それらを各要素の代替として任意の手法で共通集合を求める秘匿共通集合計算プロトコルが知られているが，照合タグは個別に再利用できるため，同一または類似の集合に対して繰り返し共通集合を求める場合において特に有効である．本稿では照合タグの応用として，キャンセラブルバイオメトリクスにおける効率的な照合方式を提案する．

Tag-Based Secure Set-Intersection Protocol and Its Application

Koji Chida†　　Dai Ikarashi†　　Katsumi Takahashi†

†NTT Information Sharing Platform Laboratories
3-9-11 Midori-Cho Musashino-Shi Tokyo 180-8585 Japan

Abstract　We propose an application of *secure set-intersection protocols* which output the intersection $X \cap Y$ of two sets X and Y without knowing information about the other parties' sets except the result. Some of existing protocols efficiently generate *matching tag*, which is a substitute of each element of sets to hide the element itself, so that each party can evaluate the intersection from the tags corresponding to all elements of sets in a usual way. Due to the reusability of tags, our application enhances template search in *cancelable biometrics*.

1　はじめに

IT サービスの多様化に伴い，利便性とプライバシーの両立は情報セキュリティ分野における重要な研究課題となっている．特に最近では，データベースに蓄積された個人の情報を分析素材として利活用する動きが各所で見られるようになり，例えば我が国では，総務省，経済産業省，及び厚生労働省の連携による “健康情報活用基盤実証事業” [1] や，経済産業省による “情報大航海プロジェクト” [2] において，個人の情報の利活用を促進するためのプライバシー保護技術が検討課題として挙げられている．

このような動きに関連して，暗号技術を用いる事で，集合 X, Y (例えば各組織が保管している顧客リスト) を互いに開示する事無く共通集合 $X \cap Y$ (例えば共通顧客のリスト) を得る**秘匿共通集合計算プロトコル** (Secure Set-Intersection Protocol)，及びその単純な変形として $X \cap Y$ の要素数 $|X \cap Y|$ (例えば共通顧客数) を得る**秘匿共通要素数計算プロトコル** (Secure Set-Intersection Cardinality Protocol) に関する研究が活発に行われている [4, 9, 13, 14, 26, 11, 24, 16, 10, 18, 19, 23, 27, 28, 21, 29]. そしてこれらの研究は特に，機能や安全性を制限して効率性を重視するアプローチと，機能拡張や安全性を重視するアプローチに大別できる．

共通集合やその要素数の計算は，リストの突き合わせといった直接的な利用のほか，データマイニングの要素演算としても広く用いられるため，秘匿共通集合計算プロトコルや秘匿共通要素数計算プロトコルは，[1] や [2] 等に代表さ

れる，患者データや顧客データといった開示が望ましくない情報をデータマイニング等の分析素材として利活用する際に有用であるといえる．

本稿では既存の秘匿共通集合計算プロトコルや秘匿共通要素数計算プロトコルについて，**照合タグ**を用いた手法に着目する．具体的には，集合の各要素を推定困難な一意のデータに効率良く変換し，それらを各要素の代替として任意の手法で共通集合やその要素数を求める効率重視のプロトコルである．本稿ではこの推定困難なデータを照合タグと呼ぶ．

以降，2 節で既存のいくつかの秘匿共通集合計算プロトコルや秘匿共通要素数計算プロトコルについて述べた後，3 節で照合タグの形式的な定義を与え，既存の手法が本稿で定義した照合タグの性質を満たしているか考察する．なお照合タグは個別に再利用できるため，同一または類似の集合に対して繰り返し共通集合やその要素数を求める場合において特に有効である．そこで 4 節では照合タグの応用として，キャンセラブルバイオメトリクスにおける効率的な照合方式を提案する．

2 関連研究

前節で述べたように，集合を直接開示せずに集合計算 (Set Operation) の結果のみを得る**セキュア集合計算**に関する様々な研究が行われている．代表的には二つの集合を直接開示する事無く共通集合やその要素数のみを求める秘匿共通集合計算プロトコル及び秘匿共通要素数計算プロトコルが挙げられるが，その他にも共通集合の要素数の閾値判定 [9]，和集合の要素数計算 [4] やその閾値判定 [14]，そして集合が互いに疎かどうかの判定 [13, 11, 29] といったセキュア集合計算が提案されている．また前節では二つの集合に特化して例示したが，既存方式の多くは三つ以上の集合におけるセキュア集合計算についても検討がなされている．これについては次節で考察を与える．

[4, 26] では一方向性を持つ可換の関数 (Commutative Function) $f_{\mathcal{C}}, f_{\mathcal{S}}$ を用いて，集合の各要素を推定困難な一意のデータに効率良く変換し，それらを各要素の代替として任意の手法で共通集合やその要素数を求める事のできるセキュア集合計算が提案されている．すなわち集合 X を保持する $\mathcal{C}$，及び Y を保持する $\mathcal{S}$ が計算可能な関数をそれぞれ $f_{\mathcal{C}}$, $f_{\mathcal{S}}$ としたとき，先ず $\mathcal{C}, \mathcal{S}$ はそれぞれ $f_{\mathcal{C}}(x_i)$ $(x_i \in X)$, $f_{\mathcal{S}}(y_j)$ $(y_j \in Y)$ を計算して互いに送信し，次にそれぞれ $f_{\mathcal{C}}(f_{\mathcal{S}}(y_j))$, $f_{\mathcal{S}}(f_{\mathcal{C}}(x_i))$ を計算して，これらを照合する．すると $f_{\mathcal{C}}, f_{\mathcal{S}}$ の可換性より $f_{\mathcal{C}} \circ f_{\mathcal{S}} = f_{\mathcal{S}} \circ f_{\mathcal{C}}$ $(= T)$ であるから，$x_i = y_j$ ならば $T(x_i) = T(y_j)$ が成り立つ．ただし逆は $f_{\mathcal{C}}, f_{\mathcal{S}}$ の取り方に依存する．Agrawal らは具体的に次のようなプロトコル (AES03 プロトコル) を提案している [4].

g を生成元とした位数 q の巡回群を $G = \langle g \rangle$ とし，$\mathcal{H}$ を G へ写すハッシュ関数とする．$u, v \in_R \{1, 2, \ldots, q-1\}$ とする．このとき集合 $X, Y \subseteq S$ について $T : S \to \mathcal{H}(S)^{uv}$ とし，以下の手順により $T(s)$ $(s \in S)$ を生成する．

1. ある集合 S について $\mathcal{C}$, $\mathcal{S}$ はそれぞれ $X = \{x_1, \ldots, x_{k_{\mathcal{C}}}\} \subseteq S$, $Y = \{y_1, \ldots, y_{k_{\mathcal{S}}}\} \subseteq S$ を入力する．

2-$\mathcal{C}$. $\mathcal{C}$ は $\mathcal{H}(x_i)^u$ を計算して $\mathcal{S}$ に送る．

2-$\mathcal{S}$. $\mathcal{S}$ は $\mathcal{H}(y_i)^v$ を計算して $\mathcal{C}$ に送る．

3. $\mathcal{C}, \mathcal{S}$ はそれぞれ $T(y_i) = (\mathcal{H}(y_i)^v)^u$, $T(x_i) = (\mathcal{H}(x_i)^u)^v$ を計算する．

上記の AES03 プロトコルについて，$T(x_i)$, $T(y_j)$ の順序を記憶して照合すれば秘匿共通集合計算プロトコル，順序を撹乱して照合すれば秘匿共通要素数計算プロトコルとなる．

AES03 プロトコルはランダムオラクルモデル [5] の下で $\mathcal{C}$, $\mathcal{S}$ が Semi-Honest である事を仮定している．また計算量が $O(|S|)$ のソートアルゴリズムがいくつか知られているため [15, 364 ページ]，$T(x_i)$, $T(y_j)$ を直接参照した共通集合計算の計算量は $O(k)$ とできる ($k = \max(k_{\mathcal{C}}, k_{\mathcal{S}})$)．通信量は $O(k|G|)$ ビットとなる．

野島らは $\mathcal{C}$ のみが $X \bigcap Y$ を得る事を前提として，[4, 26] とは異なる構成方法を示した [19]．基本的なアイデアは，$\mathcal{S}$ のみ計算可能な一方向性を持つ関数 $f_{\mathcal{S}}$ を用いて，$\mathcal{C}$ は $\mathcal{S}$ に $x_i \in X$ を知られる事無く $f_{\mathcal{S}}(x_i)$ を得るものであり，具体的にブラインド署名 [7, 8] を用いて構成している．その構成方法について簡単に説明する．$f_{\mathcal{S}}$ を $\mathcal{S}$ の署名生成関数とする．このとき $\mathcal{C}$ は $\mathcal{S}$ とブラインド署名のプロトコルを実行して x_i に対する一意の署名 $f_{\mathcal{S}}(x_i)$ を得る．そして $\mathcal{C}$ はハッシュ関数 $\mathcal{F}$ を用いて $T(x_i) = \mathcal{F}(x_i, f_{\mathcal{S}}(x_i))$ を計算する．一方 $\mathcal{S}$ は y_j の署名 $f_{\mathcal{S}}(y_j)$ を生

成し $T(y_j)=\mathcal{F}(y_j, f_{\mathcal{S}}(y_j))$ を計算して $\mathcal{C}$ に送る．最終的に $\mathcal{C}$ は $T(x_i), T(y_j)$ を照合する．

Freedman らは加法準同型暗号に基づく紛失多項式評価 (Oblivious Polynomial Evaluation) [17] を利用した秘匿共通集合計算プロトコル (FNP04 プロトコル) を提案した [9]．紛失多項式評価は数値を明かす事無く多項式計算を行うプロトコルである．また加法準同型暗号は平文 $a, \tilde{a}$ の暗号文 $E(a), E(\tilde{a})$ 及び定数 c を入力として，復号鍵を知る事無く効率良く $E(a+c\tilde{a})$ を求める事のできる暗号系であり，Paillier 暗号 [20] 等が知られる．Paillier 暗号では $\mathbb{Z}/N^2\mathbb{Z}$ 上で $E(a)E(\tilde{a})^c = E(a+c\tilde{a})$ が成り立ち (N は素数 p, q の積)，平文は $\mathbb{Z}/N\mathbb{Z}$ の元となる．

以下に $\mathcal{C}$ が $X \bigcap Y$ を得るための基本的な FNP04 プロトコルを説明する．なお $E_{\mathcal{C}}$ を $\mathcal{C}$ が復号可能な加法準同型暗号の暗号化関数とする．

1. $\mathcal{C}, \mathcal{S}$ はそれぞれ $X=\{x_1, \ldots, x_{k_{\mathcal{C}}}\}$, $Y=\{y_1, \ldots, y_{k_{\mathcal{S}}}\}$ を入力する．
2. $\mathcal{C}$ は $P_{\mathcal{C}}(x)=\prod_{i=1}^{k_{\mathcal{C}}}(x-x_i)=\sum_{i=0}^{k_{\mathcal{C}}-1} a_i x^i$ となる $a_i \in \mathbb{Z}/N\mathbb{Z}$ の暗号文 $E_{\mathcal{C}}(a_i)$ を計算して $\mathcal{S}$ に送る．
3. $\mathcal{S}$ は紛失多項式評価により $E_{\mathcal{C}}(a_i)$ から $E_{\mathcal{C}}(r_j(\sum_{i=0}^{k_{\mathcal{C}}-1} a_i y_j^i)+y_j)$ $(j=1, \ldots, k_{\mathcal{S}})$ を求め $\mathcal{C}$ に送る (r_j は乱数)．
4. $\mathcal{C}$ は $\mathcal{S}$ から受け取った暗号文を復号し，復号結果が自身の入力に含まれている場合その復号結果を出力する．

上記ステップ 3 の暗号文を $E_{\mathcal{C}}(r_j(\sum_{i=0}^{k_{\mathcal{S}}-1} a_i y_j^i))$ とし，ステップ 4 では復号結果が 0 となる個数を計算する事で秘匿共通要素数計算プロトコルとできる．なお上記の FNP04 プロトコルは $\mathcal{C}, \mathcal{S}$ が Semi-Honest である事を仮定しているが，[9] では $\mathcal{C}, \mathcal{S}$ が Malicious，すなわちプロトコルを不正に実行する状況を考慮した方式も検討されている．FNP04 プロトコルを単純に実装した場合，ステップ 3 より計算量は $O(k^2)$ となり，通信量は全体で $O(k|N^2|)$ ビットとなる．ただし [9] では計算量を $O(k \ln \ln k)$ とする改良方式が提案されている．

3 考察

2 節では既存の秘匿共通集合計算プロトコルが技術特性の軸から二つに分類される事をみた．一つは [4] に代表される，集合の各要素を推定困難な一意のデータに効率良く変換し，それらを各要素の代替として任意の手法で共通集合やその要素数を求めるアプローチである．もう一つの分類は，[9] に代表される，紛失多項式評価によって集合の各要素を根に持つような多項式を扱うアプローチによるものである．

先述のとおり，本稿では前者のアプローチにおける各要素の変換データを照合タグと呼ぶ．以下に照合タグの形式的な定義を与える．

定義 1 (識別困難性 [4])．$\Omega_\kappa \subseteq \{0,1\}^\kappa$ を κ ビットの有限領域とし，$\mathcal{D}_1 = \mathcal{D}_1(\Omega_\kappa)$, $\mathcal{D}_2 = \mathcal{D}_2(\Omega_\kappa)$ を Ω_κ 上の分布とする．$\mathcal{A}_\kappa(x)$ を，$x \in \Omega_\kappa$ が与えられたとき true または false を返すようなアルゴリズムとする．$\Pr[\mathcal{A}_\kappa(x)]$ を $\mathcal{A}_\kappa(x)$ が true を返す確率とする．このとき κ の任意の多項式アルゴリズム $\mathcal{A}_\kappa(x)$, 任意の多項式 $p(\kappa)$, 及び十分大きな κ について

$$|\Pr[\mathcal{A}_\kappa(\mathcal{D}_1)] - \Pr[\mathcal{A}_\kappa(\mathcal{D}_2)]| < \frac{1}{p(\kappa)}$$

であるとき，$\mathcal{D}_1$ と $\mathcal{D}_2$ は計算量的に識別困難であるという．

定義 2 (照合タグ)．ある集合 S, S' 及び $T : S \to S'$ について，$T(s) \in S'$ $(s \in S)$ を計算する手続きを $\Pi_T \subseteq \{0,1\}^*$ とする．このとき $\Pi_{\mathcal{A}} \subseteq \Pi_T$ について以下を満たすとき，$T(s)$ を $(S, S', \theta, \kappa, \Pi_{\mathcal{A}})$ における s の照合タグと呼ぶ．

(完全性) T は κ の多項式時間で計算でき，任意の $s, \tilde{s} \in S$ について θ 以上の確率で $s=\tilde{s} \Leftrightarrow T(s)=T(\tilde{s})$ が成り立つ．

(秘匿性) $s_i \in S$, $m=poly(\kappa)$, $z_m \in_R S$ について

$$\mathcal{D}_1 = \begin{pmatrix} s_1 & \cdots & s_{m-1} & s_m \\ T(s_1) & \cdots & T(s_{m-1}) & T(s_m) \end{pmatrix},$$

$$\mathcal{D}_2 = \begin{pmatrix} s_1 & \cdots & s_{m-1} & s_m \\ T(s_1) & \cdots & T(s_{m-1}) & z_m \end{pmatrix}$$

としたとき，$\Pi_{\mathcal{A}}$ を与えた κ の任意の多項式アルゴリズム $\mathcal{A}_\kappa$ について，$\mathcal{D}_1$ と $\mathcal{D}_2$ は計算量的に識別困難である．

次に既存の秘匿共通集合計算プロトコルの技術特性について考察する．現在までの研究動向として，紛失多項式評価を用いたアプローチには次の特徴がみられる．

- 一般に 3 以上の集合についても適用可能
- Malicious な攻撃者の存在を仮定
- 情報理論的安全性に向けた拡張

- 閾値判定等の応用

一方，照合タグを用いたアプローチには上記の特徴はみられず，機能拡張や安全性の向上が比較的困難であるといえる．実際 3 以上の集合についてそれぞれ照合タグを生成して開示した場合は，部分集合の共通集合やその要素数も照合タグから識別できてしまう問題がある．すなわち例えば $\mathcal{C}, \mathcal{S}, \mathcal{M}$ がそれぞれ集合 X, Y, Z を保持するとき，$Y \bigcap Z$ 等も分かってしまう事になる．また [4, 19] の具体例ではランダムオラクルモデルに制限されている．しかし照合タグは処理効率に優れる事に加え，個別に再利用できる特徴から，同一または類似の集合に対して繰り返し共通集合やその要素数を求める場合において特に有効である．例えば $X_i = \{x_i\}$, $Y = \{y_1, \ldots, y_{k_\mathcal{S}}\}$ として $i = 1, 2, \ldots$ について繰り返し Y と照合を行う場合，事前に Y を照合タグに変換しておけば，その後は x_i の照合タグを生成し Y の照合タグと照合すれば良い．先述のとおり照合は $O(\ln k_\mathcal{S})$ とできるため，一度の照合の計算量は $O(\ln k_\mathcal{S})$ となる．一方，紛失多項式評価を用いた場合は，事前に Y を根に持つ多項式の係数を暗号化しておいたとしても，一度の照合 (紛失多項式評価) に $O(k_\mathcal{S})$ の計算が必要となる．また Y の要素の一部を変更する場合，照合タグは変更部分のみの操作で済むが，紛失多項式評価を用いた場合は多項式を生成し直す必要がある．

最後に，定義 2 で与えた照合タグの定義と既存方式の適合性を AES03 プロトコルを例に検証する．先ず完全性については以下がいえる．

補題 1. κ をセキュリティパラメータとし，$\max(k_\mathcal{C}, k_\mathcal{S}) = poly(\kappa)$ とする．g を生成元とした位数 q の巡回群 $G = \langle g \rangle$ について $\mathcal{H}$ を G の置換関数とする．このとき AES03 プロトコルは κ の多項式時間で $T(x_i), T(y_i)$ を計算する事ができる．

補題 2. κ をセキュリティパラメータとし，$\max(k_\mathcal{C}, k_\mathcal{S}) = poly(\kappa)$ とする．g を生成元とした位数 q の巡回群 $G = \langle g \rangle$ について $\mathcal{H}$ を G の置換関数とする．$X = \{x_1, \ldots, x_{k_\mathcal{C}}\}$, $Y = \{y_1, \ldots, y_{k_\mathcal{S}}\}$ $(X, Y \subseteq G)$ とする．$u, v \in_R \{1, 2, \ldots, q-1\}$ とする．このとき任意の x_i, y_j について以下が成り立つ．

$$x_i = y_j \Leftrightarrow \mathcal{H}(x_i)^{uv} = \mathcal{H}(y_j)^{uv}$$

次に秘匿性について考察する．以下では AES03 プロトコル同様，Semi-Honest モデルを前提とする．また AES03 プロトコルでは $\mathcal{C}$ 及び $\mathcal{S}$ の手続きは対称的であるため，以下では $\mathcal{C}$ の任意の情報を取得できるような攻撃者のみ考える．

補題 3. κ をセキュリティパラメータとし，$\max(k_\mathcal{C}, k_\mathcal{S}) = poly(\kappa)$ とする．g を生成元とした位数 q の巡回群 $G = \langle g \rangle$ について $\mathcal{H}$ を G のランダムオラクルとする．$X = \{x_1, \ldots, x_{k_\mathcal{C}}\}$, $Y = \{y_1, \ldots, y_{k_\mathcal{S}}\}$ $(X, Y \subseteq G)$ とする．$b \in_R \{0, 1\}$, $rand \in_R G$ とする．AES03 プロトコル及びその中で $\mathcal{C}$ が得る情報をそれぞれ Π_T, $\Pi_\mathcal{C}$ $(\Pi_\mathcal{C} \subseteq \Pi_T)$ とする．ただし

$$z = \begin{cases} T(y_{k_\mathcal{S}}) & (b = 0) \\ rand & (b = 1) \end{cases}$$

とし，$T(y_{k_\mathcal{S}}) \in \Pi_\mathcal{C}$ を z に置き換える．

$$\mathcal{D}_1 = \begin{pmatrix} y_1 & \cdots & y_{k_\mathcal{S}-1} & y_{k_\mathcal{S}} \\ T(y_1) & \cdots & T(y_{k_\mathcal{S}-1}) & T(y_{k_\mathcal{S}}) \end{pmatrix},$$

$$\mathcal{D}_2 = \begin{pmatrix} y_1 & \cdots & y_{k_\mathcal{S}-1} & y_{k_\mathcal{S}} \\ T(y_1) & \cdots & T(y_{k_\mathcal{S}-1}) & rand \end{pmatrix}$$

とする．このとき，$(\Pi_\mathcal{C}, \mathcal{D}_1, \mathcal{D}_2)$ を入力とした κ の多項式時間攻撃者 $\mathcal{A}_\kappa$ について

$|\Pr[\mathcal{A}_\kappa(\mathcal{D}_1)] - \Pr[\mathcal{A}_\kappa(\mathcal{D}_2)]| > \frac{1}{p(\kappa)}$

となるとき，DDH 仮定 [6] を破る多項式時間アルゴリズムが存在する．

(証明の概略) $\mathcal{C}$ が $\mathcal{S}$ から得る情報は $\mathcal{H}(x_i)^{uv}$, $\mathcal{H}(y_i)^{uv}$ であり，$b \in_R \{0, 1\}$, $k = k_\mathcal{C} + k_\mathcal{S}$,

$$z_i = \begin{cases} x_i & (1 \leq i \leq k_\mathcal{C}) \\ y_{i-k_\mathcal{C}} & (k_\mathcal{C} < i \leq k) \end{cases}$$

としたとき

$$\mathcal{D}'_b = \begin{pmatrix} z_1 & \cdots & z_{k-1} & z_k \\ T(z_1) & \cdots & T(z_{k-1}) & T(z_k) \end{pmatrix}$$

$$\mathcal{D}'_{1-b} = \begin{pmatrix} z_1 & \cdots & z_{k-1} & z_k \\ T(z_1) & \cdots & T(z_{k-1}) & rand \end{pmatrix}$$

の分布を考えれば良い．ただし $y_{k_\mathcal{S}} \notin X$．ここで複数基底による DDH のインスタンス $(g, g_1^x, \ldots, g_{k-1}^x, g'_0, g'_1)$ に対し $(g_i, g'_{1-b'} \in_R G$, $g'_{b'} = g^x$, $b' \in_R \{0, 1\}$, $x \in_R \mathbb{Z}/q\mathbb{Z})$,

$$\mathcal{D}''_{b''} = \begin{pmatrix} r_1 & \cdots & r_{k-1} & g \\ g_1^{x\tilde{u}} & \cdots & g_{k-1}^{x\tilde{u}} & g_{b''}^{\tilde{u}} \end{pmatrix},$$

$$\mathcal{D}''_{1-b''} = \begin{pmatrix} r_1 & \cdots & r_{k-1} & g \\ g_1^{x\tilde{u}} & \cdots & g_{k-1}^{x\tilde{u}} & rand' \end{pmatrix}$$

を計算する．ただし r_i は $1 \leq i \leq k_\mathcal{C}$ であれば X の分布，$k_\mathcal{C} < i \leq k$ であれば Y の分布から取るものとし，$\tilde{u} \in_R \{1, 2, \ldots, q-1\}$,

$rand' \in_R G$ とする．そして
$b' = \{b'' \mid \max(\Pr[\mathcal{A}_\kappa(\mathcal{D}''_{b''})], \Pr[\mathcal{A}_\kappa(\mathcal{D}''_{1-b''})])\}$
を複数基底による DDH の出力とする．

以上，補題 1, 2, 3 から次の定理がいえる．

定理 1. κ をセキュリティパラメータとし，$\max(k_\mathcal{C}, k_\mathcal{S}) = poly(\kappa)$ とする．g を生成元とした位数 q の巡回群 $G = \langle g \rangle$ について $\mathcal{H}$ を G のランダムオラクルとする．$X = \{x_1, \ldots, x_{k_\mathcal{C}}\}$, $Y = \{y_1, \ldots, y_{k_\mathcal{S}}\}$ $(X, Y \subseteq G)$ とする．AES03 プロトコル Π_T によって $\mathcal{C}$ または $\mathcal{S}$ が得る情報を $\Pi_\mathcal{A}$ $(\subseteq \Pi_T)$ とする．このとき，$\Pi_\mathcal{A}$ を取得できる多項式時間攻撃者 $\mathcal{A}$ に対して，AES03 プロトコルは $(G, G, 1, \kappa, \Pi_\mathcal{A})$ における $s \in S$ の照合タグを生成する．

4 応用

照合タグは個別に再利用できるため，同一または類似の集合に対して繰り返し共通集合やその要素数を求める場合において特に有効である．本節ではそのような例として**キャンセラブルバイオメトリクス** (Cancelable Biometrics) [22] への適用について述べる．

キャンセラブルバイオメトリクスは，生体認証に用いる登録情報からの生体情報の復元を防止し，更に登録情報を更新可能とする技術である [22]．生体情報は一般に変更が利かない事から，登録情報の更新はなりすまし防止に有効とされる．Schoenmakers らは，二台の検証サーバの協調動作によって登録済みの生体情報の暗号化データ及び照合時に生成した生体情報の暗号化データを復号する事無く照合可能とするプロトコルを提案した [25]．二台の検証サーバの内部データを突き合わせない限りは生体情報の復元は計算量的に困難であり，任意の照合アルゴリズムに適用できる汎用性を持つ．

しかし [25] やその類似による方法は処理効率が良くないため，例えば大量に登録された生体情報の何れかと照合に成功すれば良い場合，各々の登録情報について照合する方法は現実的では無い．そこで生体情報から (一意の) 鍵データを生成し [12, 3]，その鍵データの照合タグを生成して生体情報の暗号化データに対応付ける事を考える．先ず登録時に生体情報から鍵データの照合タグを生成して生体情報の暗号化データに対応付けておき，照合時には新たに生成した生体情報から鍵データの照合タグを生成して登録済みの照合タグとの照合を行う．これにより照合すべき登録済みの生体情報の暗号化データを効率良く探索できる．なお照合タグの特徴により，照合タグから鍵データの抽出は困難であるため，鍵データからの生体情報の漏洩を防ぐ事ができる．

上記の方法は，n 個の生体情報が登録されているとき，単純に [25] を用いて照合する場合は最悪 $O(n)$ の計算量となるが，照合タグから直接照合する場合は $O(\ln n)$ とできる．また特に，事前に生体情報の照合タグを生成しておけば，照合時は [25] のプロトコルを一対の生体情報の暗号化データの照合に対してのみ実行すれば良い．

一方，複数の照合タグと一致した場合であっても，一致した各々の照合タグに対応した生体情報の暗号化データについて順次照合を行えば良く，一致した照合タグの総数が全体数と比べ十分小さければ処理効率の向上が見込めるしたがって，生体情報から鍵データへの変換について，本人拒否率 (FRR: False Rejection Rate) は低く設定する必要があるが，他人受入率 (FAR: False Acceptance Rate)，すなわち異なる生体情報から同一の鍵データが生成される確率は緊密にしなくて良いため，生体情報から鍵データを生成するアルゴリズムの自由度は高いといえる．

5 まとめ

集合の各要素を推定困難なデータに変換し，それらを各要素の代替として利用し共通集合やその要素数を求めるプロトコルについて，その特性や既存方式の考察を行った．特にその変換データを「照合タグ」として形式的に定義し，既存方式の一例について適合性を検証した．また照合タグを用いた集合計算は，同一または類似の集合に繰り返し適用する場合において特に有効である事に着目し，キャンセラブルバイオメトリクスへの適用を例にその有効性を明らかにした．

参考文献

[1] 経済産業省, 健康情報活用基盤構築のための標準化及び実証事業, https://microsite.accenture.com/meti/Pages/.

[2] 経済産業省, 情報大航海プロジェクト, http://www.meti.go.jp/policy/it_policy/daikoukai/.

[3] 柴田, 三村, 高橋, 中村, 曽我, 西垣, メカニズムベース PKI —指紋からの秘密鍵動的生成, 情報処理学会論文誌, Vol. 45, No. 8, pp. 1833–1844, 2004.

[4] R. Agrawal, A. Evfimievski, and R. Srikant, Information sharing across private databases, ACM SIGMOD 2003, pp. 86–97, 2003.

[5] M. Bellare and P. Rogaway, Random oracles are practical: A paradigm for designing efficient protocols, ACM CCS '93, pp. 62–73, 1993.

[6] D. Boneh, The decision Diffie-Hellman problem, ANTS '98, LNCS 1423, pp. 48–63, Springer-Verlag, 1998.

[7] D. Chaum, Blind signatures for untraceable payments, CRYPTO '82, pp. 199–203, Springer-Verlag, 1983.

[8] D. Chaum, Security without identification: Transaction systems to make big brother obsolete, Communications of the ACM, Vol. 28, No. 10, pp. 1030–1044, 1985.

[9] M. J. Freedman, K. Nissim, and B. Pinkas, Efficient private matching and set intersection, EUROCRYPT 2004, LNCS 3027, pp. 1–19, Springer-Verlag, 2004.

[10] C. Hazay and Y. Lindell, Efficient protocols for set intersection and pattern matching with security against malicious and covert adversaries, TCC 2008, LNCS 4948, pp. 155–175, Springer-Verlag, 2008.

[11] S. Hohenberger and S. A. Weis, Honest-verifier private disjointness testing without random oracles, PET 2006, LNCS 4258, pp. 277–294, Springer-Verlag, 2006.

[12] A. Juels and M. Wattenberg, A fuzzy commitment scheme, ACM CCS '99, pp. 28–36, 1999.

[13] A. Kiayias and A. Mitrofanova, Testing disjointness and private datasets, FC 2005, LNCS 3570, pp. 109–124, Springer-Verlag, 2005.

[14] L. Kissner and D. Song, Privacy-preserving set operations, CRYPTO 2005, LNCS 3621, pp. 241–257, Springer-Verlag, 2005.

[15] D. E. Knuth (有澤, 和田 監訳), The art of computer programming 3 Sorting and searching (Second Edition 日本語版), 株式会社アスキー, 2006.

[16] R. Li and C. Wu, An unconditionally secure protocol for multi-party set intersection, ACNS 2007, LNCS 4521, pp. 226–236, Springer-Verlag, 2007.

[17] M. Naor and B. Pinkas, Oblivious transfer and polynomial evaluation, STOC '99, pp. 245–254, 1999.

[18] R. Nojima and Y. Kadobayashi, On the security and the performance of a practical two-party set-intersection protocol, SCIS 2008, 2008.

[19] R. Nojima and Y. Kadobayashi, On the construction of the set-intersection protocol from blind signatures, IEICE Technical Report, ISEC2008-9 (2008-5), pp. 57–60, 2008.

[20] P. Paillier, Public-key cryptosystems based on composite degree residuosity classes, EUROCRYPT '99, LNCS 1592, pp. 223–238, Springer-Verlag, 1999.

[21] A. Patra, A. Choudhary, and C. P. Rangan, Information theoretically secure multi party set intersection re-visited, SAC 2009, 2009.

[22] N. K. Ratha, J. H. Connell, and R. M. Bolle, Enhancing security and privacy in biometric-based authentication systems, IBM System Journal, Vol. 40, No. 3, 2001.

[23] Y. Sang and H. Shen, Privacy preserving set intersection based on bilinear groups, ACSC 2008, pp. 47–54, 2008.

[24] Y. Sang, H. Shen, Y. Tan, and N. Xiong, Efficient protocols for privacy preserving matching against distributed datasets, ICICS 2006, LNCS 4307, pp. 210–227, Springer-Verlag, 2006.

[25] B. Schoenmakers and P. Tuyls, Efficient binary conversion for Paillier encrypted values, EUROCRYPT 2006, LNCS 4004, pp. 522–537, Springer-Verlag, 2006.

[26] J. Vaidya and C. Clifton, Secure set intersection cardinality with application to association rule mining, Journal of Computer Security, Vol. 13, No. 4, pp. 593–622, 2005.

[27] Q. Ye, H. Wang, and J. Pieprzyk, Distributed private matching and set operations, ISPEC 2008, LNCS 4991, pp. 347–360, Springer-Verlag, 2008.

[28] Q. Ye, H. Wang, and C. Tartary, Privacy-preserving distributed set intersection, ARES 2008, pp. 1332–1339, 2008.

[29] Q. Ye, H. Wang, J. Pieprzyk, and X. M. Zhang, Unconditionally secure disjointness tests for private datasets, International Journal of Applied Cryptography, Vol. 1, No. 3, 2009.