RL-003

準同型暗号を用いた複数企業間の顧客情報分析

A customer information analysis between enterprises using homomorphic encryption

安田 雅哉 † , 下山 武司 †, 横山和弘 ‡ , 小暮 淳 †
Masaya Yasuda, Takeshi Shimoyama, Kazuhiro Yokoyama, Jun Kogure,

1 まえがき

近年、スマートフォンやタブレット端末等が広く普及し、いつでもどこでもメールアクセスや情報検索などが出来るようになった。これらの背景には、ユーザからの要求を即座に処理可能な大規模計算機、いわゆるクラウドコンピュータやこれまでに蓄積された莫大な情報、いわゆるビッグデータの存在が大きく関わっている。複雑な演算処理やビッグデータの管理をクラウド上で一元化することで情報処理コストを下げるだけなく、収集データから有益な情報を導き出すことが可能となり、更に便利な情報ツールへと進化することが期待されている。特に最近では、分野の異なる企業が複数協業し、顧客需要を喚起する動きが加速していることから、複数企業間の情報分析を行うための新しいビジネスツールとしてのクラウド利活用が強く期待されている。

しかし一方で、クラウドを用いたビジネスに対しては、懸念される点も依然として多く、特に企業利用においては、クラウド上のデータセキュリティが確保されるのかという事が常に問題視されている。顧客情報の流出事故等は企業の存亡を揺るがす大きな問題へと発展する恐れがあるため、企業における情報管理はより慎重かつ厳重にならざるを得ない。このように、企業利用において、データをクラウド上に預ける際には、安全性対策が必須であり、特にデータの暗号化は有用な対策の 1 つである。データを暗号化することで、鍵を適切に管理すればデータの秘匿性を保つことができるため、データ保護の観点から有効であるが、一方で、クラウドに集められた情報の価値を最大限利活用しようとした場合、統計分析や検索時には、通常元データが必要となるため暗号データのまま処理することは難しい。同一企業内のデータであれば、分析に必要なデータを企業内に引き戻し組織内で復号することで、元データに対する処理が可能となるが、もし複数の企業に関係する情報を抽出しようとした場合には、このような手法をとることはできない。複数企業のデータをクラウド上で分析処理する際に考えられる素朴な方法としては、クラウド内で暗号文を復号し、平文の状態に戻してデータ処理を進めることであるが、この場合は鍵管理や復号データの利用時のクラウドへのアクセス制御等、セキュリティを高めるためのシステム管理上の対策が必要になり、クラウド利点の一つであるコストの面で不利となってしまい、決して有用な方法ではない。

上記の課題を克服する方法として、データを暗号化したまま分析処理等を行うことが出来る準同型暗号や、複数の計算機が安全な手順に従いデータを秘匿したまま協調して処理を行うマルチパーティー計算 (MPC) 等のプライバシー保護利活用技術が近年注目されている (文献 [18, 19] 等を参照)。特に、準同型暗号は、例えば暗号化されたテスト結果に対し、暗号化したまま集計可能で、各テスト結果を秘匿したまま集計結果のみを計算できるといった技術であり、暗号化したまま乗算が可能な RSA 暗号 [15]、暗号化したまま加算が可能な additive ElGamal 暗号 [6] や Paillier 暗号 [14] などが知られていた。これらの暗号は、電子投票・電子現金に応用可能だが、暗号加算のみ、あるいは暗号乗算のみといった処理しかできないため、適用先がかなり限定されるといった課題があった。これらの課題に対し、2009 年 Gentry によって完全準同型暗号の具体的な構成法が提案されて以降 [8]、理論的に任意の論理演算が暗号化されたまま処理できるようになった。しかし、処理コストと暗号データサイズの両面で大きな課題があり、完全準同型暗号の実用化にはさらなるブレークスルーが必要であると考えられている。

その一方で、完全準同型暗号の基礎構成要素であり、演算回数が限定的ではあるが暗号加算と暗号乗算の両方が可能な準同型暗号 (SHE 準同型暗号と呼ぶ) の応用研究が近年注目されつつある (文献 [12] 等を参照)。SHE 準同型暗号は、暗号演算回数が限定されるものの、完全準同型暗号よりも処理コスト・暗号データサイズ共にかなり小さく、幅広い適用先に対し実用的に利用できると期待されている。2012 年には、我々は文献 [21] でイデアル格子ベースの SHE 準同型暗号 [8, 9] を用いた複数企業間におけるクラウド上の秘匿購買履歴データ分析を提案した。文献 [21] では、ベクトル化された購買履歴データに対して、ベクトルの各成分の暗号化を行っていたため、暗号データサイズと秘匿計算コストの肥大化という課題があった。その課題に対し、本論文では Ring-LWE ベースの SHE 準同型暗号 [3, 4, 12] を用いて、ある長さのベクトルを 1 つの暗号文に集約する一括暗号化方式という実用化技術を提案する。本論文では、処理性能と安全性の観点から 2048 次元の Ring-LW ベースの SHE 準同型暗号を実装した。これにより、2048 次元のベクトルを一括暗号化でき、さらに一括暗号文上で高速に秘匿内積を計算する手法を紹介すると共に、複数企業間の秘匿購買履歴データ分析に対する具体的な適用方法を示す。今回の提案技術は、医療データ分析や生体認証照合などのプライバシー保護とその情報利活用が求められる分野に広く適用できる。

†富士通研究所, ソフトウェアシステム研究所
‡立教大学 理学部数学科

2　準同型暗号と秘匿購買履歴分析

ここでは、まず購買履歴データ分析の一般的手法について紹介する。次に、プライバシー保護データマイニング手法の中で、本適用先に対して準同型暗号が適切であることを説明すると共に、文献 [21] で提案した準同型暗号を利用した秘匿購買履歴データ分析について紹介する。

2.1　一般的な購買履歴データ分析手法

本論文では A, B 企業における 2 企業間の購買履歴データ分析のみを考える。2 企業間の購買履歴データ分析において、まず A 企業と B 企業の間で共通の顧客 ID を作成する。次に、A 企業と B 企業の購買履歴データが、それぞれ表 1 のように与えられたとする (簡単のため、A 企業の商品を A 商品、B 企業の商品を B 商品と表記する)。購買履歴データ分析において、A 商品と B 商品の関連性を分析するために、商品間の類似度を以下のように計算する: まず、表 2 のように顧客 ID ごとに A 商品と B 商品の購買履歴データをまとめる。次に、表 2 から「A 商品と B 商品の両方を購入した人数 a」などの購買履歴集計データを表 3 のように計算する。さらに、表 3 で得られた $a, b, c, d, R_1, R_2, C_1, C_2, n$ の数値データを図 1 で示した類似度指標の計算式に代入することで、A 商品と B 商品の間のさまざまな類似度を計算することが可能である。

表 1: A 企業と B 企業の購買履歴データの数値例 (購入有りを 1、購入無しを 0 で表す)

顧客 ID	A 商品
123	1
124	1
126	0
129	1
130	0
135	1
137	1

顧客 ID	B 商品
123	0
124	1
129	1
131	0
132	0
135	1
136	0

表 2: 顧客 ID ごとにまとめた購買履歴データ (購買データがない場合は、“–” と表し購入無し 0 とみなす)

顧客 ID	A 商品	B 商品
123	1	0
124	1	1
126	0	–
129	1	1
130	0	–
131	1	0
132	–	0
135	1	1
136	–	0
137	1	–

表 3: A 商品と B 商品の購買履歴の集計データ (具体的な数値は表 2 から算出)

		B 商品		合計
		購入	×	
A 商品	購入	$a = 3$	$b = 3$	$R_1 = 6$
	×	$c = 0$	$d = 4$	$R_2 = 4$
合計		$C_1 = 3$	$C_2 = 7$	$n = 10$

	名称	類似度 s_{ij}	非類似度　d_{ij}
M1	交互作用統計量	$\frac{ad-bc}{\sqrt{R_1R_2C_1C_2/n}}$	$\max_{i,j}\{s_{ij}\} - s_{ij}$
M2	Cohen's Kappa	$\frac{2(ad-bc)}{2(ad-bc)+n(b+c))}$	$\frac{n(b+c)}{2(ad-bc)+n(b+c)}$
M3	Psi	$\frac{\lvert ad-bc\rvert}{\sqrt{R_1R_2C_1C_2}}$	$\max_{i,j}\{s_{ij}\} - s_{ij}$
M4	Sokal and Sneath 2	$a/(a+2b+2c)$	$(2b+2c)/(a+2b+2c)$
M5	Jaccard	$a/(a+b+c)$	$(b+c)/(a+b+c)$
M6	Czesanowski	$2a/(2a+b+c)$	$(b+c)/(2a+b+c)$
M7	Kulczynski	$a/(b+c)$	$\max_{i,j}\{s_{ij}\} - s_{ij}$
M8	Ochiai	$a/\sqrt{(a+b)(a+c)}$	$1 - a/\sqrt{(a+b)(a+c)}$
M9	Yule's Q	$(ad-bc)/(ad+bc)$	$2bc/(ad+bc)$
M10	Russel and Rao	a/n	$(b+c+d)/n$
M11	カイ二乗距離	未定義	$\sum_{k=1}^{n} \frac{n}{n_{k\bullet}}(\frac{n_{ki}}{n_{\bullet i}} - \frac{n_{kj}}{n_{\bullet j}})^2$
M12	Rogers and Tanimoto	$(a+d)/(a+b+2c+2d)$	$(b+2c+d)/(a+b+2c+2d)$
M13	Sokal and Sneath 1	$2(a+d)/(2a+2d+b+c)$	$(b+c)/(2a+2d+b+c)$
M14	Simple matching	$(a+d)/n$	$(b+c)/n$
M15	Hamann	$(a+d-b-c)/n$	$(2b+2c)/n$
M16	主効果の相乗平均	$\sqrt{(a+b)(a+c)}/n$	$1 - \sqrt{(a+b)(a+c)}/n$

図 1: 購買履歴データ分析で利用される主な類似度指標 [17] (表 3 で得られた $a, b, c, d, R_1, R_2, C_1, C_2, n$ の数値データを代入することで、A 商品と B 商品の間の類似度を計算することが可能)

2.2　プライバシー保護データマイニング

近年、プライバシー情報の保護とその利活用のバランスを適切に管理しながら、利用価値の高い情報を安全かつ有効に活用するプライバシー保護データマイニングの研究が各方面で盛んに行われている。現在知られているプライバシー保護データマイニング手法には、大きく分けて匿名化・ランダム化・MPC・準同型暗号の 4 つのアプローチがある [18]。匿名化は個人識別子を加工して分析する手法であるが、§3.4 で説明した購買履歴データ分析において、顧客 ID を各企業ごとに加工すると企業間分析が困難となり、匿名化は本適用先には適さない。次に、ランダム化は各データにランダムノイズを付加することでプライバシーを保護する方法であるが、大域的な統計・分析情報しか得られないために、図 1 に示した商品間の類似度指標を正確に求めたい本適用先には有効ではない。最後に、MPC と準同型暗号の両アプローチは、ランダム化とは異なり、データを秘匿したまま正確な分析情報を計算可能なため本適用先に適している ([18, 表 1] を参照)。両アプローチの本質的な差異は、MPC は汎用の関数の安全な計算を目標とするのに対し、準同型暗号は汎用性を犠牲にして特定の関数の安全な計算を効率的に行うことを目的とする点である (詳細は [18, 2.4 節] を参照)。そのため、本適用先の購買履歴データ分析に対しては計算関数が特定的であり、計算効率の観点から準同型暗号の適用が最適だと考える。

2.3　準同型暗号

準同型暗号は、暗号化したままデータ操作可能な公開鍵暗号で、表 4 で示すように処理可能な暗号操作により以下の 3 種類に大別できる:

表 4: 準同型暗号の種類 (処理可能な暗号操作別に分類)

スキーム	暗号操作	処理性能	応用	代表的な方式
Add. HE	加算のみ	高速	秘匿集計 (電子投票, 電子現金など)	Paillier 方式 [14], additive ElGamal 方式 [6]
SHE	加算＋乗算 m 回	m 増大⇒遅くなる (パラメータ設定に大きく依存)	高度な秘匿集計 (秘匿共通集合計算 [21] 標準偏差などの統計処理 [12])	BGN 暗号 [1] (暗号乗算は深さ $m = 1$ 回のみ) イデアル格子ベース [8, 9]
(leveled) FHE	任意	かなり遅い [5, 9] (再暗号化:数十分)	任意の秘匿集計 (秘匿検索 [11], 秘匿 AES 回路 [10])	整数ベース [5, 7] (Ring-)LWE ベース [2, 3, 4]

- Add. HE(Additively Homomorphic Encryption)[1]
  暗号加算のみ可能な暗号方式で、加算準同型暗号とも呼ばれ、Paillier 暗号 [14] や additive ElGamal 暗号 [6] が代表的。暗号加算のみしかできないが、RSA 暗号 [15] と同程度の処理性能を持ち、電子投票や電子現金などへの応用が期待されている。

- SHE(Somewhat Homomorphic Encryption)
  限定回の暗号加算と暗号乗算の両方が可能な暗号方式で、以下で説明する FHE スキームに比べ暗号操作の種類は限定的になるものの、処理性能や暗号データサイズがかなり小さくてすむため、さまざまな統計処理への応用が期待されている。

- FHE(Fully Homomorphic Encryption)
  任意の暗号操作が可能な暗号方式で、完全準同型暗号と呼ばれることが多い。2009 年、Gentry[8] がイデアル格子を用いた FHE の構成法を示して以降、FHE の適用や構成法に関する研究が急速に加速した。これまでの実装報告からわかるように [5, 9][2]、実用的な処理性能や暗号データサイズを実現するにはさらなる改良・研究開発が必要である。近年では、FHE を特定の応用先にカスタマイズする leveled FHE の概念が提案され、秘匿 AES 回路実装の研究 [10] などもある (秘匿 AES 処理に約 1 日半かかるので、leveled FHE の実用化ですらさらなるブレークスルーが必要)。

特に、2009 年に FHE の構成が示されて以降、準同型暗号の幅広い適用研究が盛んになりつつあり、特にクラウドコンピューティング分野への適用が強く期待されている。次節以降、加算準同型暗号よりも応用性が高く、かつ FHE よりも処理性能・暗号データサイズの面で実用性が高い SHE 準同型暗号に着目していく。

2.4 文献 [21] による SHE 準同型暗号を利用したクラウド秘匿集計 (ビット暗号化)

まず、図 1 で示した類似度指標を求めるには、表 3 の購買履歴の集計データを計算すればよいことに注意する。さらに、表 3 の集計データを得るためには、表 3 内の a, R_1, C_1, d のデータを求めれば十分である (顧客 ID 数 n は既知とする)。特に、A, B 商品購入者数 a は、表 2 から得られる A 商品と B 商品の 2 つのバイナリ化購買履歴データに対する内積計算により求めることができる。特に、本論文では、A 商品と B 商品の購買履歴データを準同型暗号で暗号化したまま内積計算 a を求める方法を説明していく。まず、適用する準同型暗号の種類について、加算準同型暗号では暗号加算と暗号乗算の両方が必要な内積計算は計算できない[3]。そこで、秘匿内積計算が可能で、処理性能・暗号データサイズとも FHE より実用的な SHE 準同型暗号が本適用先には最適である。

文献 [21] で我々は初めて、イデアル格子ベースの SHE 準同型暗号 [8, 9] を用いた購買履歴データのクラウド秘匿集計モデルを提案した。SHE 準同型暗号方式の 1 つであり、ペアリング暗号ベースの BGN 準同型暗号 [1] でも秘匿計算可能であるが、文献 [12, Section 1.2] で詳細に比較されているように、BGN 暗号は高々1 回の暗号乗算に楕円曲線上のペアリング操作が必要で、イデアル格子ベースや §3 で説明する Ring-LWE ベースの格子準同型暗号に比べ、秘匿内積計算の処理コストが非常に高い。よって、効率性の観点から本適用先には格子準同型暗号が有用であることに注意しておく。文献 [21] で示した秘匿集計モデルでは、A 企業と B 企業の他に、秘匿計算の委託先であるクラウド、購買履歴データを分析する分析者による 4 者による集計モデルを考える。我々が文献 [21] で提案した秘匿集計の手順を以下に示しておく:

1. 分析者が SHE 準同型暗号の公開鍵と秘密鍵を生成し、公開鍵のみを A, B 企業とクラウドに公開しておく。

2. A, B 企業はそれぞれ、公開鍵を用いてバイナリ化購買履歴ベクトル $(a_1, \ldots, a_n)$, $(b_1, \ldots, b_n)$ を各成分ごとに暗号化し、クラウドに顧客 ID 付きの暗号化購買履歴データを送信する。

3. クラウドは、A, B 企業から送信された暗号化購買履歴データを表 2 のように顧客 ID ごとにまとめ、秘匿集計を行う。A, B 商品の暗号化購買履歴データを

$$(E(a_1), \ldots, E(a_n)),\ (E(b_1), \ldots, E(b_n)) \qquad (1)$$

としたとき (E は SHE 準同型暗号化関数)。、暗号化された A, B 商品購入者数 $E(a)$ は、暗号化データ上の内

[1] RSA 暗号 [15] などの乗算準同型暗号については言及しない。
[2] 任意の暗号化処理を可能にするために、完全準同型暗号では再暗号化処理を行う必要があるが、汎用的な PC 上で数十分程度かかると実装報告されている。
[3] Paillier 暗号などの加算準同型暗号を利用した秘匿内積計算が文献 [20] などで紹介されているが、2 つのベクトルの内 1 つのベクトルは準同型暗号で保護しない計算方法のため、本適用先のような 2 つの企業が持つ購買履歴データを両方暗号化したい場合には適用できない。

積計算

$$E(a) = \sum_{i=1}^{n} E(a_i) \cdot E(b_i) \quad (2)$$

で求めることができる。同様に、その他の暗号化集計データ $E(R_1), E(C_1), E(d)$ も計算することができる。クラウドは、暗号化集計結果のみを分析者に送信する。

4. 分析者は、秘密鍵を用いて暗号化集計結果を復号し、A, B 企業が持つ個々の購買履歴データを知ることなく、a, R_1, C_1, d の集計データを求めることができる。次に、a, R_1, C_1, d の値から、表 3 の各データを計算し、図 1 で示した類似度の計算式に代入することで、A, B 商品の間の類似度を求める。最後に、分析結果である A, B 商品間の類似度を A 企業と B 企業に公開する。

上記で説明したクラウド秘匿集計を利用することで、A 企業と B 企業が持つ購買履歴データをお互いに公開することなく、A 商品と B 商品の類似度を計算できる。さらに、準同型暗号を利用しているため、購買履歴データをクラウドに秘匿したまま集計可能である。

3 一括暗号化手法の提案

ここでは、§2.4 で説明したビット暗号化による秘匿集計の問題点を挙げると共に、その問題点を解決する一括暗号化と呼ぶ実用化技術を提案し、その効果を説明する。

3.1 ビット暗号化の問題点

§2.4 における秘匿集計モデルでは、A, B 商品の購買履歴データに対して、各成分の暗号化データ (1) を考え、秘匿内積 (2) を計算する。この場合、以下の問題が生じる:

- 暗号データサイズの肥大化: n 次元のバイナリ化購買履歴データの暗号データとして、準同型暗号文が n 個必要となり、例えば、顧客 ID 数 $n = 10,000$ 件に対して、文献 [21] で利用したイデアル格子ベースの SHE 準同型暗号 E に対し、秘匿内積計算を可能とする最適なパラメータ設定では ([21, 表 5] を参照)[4]、1 つの暗号文データサイズが約 20KByte で、それが $n = 10,000$ 件必要なので、約 200MByte という巨大な暗号データサイズが必要になる。

- 秘匿内積計算処理コストの増大: 一方、ビット暗号化による秘匿内積計算 (2) の処理コストして、n 回の暗号乗算が必要となり、例えば、顧客 ID 数 $n = 10,000$ の場合、$10,000$ 回の暗号乗算が必要となる。イデアル格子準同型暗号に対する我々の実装では、1 回の暗号乗算に対し約 9.00ms（実装環境 : Intel Xeon X3480, 3.7 GHz, 16 GByte memory）かかったので、ビット暗号化による秘匿内積計算 (2) に約 $9.00 \times 10,000 \approx 90$ 秒かかるため、高速処理とは言い難い (それでも BGN 暗号よりはかなり効率的ではあるが...)。

[4] イデアル格子次元 4096 で、秘密鍵行列の各成分サイズが 40 ビットの鍵パラメータを選択した場合 (格子ベース SHE 準同型暗号はパラメータ設定により、処理できる秘匿計算の種類が大きく変わる)。

そこで、これらの問題点を省みて、格子ベース SHE 準同型暗号を利用した秘匿購買履歴分析において、暗号文データサイズと秘匿計算処理コストの両方を大幅に削減する一括暗号化方式という新しい手法を以下で説明する。

3.2 Ring-LWE ベースの SHE 準同型暗号

まず、今回の秘匿購買履歴データ分析で利用する SHE 準同型暗号方式を説明する [5]。以下で説明する SHE 準同型暗号は、Lauter-Naehrig-Vaikuntanathan[12] によって紹介されたもので、Brakerski-Vaikuntanathan が最初に提案した Ring-LWE 準同型暗号方式 [3, 4] に基づく [6]。暗号方式構成のために、次のパラメータが必要である：

- N: 2 冪整数で、円分多項式 $f(x) = x^N + 1$ を定める (格子次元とも呼ばれる)。また、今回の準同型暗号方式で基本となる環 $R = \mathbb{Z}[x]/(f(x))$ を定める。

- q: 条件 $q \equiv 1 \bmod 2N$ を満たす素数で、暗号文空間の基本環 $R_q = R/qR = \mathbb{F}_q[x]/(f(x))$ を定める（R_q の元は係数が有限体 $\mathbb{F}_q$ の $(N-1)$ 次多項式で表現)。

- t: 条件 $t < q$ を満たす（素数とは限らない）正整数で、平文空間の環 $R_t = R/tR = (\mathbb{Z}/t\mathbb{Z})[x]/(f(x))$ を定める（R_t の元は係数が剰余環 $\mathbb{Z}/t\mathbb{Z}$ の $(N-1)$ 次多項式で表現)。

- σ: 標準偏差が σ となる n 次元整数格子 $\mathbb{Z}^N$ 上の離散ガウス分布 $\chi = D_{\mathbb{Z}^N,\sigma}$ を定める（具体的には、N 個のガウス分布 $N(0,\sigma)$ によるサンプル元の最近似整数への丸め込み写像で実現)。主に、暗号文生成時における乱数要素として用いられる。

$\mathbb{Z}$ 加群同型写像 $R \ni v_0 + v_1 x + \cdots + v_{N-1} x^{N-1} \mapsto (v_0, v_1, \cdots, v_{N-1}) \in \mathbb{Z}^N$ により、$(N-1)$ 次整数係数多項式を N 次元整数ベクトルと同一視することができ、同様に、R_q と R_t の元をそれぞれ $\mathbb{F}_q$ と $\mathbb{Z}/t\mathbb{Z}$ 成分の N 次元ベクトルとして同一視できる。次に、上記パラメータ (N, q, t, σ) を用いた Ring-LWE ベースの SHE 準同型暗号の構成を説明する (詳細は [12, Section 3] を参照):

3.2.1 鍵生成

まず、離散ガウス分布 χ からサンプリングされた元 $R \ni s \leftarrow \chi$ を 1 つ固定する。次に、R_q から乱数 $p_1 \in R_q$ と離散ガウス分布 χ から得られる元 $R \ni e \leftarrow \chi$ をとる。そこで、$p_0 = -(p_1 \cdot s + t \cdot e) \in R_q$ と計算し、公開鍵 $\mathsf{pk} = (p_0, p_1)$ と秘密鍵 $\mathsf{sk} = s$ を生成する。

3.2.2 暗号化

平文 $m \in R_t$ と公開鍵 $\mathsf{pk} = (p_0, p_1)$ に対して、まず離散ガウス分布 χ から 3 つの元 $R \ni u, f, g \leftarrow \chi$ をサンプ

[5] 文献 [21] で利用したイデアル格子準同型暗号に対しても、今回提案する一括暗号化方式を導入することは可能であるが (文献 [22] を参照)、構成のしやすさから本論文では Ring-LWE 準同型暗号を選択。

[6] ここで与える準同型暗号は、Ring-LWE 問題困難性 (詳細は [12, Section 3.1] を参照) の下で、IND-CPA 安全性を持つことが知られている [12, Lemma 3.4]。

リングし、平文 m の “fresh”暗号文 [7] を

$$\begin{aligned}\mathsf{Enc}(m,\mathsf{pk}) &= (c_0,c_1)\\ &= (p_0u+tg+m, p_1u+tf)\in(R_q)^2\end{aligned}$$

と生成する (条件 $t<q$ から元 $m\in R_t$ を自然に R_q の元としてみなす必要がある)。

3.2.3 準同型暗号操作 (暗号加算・暗号乗算)

fresh 暗号文とは限らない 2 つの暗号文 [8] である $\mathsf{ct}=(c_0,c_1,\ldots,c_\xi)\in(R_q)^{\xi+1}$, $\mathsf{ct}'=(c'_0,c'_0,\ldots,c'_\eta)\in(R_q)^{\eta+1}$ に対し、準同型暗号加算 “$\dot{+}$”を

$$\mathsf{ct}\dot{+}\mathsf{ct}' := (c_0+c'_0, c_1+c'_1,\ldots,c_{\max(\xi,\eta)}+c'_{\max(\xi,\eta)})$$

と定める。さらに、準同型暗号乗算 “$*$”を

$$\mathsf{ct}*\mathsf{ct}' := (\hat{c}_0,\hat{c}_1,\ldots,\hat{c}_{\xi+\eta})$$

と定める。ただし、各成分 $\hat{c}_i$ は z を不定元とした多項式環 $R_q[z]$ 上の以下の乗算で定まる [9]:

$$\sum_{i=0}^{\xi+\eta}\hat{c}_iz^i := \left(\sum_{i=0}^{\xi}c_iz^i\right)\cdot\left(\sum_{i=0}^{\eta}c'_iz^i\right)\in R_q[z].$$

3.2.4 復号

fresh 暗号文とは限らない暗号文 $\mathsf{ct}=(c_0,c_1,\ldots,c_\xi)\in(R_q)^{\xi+1}$ に対して、秘密鍵 $\mathsf{sk}=s$ による復号処理は

$$\mathsf{Dec}(\mathsf{ct},\mathsf{sk}) := [\tilde{m}]_q \bmod t\in R_t$$

と定義される。ただし、$\tilde{m}=\sum_{i=0}^{\xi}c_is^i\in R_q$ とし、$[\tilde{m}]_q$ は範囲 $[-q/2,q/2)$ への素数 q による剰余演算とする (通常の剰余演算 $\bmod q$ は範囲 $[0,q)$ への射であることに注意)。

3.3 一括暗号化方式

ここでは、上記で説明した Ring-LWE ベースの SHE 準同型暗号を利用した一括暗号化方式の具体的な構成法を説明する。我々の一括暗号化方式のアイデアは、N 次元ベクトルを環 R_t 上の多項式に変換し、その環 R_t 上の変換多項式を暗号化することで、N 次元ベクトルを 1 つの暗号文 (=一括暗号文) に集約する。さらに、秘匿内積計算については、上記で説明した Ring-LWE 準同型暗号が持つ環 R_t 上の準同型性 (詳細は後述の §3.4 を参照) を利用することで、一括暗号文のまま効率的に処理可能である [10]。本論文の適用先である秘匿購買履歴データ分析に対する一括暗号化方式は以下のように行う:

顧客 ID 数 n と格子次元パラメータ N に対して、$\ell=\lceil n/N\rceil$ とおく。最初に、2 つのバイナリ化購買履歴データ $(a_1,\ldots,a_n),(b_1,\ldots,b_n)$ をそれぞれ N 次元のバイナリ化部分ベクトルに分割する。つまり、

$$\begin{aligned}(a_1,\ldots,a_n) &= \left(A_1^{(1)},\ldots,A_N^{(1)}\mid\cdots\mid A_1^{(\ell)},\ldots,A_N^{(\ell)}\right)\\ (b_1,\ldots,b_n) &= \left(B_1^{(1)},\ldots,B_N^{(1)}\mid\cdots\mid B_1^{(\ell)},\ldots,B_N^{(\ell)}\right)\end{aligned}$$

と ℓ 分割しておく ($A_i^{(j)}=a_{i+j(N-1)}$, $B_i^{(j)}=b_{i+j(N-1)}$ とおく。ただし　$n<\ell N$ の場合は、足らない成分として 0 を付加しておく)。以下で、N 次元ベクトルを 1 つの暗号文にまとめる一括暗号化方式を定義する：

一括暗号化方式の構成

2 つの N 次元部分ベクトル $A^{(j)}=\left(A_1^{(j)},\ldots,A_N^{(j)}\right)$, $B^{(j)}=\left(B_1^{(j)},\ldots,B_N^{(j)}\right)$ を考える。

- A 商品に関する N 次元部分ベクトル $A^{(j)}$ に対して、ベクトル成分が昇順となる多項式

$$m_1\left(A^{(j)}\right)=\sum_{i=0}^{N-1}A_{i+1}^{(j)}x^i\in R_t \qquad (3)$$

を考え、この多項式を §3.2.2 の暗号化を用いて、

$$\mathsf{pEnc}_1\left(A^{(j)},\mathsf{pk}\right) := \mathsf{Enc}\left(m_1(A^{(j)}),\mathsf{pk}\right)$$

と定義する。これをタイプ 1 の一括暗号化 (packed encryption of type 1) と呼ぶ。

- さらに、B 商品に関する N 次元部分ベクトル $B^{(j)}$ に対しては、タイプ 1 とは異なり、ベクトル成分が降順となる多項式 (−符号もつく)

$$m_2\left(B^{(j)}\right)=-\sum_{i=0}^{N-1}B_{i+1}^{(j)}x^{N-i}\in R_t \qquad (4)$$

を考え、この多項式を §3.2.2 の暗号化を用いて、

$$\mathsf{pEnc}_2\left(B^{(j)},\mathsf{pk}\right) := \mathsf{Enc}\left(m_2(B^{(j)}),\mathsf{pk}\right)$$

と定義する。これをタイプ 2 の一括暗号化 (packed encryption of type 2) と呼ぶ。

上記で構成した一括暗号化方式では、R_t 上の変換多項式に対し通常の暗号化を行うだけなので、Ring-LWE 準同型暗号が持つ暗号の安全性は全く変えないことに注意する。n 次元の 2 つのバイナリ化購買履歴データ $(a_1,\ldots,a_n)$,

[7] 以下で定義する準同型暗号操作により暗号文上の加算・乗算が可能であるが、平文データを暗号化した fresh な状態の暗号文のことを fresh 暗号文と呼ぶ。

[8] fresh 暗号文は環 R_q の 2 つの成分で構成されるが、以下で定義する準同型暗号乗算により、成分の個数が 1 つずつ増加していくことに注意しておく。そのため、ここでは任意の成分個数を持つ暗号文に対して、準同型暗号操作を定義する必要がある (次に定義する復号処理についても同様)。

[9] 準同型暗号操作により暗号文内の乱数成分が次第に増大し、乱数成分のサイズが素数 q で決まる復号範囲を超えたとき準同型性が成立しなくなる (よって、この暗号方式は SHE 準同型暗号に属する)。

[10] 我々の一括暗号化方式では環 $R=\mathbb{Z}[x]/(f(x))$ 上の多項式演算を利用しているので、ペアリング暗号ベースの BGN 暗号方式には適用できない。

$(b_1, \ldots, b_n)$ の暗号文として、以下の一括暗号文を考える：

$$\left(\mathsf{pEnc}_1(A^{(1)}, \mathsf{pk}) \mid \cdots \mid \mathsf{pEnc}_1(A^{(\ell)}, \mathsf{pk})\right) \quad (5)$$

$$\left(\mathsf{pEnc}_2(B^{(1)}, \mathsf{pk}) \mid \cdots \mid \mathsf{pEnc}_2(B^{(\ell)}, \mathsf{pk})\right) \quad (6)$$

つまり、n 次元ベクトルを暗号化するのに、$\ell = \lceil n/N \rceil$ 個の暗号文が必要となる。これにより、§2.4 で説明したビット暗号化では n 個の暗号文が必要だったのに対し、一括暗号化を用いることで約 $1/N$ 倍の暗号データサイズの削減が可能となる (表 5 を参照)。

3.4 一括暗号文上の秘匿分析

一括暗号文 (5), (6) に対して、A, B 商品購入者数 $a = \sum_{i=1}^{n} a_i b_i$(=秘匿内積計算) などを暗号化したまま計算する方法を説明する。まず、§3.2 で説明した準同型暗号は、平文空間の環 R_t に対して準同型性を持つことに注意する (詳細は、[12, Section 3.2] を参照)。つまり、2 つの元 $m, m' \in R_t$ の暗号文 $\mathsf{ct} = \mathsf{Enc}(m, \mathsf{pk}), \mathsf{ct}' = \mathsf{Enc}(m', \mathsf{pk})$ に対して、

$$\begin{aligned} \mathsf{Dec}(\mathsf{ct} \dot{+} \mathsf{ct}', \mathsf{sk}) &= m + m' \in R_t \\ \mathsf{Dec}(\mathsf{ct} * \mathsf{ct}', \mathsf{sk}) &= m \cdot m' \in R_t \end{aligned}$$

を満たす。そこで、§3.3 で定義した 2 つの一括暗号文 $\mathsf{ct}_1 = \mathsf{pEnc}_1(A^{(j)}, \mathsf{pk}), \mathsf{ct}_2 = \mathsf{pEnc}_2(B^{(j)}, \mathsf{pk})$ 上の準同型暗号乗算 $\mathsf{ct}_1 * \mathsf{ct}_2$ の復号結果は

$$\begin{aligned} \mathsf{Dec}(\mathsf{ct}_1 * \mathsf{ct}_2, \mathsf{sk}) &= m_1(A^{(j)}) \cdot m_2(B^{(j)}) \in R_t \\ &= \left(\sum_{i=0}^{N-1} A_{i+1}^{(j)} x^i\right) \cdot \left(-\sum_{i=0}^{N-1} B_{i+1}^{(j)} x^{N-i}\right) \\ &\equiv \sum_{i=1}^{N} A_i^{(j)} B_i^{(j)} \bmod xR_t \end{aligned}$$

となる (環 R_t 上では、$x^N = -1$ となることに注意, また xR_t は x で生成される単項イデアルとする)。よって、準同型暗号乗算 $\mathsf{ct}_1 * \mathsf{ct}_2$ の復号結果から得られる R_t 上の多項式の定数成分が 2 つの N 次元部分ベクトル $A^{(j)}, B^{(j)}$ の内積計算 $\sum_{i=1}^{N} A_i^{(j)} B_i^{(j)}$ と一致する [11]。ゆえに、A, B 商品購入者数 $a = \sum_{i=1}^{n} a_i b_i$ に対応する秘匿計算として、

$$\sum_{i=1}^{\ell} \mathsf{pEnc}_1\left(A^{(i)}, \mathsf{pk}\right) * \mathsf{pEnc}_2\left(B^{(i)}, \mathsf{pk}\right) \quad (7)$$

を計算し、復号結果として得られる R_t の元の定数成分のみを出力すればよいことが分かる。特に、一括暗号文上の秘匿内積 (7) の計算処理コストとして、ℓ 回の準同型暗号加算と暗号乗算が必要となり、§2.4 のビット暗号化より約 N 倍の高速化が可能となる (表 5 を参照)。

[11] §3.3 で定義した環 R_t 上の昇順多項式 $m_1\left(A^{(j)}\right)$ と降順多項式 $m_2\left(B^{(j)}\right)$ は、その 2 つの R_t 上の乗算結果の定数項が 2 つのベクトル $A^{(j)}, B^{(j)}$ の内積計算結果になるように構成してある。

表 5: 従来法と提案手法との比較まとめ

	従来法 [21]	提案手法
技術の特徴	ビット暗号化 (§2.4 参照)	一括暗号化 (§3.3 参照)
暗号データサイズ	n 個の暗号文	$\ell = \lceil n/N \rceil$ 個の暗号文 † (約 $1/N$ 倍の削減効果)
秘匿内積計算処理コスト	n 回の暗号乗算 (式 (2) 参照)	ℓ 回の暗号乗算 (約 N 倍の高速化効果)

† N は Ring-LWE ベースの SHE 準同型暗号の格子次元パラメータ (詳細は、§3.2 を参照)。特に、今回の実装では格子次元 $N = 2048$ を利用 (§4 を参照)。

さらに、A 商品の購入者数 $R_1 = \sum_{i=1}^{n} a_i$ と B 商品の購入者数 $C_1 = \sum_{i=1}^{n} b_i$ の一括暗号文上の秘匿計算として、R_q 上の 2 つの元

$$C_1 = -\sum_{i=0}^{N-1} x^{n-j}, C_2 = \sum_{i=0}^{N-1} x^i = 2 - C_1 \in R_t$$

を用意し、R_1 と C_1 に対応する秘匿計算として

$$\sum_{i=1}^{\ell} C_1 * \mathsf{pEnc}_1(A^{(i)}, \mathsf{pk}) \quad (8)$$

$$\sum_{i=1}^{\ell} C_2 * \mathsf{pEnc}_2(B^{(i)}, \mathsf{pk}) \quad (9)$$

を計算し、復号結果として得られる R_t の元の定数成分のみを出力すればよい。また、A, B 商品未購入者数 d に対する秘匿計算については、上記の秘匿計算式 (7), (8), (9) を組み合わせることで計算可能である。以上の議論から、購買履歴データ分析で必要な表 3 内のデータ a, R_1, C_1, d に対して、非常に効率的な一括暗号化文上での秘匿計算が可能であることを示した。

4 実装報告

ここでは、顧客 ID 数 $n = 10,000$ に対して、§3 で提案した一括暗号化方式を利用した秘匿購買履歴データ分析の実装結果について報告する。

4.1 SHE 準同型暗号のパラメータ設定

今回の実装で用いた Ring-LWE ベースの SHE 準同型暗号のパラメータ (N, q, t, σ) について簡単に説明しておく。パラメータ設定方法は文献 [12] に基づいており、暗号方式の安全性を支える Ring-LWE 問題 [12, Lemma 3.4] に対する格子縮約攻撃などを利用した最良の攻撃法に対して、80-bit security を持つように設定した (一般的な LWE 問題に対する攻撃法とその見積もり攻撃時間の解析については、文献 [13] を参照のこと)。今回利用した具体的なパラメータは

$$(N, q, t, \sigma) = (2048, 63\text{-bit 素数}, 2^{14}, 8)$$

である。特に、σ については文献 [12] と同様に選択し、t については $n = 10,000$ 件の顧客数に対して、秘匿計算結

果 (7), (8), (9) が得られるように $t > n$ を満たす実装上有効な 2 冪整数を選んだ。また、SHE 準同型暗号は、暗号化の際に付加した乱数要素が準同型暗号操作によって増大し、復号範囲に収まらなくなると復号に失敗する特性を持つ。今回利用した SHE 準同型暗号における復号範囲は q に大きく依存し、$n = 10,000$ 件に対する秘匿購買履歴データ分析が可能なように q の bit サイズを選択した。最後に、格子次元 N については、SHE 準同型暗号の安全性に決定するパラメータであり、現在よく利用される公開鍵暗号の安全性レベルである 80-bit security を持つように、$N = 2048$ 次元を選択した (表 5 から分かるように、このパラメータ設定下で、一括暗号化を用いることでビット暗号化よりも暗号データサイズ約 1/2000 倍削減、秘匿内積計算コスト約 2000 倍の高速化効果が得られる)。

4.2 一括暗号化による秘匿集計

まず、図 2 に今回提案した一括暗号化を用いた秘匿購買履歴データ分析のデモ画面を示す (デモ画面内の暗号化データは 16 進数で表記している)。一括暗号化による秘匿集計は、§2.4 で説明したビット暗号化における秘匿集計同様、4 者集計モデルでその計算手順は以下である:

1. まず、分析者が §4.1 で設定したパラメータを用いて、準同型暗号の公開鍵 pk と秘密鍵 sk を生成し、公開鍵 pk のみを A 企業・B 企業とクラウドに公開しておく。

2. A, B 企業はそれぞれ、自身が持つ $n = 10,000$ 次元のバイナリ化購買履歴データ $(a_1, \ldots, a_n)$, $(b_1, \ldots, b_n)$ に対して、公開鍵 pk を用いて、$\ell = \lceil n/N \rceil = 5$ 個の一括暗号文 (5), (6) を生成し、クラウドに顧客 ID 付きの一括暗号文を送信する (デモ画面図 2 の A, B 企業の暗号化データが共に 5 個ずつであることに注意)。

3. クラウドは、A 企業と B 企業から送信された一括暗号文をデモ画面図 2 のように顧客 ID ごとにまとめる。次に、一括暗号文上の秘匿計算式 (7), (8), (9) を用いて、暗号化したまま購買履歴データ分析に必要な表 3 内のデータ a, R_1, C_1, d を計算し、秘匿集計結果のみを分析者に送信する。

4. 最後に、分析者は秘密鍵 sk を用いて秘匿集計結果を復号することでデータ a, R_1, C_1, d を求め、表 1 の購買履歴データ分析における類似度指標を計算し、それらの分析結果を A 企業と B 企業に送信する。

4.3 実装環境と実装結果

今回 §4.1 で設定したパラメータに対して、SHE 準同型暗号方式と一括暗号化による秘匿計算を実装した (ビット暗号化における実装結果 §3.1 と比較して、その効果を実感してもらいたい)。実装で用いた PC は Intel Xeon X3480, 3.07 GHz, 16GByte memory で、x86_64 アセンブリ言語による独自のソフトウェアライブラリを利用した。特に、暗号文空間の基礎となる環 R_q 上の効率的な乗算として、Karatsuba 乗算とモンゴメリ乗算法を利用した。

- 公開鍵サイズは約 31KByte で、秘密鍵サイズは約 16KByte となる。また、1 つの fresh 暗号文サイズが約 31KByte となり、秘匿内積計算 (7) 後の暗号文は環 R_q の成分が 3 つとなるので、そのサイズは約 46.5KByte となる。

- 処理性能に関しては、鍵生成に約 1.89ms、$N = 2048$ 次元ベクトルの一括暗号化に約 3.65ms、準同型暗号加算に約 0.001ms、準同型暗号乗算に約 5.31ms、復号処理に約 3.47ms かかった ($N = 2048$ 次元ベクトルの秘匿内積計算は約 5.31ms で処理可能)。

以上のことから、$n = 10,000$ 次元のバイナリ化購買履歴データを一括暗号化するのに、$\ell = 5$ 個の一括暗号文 (8) または (9) が必要となり、そのサイズは約 5×31 = 155KByte となる。さらに、一括暗号文上の秘匿内積計算 (7) は、$\ell = 5$ 回の準同型暗号加算と準同型乗算が必要となるので、その処理に約 $5 \times 5.31 = 26.55$ms かかると見積もれる (デモ画面図 2 における処理時間表記は、ファイルアクセスやデータ表示などで実際の暗号処理よりもかなり遅くなっていることに注意)。

5 まとめ

本論文では、クラウドコンピューティング分野への適用が期待されている暗号技術の 1 つである準同型暗号について、その実用化に向けた課題であった暗号データサイズと処理性能の両方の大幅削減に対し、Ring-LWE 問題をベースとした SHE 準同型暗号上の一括暗号化方式を提案することで、2048 次元ベクトルを 1 つの暗号文に集約可能とすると共に、2048 次元ベクトルの秘匿内積計算が汎用 PC 上で 6ms 以下で処理可能であることを示した (それにより、1 万件データの秘匿内積計算に 30ms 以下で処理可能)。これにより、本提案方式を用いることで、各企業が持つ購買履歴データを準同型暗号により秘匿したまま、実用的にマーケティング分析で利用される類似度指標を求めることが可能であることを示した。今回の提案方式は、複数医療データの相関分析、生体認証における照合距離計算などといったプライバシー保護とその情報利活用が求められる多様な分野にも適用範囲を広げることが可能であり、幅広い適用が期待される。

6 今後の課題

本論文では、2 企業間における購買履歴データ分析で利用できる一括暗号化方式を提案したが、本提案方式では 3 企業間以上の分析に利用することができない。したがって、3 企業間以上の分析に利用可能な方式を研究開発することが今後の課題である。また、本論文とは全く異なる生体認証における秘匿照合計算への適用も今後の課題であるが、生体認証への具体的な適用方法と処理性能などの効果については文献 [16] を参照してほしい（利用する暗号方式がイデアル格子ベースの SHE スキームで本論文の方式とは異なる）。

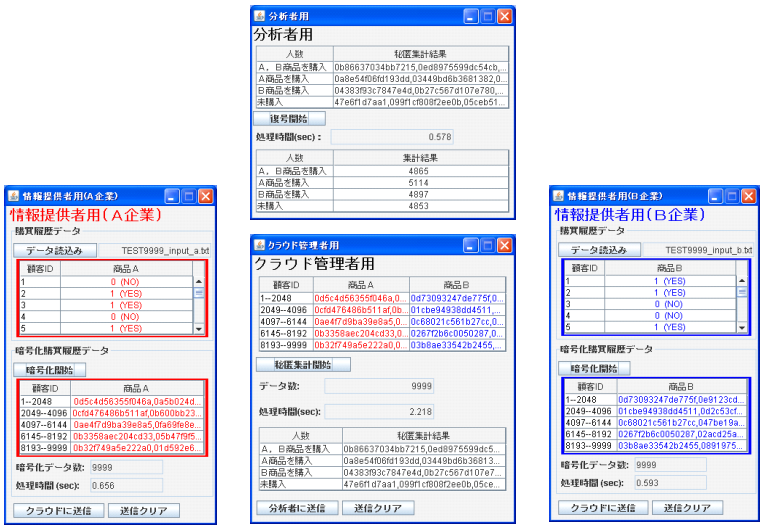

図 2: 一括暗号化を用いた秘匿購買履歴データ分析のデモ画面

参考文献

[1] D. Boneh, E. -J. Goh and K. Nissim, “Evaluating 2-DNF formulas on ciphertexts”, In *Theory of Cryptography – TCC 2005*, Springer LNCS 3378, 325-341, 2005.

[2] Z. Brakerski, C. Gentry and V. Vaikuntanathan, “(Leveled) fully homomorphic encryption without bootstrapping”, In *Innovations in Theoretical Computer Science – ITCS 2012*, 309-325, 2012.

[3] Z. Brakerski and V. Vaikuntanathan, “Fully homomorphic encryption from ring-LWE and security for key dependent message”, In *Advances in Cryptology – CRYPTO 2011*, Springer LNCS 6841, 505-524, 2011.

[4] Z. Brakerski and V. Vaikuntanathan, “Efficient fully homomorphic encryption from (standard) LWE”, In *Foundations of Computer Science – FOCS 2011*, 97-106, 2011..

[5] J. -S. Coron, A. Mandal, D. Naccache and M. Tibouchi, “Fully homomorphic encryption over the integers with shorter public-keys”, In *Advances in Cryptology – CRYPTO 2011*, Springer LNCS 6841, 487-504, 2011.

[6] R. Cramer, R. Gennaro and B. Schoenmakers, “A secure and optimally efficient multi-authority election scheme”, In *Advances in Cryptology – EUROCRYPT 1997*, Springer LNCS 1462, pp. 103 - 118, 1997.

[7] M. van Dijk, C. Gentry, S. Halevi and V. Vaikuntanathan, “Fully homomorphic encryption over the integers”, In *Advances in Cryptology – EUROCRYPT 2010*, Springer LNCS 6110, 24-43, 2010.

[8] C. Gentry, “Fully homomorphic encryption using ideal lattices”, In *Symposium on Theory of Computing – STOC 2009*, ACM, 169-178, 2009.

[9] C. Gentry and S. Halevi, “Implementing Gentry’s fully-homomorphic encryption scheme”, In *Advances in Cryptology – EUROCRYPT 2011*, Springer LNCS 6632, 129-148, 2011.

[10] C. Genry, S. Halevi and N. P. Smart, “Homomorphic evaluation of the AES circuit”, In *Advances in Cryptology – CRYPTO 2012*, Springer LNCS 7417, 850-867, 2012.

[11] IBM Press release, available at `http://www-03.ibm.com/press/us/en/pressrelease/27840.wss`.

[12] K. Lauter, M. Naehrig and V. Vaikuntanathan, “Can homomorphic encryption be practical?”, In *ACM workshop on Cloud computing security workshop – ACM CCSW 2011*, 113-124, 2011.

[13] R. Lindner and C. Peikert, “Better key sizes (and attacks) for LWE-based encryption”, In *International conference on Topics in cryptology – CT-RSA 2011*, Springer LNCS 6558, 319-339, 2011.

[14] P. Paillier, “Public-key cryptosystems based on composite degree residuosity classes”, In *Advances in Cryptology – EUROCRYPT 1999*, Springer LNCS 1592, 223-238, 1999.

[15] R. Rivest, A. Shamir and L. Adelman, “A method for obtaining digital signatures and public-key cryptosystems”, Communications of the ACM 21, 120-126, 1978.

[16] M. Yasuda, T. Shimoyama, J. Kogure, K. Yokoyama and T. Koshiba, “Packed homomorphic encryption based on ideal lattices and its application to biometrics”, to be presented at Workshop on Modern Cryptography and Security Engineering (MoCrySEn 2013).

[17] 石田実, 西尾チヅル, 佐藤忠彦, 「交互作用距離による音楽 CD の購買予測」, 日本マーケティング・サイエンス学会 第 84 回研究大会, 2008.

[18] 佐久間淳, 小林重信, 「プライバシー保護データマイニング」, 人工知能学会論文誌 Vol. 24 (No. 2), 2009.

[19] 佐久間淳, 高橋克己, 「 クラウドストレージにおける個人情報の利活用とプライバシー保護」, 情報処理 Vol.52, 2011.

[20] 佐藤智貴, 菊池浩明, 佐久間淳, 「傾向性の検定における秘匿疫学調査プロトコル」, SCIS2013, 3C1-4, 2013.

[21] 安田雅哉, 矢嶋純, 下山武司, 小暮淳, 「複数企業が持つ購買履歴データのクラウド秘匿集計」, SCIS2012, 3D2-5, 2012.

[22] 安田雅哉, 下山武司, 横山和弘, 小暮淳, 「イデアル格子準同型暗号を用いた秘匿内積計算」, SCIS2013, 2A3-2, 2013.