様式 C-19

科学研究費補助金研究成果報告書

平成２３年05月31日現在

機関番号：１７１０２
研究種目：基盤研究（Ｂ）
研究期間：2008～2010
課題番号：20300005
研究課題名（和文）プライバシ―データマイニングのための暗号プロトコルの設計と安全性評価

研究課題名（英文） Design and Security Analysis of Cryptographic Protocol for Privacy-Preserving Data Mining

研究代表者
櫻井　幸一（SAKURAI KOUICHI）
九州大学・システム情報科学研究院・教授
研究者番号：60264066

研究成果の概要（和文）：
データマイニング処理過程に医療情報や個人の好き嫌いなどの情報が介在し、プライバシーが侵害される、という問題がある。これを解決するため。データに雑音を入れたり、データの構造を維持したまま必要な演算や通信を行うプライバシーデータマイニングの研究が行われている。本研究では、安全性が理論的に証明できるプロトコルの設計を行なった。安全性の基準としては、安全な複数のプロトコルを組み合わせた場合にでも、安全性を保証できる汎用結合性(Universal Composability、UC)を用いる。このために、ＵＣ以前の安全性でしか証明されていなかった紛失通信など古典的な道具を、データマイ二ングに特化した形でＵＣ安全なものに再設計した。

研究成果の概要（英文）：
In the process of data mining, there exists a problem about privacy breach because we need to deal with health data, personal preferences, etc. There are approaches to privacy preserving data mining in which random noisy data are added to the original data or computation and communication are performed without breaking the structure of original data. In this research, we designed cryptographic protocols whose security can be proved in a theoretical way. For the security model, we use Universal Composability (UC) in which combining several protocols can be done securely. We designed protocols such as oblivious transfer for data mining whose security was proved only in the previous security model, and proved the security in the notion of UC.

交付決定額

（金額単位：円）

	直接経費	間接経費	合　計
２００８年度	3,500,000	1,050,000	4,550,000
２００９年度	2,600,000	780,000	3,380,000
２０１０年度	2,300,000	690,000	2,990,000
年度			
年度			
総　計	8,400,000	2,520,000	10,920,000

研究分野：暗号理論
科研費の分科・細目：情報学・情報学基礎
キーワード：暗号，データマイニング，プライバシー保護，安全性検証，普遍安全性

１．研究開始当初の背景

[80-90 年代] 暗号プロトコルは、1980 年代後半から M.Blum（UC バークレー）のグループに始まり、その弟子だちの MIT を中心に盛んに研究されてきた。手法としては、計算機科学と融合し、安全性が理論的に証明できるプロトコルの設計と解析が主流であった。1986 年、A.C Yao が安全な２者間計算”入力 x を持つ人物 A と入力 y を持つ人物 B が、ある関数 f(x,y)を協力して計算する、ただし、お互いに持っている入力が、相手に知らずに、実現する方法を提案した[Yao FOCS86]。この結果は、 Goldreich ら[STOC87]により、一方向性落とし戸関数を用いて、２者間から複数者間一般に拡張され、その後の暗号プロトコル研究の大きな流れとなる。

1990 年代前半の一般的な枠組みに対する理論計算機科学的な研究が、90 年代後半からは、インターネットの発達と現実のサービスの要求により、実際に実装・利用できる暗号プロトコル設計にかわっていく。特に、Yao の回路を利用した基本手法は、現実の実装に適用することは複雑であり、GMW らの一般的構成を、個別のプロトコルに適用するには、すべての暗号機能・部分プロトコルルーチンなどを具現する課題が残り、単純な実装では、非現実的な計算と通信を必要としたため、新たな暗号技術の研究開発が盛んになった。

[普遍結合安全性] 安全なプロトコルを組み合わせた場合に、安全性が必ずしも維持されないことは、(単純な並列処理の場合ですら)以前から知られていた。いかなる安全性を定義し証明すれば、暗号プロトコルが安全であるという良い性質が維持できるかという課題は、90 年代の未解決問題であった。これに対して、Canneti は汎用結合性(Universal Composability、UC)という概念を定義し解決の糸口を作った[Canetti, “Universally Composable Security: A New Paradigm for Cryptographic Protocols.” FOCS 2001]。従来の暗号の理論的安全性は、攻撃者をモデル化した議論であったが、Canetti は、プロトコルが利用される環境に対応するパラメータも加えた新しい UC 安全性を定義し、この定義を満たすプロトコルの組み合わせであれば、安全性が維持されるという理想の定理を確立した。現在の研究は、80 年第～90 年代に設計され安全と証明されていたプロトコルを見直し、UC 安全のもとで証明できるように再設計する研究が活発である。こうして改良され、UC 安全と証明されたプロトコル群であれば、自由に組み合わせても UC 安全ということで、その効果と応用は計り知れない期待がある。ただし、UC 安全を満たす基本プロトコルをどのように設計し、いかに証明を与えるかは自明でなく、現在もホットな研究が続けられている。

[研究代表者のグループ] 研究代表者のグループでは、プライバシー保護データマイニング(Privacy Preserving Data Mining, PPDM or PDM)を 2000 よりはじめた。データマイニングは、ここ１０年くらい活発に研究され、すでに多くの場面で実用化・商用化され、新聞などの大規模データに対しても、その効果が認識されている。

本研究でとりあげる課題は、データマイニング処理過程に医療情報や個人の好き嫌いなどの情報が介在し、プライバシーが侵害される、という問題である。このため、対象とするデータのプライバシーを維持したまま、必要な演算や通信を行うプライバシーデータマイニングの研究が、暗号研究者を中心に行われている。研究代表者のグループもクラスタリングをはじめとする代表的なデータマイニング・プロトコルをプライバシ―維持型に、暗号技術を用いて変形する研究を行ってきた。

２．研究の目的

本研究では、マルチパーティプロトコルの中でも、一番現実的応用性の要求のあるプライバシーデータマイ二ングをとりあげ、安全性が理論的に証明できるプロトコルの設計をめざし、このために、ＵＣ以前の安全性でしか証明されていなかった紛失通信など古典的な道具を、データマイ二ングに特化した形でＵＣ安全なものに再設計する。

暗号技術を用いずに、データにノイズをいれるなどの手法による PPDM は、Agrawal らによって提案され[D. Agrawal and C.C. Aggarwal,“On the Design and Quantification of Privacy Preserving Data Mining Algorithms,” SIGMOD2001]、すでにＩＢＭなどでも商用化されている。しかし、暗号プロトコルを利用したシステムは、守秘性の強度の立場からは優れているものの、計算量・通信量が多いため、未だ実用化されていない。それどころか、学会レベルでもほとんど実装の報告がない。本研究では、既存の方式の計算量・通信量の問題も解消し、実用化できるレベルのシステム設計を計画する。

３．研究の方法

[k 平均クラスター] 本研究では、データマイニングの中でも、もっと基本であり、よく利用されるクラスターアルゴリズムを取り上げる。K-means 法は一般には以下のような手順で計算を行う:

Step 1：K 個のクラスターの中心をランダムに設定する。

Step 2：距離計算によって、それぞれの個体を最も近い中心に割り当てる。

Step 3：クラスターごとに中心を計算しなおす。

すべてのクラスター中心が変化しなければ終了、それ以外は Step 2 へ戻る。プライバシー保護した k-means クラスタリングでは、データが複数サイトに分散している状況において、そのデータのプライバシーを保護したまま共同クラスタリングの結果のみを共有しなけれなならない。

[関連研究] 現在までのプライバシーを保護した K-means クラスタンリング研究は多数ある[Vaidya and Clifton. Proc. 9th ACM SIGKDD 2003] [Jagannathan and Wright. Proc. 11th ACM SIGKDD 2005] [Jha, Kruger, and McDaniel. 10th European Symposium On Research In Computer Security, 2005] [Jagannathan et al., Proc. 2006 SIAM International Conference on Data Mining ,2006.]が、どれも、汎用結合(UC)以前の安全性の議論であることに注意する。これらの既存提案では, 他パーティに自分のデータを明かさずにデータのクラスタリング実行するために, 必要な計算(加算と積算)を暗文化したまま行う、準同型性公開鍵暗号系を基本に用いる。また、クラスターアルゴリズムの STEP-2 と STEP-3 の処理は、一般的に知られる安全なマルチパーティ計算の結果を応用する。特に、既存提案では Step2 は スカラー計算プロトコルを対話型に利用し距離計算を実現する。Step3 では、古典的な回路を利用した一般的手法[Yao86]で実現している。

[問題点]しかしながら、既存の提案は以下の問題点が存在している：

1. 要求される計算量と通信量の膨大化より、実用化が不可能: 今までの研究論文の既存提案では、Yao の garbled gate や GMW モデルに基づく手法を基本に、UC 安全な方式を設計したので、効率改善に関する議論がとどんとおこなわれていない。

2. データの標準化を考えていない: 実際の応用で、異なるデータベースからデータの変数やばらつきなど偏差があるので、 それを克服するために、データの標準化を行わなければならない。また、従来の提案は、データ間の距離を計算する時に、それぞれのパーティに最初に k 個のクラスター中心を選ばせ、それぞれの点に対する対話式距離計算暗号プロトコルを実行する。 このため、計算量と通信量は多大である。

3. 安全性に関する議論と証明が不完全: k-means クラスタリングは通常のマルチパーティプロトコルと違って、暗号部分プロトコルを繰り返して使用する。この繰り返し処理において、次の計算結果は前の計算状態に依存している。このため、従来の Yao の構成や GMW の手法での安全性が保障できない。

本研究では、これらの問題を解決し、安全性の証明が付いたクラスターデータマイニングを提案する。

４．研究成果

4.1 現在までのプライバシーを保護した K-meansクラスタンリング研究は複数ある[Vaidya and Clifton. Proc.9th ACM SIGKDD2003. Jagannathan et al., Proc.2006SIAM International Conference on Data Mining2006.]が、どれも、UC以前の安全性の議論である。これらに対して、研究代表者のグループは、個別に問題点を指摘し、国際会議で発表した[Sa3]。しかし、この論文では、具体的詳細な問題点、とくにどのような機密漏洩が起こるか、の記述が不十分であった。

今回は、研究代表者がこれまで得ている結果を基に、従来方式の問題点の解析、理論解析を意識したプロトコルの設計と評価、計算機実験による性能評価および実験結果に基づく、理論の再検討を中心に、以下の研究を行った。

(a)プライバシーの概念を汎用結合性の概念に基づいて記号論的に定式化し、プライバシー保存プロトコルに関する安全性検証の方法を検討した。特に検証の基本方式の実例による正当性検証を考察した。

(b)乱数行列を利用した情報拡散方式の問題点解消を実験的に検討した。この対応手法の理論的解析を試みた。

(c)データマイニングプロトコルの実装の際に、データ正規化が安全性に関係することを指摘した。既存の安全性の証明が議論されたプロトコルでも、実装次第では安全ではなくなる、本研究では、改良法も提案した。

4.2 汎用結合性に関する結果

[A]K平均クラスタリングアルゴリズムに対して、本研究では2者間のマルチパーティ計算に特化し，それぞれの参加者が秘密のデータベース情報を保持している状況で，自分のデータベース情報を相手に秘匿したままで2者の結合したデータベースに対するクラスタリングを実行する手法を提案した．またこの提案手法に対し，過去の方式とは異なり，汎用結合性のモデルにおいて安全性証明を与えた．

[B]データマイニングでの基礎計算に有用な技術として秘密多項式評価法がある．これは多項式の入力となる a を持つ参加者と，多項式f(x)を入力として持つ参加者がお互いにaとfを秘密にしながらf(a)を計算するというものである．本研究では汎用結合性において安全性証明が可能な秘密多項式評価法を実現する

ための条件として準同型性Non-committingと呼ばれる2つの性質を満たす公開鍵暗号方式が必要であることを明らかにした.

4.3　UC コミットメントに関する成果
プライバシーデータマイニングで利用する暗号プロトコルをより強力な攻撃者に対して耐性があるように設計する道具としてコミットメントがある．通常 UC 安全なコミットメントの設計は非常に困難であるが，ある種の Trusted Setup を仮定することで，本研究では Paillier 暗号と呼ばれる準同型性を持つ暗号を使用し，UC 安全性を持つコミットメントを構成した[10, Zhu, Araragi, Nishide, and Sakurai, "Adaptive and Composable Non-interactive String-Commitment Protocols", International Conference on Security and Cryptography (SECRYPT), 2010].
また UC 安全なマルチパーティプロトコルを実現するための基礎技術である Non-committing 暗号の提案も行った[9, Zhu, Araragi, Nishide, and Sakurai, "Adaptive and Composable Non-committing Encryptions", 15th Australasian Conference on Information Security and Privacy (ACISP), 2010.], [11, Zhu, Araragi, Nishide, and Sakurai, "Universally Composable Non-committing Encryptions in the Presence of Adaptive Adversaries", International Conference on Security and Cryptography (SECRYPT), 2010]. Non-committing 暗号は通信が必要になるが UC 安全の証明を可能とする技術で，本研究では Decision Diffie-Hellman 仮定と呼ばれる広く受け入れられた計算困難性仮定に基づいた比較的シンプルな構成を提案した.

蘇春華（博士課程・学振特別研究員）は、共同研究者として、次の内容で学位を取得した：A Cryptographic Study of Privacy-Preserving Data Mining in Distributed Environment（分散環境におけるプライバシー情報保護を備えたデータマイニングの暗号手法に関する研究）
(蘇-1)すでに k 平均クラスタリングに対する手法は、いくつか知られているが、これら既存プロトコルを分析し、実行時の部分情報の漏れを具体的に解析し、また出力結果が正しくない場合も生じるという欠点を指摘した。さらに、既存提案を実装する場合には、異なるデータベース上でデータ変数などにばらつきがある場合、計算誤差を克服するためにデータの標準化が必要となる。蘇は相手に自分のデータを公開せず、標準化用パラメータを算出し利用する共同データ標準化手法を新たに開発し、この誤差を減らすことに成功した。
(蘇-2)構造化されていないテキストデータに対する凝集型ドキュメントクラスタリングのプライバシー保護化研究対象とした。まず、蘇は既存の単一暗号文に対する検索手法を複数の暗号化されたドキュメントに拡張するために、各パーティのドキュメントを二進数列で表記し、同じ単語を持つドキュメントとその単語を組み合わせる手法を開発した。つぎに各パーティのデータベースに格納しているドキュメントの内容を公開せずに、その中の単語出現頻度や相関関係などによって類似する暗号化ドキュメントを検出し、クラスタを生成するステップを導入した。このクラスタの中で重複しているクラスタの合併も、準同型暗号を用いて、安全に実現できることを示した。
(蘇-3)相関ルールの抽出手法におけるプライバシー問題を取り上げた。従来提案はアプリオリアルゴリズムを可換暗号を用いて実現したものであり、頻出アイテムの生成には、データベースを頻繁にスキャンし、膨大な量の候補アイテム生成が必要となる。さらに候補アイテムを生成するたびに、暗号化と各参加者への送信を要するという課題もある。これらの問題を解決するため、蘇は頻出パターン木を導入することで、候補アイテムの生成を不要にし、データベースに対するスキャンが２回だけで済む方法を開発し、準同型暗号を用いて安全なプロトコルの実現に成功した。
(蘇-4)密度推定データクラスリング分散計算法を提案した。既存方式では信頼第三者の存在を仮定し、この第三者が各クライアントの情報を集める。第三者はデータを悪用する可能性があり、また信頼第三者の存在を仮定できない応用場面もある。蘇は検証可能な秘密分散方式を導入することで第三者の協力なしに、複数のクライアントが同時にランダムノイズを生成し、ランダム化されたデータからデータクラスリング分散計算を行う手法を設計した。さらに、フィルタリング攻撃耐性の評価として、ランダムデータ行列の固有値分布と攻撃成功確率を実験的に分析し、提案方式の安全性限界も与えた。

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　13 件）
1. Fangming Zhao, Takashi Nishide, and Kouichi Sakurai, "Realizing Fine-grained and Flexible Access Control to Outsourced Data with Attribute-based Cryptosystems", 7th

Information Security Practice and Experience Conference (ISPEC), LNCS 6672, pp.83--97, Springer-Verlag, 2011(査読有)

2. Takashi Nishide and Kouichi Sakurai, "Distributed Paillier Cryptosystem without Trusted Dealer", 11th International Workshop on Information Security Applications (WISA2010), LNCS 6513, pp.44-60, 2011. (査読有)
3. 志村正法, 宮崎邦彦, 西出隆志, 吉浦裕, "秘密分散データベースの構造演算を可能にするマルチパーティプロトコルを用いた関係代数演算", 情報処理学会論文誌, Vol.51, No.9, pp.1563--1578, 2010. (査読有)
4. Huafei Zhu, Tadashi Araragi, Takashi Nishide, and Kouichi Sakurai, "Adaptive and Composable Non-interactive String-Commitment Protocols", International Conference on Security and Cryptography (SECRYPT), pp.354--361, SciTePress, 2010. (査読有)
5. Huafei Zhu, Tadashi Araragi, Takashi Nishide, and Kouichi Sakurai, "Universally Composable Non-committing Encryptions in the Presence of Adaptive Adversaries", International Conference on Security and Cryptography (SECRYPT), pp.389-398, SciTePress, 2010. (査読有)
6. Huafei Zhu, Tadashi Araragi, Takashi Nishide, and Kouichi Sakurai, "Adaptive and Composable Non-committing Encryptions", 15th Australasian Conference on Information Security and Privacy (ACISP), LNCS 6168, pp.135--144, 2010. (査読有)
7. Amril Syalim, Takashi Nishide, and Kouichi Sakurai, "Preserving Integrity and Confidentiality of a Directed Acyclic Graph Model of Provenance", 24th IFIP WG11.3 Working Conference, Data and Applications Security and Privacy (DBSec), LNCS 6166, pp.311-318, 2010. (査読有)
8. 蘇春華, "A Cryptographic Study of Privacy-Preserving Data Mining in Distributed Environment (分散環境におけるプライバシー情報保護を備えたデータマイニングの暗号手法に関する研究)", 博士論文(九州大学), 2009. (査読無)
9. Chunhua Su, Feng Bao, Jianying Zhou, Tsuyoshi Takagi, and Kouichi Sakurai. "Security and Correctness Analysis on Privacy-Preserving k-means Clustering Schemes". IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, Vol.E92-A, No.4, 1246-1250, April 2009. (査読有)
10. Chunhua Su, Justin Zhan and Kouichi Sakurai "Importance of Data Standardization in Privacy-Preserving K-means Clustering", International Workshop on Privacy-Preserving Data Analysis in conjunction with Database Systems for Advanced Applications (DASFAA 2009), Springer LNCS 5667, pp.276-286, April, 2009. (査読有)
11. Chunhua Su, Feng Bao, Jianying Zhou, Tsuyoshi Takagi, Kouichi Sakurai. "Distributed Noise Generation for Density Estimation Based Clustering without Trusted Third Party". IEICE Transactions 92-A(8), 1868-1871, 2009 (査読有)
12. Chunhua Su and Kouichi Sakurai. "A Distributed Privacy-Preserving Association Rules Mining Scheme Using Frequent-Pattern Tree", The Fourth International Conference on Advanced Data Mining And Applications(ADMA2008), Springer LNAI5139, pp. 170-181, Chengdu, China, October, 2008. (査読有)
13. Chunhua Su, Feng Bao, Jianying Zhou, Tsuyoshi Takagi, and Kouichi Sakurai. "A New Scheme for Distributed Density Estimation based Privacy-Preserving Clustering". Proceedings of 2008 International Conference on Availability, Reliability and Security, (AReS'08) pp. 48-57, IEEE Computer Society, Barcelona, Spain, March 2008. (査読有)

〔学会発表〕(計 12 件)

1. 西出隆志, Amril Syalim, and 櫻井幸一, "IFIP WG 11.3 Working Conference on Data and Applications Security and Privacy 2010 会議参加報告", コンピュータセキュリティシンポジウム(CSS), 6pages, 岡山, 10 月 21 日, 2010.
2. Amril Syalim, Takashi Nishide, and Kouichi Sakurai, "Preserving Integrity and Confidentiality of a Directed Acyclic Graph Model of

Provenance”，情報科学技術フォーラム(FIT)，2pages，福岡，9月9日，2010.
3. 志村正法，西出隆志，吉浦裕，″(k,n)秘密分散上での関係データベースの構造演算の実現”，暗号と情報セキュリティシンポジウム(SCIS)，6pages，香川，1月21日，2010.
4. Takashi Nishide and Kouichi Sakurai，″Distributed Key Generation for Paillier Cryptosystem”，暗号と情報セキュリティシンポジウム(SCIS)，6pages，香川，1月19日，2010.
5. Chunhua Su，Tadashi Araragi，Takashi Nishide，and Kouichi Sakurai，″Efficient Adaptively Secure Oblivious Polynomial Evaluation with Universal Composability: If Homomorphic and Non-committing Encryption Exists”，暗号と情報セキュリティシンポジウム(SCIS)，6pages，香川，1月19日，2010.
6. 蘇 春華，櫻井幸一，″線形変換したデータ行列の固有値分布と安全性評価″，CDROM 3F4-4，5ページ，滋賀，2009年1月22日.
7. 蘇 春華，櫻井幸一，″プライバシー保護したK平均クラスタリングにおけるデータ標準化問題″，CDROM 3F4-5，5ページ，滋賀，2009年1月22日.
8. 蘇 春華，櫻井 幸一，″分散環境におけるテキストマイニングのプライバシー保護手法と実験分析″，コンピュータセキュリティシンポジウム 2008 (CSS 2008)，CDROM C5-1，6ページ原稿，沖縄，2008年10月9日.
9. 蘇 春華，櫻井幸一，″頻出パターンを利用した安全な相関ルール発見手法″，信学技報，vol. 108，no. 162，ISEC2008-59，pp. 177-182，2008年7月25日.
10. Chunhua Su，Feng Bao，Jianying Zhou，Tsuyoshi Takagi，Kouichi Sakurai，″Security Problems in Existing Privacy-preserving K-means Clustering Schemes″，電子情報通信学会全国総合大会，北九州学研都市，2008年3月19日.
11. 蘇 春華，櫟 粛之，櫻井幸一，″Universally Composable Oblivious Polynomial Evaluation″，電子情報通信学会全国総合大会，北九州学研都市，2008年3月18日.
12. 蘇 春華，櫟粛之，櫻井幸一，″汎用的結合可能なK平均クラスタリングプロトコル″，2008 年 暗号と情報セキュリティシンポジウム (SCIS 2007)，3E-3，宮崎，2008年1月24日.

〔図書〕(計　0件)

〔産業財産権〕
○出願状況（計1件）
名称：理想機能実現装置、方法、プログラム
発明者：櫻井幸一
権利者：櫻井幸一
種類：特許
番号：特願2010-221865
出願年月日：2010-09-30
国内外の別：国内
○取得状況（計0件）
名称：
発明者：
権利者：
種類：
番号：
取得年月日：
国内外の別：

〔その他〕
ホームページ等
http://itslab.csce.kyushu-u.ac.jp/index-j.html

６．研究組織
(1)研究代表者
櫻井 幸一　(SAKURAI　KOUICHI)
九州大学・システム情報科学研究院・教授
研究者番号：60264066

(2)研究分担者
堀 良彰　(HORI YOSHIAKI)
九州大学・システム情報科学研究院・准教授
研究者番号：90264126
高橋 健一　(TAKAHASHI KENICHI)
鳥取大学・鳥取大学大学院工学研究科 情報エレクトロニクス・准教授
研究者番号：30399670
橋本 康史　(HASHIMOTO YASUFUMI)
財団法人九州先端科学技術研究所・研究員
研究者番号：30452773
西出 隆志　(NISHIDE TAKASHI)
九州大学・システム情報科学研究院・助教
研究者番号：70570985

(3)連携研究者
櫟 粛之　(ARARAGI TADASHI)
NTT・コミュニケーション科学基礎研究所・主任研究員
研究者番号：00396136
蘇 春華　(CHUNHUA SU)
九州大学・大学院システム科学府・博士(学振特別研究員)
（現在シンガポール Institute for Infocomm Research 勤務）
研究者番号：なし
EOF