別紙 1

論文の内容の要旨

論文題目	サービス事業者間データ連携における分散匿名化手法の提案
学位申請者	竹之内　隆夫

第１章：序論

近年，複数のサービス事業者が保持するユーザのプライバシ情報を含むパーソナル情報を連携し，新たな知見を得ることによるサービス創出が期待されている．そのため，パーソナル情報を必要最小限の開示に留めながら結合し，個人が特定されない形に加工した結合匿名テーブルを生成・開示する分散匿名化手法が求められている．しかし，既存の分散匿名化手法では，双方のサービス事業者のユーザ集合が一致しない場合にユーザのパーソナル情報がそのサービス事業者に保持されているか否かというユーザ存在情報が他方のサービス事業者に漏洩する問題があった．

そこで本論文では，双方のサービス事業者のユーザ集合が一致しない場合にユーザ存在情報が漏えいする問題を解決した新たな分散匿名化手法を提案する．

第２章：関連研究

本章では，本論文が提案するユーザ存在情報の漏洩を軽減した分散匿名化手法に関連した研究について説明する．まず，k-匿名化について説明し，その後ユーザ存在情報を隠蔽した既存の集中型の匿名化手法について説明する．続いて，分散環境における匿名化である分散匿名化の既存研究について説明する．さらに，分散匿名化で利用されるセキュア計算とMulti Party Computationについてと，プライバシを保持したデータマイニング手法であるPrivacy Preserving Data Miningについて説明する．

第３章：分散匿名化におけるユーザ存在情報の漏洩の課題

本論文で解決しようとしている分散匿名化におけるユーザ存在情報が漏えいする問題の詳細を説明する．この問題は「(1)結合匿名テーブルによるユーザ存在情報の漏洩問題」と「(2)ユーザID通知によるユーザ存在情報の漏洩問題」の２つに分割することができ，それぞれの課題の詳細を説明する．

第４章：ユーザ存在情報の漏洩を軽減した分散匿名化手法の提案

分散匿名化におけるユーザ存在情報が漏洩する問題を軽減するために，新たにδ-site-presenceというプライバシ指標を提案する．この指標は，ユーザ存在情報が漏洩する可能性の許容範囲を示した指標である．

そして，提案した指標を満たしつつ，データマイニング等での有用性を保った結合匿名テーブルを生成するための新たな分散匿名化のプロトコルを提案する．このプロトコルは，存在するユーザと存在しないユーザの区別を困難にさせるダミーユーザを導入することで，ユーザ存在情報の漏洩を軽減している．

第５章：評価実験

提案手法を米国の国勢調査をもとに作成された評価データと実際のレセプトデータ（診療報酬明細情報）を用いて評価した結果を説明する．提案手法と既存の分散匿名化手法との実行結果を比較した結果，一定の条件下において提案手法は既存手法よりも大幅にデータの有用性を保った匿名化が行えることを確認した．また，提案手法と既存の集中型のユーザ存在隠蔽の匿名化手法との比較を行い，既存手法は既存手法とほぼ同等に有用な匿名化が行えることを確認した．

さらに，複数の医療機関が保持する医療データを結合・分析する場面での利用を想定し，データ分析を行った際の集計誤差を計測した結果，提案手法はユーザ存在情報の漏えいを軽減しながらも相対誤差15%以下でデータ分析が可能であることがわかった．これは，近年言われている医療の効率化や医療サービスの質向上のための医学研究に適用できると考えられる．

第６章：計算量・通信量と安全性の評価

提案手法の計算量・通信量の評価を行い，双方の事業者が持つ情報を開示せずに単純な関数計算を行う既存のセキュア計算の計算量・通信量と比較した．その結果，提案手法の計算量・通信量は既存のセキュア計算の計算量・通信量と比較して，大幅な増加がないことを確認した．また，提案手法のプロトコルの安全性を暗号理論で用いられるシミュレータを用いた評価手法によって証明し，プライバシ性の高いパーソナル情報やユーザ存在情報が漏洩しないことを確認した．

第７章：結論

本論文では，双方のサービス事業者のユーザ集合が一致しない場合にユーザ存在情報が漏えいしてしまう問題を解決した新たな分散匿名化手法を提案した．提案手法を用いることによって，国勢調査データや医療データにとどまらず，様々な種類のパーソナル情報をサービス事業者間で安全にデータ連携することができ，新たなサービスが創出されることが期待できる．今後は，適切な並列化を行うことによるスケーラビリティの向上や，更なる有効性の向上などが望まれる．

別紙２

論文審査の結果の要旨

学位申請者氏名　　竹之内　隆夫

審査委員主査　　大須賀　昭彦

委員　　田中　健次

委員　　小池　英樹

委員　　大森　匡

委員　　川村　隆浩

近年，複数のサービス事業者が保持するユーザのプライバシ情報を含むパーソナル情報を連携(データ連携)し，新たな知見を得ることによるサービス創出が期待されている．データ連携におけるプライバシ上の問題として，「(問題1)結合したデータの個人が特定されてしまう問題」とデータを結合する際に「(問題2)必要以上にデータを開示してしまう問題」が従来から指摘されている．そして，これらの問題を解決する既存技術として分散匿名化手法が知られている．分散匿名化手法とは，パーソナル情報を必要最小限の開示に留めながら結合し，個人が特定されない形に加工した結合匿名テーブルを生成・開示する手法である．

しかし，既存の分散匿名化手法では，双方のサービス事業者のユーザ集合が一致しない場合にユーザのパーソナル情報がそのサービス事業者に保持されているか否かというユーザ存在情報が他方のサービス事業者に漏洩する問題があった．この問題は，分散匿名化手法を実際のアプリケーションに適用することを考えると，多く発生すると考えられる．

そこで本研究では，分散匿名化手法を実際のアプリケーションに適用することを目指し，既存の分散匿名化手法が解決している「(問題1)結合したデータの個人が特定されてしまう問題」と「(問題2)必要以上にデータを開示してしまう問題」だけでなく，さらに「(問題3)ユーザ存在情報が漏えいしてしまう問題」も解決する新たな分散匿名化手法を確立することを目的としている．

第２章では，本論文で提案するユーザ存在情報の漏洩を軽減した分散匿名化手法に関連する既存研究について述べられている．

第３章では，分散匿名化におけるユーザ存在情報が漏えいする問題の詳細を説明している．この問題は「(問題3-1)結合匿名テーブルによるユーザ存在情報の漏洩問題」と「(問題3-2)ユーザID通知によるユーザ存在情報の漏洩問題」の２つに分割することができ，それぞれの問題の詳細が述べられている．

第４章では，ユーザ存在情報が漏洩する問題を軽減するために，新たにδ-site-presenceというプライバシ指標を提案している．この指標は，ユーザ存在情報が漏洩する可能性の許容範囲を示した指標である．
さらに，提案した指標を満たしつつ，データマイニング等での有用性を保った結合匿名テーブルを生成するための新たな分散匿名化のプロトコルを提案している．このプロトコルは，存在するユーザと存在しないユーザの区別を困難にさせるダミーユーザを導入することで，ユーザ存在情報の漏洩を軽減するというものである．

第５章では，提案手法の有用性を評価するために，複数の医療機関が保持する医療データを結合・分析する場面での利用を想定し，データ分析を行った際の集計誤差を計測している．評価に用いたデータは，実際のレセプトデータ（診療報酬明細情報）と匿名化の研究でベンチマークとして利用されている米国の国勢調査をもとに作成された評価データである．評価の結果，一定の条件下において提案手法は既存手法よりも大幅にデータの有用性を保った匿名化が行えることを確認している．また，提案手法と既存の集中型のユーザ存在隠蔽の匿名化手法との比較を行い，既存手法は既存手法とほぼ同等に有用な匿名化が行えることを確認している．さらに，提案手法が対応可能なユーザ数の評価を行い，10000人以下のユーザ数であれば十分有用であることを確認している．このことから，提案手法は，近年言われている医療の効率化や医療サービスの質向上のための医学研究に適用できると考えられる．

第６章では，提案手法のプロトコルの安全性を暗号理論で用いられるシミュレータを用いた評価手法によって証明し，プライバシ性の高いパーソナル情報やユーザ存在情報が漏洩しないことを確認している．さらに，提案手法の計算量・通信量の評価を行い，双方の事業者が持つ情報を開示せずに単純な関数計算を行う既存のセキュア計算の計算量・通信量と比較している．その結果，提案手法の計算量・通信量は既存のセキュア計算の計算量・通信量と比較して，大幅な増加がないことを確認している．

第７章では，提案内容と評価結果をまとめ，提案手法を用いることで国勢調査データや医療データにとどまらず，様々な種類のパーソナル情報をサービス事業者間で安全にデータ連携することができ，新たなサービスが創出されることが期待できることを示している．

以上で述べてきたように，本研究成果はオリジナリティに富み，実用面でも高い価値を持つものである．よって，本論文は博士(工学)の学位論文として十分な価値を有するものと認める．

以上