公募企画

公募企画シンポジウム6
安心・安全なビッグデータの流通プラットホームとセキュリティ基盤技術

2017年11月22日(水) 08:45 〜 10:45 C会場 (10F 会議室1001)

[3-C-1-PS6-5] セキュリティ基盤技術による診療情報の統合的利活用

田中 勝弥[1], 山本 隆一[2] （1.東京大学大学院医学系研究科, 2.医療情報システム開発センター）

これまでに複数の大規模な医療情報データベースが構築・運用されてきており、医療機関の外部と情報を共有・収集し解析する多施設間の研究利用も盛んに行われている。また、次世代医療基盤法の成立にみられるように、集積された診療情報の研究開発や公益目的への二次利用は今後さらに促進されることが期待される。一方で、改正個人情報保護法の施行や医学系研究に関する倫理指針の改正に見られるプライバシー保護や患者同意の管理への対応は必須の課題でもある。平成26年度より開始された JST CREST「ビッグデータ統合利活用のための次世代基盤技術の創出・体系化」研究プロジェクトでは大規模データの二次利用とプライバシー保護への課題に対し、クラウド上に電送・保管される診療情報の安全管理、分析・二次利用に対するプライバシーリスク評価を中心にその要素技術とシステムイメージについて検討を重ねてきた。本稿では、1) ORAM（ Oblivious RAM）技術を応用した暗号・分散ストレージによる医療機関等の間での放射線画像送受システム、2) PSI（ Private Set Intersection）を利用した SS-MIX2からの安全な診療データ収集、3) 二次利用シーンでの解析対象データセットに対するプライバシーリスクアセスメント、の3つのおもな医療情報分野での取り組みについて紹介する。

セキュリティ基盤技術による診療情報の統合的利活用

田中　勝弥[*1]、山本　隆一[*2]
*1 東京大学大学院医学系研究科、*2 医療情報システム開発センター

Integrated utilization of clinical information using security infrastructure technology

Katsuya Tanaka[*1], Ryuichi Yamamoto[*2]
*1 Graduate School of Medicine, The University of Tokyo, *2 Medical Information System Development Center

With the progress of information technology in healthcare, secondary use which utilizes medical information to public interest purposes also come to be promoted. Large-scale medical information databases have been built in place, and electronically transferring the information of the medical institution to the outside, research use to store also have been actively carried out. Along with this, for resolving conflicting problems of public interest use and privacy, which is an important factor that must be subsequently considered. On the other hand, spills of personal information data by targeted attacks, etc., have been occurred but it is a situation that people must strictly carry out safety management for the operation of medical information systems, even when technical measures are not established. Risk reduction measures have been taken by many of human measures. It is important to establish information infrastructure technologies for utilization of medical information on the principles of privacy protection. In this paper, the research project, "Creation and Systematization of the Next Generation Platform Technology for Big Data Integrated Utilization" in the JST Crest is described. With respect to issues of privacy protection problems as described above, for the realization of secure transmission and storage of medical information to the cloud services and more privacy risk management in the analytical use of medical information, elemental technologies and concepts of experimental systems are introduced.

Keywords:　Privacy Protection, Information Security, Electronic Medical Records

1. はじめに

これまでに複数の大規模な医療情報データベースが構築・運用されてきており、医療機関の外部と情報を共有・収集し解析する多施設間の研究利用も盛んに行われている。また、次世代医療基盤法の成立にみられるように、集積された診療情報の研究開発や公益目的への二次利用は今後さらに促進されることが期待される。一方で、改正個人情報保護法の施行や医学系研究に関する倫理指針の改正に見られるプライバシー保護や患者同意の管理への対応は必須の課題でもある。平成 26 年度より開始された JST CREST「ビッグデータ統合利活用のための次世代基盤技術の創出・体系化」研究プロジェクトでは大規模データの二次利用とプライバシー保護への課題に対し、クラウド上に電送・保管される診療情報の安全管理、分析・二次利用に対するプライバシーリスク評価を中心にその要素技術とシステムイメージについて検討を重ねてきた。医療分野に関するテーマとしては、医療情報システムが保有し、二次利用が期待される診療情報に対して、おもに組織外部への情報の電送や格納、臨床研究目的の解析をターゲットとして、

- データの消失や流出といった、データの安全管理にかかわるセキュリティ対策に寄与する技術要素を実装したパイロットシステムを示すことにより、
- 統計解析などの二次利活用における匿名加工に対して、プライバシーリスク評価可能な方法を提案し、実証的に示すこと、

に主眼を置いている。

2. セキュリティ基盤技術

診療データの情報伝送や外部ストレージへの格納に対しては、安全管理上の対策や、二次利用に対するプライバシー保護の課題があり、双方を満たす技術的対策と、実用可能性を実証する必要がある。いくつかのセキュリティ技術を実装した、実証システムによりその実用可能性の評価を進めている。個々のセキュリティ対策における基本的なセキュリティ関連ライブラリは、本研究プロジェクトの大阪大学宮地研究室、および KDDI 研究所清本らにより開発されたソフトウェア群を必要に応じて拡張しながら適用する。

本稿では、1) ORAM（Oblivious RAM）技術を応用した暗号・分散ストレージによる医療機関等の間での放射線画像送受システム、2) PSI（Private Set Intersection）を利用した SS-MIX2 からの安全な診療データ収集、3) 二次利用シーンでの解析対象データセットに対するプライバシーリスクアセスメント、の3つのおもなテーマの取り組みについて紹介する。

2.1 Oblivious RAM(ORAM)

Oblivious RAM（ORAM）は、サーバ内に保持されたデータおよび、保持されるデータへのアクセスパターンを秘匿するための技術であり、クラウドサービスへの一時的なデータ保管に際し、サーバに保管されたデータやデータへのアクセスパターンを第三者に対して秘匿する技術である。たとえば、診療情報の医療機関間の送受を行うための、クラウドストレージを民間事業者が運用管理する場合を想定すると、ORAM により、ストレージに保持されるデータやストレージへのアクセスパターンを事業者や悪意を持った攻撃者に対して秘匿可能な診療情報送受信システム構築への寄与が期待できる。本稿ではこれを応用した PDI 形式の放射線検査データの送受信プロトタイプシステムについて記載する。

プロトタイプシステムの概要を図 1 に示す。CD-R 等に書き込まれて可搬媒体としてやりとりされている PDI 形式データをパブリッククラウド上で送受信可能なシステムである。本システムが実現されることにより、CD-R 等への検査データの書き込

みや受け取り側での PACS への取り込み作業に要する時間が軽減されることが期待される。

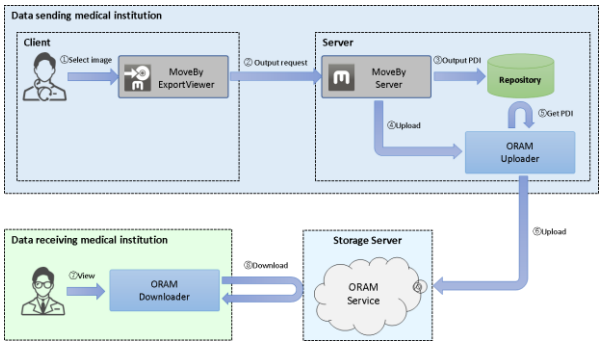

図 1　ORAM によるデータ転送システムの概要

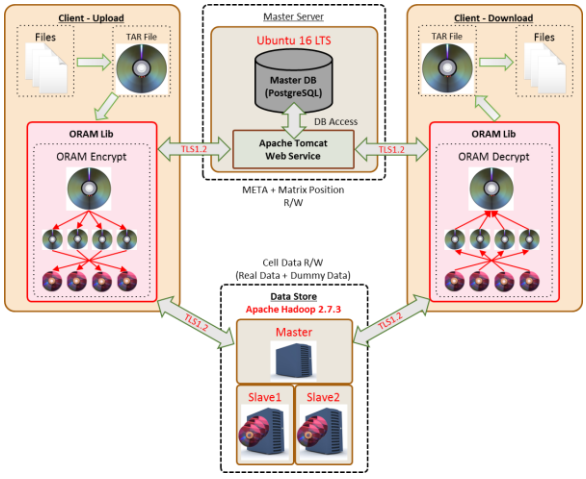

図 2　プロトタイプシステムにおけるデータ処理の概要

クラウド上のストレージおよび送受信インターフェイスに宮地らが開発した Matrix ベース ORAM 機能 [1]を付与しており、CD イメージ相当のデータを分割後暗号化し、分散ファイルシステム内に管理する。

サービスへのアクセスのための認証やデータの暗号化には公開鍵基盤を使用し、認証と暗号化で鍵ペアを分けることとし、サービス認証は TLS1.2 によるクライアント認証とする。また、分割後データの暗号化には CMS（Cryptographic Message Syntax）を採用した。クラウドストレージとしては、Apache Hadoop による実装とした。なお、非 VPN 環境下のインターネット上での利用を想定した点が特徴である。

結果、ORAM によるダミーデータ挿入による通信冗長処理や分割後データのシャッフルによるストレージトランザクション上の冗長処理が介在するもののアップロード・ダウンロードに対してそれぞれ、228Mbps/178Mbps の処理性能が確認できた。試作システムをベースとしたオンライン放射線画像システムを東京大学医学部附属病院と同病院医療連携機関登録制度に参加する数病院への導入に向けて検討、調整を開始している。

2.2 PSI(Private Set Intersection)

宮地らの PSI [2]は、各施設に分散されたデータセット群を突合し、識別子を突合しながら、異なるデータセット間で必要項目を連結したり、サブセットを抽出するなどの、集合演算が可能な機能を提供する。ここでは、多施設間に配置された SS-MIX2 標準化ストレージ上のデータを PSI に適合させる試みについて記述する。SS-MIX2 標準化ストレージそのものを、FUSE（Filesystem in Userspace）と RDMBS（Relational Database Management System）上に実装することで、SS-MIX2 メッセージ内のセグメント、フィールドをデータ項目単位で検索可能とするシステムを試作した。システムの概要を図 3 に示す。

ファイルシステム上に生成された HL7 v2 形式のメッセージは、ファイルの書き込みと同時に RDB 内の表領域に LOB として格納される。メッセージファイルの追加・削除等の情報が DBMS 内でリアルタイムに追跡され管理テーブルに保持される。ファイルの管理情報をもとに、定期的に LOB に格納された HL7v2 メッセージを PL/SQL により作成したモジュールにより解析し、別のテーブルへパース結果を反映する。外部システムからこのテーブルを検索することで、SS-MIX2 標準化ストレージ内のデータを患者横断的に検索する仕組みである。

処方オーダメッセージ（OMP-01）をパースするための PL/SQL を作成し、検証用の OMP-1 メッセージファイル群（ファイル数：22004、総容量：199MB）に対する処理時間の測定を行い、表1の結果を得た。

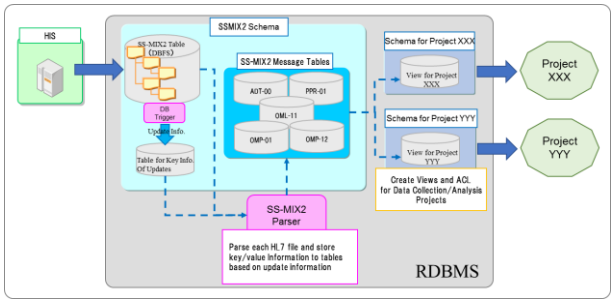

図 3 SS-MIX2 検索プロトタイプシステムの概要

表 1　PSI 向けストレージ性能測定結果

コピー：HDD→DBFS	1902 sec (Ave: 837 Mbps)
コピー：DBFS→HDD	147 sec (Ave: 10.8 Mbps)
削除：DBFS	858 sec (Ave: 25.6 files/sec)
メッセージパース	101 sec (Ave: 218 files/sec)

SS-MIX2 標準化ストレージのデータを RDB と連携させ管理する実装が可能となったことで、RDB としてのユーザ、スキーマ定義、ACL により、データ項目単位の抽出制御、外部プロジェクトからのアクセス制御などをワンストップで実現可能な段階になったと考える。PSI 機能の適用については PSI 自身の RDB への対応を含め、今後検討を進める。

2.3 リスク評価機能付き匿名化

現在、多施設から SS-MIX2 標準化ストレージのデータをバックアップし、匿名加工したのちに、統計処理・データマイニングを行う基盤の開発が「SS-MIX2 を基礎とした大規模診療データ収集と利活用に関する研究」（代表者：山本隆一）において進められている（図4）。SS-MIX2 ストレージから検索、抽出されたデータセットにより、一定程度匿名加工されたデータのセットを作成することが可能であるが、個人識別可能性の低減を優先した場合、k-匿名化手法では、素データの精度を

低減させ、解析結果に影響を及ぼす可能性がある。

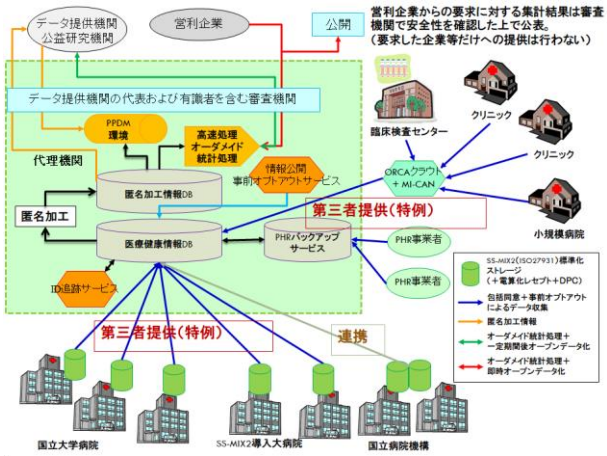

図 4 大規模診療データの収集と利活用基盤の概要

また、収集、加工後に抽出されたデータセットは、個別提供ないしはオープンデータとして他へ流通される想定であり、プライバシー侵害が起きないように抽出後のデータセットに対する安全管理措置にも留意する必要があるが、それぞれの利用目的に応じてデータセットへの個人特定性を低減するための匿名加工方法は異なることが想定される。

本研究では、抽出されたデータセット（匿名化サブセット）に対して、プライバシーリスク評価可能な手法を適用し、抽出したデータセットに対して、プライバシーリスクに関する評価指標付きで流通上の制限を設ける方針として、清本らが開発したリスク評価ツールの試行を行っている。

3. おわりに

本稿では、JST CREST「ビッグデータ統合利活用のための次世代基盤技術の創出・体系化」における技術開発要素に対して、電子化された診療情報の利用場面におけるセキュリティ課題に対する取り組み、現在の状況について記載した。

おもにセキュリティ技術としてのクラウドストレージ、分散データの秘匿検索・集合演算、プライバシー保護を前提としたリスクアセスメントへの統合・運用を目指すものであり、本プロジェクトで開発したプロトタイプが、診療情報における収集、解析、利活用を促進するためのセキュアな情報基盤の創出、実用化を目指すものである。

なお、本研究は、JST CREST グラント番号 JPMJCR1404「ビッグデータ統合利活用促進のためのセキュリティ基盤技術の体系化」により実施した。

参考文献

1) Gordon S, Miyaji A, Su C, Sumongkayothin K. A Matrix Based ORAM: Design, Implementation and Experimental Analysis. IEICE Transactions on Information and Systems. 2016;E99.D(8):2044-55.

2) A. Miyaji, K. Nakasho, and S. Nishida, "Privacy-Preserving Integration of Medical Data," Journal of Medical Systems, journal article vol. 41, no. 3, p. 37.