公募シンポジウム

公募シンポジウム4
医療分野の研究開発に資するための匿名加工医療情報に関する法律施行後の状況

2018年11月23日(金) 16:00 〜 18:00 E会場 (5F 501)

[2-E-3-3] 分散型ブロックチェーン基盤とデータ流通プラットフォームへの応用

○清本 晋作 （株式会社KDDI総合研究所）

本発表では、同意取得もしくは匿名加工によるデータ流通において、その履歴をセキュアに検証可能とするためのブロックチェーン基盤の活用方法と、スケーラブルなアーキテクチャを実現するための、分散型の構成方法について述べる。

分散型ブロックチェーン基盤とデータ流通プラットフォームへの応用

清本　晋作[*1]
*1 KDDI 総合研究所

Distributed Blockchain Architecture and Its Application to Data Distribution Platform

Shinsaku Kiyomoto[*1]
*1 KDDI Research Inc.

In this paper, we discuss a conceptual platform for privacy-respecting data distribution and application of blockchain technologies in the platform. We clarify functional requirements for the platform in terms of privacy protection. We introduce a distributed blockchain system and consensus algorithms in the system as well.

Keywords: Blockchain, Privacy, Data Platform, Distributed Architecture. Privacy Preference

1. はじめに

情報通信技術の発展により、多種多様かつ大量のデータの収集・分析が可能となってきており、それにより新たなサービスが生み出されてきている。パーソナルデータの利活用を通して、社会変革を進める試みは次の段階へと移行し、プライバシセンシティブな個人の情報を集積し積極的に活用しようという方針のもと様々な検討が進められている。医療情報分野も例外ではなく、より良い社会の実現に向けて様々な試みが検討されている。一方で、医療情報分野で取り扱うデータはプライバシセンシティブであることも多く、そうした情報の利活用においては、プライバシ・バイ・デザインのもとシステムの設計を行うとともに、プライバシに配慮した適切な運用が求められる。特に、欧州においては、個人のデータは個人のコントロールに委ねるという考えのもと、様々な施策の実施が進められている。こうした取り組みにおいては、データの利活用における一連のオペレーションのコントロール権の確保ならびに処理プロセスにおける信頼性と透明性確保を技術的に実現することが課題となっている。データ流通プラットフォームとは、組織間あるいは個人と組織において、こうしたプライバシセンシティブなデータをやり取りする共通基盤として検討される概念モデルであり、プライバシ保護の視点から様々な機能要件が定義されなければならない。本稿では、プライバシに配慮しつつ、データの利活用を推進するデータ流通プラットフォームとその信頼性・透明性確保に寄与するブロックチェーン基盤技術の活用について述べる。

2. データ流通プラットフォーム

本稿では、データ流通プラットフォームを「組織間あるいは個人と組織において、プライバシセンシティブなデータをやり取りする共通基盤」と定義する。なお、匿名加工情報や統計情報などの同意取得によらず流通可能なデータは本稿の議論の対象外とする。この時、プライバシ保護の観点から、以下の機能要件を満足することが必要となる。

1) データに対する安全管理措置
データの保存および流通の過程において、当該データが不正に漏洩、取得、あるいは改ざんされないように、適切に保護されなければならない。

2) 同意とプライバシプリファレンスに基づいた適切なデータ流通管理
データの提供は、データの所有者の同意のもとに行われなければならない。また、データ所有者のプライバシプリファレンスに基づいて、提供されるデータは選択されなければならない。

3) データ流通の一連の処理の信頼性および透明性の確保
データの流通の一連の処理は、データ所有者が確認可能でなければならない。また、データ流通における紛争を回避するため、一連の処理の正当性を第三者が検証可能であることは望ましい。

以上の要件を満足するデータ流通プラットフォームの機能要件について考察する。まず、データに対する安全管理措置については、暗号化等の適切な技術を使って安全管理を徹底することが求められる。また、より高度な安全管理措置として、提供データの選択等のデータ提供に係る一連の処理をデータが暗号化されたまま実行するような方式も適用できる可能性がある。

次に、同意とプライバシプリファレンスに基づいた適切なデータ流通管理については、データ流通プラットフォームとして、同意取得を行う機能ならびにプライバシプリファレンスをシステムに入力できる機能を提供し、それらの入力された情報に基づいてデータの提供可否を制御できる機能を備える必要がある。本機能の実現事例としては、oneM2M における PPM を用いられるデータ流通制御 [1]などが挙げられる。

一方、データ流通の一連の処理の信頼性および透明性確保については、一連の処理のログ等を適切に保管しておいて、データ所有者が閲覧できるようにすることが考えられる。また、保管されたログを第三者が検証可能とする仕組みを提供するケースも想定される。本稿では、このような透明性確保の要件に対してブロックチェーンの適用を試みる。

3. ブロックチェーン基盤の適用

ブロックチェーン [2]とは、分散型のデータ共有基盤であり、一旦登録されたデータは改ざん困難であること、第三者が登録されたデータの完全性（および存在証明）を検証できること、といった特徴を有する。本稿では、ブロックチェーンの 1 つのユースケースとして、データ流通プラットフォームを対象とし、

その適用方法について考察する。ブロックチェーンを用いることで、信頼の基点となる中央集権的なシステムを構築することなく、参加組織間で信頼性を保証するコストを分担するようなアーキテクチャとすることができる。これは、複数の組織が相互連携する際に有効なアーキテクチャである。

3.1 対象とするデータ

まず、2章で挙げたデータ流通プラットフォームの機能要件から、ブロックチェーンが対象とするデータについて列挙する。データ流通プラットフォームにおいては、少なくとも以下のデータをブロックチェーンへの登録対象とする必要がある。

・流通対象データ（のハッシュ値）
・データ流通のトランザクションログ（時刻、送信者、受信者、送信データの識別子、送信データのハッシュ値）
・同意取得情報（のハッシュ値）
・プライバシプリファレンス設定情報（のハッシュ値）

データ流通のトランザクションログについては、ブロックチェーンを介してデータ流通プラットフォームに参加している全ノード間で共有しても良いと考えられるが、組織間のデータ流通を秘匿した場合などは、ハッシュ値で置き換える必要がある。ただし、データの所有者に対しては、閲覧できるように開示制御をかけなければならない。

一方、流通対象データ、同意取得情報、プライバシプリファレンス設定情報は、当該時刻におけるデータの完全性を保証出来れば良いため、ハッシュ値の登録で良い。なお、流通対象データとは、当該時刻において、流通させる可能性があるデータとして、当該ノードが保有するデータを表す。適切な処理で削除が行われたデータは、当然ながら対象から外れる。データは膨大な量となることが想定されるため、ハッシュ値については階層木を構築するなど、効率化手法を導入する必要がある。また、検索処理等を考慮して、流通対象データのカタログ情報をブロックチェーンに登録し、共有しても良い。

同意取得やプライバシプリファレンスの情報を参加ノード間で共有可能な（データの所有者から同意を得ている）場合には、ブロックチェーンに登録して共有することで、1 つのノードに登録するだけで、参加ノードすべてに適用することが可能となる。

3.2 分散型ブロックチェーン基盤

ブロックチェーンは、参加するノード数が増加するにしたがって処理遅延が増加し、またその運用コストが増大するという懸念がある。これを解決する手段として、参加ノードが多い場合には全体をいくつかのローカルチェーンに分割して構成する手法がある。ローカルチェーンとは、参加機関のいくつかが集まって構成されたブロックチェーンである。例えば、ある地域ごとにローカルチェーンを準備し、複数のローカルチェーンが相互に参照することで、全体のシステムを構成するような分散型のアーキテクチャを構築することができる。この時、ローカルチェーンをまたがったデータ流通が必要なケースのみ、双方のローカルチェーンにデータを書き込むようにすることで、処理の効率化を図ることができる。このようなデータ流通は、Inter-Ledger Protocol （ILP）と呼ばれる技術を応用することで実現可能である。

一方で、ローカルチェーン単体では、全体でブロックチェーンを構成した場合と比較して、信頼の保証が十分ではない。従って、複数のローカルチェーンが連携したプラットフォーム全体で信頼性を保証するためには、各ローカルチェーンのコンセンサス構築とブロックの生成において、相互に連携する必要がある。次節において、分散型ブロックチェーン基盤に適用するコンセンサス構築およびブロック生成手順について述べる。

3.3 コンセンサスの手法

複数のローカルチェーン間で相互に参照することで、全体でコンセンサスを構築する手法を述べる。方式の概要は以下の通りである。

各ローカルチェーンの新規ブロックに入力する値として、ローカルチェーンに入力されたデータ（台帳情報）のルートハッシュ値に加え、自身の 1 つ前のブロックのハッシュ値、および選択関数によって選択された他のローカルチェーンのブロックのハッシュ値の三種類を選択する。そして、生成されたブロックは、他のローカルチェーンに対して送信される。各ローカルチェーンは受信したブロックをブロックバッファと呼ばれる領域に保管しておき、そこから選択関数によって選択されたブロックのハッシュ値が新たなブロックの入力として選択される、という処理を繰り返すことでローカルチェーンのブロックを生成していく。なお、ブロックバッファは、一杯になった段階で First-In-First-Out（FIFO）方式に従って、古いものから順に削除される。また、選択関数によって選択され、新規ブロックに入力されたブロックについても削除される。

上記方式において、ブロックバッファを導入することにより、各ローカルチェーンにおける新規ブロックの生成処理を非同期とすることができる。このことは、全体のスケーラビリティ向上に寄与する。一方で、ブロックバッファ導入に伴って、安全かつ公平な選択関数が必要となる。選択関数としては、以下の方式が考えられる。

1） ブロックバッファにあるすべてのブロックを入力とする全選択方式
2） ブロックバッファにある古いブロックを N 個選択する順次選択方式
3） 当該ローカルチェーンにおける直前のブロックのハッシュ値を計算し、そのハッシュ値にもとづいてブロックバッファ内のブロックを選択する疑似ランダム選択方式

安全で公平な選択関数の設計において重要な点は、その選択方法（アルゴリズム）が、意図的に変更不可能であり、また選択方法が第三者にとって検証可能でなければならない、ということである。第三者が検証可能で無い場合、ローカルチェーンのコンセンサスに参加したノードが意図的にブロックを選択することで、ブロックの偽造につながる不正を行う余地が生じる。

3.4 選択関数の評価

本節では、3.3 節で述べた選択関数の評価を行う。まず方式（1）では、毎回バッファにあるすべてのブロックを対象とするため、ブロックを取りこぼすことなくローカルチェーンに取り込むことが可能となっている。従って、ほとんどすべてのローカルチェーンにおいて、結託して不正を行うノードの数が閾

値を超えない限り、ブロックの偽造や不正なトランザクションの挿入は不可能であると考えることができる。各ローカルチェーンの閾値を d、連携するローカルチェーンの数は l とした場合、dl/2 以上のノードが結託しなければ不正は難しいと考えられる。一方、それぞれのローカルチェーンで、ブロックを生成し送付するタイミングは異なるため、バッファに存在するブロックの数は一定とはならない。従って、バッファからローカルチェーンに取り込むブロックの数は、一定とはならないという課題が生じる。この課題を解決する手法として、方式(2)が考えられる。方式(2)では、バッファに格納されているブロック数によらず、一定の数のブロック(N 個)をローカルチェーンに取り込む。従って、バッファに存在するブロックの数が N よりも小さくならない限りにおいて、必ず一定の数 N のブロックをローカルチェーンに取り込むことができる。しかしながら、単位時間あたりに送付されてくるブロックの数が N よりも大きい場合には、バッファに格納されるブロックの数は増加していくため、あふれたブロックが、ローカルチェーンに取り込まれることなく破棄される、というリスクが生じる。確率 p でブロックが破棄されると仮定したとき、(元のローカルチェーンを除く)すべてのローカルチェーンに当該ブロックが取り込まれない確率は、p^{l-1} である。この時、$d(1+(1/2)*([l\Sigma(1-p)^{l-i}p^i])$ (i=0, …, l-1)以上のノードが結託しなければ不正は難しいと考えられる。なお、[x]は、x 以下の整数とする。ただし、これは攻撃者がバッファがあふれるタイミングを予測不可能であった場合に限られる。ローカルチェーンそれぞれのバッファがあふれるタイミングを攻撃者が予測可能であった場合、不正者の配置を最適化する、あるいはあらかじめ配置された不正者にとって最適のタイミングで攻撃を行うことで不正に加担しなければならないノード数を減らすことができる。従って、より攻撃がしやすい状況が生まれる。左記の攻撃への対策として、方式(3)が考えられる。方式(3)では、N 個のブロックを古いものから順に選択するのではなく、直前のブロックのハッシュ値 h(IV1, block)に基づいてバッファの中から選択する。この時、ハッシュ関数には、直前のブロックの値と共に、公開の初期ベクトル IV1 を入力する。暗号学的ハッシュ関数 h( )を疑似ランダム関数とみなすことができるのであれば、その出力ハッシュ値は乱数列とみなすことができる。従って、攻撃者が直前のブロックを正しく予測、あるいは適切なタイミングで必要なブロックを入手できなければ、不正者の配置は効率化できないこととなる。結託攻撃に対する耐性の上限は、ローカルチェーンの閾値 d で抑えられる。閾値 d は、ローカルチェーンにおけるコンセンサス手法に依存するが、例えば、Practical Byzantine Fault Tolerance(PBFT)を用いた場合、ローカルチェーンの参加ノード数 m に対して、m/3 である。従って、ローカルチェーンにおいてコンセンサス構築に参加するノード数が少ない場合、少数の不正者の結託によって攻撃することが可能となるため、信頼性が脅かされる恐れがある。そこで、m を大きくするため、他のローカルチェーン参加者から、ランダムにコンセンサス構築に招待することで、m を仮想的に大きくし、不正を抑止する手法が考えられる。この参加者の選択についても、あらかじめ参加者にシーケンシャルに ID を付与しておき、直前のブロックのハッシュ値 h(IV2, block)の出力から、ID の組み合わせを計算し、参加者を抽出する方式とする。この時、ハッシュ関数には、直前のブロックの値と共に、公開の初期ベクトル IV2 を入力する。なお、IV1 および IV2 は、後日検証を行うことができるようにするため、当該ブロックに格納される。なお、IV1 および IV2 は、攻撃者が任意の出力を得るような攻撃ができないように固定値とし、ブロックの生成ごとに変更を行わない。

4. おわりに

本稿では、プライバシに配慮しながら安全なデータ流通を実現するデータ流通プラットフォームとその構成要素となるブロックチェーン基盤について述べた。今後は、分散型ブロックチェーン基盤を実装し、評価実験を行っていく予定である。

謝辞: 本研究は、JST CREST の支援を受けたものである。

参考文献

1) oneM2M Technical Specification, Security Solutions, TS-0003-V2.12.1, March12, 2018.
2) F. Tschorsch and B. Scheuermann, Bitcoin and Beyond: A Technical Survey on Decentralized Digital Currencies, in IEEE Communications Surveys & Tutorials, vol. 18, no. 3, pp. 2084 - 2123, 2016.