秘密計算システム 算師®の試用提供

北條 裕之 †1 /山口 卓也
西山 小奈未 †1 /高橋 元 †2
宮島 麻美 †2 /廣田 啓一
西田 祥子 †2 /橋本 順子

NTT研究企画部門 †1
NTTセキュアプラットフォーム研究所 †2

NTTでは，企業秘密やパーソナルデータなど守るべきさまざまなデータの安心・安全な利活用に向け，データを暗号化したまま，実用的な速度で安全に集計・統計処理できる秘密計算システム 算師®（算師）を開発しました．データ利活用の活性化に向けた取り組みとして，秘密計算の「データを互いに開示することなく，データを暗号化したままで統合分析できる」利点を多くの方に体験いただくべく，期間限定ではありますが，算師を無償で試用提供しています．本稿ではその取り組み内容と秘密計算について紹介します．

背　景

昨今，さまざまな分野のデジタルトランスフォーメーションにより，サービス化，オープン化，ソーシャル化，スマート化への変化が進んでおり，分野横断的なデータの蓄積やデータの利活用がイノベーションを促進し，経済成長などさまざまな分野の発展につながることが期待されています．一方，データの管理に伴うインシデントリスクや社会的責任の大きさ，企業戦略等の保護の観点による，データのセキュリティ対策の必要性などがデータ利活用促進を阻害する要因となっています．

NTTはそのような要因の解消に貢献するため，データを暗号化したままデータ処理可能な秘密計算技術の研究開発に世界に先駆けて取り組んできました．秘密計算の利点は，計算結果以外は誰にも見えないデータ運用ができること（**図1**）と，これにより，今まで他組織に開示することが難しかったデータを持ち寄った新しい統合分析が可能になることにあります．これまで，多施設臨床研究データ[(1)]やゲノムデータの解析[(2)]など，さまざまな分野への適用事例を検証するとともに，演算機能の充実や高速化等の改良を加え，NTTの秘密計算システム 算師®（算師）として開発を進めてきました[(3)]．

算師の試用提供概要

NTTが開発した算師は，秘密計算の持つ「データを互いに開示することなく，データを暗号化したまま統合分析できる」という利点をシステムとして実現したものです．さまざまな分野の多くの方々にこの価値を体感していただくことを目的として，算師を無償で試用提供を開始しました．2018年8月20日より開始し，最長2019年3月まで利用いただけます．現在，ヘルスケア，製造業，SIer等，さまざまな分野のお客さまにご利用いただいています．

利用者には，クラウド上に構築した算師を用いて，データを暗号化したまま集計・統計処理を行う機能を実際に体験いただくことができます．気軽にお試しいただける代表的な分析シナリオと試用データを3種類用意しました（**表1**）．

1番目は，「同業他社との連携強化」で「データ量（行）を増やす」シナリオです．競合他社と相互に情報開示は行いたくないが，業界活性化や業界課題解決につなげることを想定しました．地域商圏を一例とし，算師上に複数事業者による購買データを安全に登

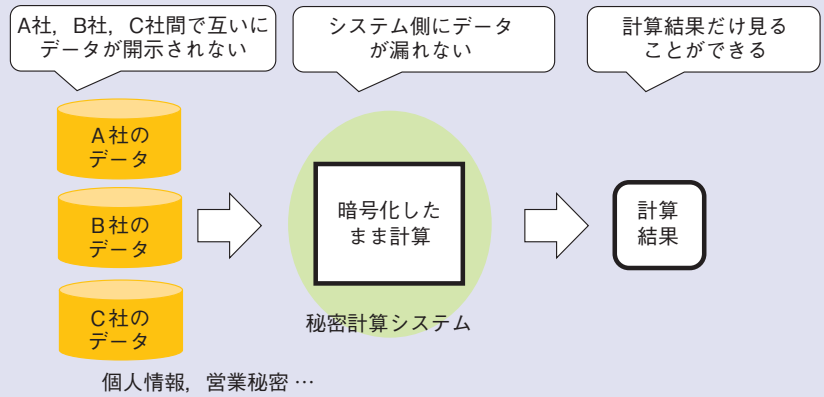

図1　秘密計算の利点

表1　試用提供システムで体験いただける代表的な分析シナリオ

項　目	シナリオ概要
同業他社との連携強化	地域商圏の活性化に向けて，地域にある複数企業の販売データを統合・分析し，地域商圏全体として品ぞろえの充実や販売機会の損失回避を図る
異業種データの連携	ネット販売会社の購買データと，健康支援アプリ提供会社が保有するバイタルデータ（BMI・歩数）を組み合わせて分析し，健康関連商品のマーケティングや商品レコメンド（広告収入拡大）に活用する
演算機能の体験	世帯属性，世帯支出，食費などのデータ群に対して，どのような演算が可能か試用いただく

録し，全事業会社のデータを結合した状態としておき，利用者（データ分析者）は商圏全体の総売上金額や年代ごとの売上上位商材などを確認できます．

2番目は，「異業種データの連携」で「データ項目（列）を増やす」シナリオです．異業種でのデータを組み合わせることで，新たな傾向発見や新たなビジネス価値を生み出すことを想定しました．一例とし，ネット販売会社の購買データと健康支援アプリ提供会社のバイタルを組み合わせ，算師上に安全に登録し，利用者（データ分析者）は健康食品を購入している顧客の年代別の歩数平均などを確認できます．

3番目は，算師がサポートする演算機能を自由に試していただけるシナリオです．公的統計情報（一般用ミクロデータ）を算師上に安全に登録し，利用者（データ分析者）は，消費支出，食費，保険医療費などのデータを用いた演算が可能です．

また，さらなる試用を希望される利用者には，自身が保有するデータに基づき，代表シナリオ以外の個別の分析シナリオ等で算師を試用するサポートもしています．

秘密計算

秘密計算とは，データを暗号化したまま，計算できる技術です．一般的な暗号ではデータの計算時には，復号する必要があるため，データが分析者やシステム運用者に漏洩するリスクがあります．一方，秘密計算は，データを暗号化したまま計算を行うことが可能であるため，分析者やシステム運用者は途中経過を含む一切のデータを見ることができません．このため，企業の秘密情報のようなデータでも，安全に利活用することが可能になります．

秘密計算は，1980年代に計算機科学・暗号理論の分野で“Secure multi-party computation”と呼ばれる理論の大枠が確立されましたが，実用上は計算に時間がかかる（遅い）ことが課題とされてきました．近年高速化・実用化研究が活性化しています．NTTでは秘密計算方式として，秘密分散をベースとした高速な秘密計算方式を開発しました．

■秘密分散による暗号化

NTTの秘密計算の暗号化の仕組みとして，秘密分散を採用しています．秘密分散はデータを複数のシェアと呼ばれる断片に分割し，機密性を高める技術です．個々のシェアから情報は漏れません．さらに，いくつかのシェアが消失してもデータを復元可能です．また，秘密分散方式として，ISO標準準拠（ISO/IEC 19592-2）仕様を用いています．この，ISO化においては，NTTはエディタとして標準化に貢献しました．

■秘密分散をベースにしたマルチパーティ計算

暗号化したまま計算する仕組みとして，秘密分散をベースにしたマルチパーティ計算を採用しています．マルチパーティ計算では，システムは複数のサーバから構成され，サーバ間でデータの交換と演算をあらかじめ決められた手順で行います．各サーバには，秘密分散されたシェアが登録され，データは常に秘密分散のシェアの状態で扱われます．

■秘密計算の安全性

秘密分散された個々のシェアから元データや計算結果を復元することは一切できません．ただし，分割したシェアを複数のサーバに各々登録しますが，一定数のサーバからシェアを不正に取得されるとデータが復元できてしまいます．このため，各サーバを正しく管理することが安全性の条件です．

■秘密計算の原理

秘密計算では，データは複数のシェアに秘密分散されます．ここでは，「2」を3つのシェアに秘密分散する例を紹介します（**図2**）．秘密分散のシェア生成では，乱数を生成し，生成した乱数を元に計算を行います．まず，乱数を2つ生成します．生成される乱数，「0」から「9」のランダムな値とします．乱数として「5」と「3」が生成された場合は，3つのシェアのうち2つのシェアを「5」と「3」とします．次に3番目のシェアをこの2つのシェアから計算して求めます．元のデータ「2」から2つのシェア「5」と「3」を足し合わせた「8」をマイナスします．この際，マイナスして得られた値「−6」は，「4」となり，3番目のシェアは「4」に決まります．

元のデータに復元する際は，3つの

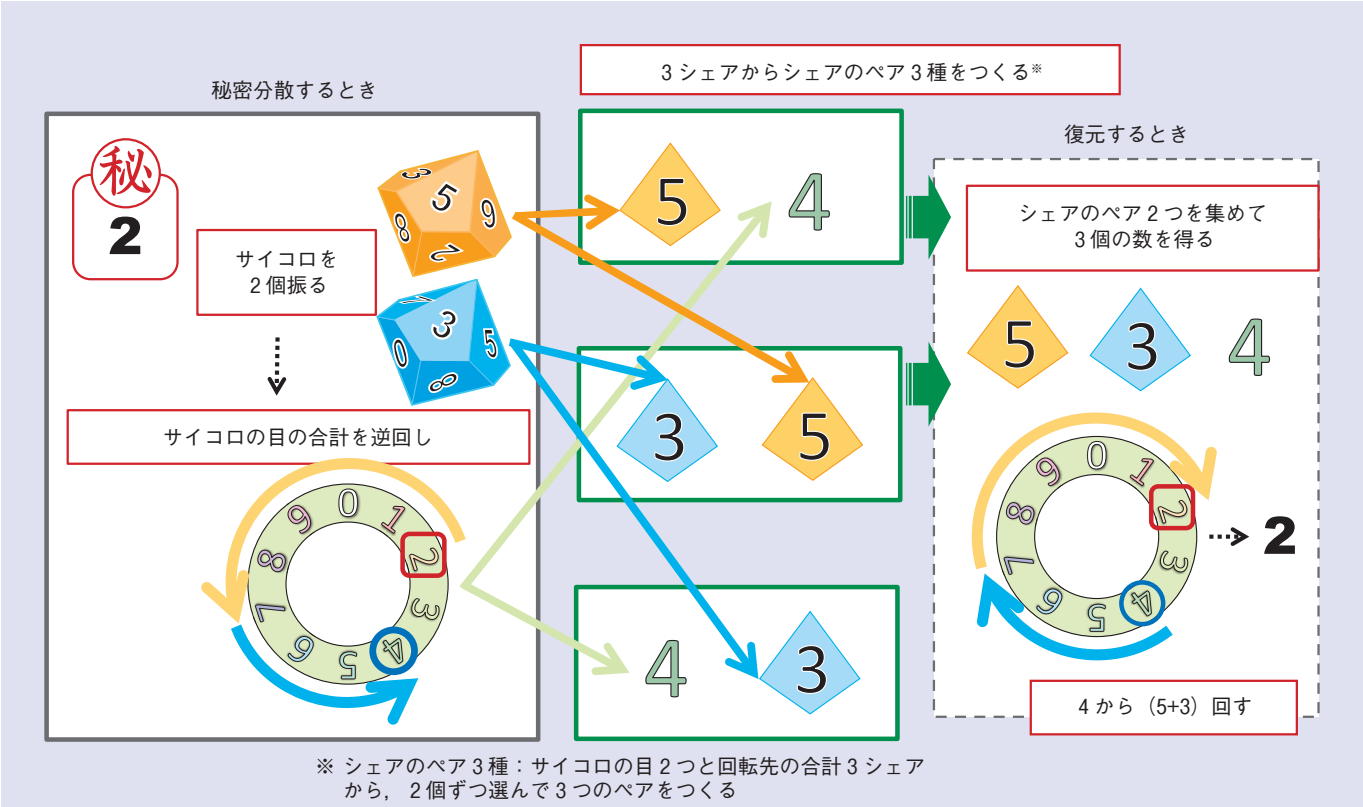

図2　秘密計算の原理

表2　NTTの秘密計算システム算師が具備する主な演算

データ操作	集　計	基本統計		検　定
		総和	最大値	t検定
テーブル結合	度数表（クロス集計）	平均	最小値	
条件によるフィルタ	数量表	分散	中央値	その他
		積和	分位数	Kaplan-Meier法

シェア「5」と「3」と「4」を集めて，それぞれ足し合わせます．この際，「4」に「3」と「5」を足し合わせた値の「12」は，「2」となり，元のデータの「2」が復元できます．

計算は，このように生成されたシェアを各サーバ上でシェアのまま計算を行います．例えば，総和を計算したい場合，各サーバでシェアの状態で総和の計算を行い，最後に，各サーバで計算した値の総和の結果を上記の方法で復元することで，総和の結果を得ることができます．

算師の特長

NTTの算師は，秘密計算の長年の技術課題であった処理速度を劇的に向上し，100属性×1000万件規模のデータの集計や統計演算を実用的な時間内に処理することができる世界最高レベルの秘密計算システムです．豊富な集計・基本統計演算処理を持ち，各演算を高速に実行することができます．

■充実した演算バリエーション

算師では，**表2**に示す演算をWebブラウザ上のGUIや統計解析ソフトウェア「R」のインタフェースでデータを見ることなく実行することができます．さらに，「R」で簡単なプログラムを作成し，回帰分析や主成分分析など，用途に応じた分析を実行することも可能です．試用提供システムでは，これらのインタフェースの一部を利用いただけます．

特に，算師が提供するテーブル結合機能（複数の表を結合キーも漏らすことなく結合できる機能）は，異なる企業間や異業種間で，互いに所有データを見せることなくデータを統合し，横断分析した結果のみを得ることを可能にします．これにより，複数の企業をまたがるサプライチェーンや顧客データの分析など，これまで一企業や一業界では成し得なかったデータ利活用の新たな価値創造に貢献できると考えています．

表 3　代表的な機能の実行時間

機　能	実行時間（ミリ秒）				
データ件数	10^3	10^4	10^5	10^6	10^7
加　算	1	1	1	2	14
乗　算	1	1	5	39	473
ソート	10	23	133	1274	12255
総　和	1	1	1	1	9
積　和	1	1	1	2	15
数量表作成	22	46	255	2252	22676
シャッフル	1	1	8	60	731
テーブル結合	19	65	518	4965	53205
条件によるフィルタ（文字列前方一致）	6	6	14	91	813
条件によるフィルタ（数値一致）	5	5	10	35	413

PC 3 台（CPU: Intel Core i7-6900K, メモリ: 32 GB, SSD: 525 GB, OS: CentOS 7.2）を10 Gbit/sネットワークで接続した環境で測定

■実用に足る高速性

算師では，秘密分散方式の採用[(4)]に加え，独自の高速化アルゴリズムと暗号実装技術により，前述の豊富な演算バリエーションの提供と処理速度の向上を両立させています．

秘密分散に基づく秘密計算では，データ処理の基本となるデータのサイズが小さい，演算を行う際に頻繁に使用される加算と乗算の両方を高速に処理できる，という圧倒的な 2 つの利点があります．そのため，準同型暗号など他の暗号方式に基づく秘密計算と比べ，さまざまな演算を高速に処理することが可能です．

加えてNTTでは，非常に小さい計算コスト・通信コストで動作する秘密計算の基本アルゴリズムを開発し，これを高度な暗号実装技術によって実装することにより，処理速度を劇的に改善し，世界最高レベルの演算速度を達成しています．

代表的な機能の実行時間を**表 3** に示します．1000万レコードの並び替え処理（ソート処理）を12.2秒で実現しています．暗号化されていない1000万レコードのデータを一般的なソートアルゴリズムでソートした場合，1 秒程度の演算時間です．秘密計算と通常のコンピュータ処理の性能比はおよそ「一桁レベル」に迫っています．

■算師のシステムイメージ

秘密計算では複数のサーバが一体となって計算を行うマルチパーティ計算を行います．算師は，秘密計算クライアント，3 台もしくは 4 台の秘密計算サーバから構成されます．データ登録を行う秘密計算クライアントはデータを秘密分散のシェアに分割して各サーバに登録します．また，データ分析を行う秘密計算クライアントは各サーバに計算（データ分析）を要求し，計算結果のみを得ます．データは，リレーショナルデータベースのようにテーブル形式で登録されており，データが格納されているテーブル名や列名を指定して，平均値や分散値等の計算を要求します．計算要求を受けた各サーバは，それぞれ協調して，マルチパーティ計算を行い，計算結果を秘密分散のシェアとしてデータ分析を行う秘密計算クライアントに応答します．秘密計算クライアントは，シェアを復元することで結果を得ます．

今後の展開

NTTは，算師の試用提供を通じて，企業秘密やパーソナルデータの安心・安全な利活用のより一層の促進をめざし，秘密計算をはじめとするデータ利活用技術の開発とグローバルを含めた普及に努めていきます．

■参考文献

(1) http://www.ntt.co.jp/news2012/1202/120214a.html
(2) http://www.ntt.co.jp/news2016/1607/160712a.html
(3) http://www.ntt.co.jp/sc/project/data-security/secure_computation.html
(4) http://www.ntt.co.jp/news2017/1710/171023a.html

（上段左から）西山　小奈未/ 北條　裕之/ 山口　卓也
（下段後列左から）宮島　麻美/ 高橋　元/ 廣田　啓一
（下段前列左から）西田　祥子/ 橋本　順子

秘密計算技術は，流通や金融，医療ヘルスケアなどさまざまな分野のデータ利活用への展開が期待されます．さまざまなパートナーの皆様と連携して，実社会への適用に向けた取り組みを進めていきます．

◆問い合わせ先
NTTセキュアプラットフォーム研究所
データセキュリティプロジェクト
E-mail　seg-product-p-ml＠hco.ntt.co.jp