［安全なデータ活用を実現する秘密計算技術］

②秘密計算の実用化に向けた研究の歴史と現在

基般

五十嵐大 | NTT セキュアプラットフォーム研究所

1982 年：秘密計算の誕生

秘密計算は“データを暗号化したまま処理できる”暗号技術です．1982 年に Yao が概念とともに最初のアルゴリズムを提案，そのアルゴリズムで任意のデータ処理ができ得ることを示しました．Yao のアルゴリズムは暗号で使われるハッシュ関数のような，“順方向の計算は簡単だが，逆関数の計算は困難”という性質の処理を利用するものでした．

Yao のアプローチは Garbled circuit と呼ばれ，現在でも盛んに研究されています．また，複数のマシン間（たとえば Garbled circuit では 2 マシン）で暗号データの通信を経て処理を行うのが大きな特徴の 1 つで，このようなアルゴリズムをマルチパーティ計算と呼ぶようになりました．

Yao 以後，1988 年には Ben-Or らが，Shamir が提唱した秘密分散技術を用いた秘密分散ベース秘密計算を提案するなど，秘密計算は計算機科学分野および暗号分野で盛んに研究されてきました．しかし当時の主な興味は“何台のマシンで，どんな安全性を達成できるか”という理論面でした．

2000 年：“Privacy Preserving Data Mining”

秘密計算の応用研究が始まった大きな契機の 1 つは，2000 年に Lindell と Pinkas が発表した“Privacy Preserving Data Mining”と題された論文です．この論文は奇しくもまったく同じ年に，異なる分野であるデータベース分野で Agrawal らがまったく同じ題名の論文を発表していることでも有名です．

“Privacy Preserving Data Mining”では，実装はされていないものの，データマイニングの基礎的なアルゴリズムである ID3 アルゴリズムが，秘密計算を用いて構成されています．それまで秘密計算の分野では“論理回路ができれば任意の処理ができる”という原理に基づき，もっぱら論理回路素子の構成が研究されていました．しかしこの論文では ID3 自体もですが，特に ID3 の実現に必要な対数関数を効率的に秘密計算で計算する方法が論じられました．これは“任意の処理が理論的にはでき得る”という理論としての認識から，“実現”のために高速なアルゴリズムを構成するという実用に向けた認識への転換でした．

2004 年：初の実装

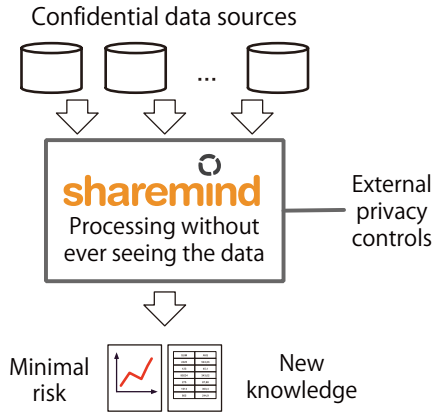
2004 年には，秘密計算の初の実装が報告されます．Malkhi らは Yao の Garbled circuit をベースとした秘密計算フレームワーク FairPlay を実装しました．FairPlay は，ID3 や対数関数のような高等な処理とはいきませんが，論理回路を自由に組むことができるインタフェースを備えていました．このときの性能は，1 論理素子あたり約 13ms という性能で，普通の PC と比較すれば 10 億倍以上の開きがありましたが，実際に秘密計算を動作させられるようになったこと，実性能が確認されたことは秘密計算分野として大きな前進でした．

国内では千田らが近い時期，2005 年に Modified ElGamal と呼ばれる暗号をベースとした秘密計算

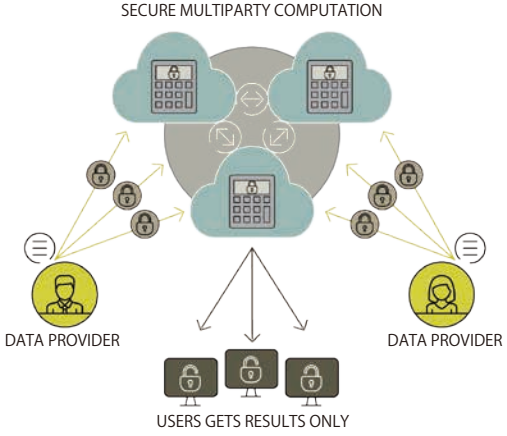
を実装し，こちらは 1 論理素子あたり約 6ms という性能でした．

2008 年：Sharemind の出現

2008 年，より高性能かつリッチなユーザインタフェースを持つ秘密計算ソフトウェアが出現します．Bogdanov らが提案し，エストニアの Cybernetica 社が今も開発し続ける，Sharemind（図 -1）です．Sharemind は FairPlay の論理回路インタフェースよりも高級な，SecreC と呼ばれる C 言語に近い独自プログラミング言語を持っていました．性能面で有利な秘密分散ベース秘密計算を採用しており，



■図 -1　Sharemind のシステムモデル
出典：https://sharemind.cyber.ee/secure-computing-platform/



■図 -2　Partisia のシステムモデル
出典：https://partisia.com/secure-simple-efficient/

FairPlay や千田らの実装と比べて 1,000 倍以上高速な，論理回路 1 素子あたり 5.7 μs（1 秒あたり約 175K 素子）を実現していました．

発表当初はオープンソースでしたが，現在は SecreC の後継である SecreC 2 インタフェースのみがオープンとなっており，オープン版では秘密計算部分は製品版での処理時間を再現するエミュレータとなっています．

2008 年：実データによる初の実験

Sharemind 登場と同じ 2008 年の 1 月 14 日，Bogetoft らはテンサイ（サトウダイコン）の競り（オークション）のシステムとして，秘密計算を適用する実験を行いました．競りには 1,229 者が参加し，25,000t のテンサイの取引が行われました．使われたのは Shamir の秘密分散を用いた秘密分散ベース秘密計算で，加算，乗算，比較が主な演算であり，このときの秘密計算の処理時間は 30 分であったということです．

この実験に関連する現在の動向として，現在，実験論文の共著者である Nielsen 氏が CEO を務めるデンマークの Partisia 社が，秘密計算を使ったオークションソリューションを手がけています（図 -2）．

2009 年：完全準同型暗号の発見

これは当時の時点では理論的な動向ですが，Yao による秘密計算自体の提唱に次ぐ大発見だったため紹介します．2009 年に Gentry によって，“完全準同型暗号”アルゴリズムが提案されました．Yao 以来，秘密計算はいくつかのアプローチがありましたがいずれも複数マシンで行うマルチパーティ計算でした[☆1]．それに対して完全準同型暗号は，1 マシンで暗号化したまま任意の処理を実行し得るアルゴリ

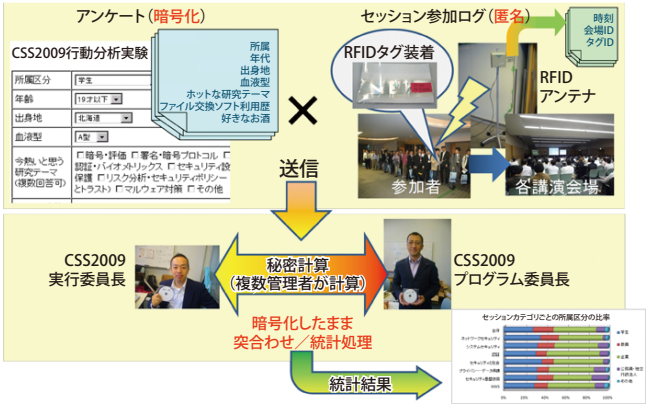
[☆1] 加算のみなど，特定の処理のみが可能な方式を除く．

ズムで，存在について研究者の間で意見が分かれていました．このGentryのブレークスルーにより，秘密計算はさらに活発に研究されることになります．

2009年：初の統合データ分析実験

Gentryの発見と同じ2009年，CSS（コンピュータセキュリティシンポジウム）2009において筆者らは，RFIDシステムによるシンポジウム参加者の行動ログと，別途行ったアンケート結果を秘密計算により統合して分析を行うという実験を行いました（図-3）．

RFIDシステムで検知された参加者数は215名で延べ2,339回，アンケート回答者は97名でした．Garbled circuitベースの方式を用いておりながらも，Garbled circuitのボトルネックである特殊な暗号処理“Oblivious Transfer☆2”を排したアルゴリズムにより，Sharemindに近い1秒あたり約110K論理素子のシステムでした．秘密計算で計算した集計を元に，単純な集計や，相関分析を行うFisherの正確確率検定を行い，たとえば“企業の参加者は認証セッションの参加率が高い”などの相関が見られました．



■図-3　CSS2009行動分析実験の概要
出典：http://www.iwsec.org/css/2009/misc.html

☆2　送信者の持つ複数の選択肢から受信者が必要なデータだけを受信し，送信者には受信者がどれを選択したか秘匿され，受信者には選択した以外のデータが秘匿される，という暗号通信．

互いのデータを見せ合えない2者のデータを統合して分析するという統合データ分析は，秘密計算の有力なアプリケーションの1つです．本実験は，真に見せたくないような情報を取得することはあえて避けたものの，実データを用いて統合データ分析を模擬した初の実験でした．

2011年：医療統計への適用

2011年，筆者らはJALSG☆3と共同で，JALSGの保持する臨床研究データを用いた，医療統計分析の実験を行いました．CSS2009行動分析は技術的な応用モデルとして有望な例を実験しましたが，こちらは医療分野という，高いセキュリティを必要とし秘密計算にマッチするであろう分野での初の試みでした．

秘密計算技術の面では，それまでの実験で使われた秘密計算が加算，乗算，大小比較のみで構成されたのに対し，本実験では中央値，t検定☆4，Kaplan-Meier法☆5などの，実際医療統計で用いられる機能が実装・実験されました．約1,000件の実症例に対して，最も処理の複雑な中央値でも約5秒と，かなり現実的に近い応答時間でした．ユーザインタフェースはMicrosoft Excelプラグインを採用しており，図-4は実際のシステムにおける，Kaplan-Meier法の出力画面例です．2つの治療法に対して，横軸を経過時間，縦軸を生存率としたグラフが出力されています．

Garbled circuitを用いたCSS2009行動分析と異なり，このとき筆者らはその高速性に気づいて秘密分散ベース秘密計算に技術方針を転換しており，論理素子の性能は1秒あたり約1M論理素子に達しました．しかし，重要な処理は論理回路ではなく，汎

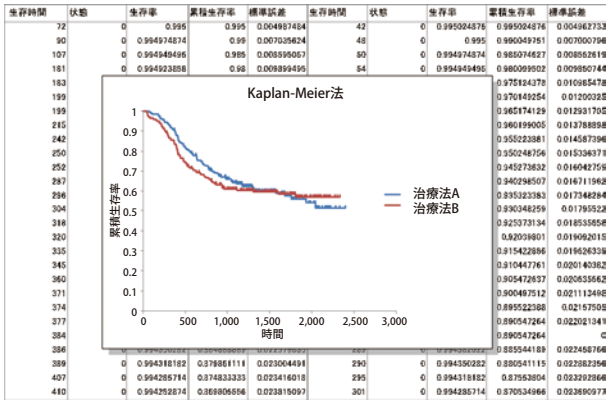
☆3　日本成人白血病治療共同研究グループ．
☆4　2組の集団の平均の間に，統計学的に有意な差があるかどうかを判断する方法．
☆5　累積生存率を計算する方法の一種．複数の治療法等の効果を比較するのに使われる．

用性を失わずに論理回路と接続できるような専用のアルゴリズムを構成するという方針をとるようになり，たとえば濱田らの提案したランダム置換（シャッフル）により Kaplan-Meier 法を，同じく濱田らの提案した秘密計算ソートによって中央値を処理していました．

2013 年：インターネット上の実験

2013 年には筆者らは，アルゴリズム・実装を刷新したシステムにより，通信を行うマルチパーティ計算としては非常に困難な，インターネット環境での医療統計の追実験を行いました．東京・金沢・長崎の 3 地点それぞれに秘密計算サーバを配置し，帯域 200Mbps（ベストエフォート，実帯域約 20Mbps），遅延は地点により約 25ms ～ 47ms でした．2011 年の実験は Gb イーサネットの LAN で，遅延が約 0.1ms でしたから，帯域 1/5 以下（ベストエフォートなので実際はもっと差がある），遅延 200 倍超という難しい環境です．しかし，新システムの効果が発揮され，LAN 上の旧システムよりむしろ高速な，全演算 1 秒以内の応答が実現されていました．

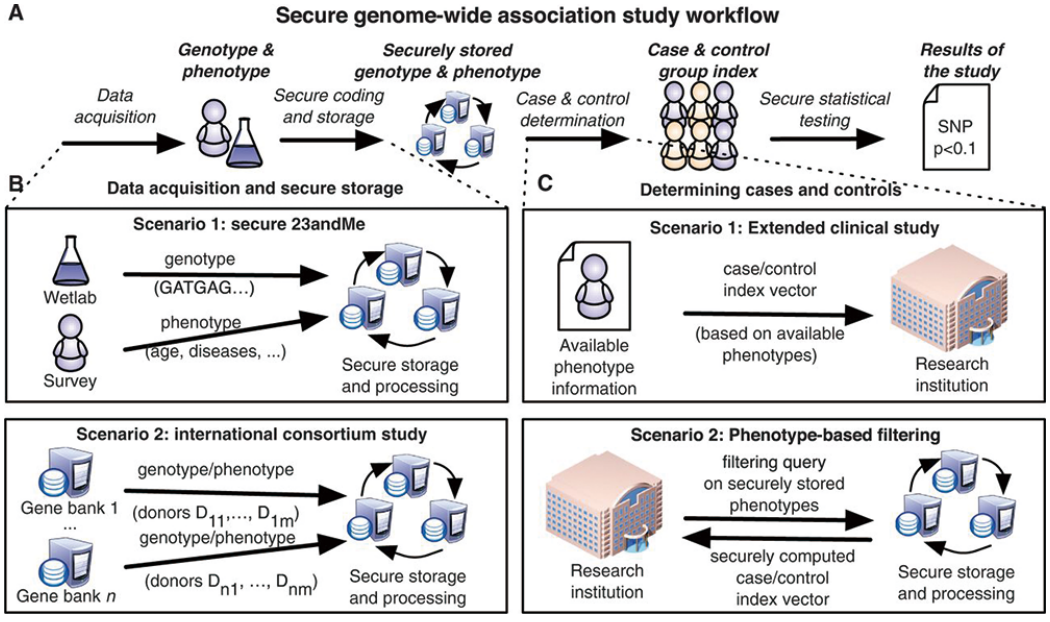
この実験により，1,000 件規模であればインターネット環境でも実用的な医療統計演算が秘密計算で処理可能なことが実証されました．



■図 -4　Kaplan-Meier 法の出力画面例

2013 年：ゲノム解析への適用

上記と同じ 2013 年に Kamm らは，Sharemind を用いたゲノム解析実験を報告しました（図 -5）．GWAS（Genome-wide association studies）と呼ば



■図 -5　Kamm らのゲノム解析実験の概要
出典：https://academic.oup.com/bioinformatics/article/29/7/886/253610

れる，ゲノムの広範囲にわたって遺伝情報と疾病等の関係を解析するものです．データ件数は 1,080 名 × 540,810SNP[☆6] で，x^2 検定[☆7]，Cochran-Armitage 検定[☆8]，TDT[☆9] の 3 つの検定法を実験しました．検定法により差はありますが，約 110 分から 480 分で解析がなされました．GWAS は秘密計算でない場合でも処理コストのため夜間バッチで実施することは普通ですので，480 分（＝8 時間）は十分実用的な処理時間だといえます．

本実験は，究極のプライバシー情報ともいわれる遺伝情報に秘密計算を初めて適用したという点において重要でした．また，技術面では，人数× SNP 数で 5 億件にもなる，それまでの実験と比較すると別次元ともいえる大規模データに対して秘密計算を実動作させた画期的な実験でした．

さらに，2016 年には長谷川らが，筆者らの開発する秘密計算ライブラリ MEVAL（Multi-party EVALuator）を用いて，ToMMo[☆10] と共同で GWAS 実験を行いました．検定としては最も正確である反面処理コストが大きい，Fisher の正確確率検定を 1,000 件× 100 万 SNP の模擬データに対して適用し，約 8 分という，GWAS としてはきわめて短時間での処理が可能なことを確認しました．

2015 年：浮動小数点演算や主要な数学的関数の実現

2015 年に Kamm らは，DARPA[☆11] の出資のもと行われた，人口衛星の所有者（アメリカやロシアなど）が互いに衛星軌道を明かさずに衛星の衝突可能性を計算するという実験について報告しました．報告では，最良のパラメータにおいて人口衛星の 1 対あたり 60 秒で衝突可能性を計算できたとされています．

報告の中では実験のために Sharemind に実装された，浮動小数点演算（加算・乗算・除算等）や，平方根，指数関数，さらには誤差関数のアルゴリズムと実装について言及されています．Lindell と Pinkas の“Privacy Preserving Data Mining”で初めて研究された秘密計算による対数関数，その双対である指数関数がついに実現されました．そして対数関数も，正弦関数も，よく使われる数学的関数はいずれも同じ，多項式近似アルゴリズムで計算されますので，指数関数の実現は多くの処理の実現可能性を意味しています．

そして，現在

性能面の状況

性能が課題であった秘密計算ですが，現在はどのような性能になっているのでしょうか．性能面で先進的な実装の 1 つである前述の MEVAL による，表 -1 のような性能が例として挙げられます．

秘密計算でない通常の処理のシングルスレッド性能と比較してみましょう．たとえば“普通に行われるコンピューティング”の一代表として，私のノート PC（CPU 3.0GHz, memory 32GB）での処理と比較してみます．この場合，256 ビット論理演算を 1 サイクルあたり 3 回処理できて 768 素子，乗算が 1 サイクルあたり 1 回処理できるとして，それぞれ 3.0GHz の CPU で 1 秒あたり約 2,300G 素

■表 -1　MEVAL の性能例 @1 億データ

演算	処理時間 [ms]	スループット [M/s]
論理回路素子	11	9090.09
乗算	446	224.22
ソート	33,700	2.97

CPU：Core i7 6900K（最大 4.0GHz），メモリ 32GB，
NW:10G（リングトポロジ）

☆6 Single Nucleotide Polymorphism．遺伝子中で，一定の頻度以上で個人差が存在する部位の塩基情報．
☆7 2 群の差異の有無を評価する検定の一種．
☆8 x^2 検定は分布に仮定を必要とするのに対し，分布の仮定を必要としないタイプの検定の一種．
☆9 伝達不平衡試験（transmission disequilibrium test）．塩基間の相関関係の存在下で親から子への遺伝の仕方の相関の有無を検定する方法．
☆10 東北メディカルメガバンク．
☆11 Defense Advanced Research Projects Agency，アメリカ国防高等研究計画局．

子，1秒あたり3G乗算です．そしてC言語のqsort関数が実測16秒です．論理回路素子で1/240，乗算で1/15，ソートで1/2のスループットと，重要な演算ごとに専用の高速アルゴリズムを用意するという前述の方針の効果により，高級な演算ほど通常の処理に近づいています．実用的なコンピューティングのボトルネックは多くの場合ソートですから，性能面ではかなり実用性が高まってきていることが分かります．

実用化動向

すでにいくつかの企業は，秘密計算によるサービスを実施しています．紹介した中ではエストニアCybernetica社のSharemind，デンマークPartisia社のPartisiaがあります．さらに“Privacy Preserving Data Mining”著者のLindellが出資者の1人である，イスラエルUnbound社は秘密計算による暗号化鍵管理サービスを展開しています．

利用可能なソフトウェア

本稿で紹介した中では，Sharemindはインタフェースの体感および，性能見積もりが可能です．FairPlay MPはFairPlayの後継です．ほかにも国際会議TPMPCのWebサイト[☆12]で多くのライブラリが紹介されています．読者の皆さんも，秘密計算に触れてみてはいかがでしょうか？

(2018年6月29日受付)

☆12 http://www.multipartycomputation.com/mpc-software

五十嵐大（正会員） ikarashi.dai@lab.ntt.co.jp
2008年東京大学大学院情報理工修士課程修了，同年NTT入社．秘密計算，匿名化，秘密分散などの研究開発に従事．PPL2008論文奨励賞，CSS2009/SCIS2011/CSS2012/CSS2013/CSS2017各論文賞，2011年度本会論文賞，2012年度山下記念研究賞，SCIS 2017イノベーション論文賞．