プライバシー保護データ解析技術の社会実装

統計数理研究所オープンハウス
“人と社会に寄り添うデータサイエンス”
2024.05.24
産業技術総合研究所
首席研究員
花岡 悟一郎

自己紹介

花岡　悟一郎（はなおかごいちろう）

（国研）産業技術総合研究所
サイバーフィジカルセキュリティ研究センター
首席研究員

兼務：内閣府SIP第3期（量子）サブPD

【専門分野】

高機能暗号技術：

「データを暗号化したまま、様々なデータ処理を実行可能」な暗号

- データだけでなく、計算を守る！

【経歴】

1997年 東京大学工学部卒
2002年 東京大学大学院博士課程修了
　　電子情報工学専攻・博士（工学）
2005年 産総研入所

1997年以降一貫して高機能暗号を研究

【受賞】

科学技術分野の文部科学大臣表彰
2018年・科学技術賞・研究
「高機能暗号およびその秘匿検索技術への応用に関する研究」

ドコモ・モバイル・サイエンス賞
2016年・先端技術部門・優秀賞
「安全性を暗号学的に保証した次世代生態認証基盤技術の先駆的研究と実用化」

The Wilkes Award
British Computer Society, 2007
The Comp. J.誌における年間最優秀論文賞

他、電子情報通信学会論文賞（2008, 2018）等

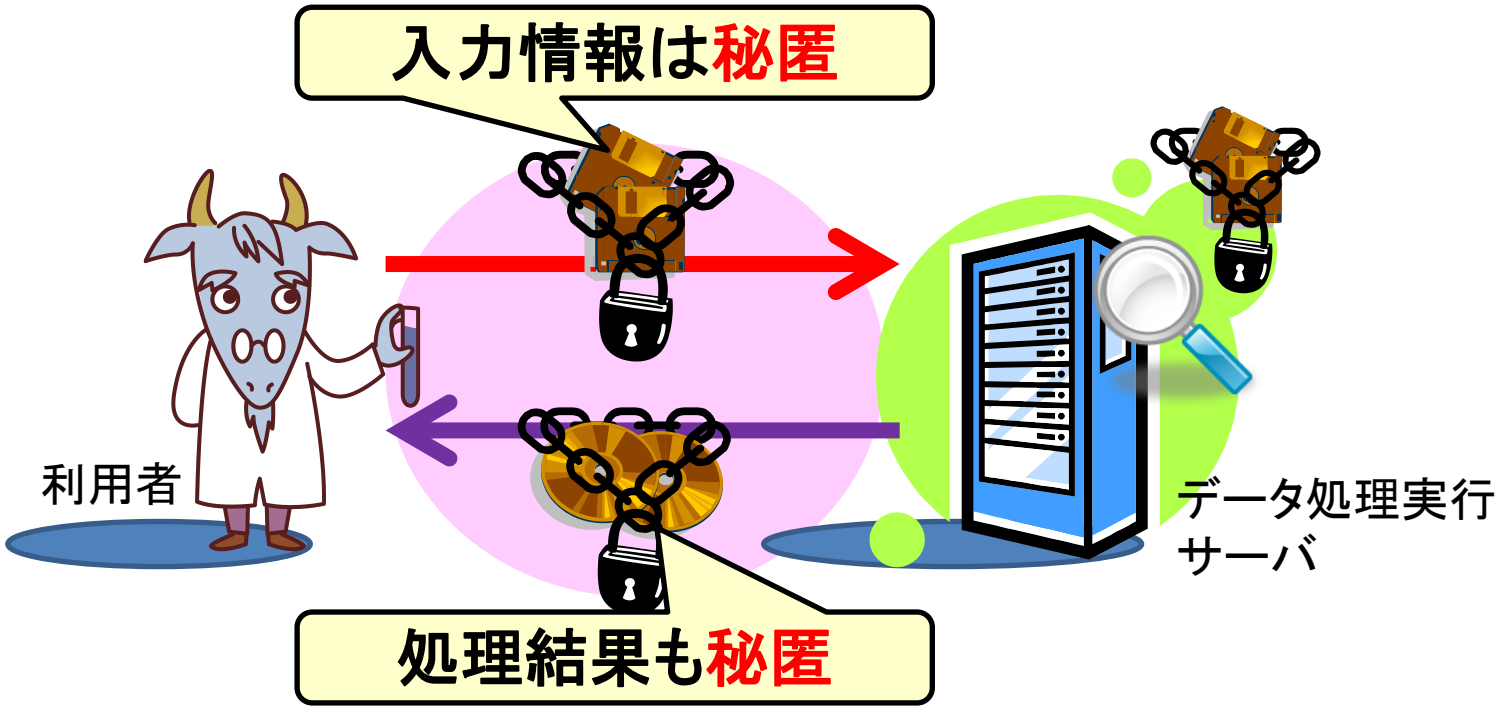
【競争的研究資金】

JST-CREST人工知能領域・加速フェーズ（研究代表者）
「プライバシー保護データ解析技術の社会実装」（H31-）
JST-AIP加速課題（研究代表者）「秘匿計算による安全な組織間データ連携技術の社会実装」（R4-）

秘密計算・秘匿計算

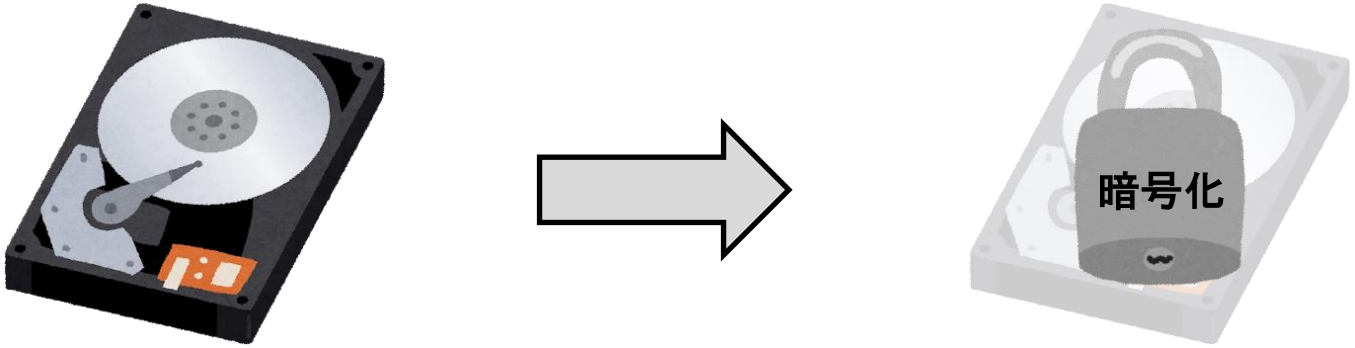
- **秘密計算（秘匿計算）**とは？

⇒ 入出力情報を隠したままデータ処理を行う技術

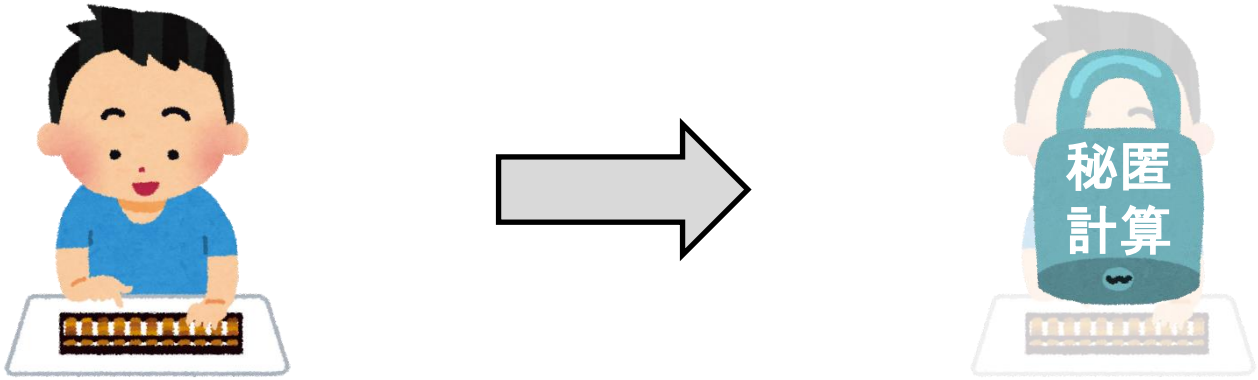


単なる暗号化との違い

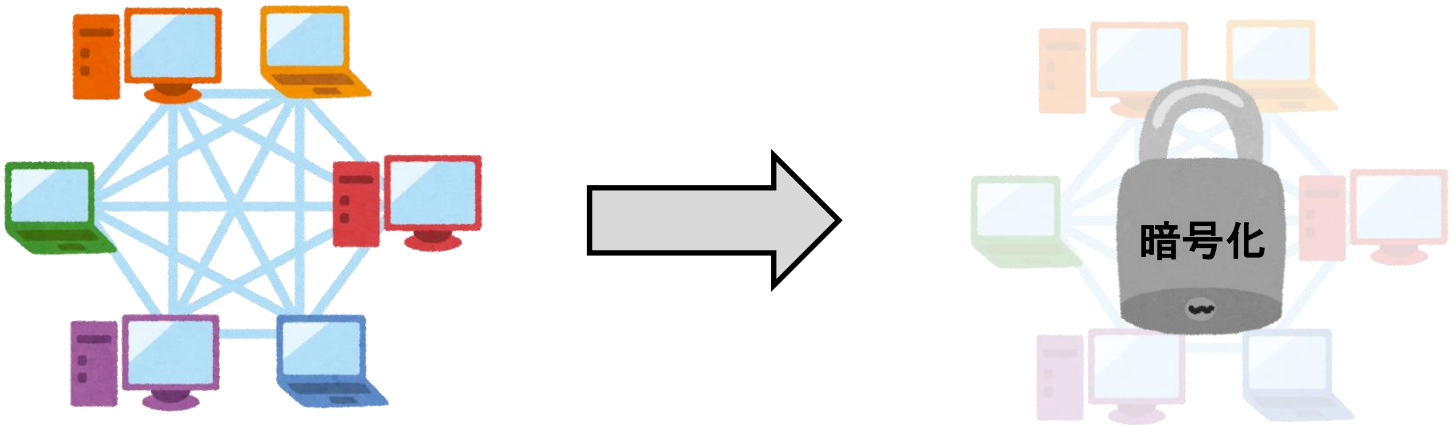
- 記録



- 計算



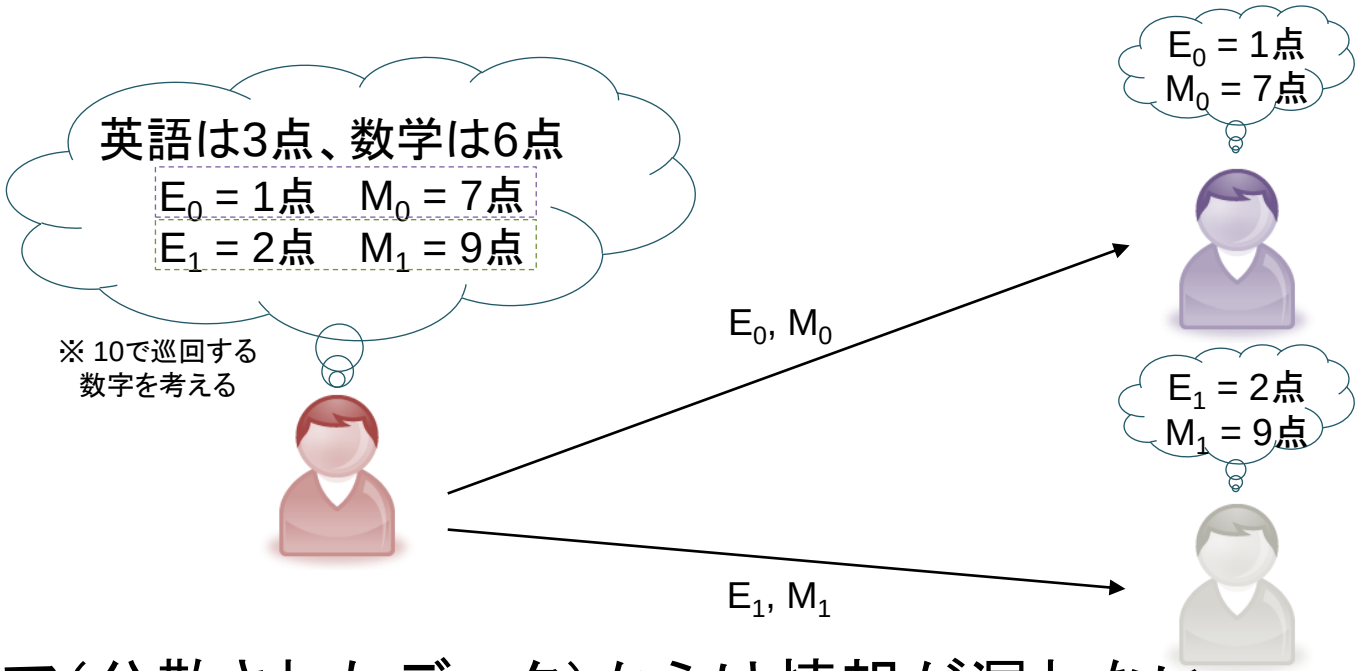
- 通信



MPCを実現するためのツール：秘密分散

- 元々はデータ紛失対策
  - 分散したデータが事前に指定した数だけ集まらないと復元できない

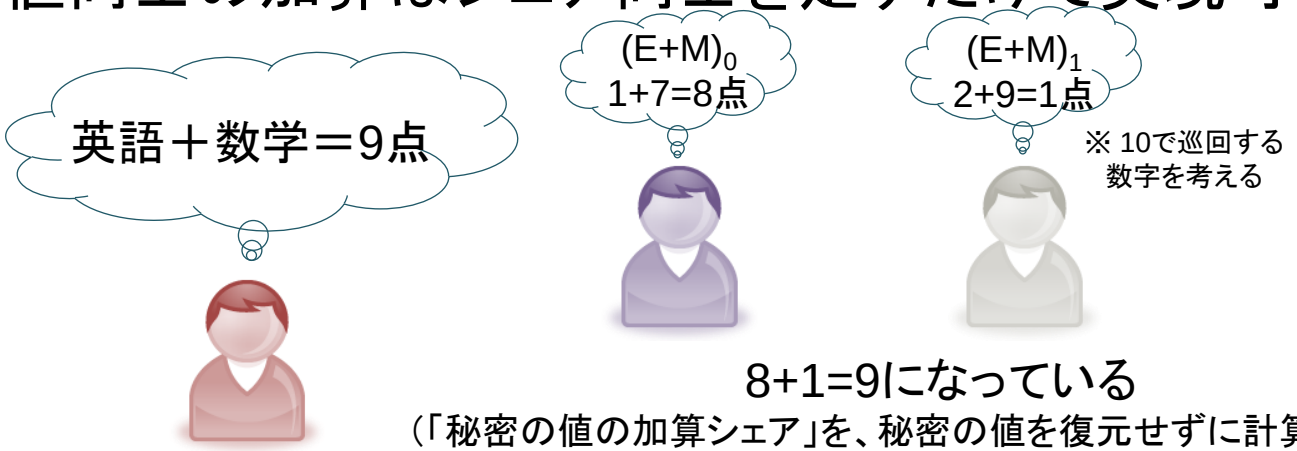
例: (2,2)-加法分散
（足したら元に戻るような2つのシェアに分散され、2つとも集まるとデータを復元できる）



- シェア（分散されたデータ）からは情報が漏れない
  - 「情報理論的に安全」という
- シェアを持つ2者が結託すると秘密が復元されてしまう
  - 分散数と復元閾値は柔軟に選べるが、MPCの都合で主流は(2,2) or (2,3)

秘密分散に基づくMPC

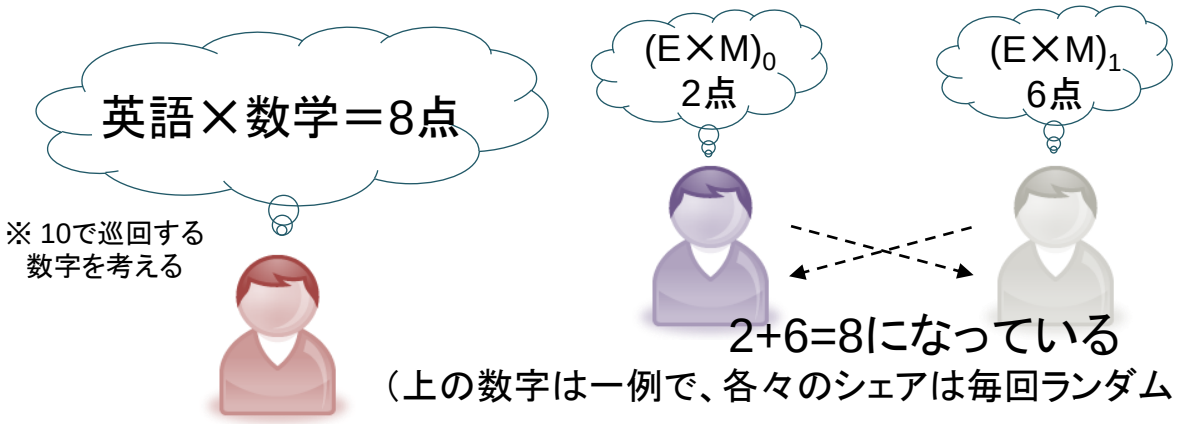
- 秘密の値同士の加算はシェア同士を足すだけで実現可能



8+1=9になっている

（「秘密の値の加算シェア」を、秘密の値を復元せずに計算できた）

- 秘密の値同士の乗算もやや複雑ではあるが可能
  - 通信が1回必要になるが、秘密の値を復元しないのは加算の場合と同じ
  - この通信がMPC実行速度のボトルネックになる場合が多い



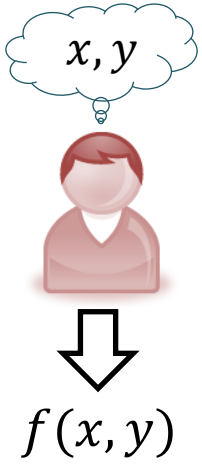
2+6=8になっている

（上の数字は一例で、各々のシェアは毎回ランダム（和は8）になる）

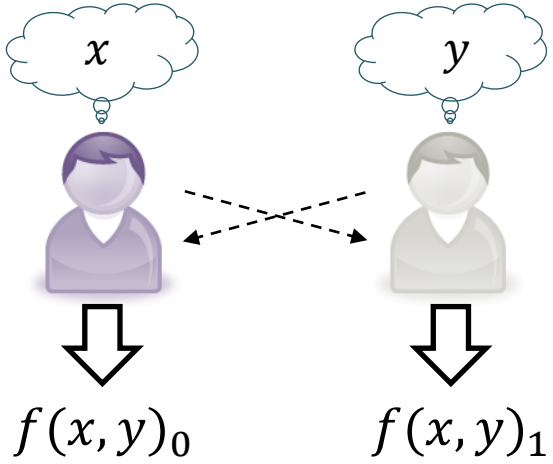
秘密分散に基づくMPC

- 加算と乗算が出来れば任意の関数fを計算可能
  - 乗算するたびに通信が1回発生する→乗算が多い関数の実行には時間がかかる
    - 特にインターネットなどの高レイテンシ環境だと厳しい
    - 非線形関数を何度も計算する必要がある深層学習は一番シビアな例

- 通常の計算



- 秘匿計算



$f(x, y)_0 + f(x, y)_1 = f(x, y)$になる

- 準同型暗号では扱えなかった関数も計算できるようになりつつある（e.g. 分散状態での深層学習訓練）
  - 良好な通信環境、潤沢な計算リソース、高度な実装によるところも大きい

さらなる研究開発要素

- 加算と乗算の秘匿計算を組み合わせることで
  任意のデータ処理の秘匿計算が（原理的には）可能

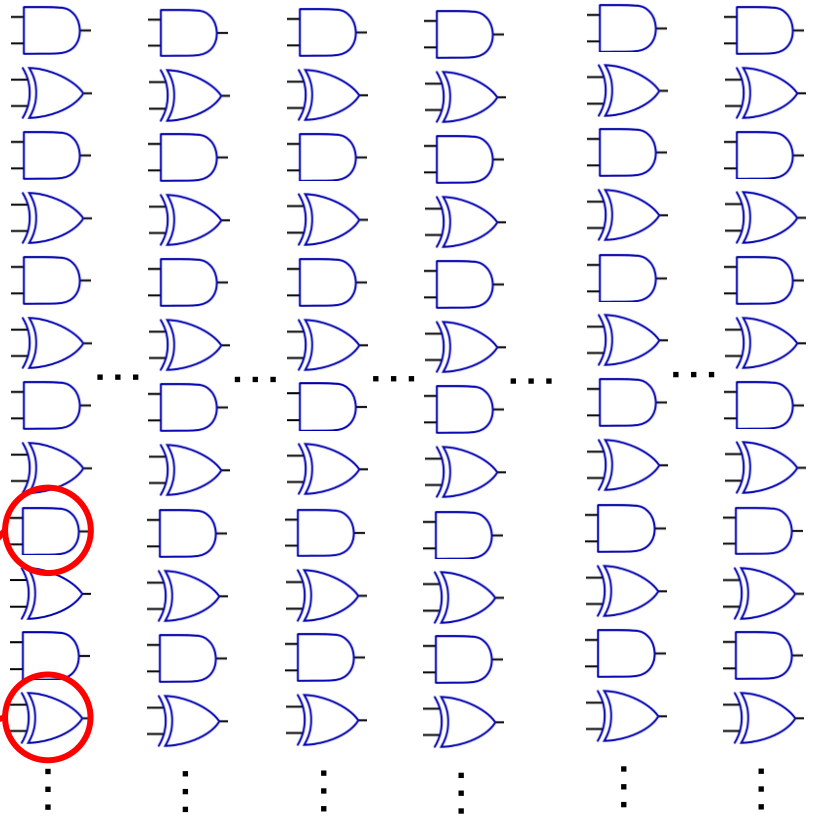
- あくまで「原理的」にはということ
- 実用的には、それ以上の研究開発が必要

これがこの30年間に積み上げられてきた

「原理的な」MPC

秘匿化したいデータ処理

AND ⇒乗算の秘密計算
XOR ⇒加算の秘密計算
で個別に秘匿化



ANDとXORによる論理回路として記述

通常、ゲート数が膨大となるため、処理時間も膨大に！

具体的な研究開発内容

- 加算と乗算の秘匿計算自体を高速化

- 加算と乗算以外の基本処理の秘匿計算の開発

- 多数の基本秘匿計算処理の具体的な活用
  - 目的とする処理をどのように基本処理に分解するか？

- 実践的な実装
  - 複数台の秘匿計算サーバを結託のないような設置・配置
  - 低級言語による高速実装・最適化

具体的な研究開発内容(おおよそ)の現状

完了

- 加算と乗算の秘匿計算自体を高速化

概ね完了

- 加算と乗算以外の基本処理の秘匿計算の開発

要研究開発

- 多数の基本秘匿計算処理の具体的な組合せ
  - 目的とする処理をどのように基本処理に分解するか？

➡ 現状は専門的研究者・開発者が職人的に対処

要研究開発

- 実践的な実装

➡ ビジネスの形態に強く依存。SIerの役割？

  - 複数台の秘匿計算サーバを結託のないような設置・配置
  - 低級言語による高速実装・最適化

➡ 高速化・最適化は突き詰められていない印象

現時点における開発の流れ

具体的に秘匿データ処理システムを作る際は、

基本秘匿
計算セット

1．秘匿化したい処理を決め
専門的開発者に伝える

2．専門的開発者が
基本処理に分解、
基本処理毎の秘匿化に
より処理全体を秘匿化

3．専門的SIerがサーバを
設置・配置し、運用方針
を決定

現時点における開発の流れ

具体的

「専門的開発者」の体制には
トップ国際会議に採択される
レベルの技量が必要

1．秘匿化したい処理を決め
専門的開発者に伝える

2．専門的開発者が
基本処理に分解、
基本処理毎の秘匿化に
より処理全体を秘匿化

3．専門的SIerがサーバを
設置・配置し、運用方針
を決定

人工知能CREST（2016.12～2022.3）
AIP加速課題（2022.4～）

JST CREST Core Research for Evolutional Science and Technology
Japan Science and Technology Agency

[人工知能] イノベーション創発に資する人工知能基盤技術の創出と統合化

← トップに戻る

領域活動についてはこちらをご覧ください。

戦略目標

「急速に高度化・複雑化が進む人工知能基盤技術を用いて多種膨大な情報の利活用を可能とする統合化技術の創出」

研究総括

栄藤　稔(大阪大学 先導的学際研究機構　教授)

プライバシー保護データ解析技術の社会実装

研究代表者

花岡　悟一郎

国立研究開発法人産業技術総合研究所

主たる共同研究者

盛合　志帆	国立研究開発法人情報通信研究機構　サイバーセキュリティ研究所　研究所長
浅井　潔	東京大学　大学院新領域創成科学研究科　教授
小澤　誠一	神戸大学　数理・データサイエンスセンター　教授
菅原　貴弘	株式会社エルテス 代表取締役

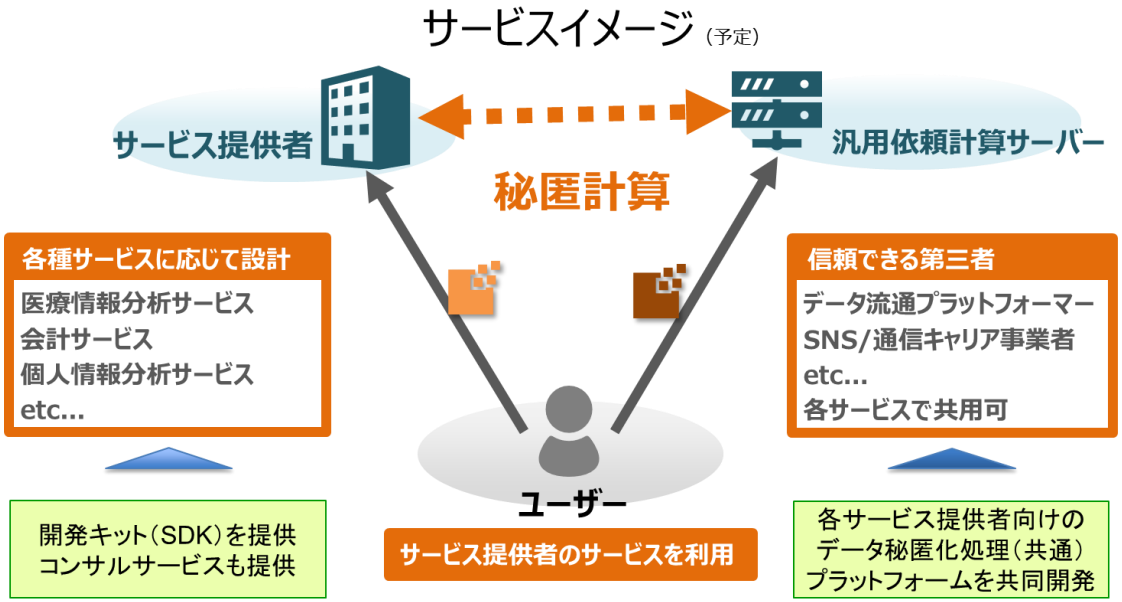
秘匿計算の社会展開を妨げる壁

■ **計算速度 (Performance)**
- 平文でできていたことが全て秘匿計算に置き換えられる訳ではない
- 最近のプロトコル改善とコンピューティング能力向上によって大きく進歩

■ **使い勝手 (Usability)**
- 広く一般で利用されるには、一般のエンジニアレベルで利用可能であることが望ましい

■ **社会への浸透 (Social Penetration)**

簡便に利用可能な汎用秘匿計算システムの開発



- 実利用を想定したモジュール化
  - ✓利用者企業へのヒアリング
  - ✓実ニーズの抽出と反映
  - ✓Python-likeな使用感
- アルゴリズムの理論的高速化
  - ✓最新の知見をもとに洗練化
  - ✓数学的安全性証明も
- 高速実装
- サンプルアプリケーションの実装
  - ✓事業展開の際の顧客向けサンプル

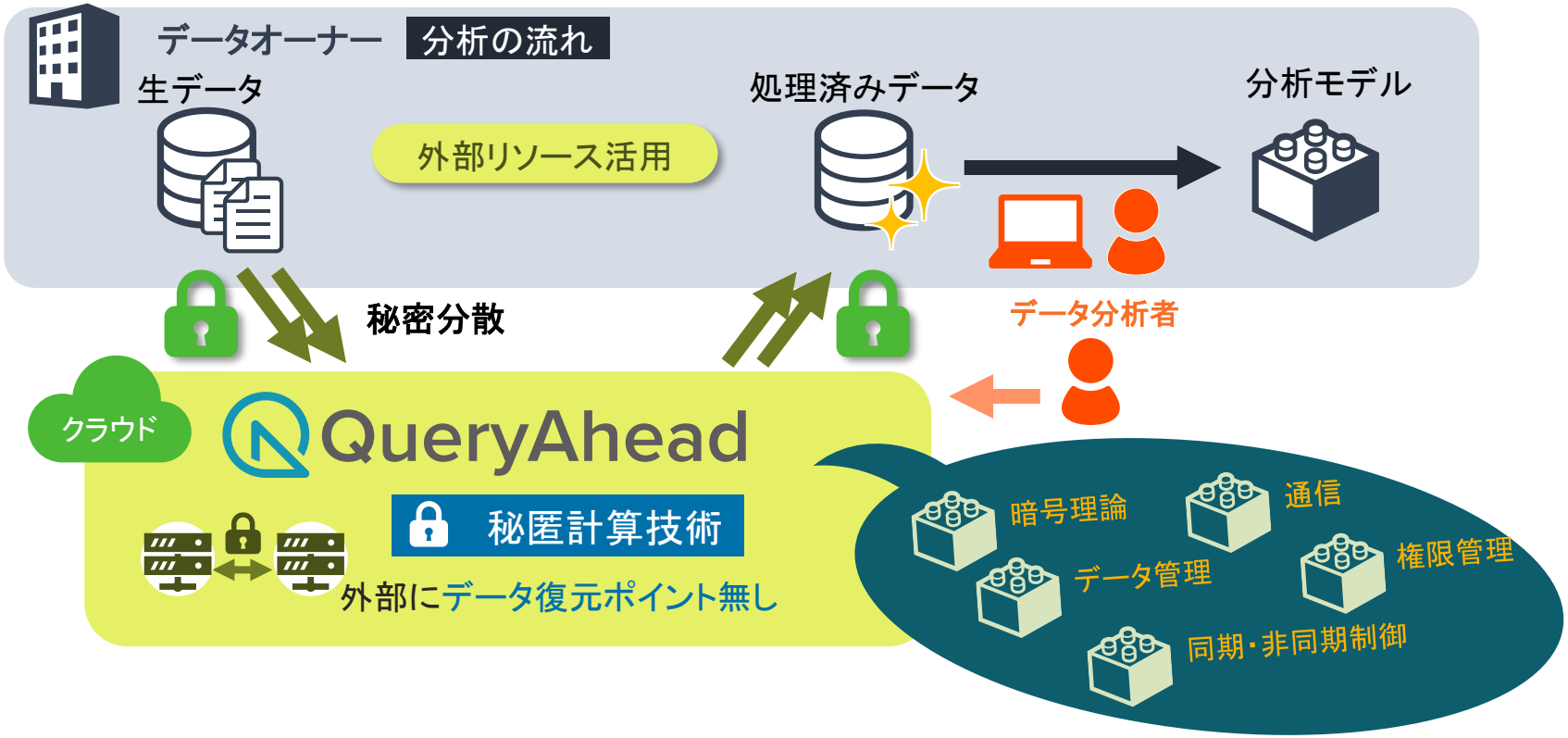
ZENMU

秘匿DBクエリーシステム QueryAhead™
概要

■ **テーブルベースのデータベース（DB）へのクエリーを秘匿計算で実行可能**
- データ分析企業からのヒアリングでテーブル情報の保護に高いニーズがあることを確認

■ **暗号の専門知識を必要としないクエリー記述が可能**
- 暗号理論、データ管理、通信制御等をパッケージ化、データ分析者がこれらを意識する必要無し

■ **データの利活用でクラウドの有効活用も可能**
- 外部にデータ復元ポイントを一切なく、データの利活用を実現



秘匿DBクエリーシステム QueryAhead™

機能 - 暗号の専門知識を必要としないクエリー記述

■ **データ分析者が慣れ親しんでいるSQL的な操作が可能**

- 基本演算 (加算 / 減算 / 乗算 / 除算)
- SELECT (列の選択)
- WHERE (各種条件での抽出)
- FROM (テーブルの選択)
- JOIN (テーブル結合)
- ORDER BY (整列)
- 基本統計演算 (平均 / 最大 / 最小 / 合計)
- GROUP BY (集約)　　など

■ **数行のPythonプログラムの記述で秘匿クエリー発行可能**

SELECT句

WHERE句

FROM句

```
q = (Query('result')
     .select(date,
             kari_code,
             kashi_code,
             kari_amount,
             kashi_amount,
             (kari_code + kashi_code) * 10)
     .where(date == _date('2019/8/27'))
     .from_('journal'))

app_client.req_query(q)
```

基本演算の記述

日付に一致する行を抽出

平文のデータ分析の感覚で秘匿クエリーを利用できる

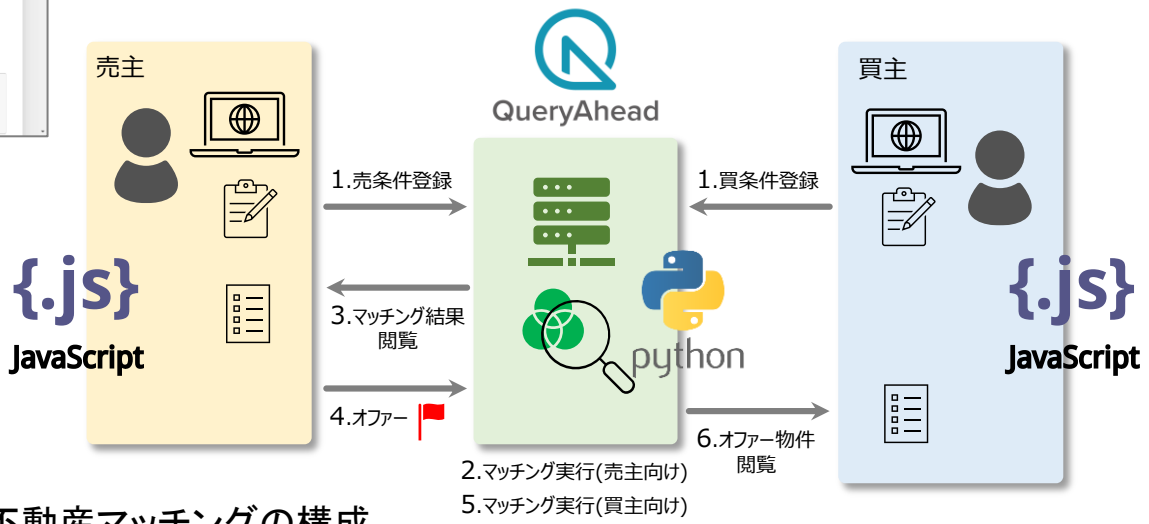
秘匿DBクエリーシステム QueryAhead™

特徴 - 事前検証から実システム構築までサポートする汎用性

- Webブラウザーから Jupyter Notebook経由でインタラクティブにクエリー可能
- サーバー側バックエンドのDBとしてシステム化
  - 実際に秘匿匿名加工、秘匿不動産マッチング、秘匿重回帰、秘匿ベイズ最適化といったところで活用

Jupyter Notebookで事前検証

システム化



秘匿不動産マッチングの構成

AIST PRESIDENT AWARD

令和４年度理事長賞
（研究）

「汎用的秘密計算技術の開発と社会実装の推進」

花岡 悟一郎、ATTRAPADUNG, Nuttapong、松田 隆宏、
大原 一真、坂井 祐介、山田 翔太
SCHULDT, Jacob、照屋 唯紀、勝又 秀一
（サイバーフィジカルセキュリティ研究センター）

「汎用的秘密計算技術の開発と社会実装の推進」

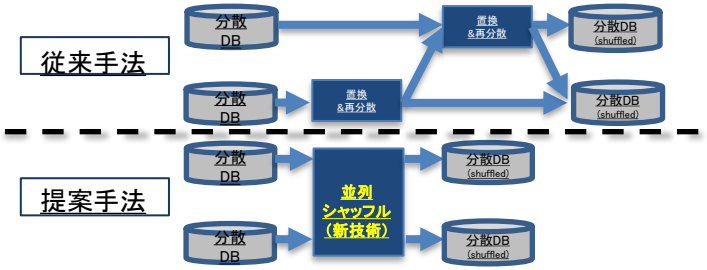
主な成果：

論文業績
（2019.4-2022.3）

論文		
88件	査読付き国際会議＆論文誌 うち、トップ国際会議が 37 件 CCS、CRYPTO、EUROCRYPT、ASIACRYPT、TCC、PKC、AsiaCCS	
発表	**書籍・総説**	**受賞**
65件	27件	12件
国内研究会等	プレプリント等	第17回情報セキュリティ文化賞（盛合） EUROCRYPT 2020 Best Paper Award（山田）

☆例えば：

秘匿シャッフル技術を従来の100倍以上高速化



・トップ国際会議 ACM-CCS 2021に採録
Google Scholar Computer Security Cryptographyカテゴリ **1位**

・QueryAhead（左記）に**導入・実装**

【再掲】秘匿計算の社会展開を妨げる壁

■ **計算速度 (Performance)**
- 平文でできていたことが全て秘匿計算に置き換えられる訳ではない
- 最近のプロトコル改善とコンピューティング能力向上によって大きく進歩

■ **使い勝手 (Usability)**
- 広く一般で利用されるには、一般のエンジニアレベルで利用可能であることが望ましい

■ **社会への浸透 (Social Penetration)**

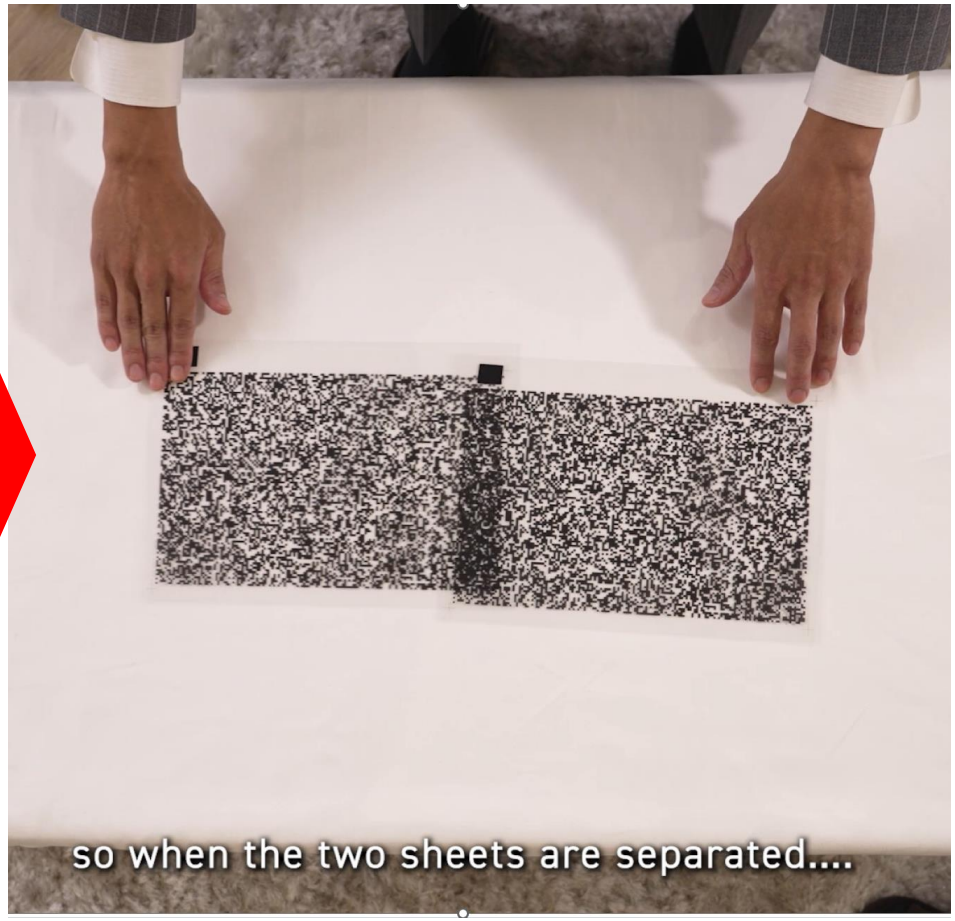
具体的な技術的な中身が腹落ちしないこと自体が最大の障壁

秘匿計算の仕組みを腹落ちさせる説明手法の研究開発

ハイライト(1)


AIST
何を守ったまま
どのような目的を、
達成したいのか明示
But you can't let others see who chose who!

ハイライト (2)



視覚化：個々のシェア情報が完全にランダム

ハイライト (3)


AIST
視覚化:シェア状態のままで処理が
進むためその過程で情報が漏れない
The shares are collected and shuffled separately
for men and women.

Highlight (4)


AIST
視覚化：対になるペア同士が同時に
ひとりの人物に集まらないようにする
The shares are collected and shuffled separately
for men and women.

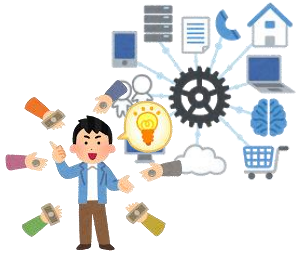
ハイライト(5)

解説編動画

解説編動画

課題解決がもたらすイノベーションの創出

- プライバシー保護データ解析技術に基づく
  ビジネス展開が可能に



- 人工知能に基づく、機微情報を用いた
  情報サービスの提供が加速

- 機微情報を用いたデータ社会の恩恵を、
  誰もが、簡単に、安全に享受

誰もが、簡単に、安全に → イノベーション