On Correctness and Privacy in Private Information Delivery with Coded Storage

栗原淳，中野光喜[1]，田中俊昭

兵庫県立大学大学院

2025-11-27 (SITA 2025)

[1]本発表は第 2 著者の博士前期課程在学中の研究成果 [NKT26] を含む．

発表の流れ

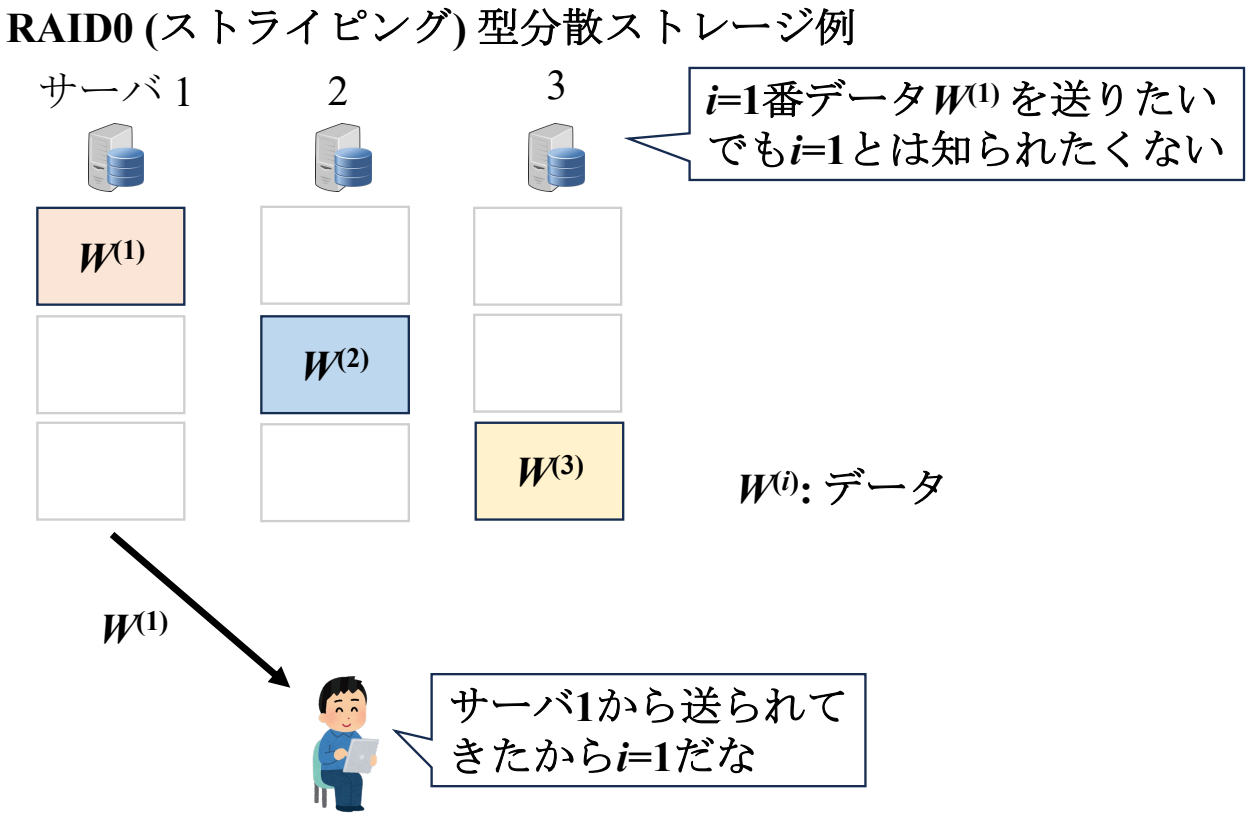
1. はじめに: PID の位置匿名性と課題
2. PID with Coded Storage の再定式化，
   Correctness と Privacy の定義
3. 主成果: Correctness と Privacy を満たす条件
4. まとめ

はじめに

Private Information Delivery (PID) [Sun20]

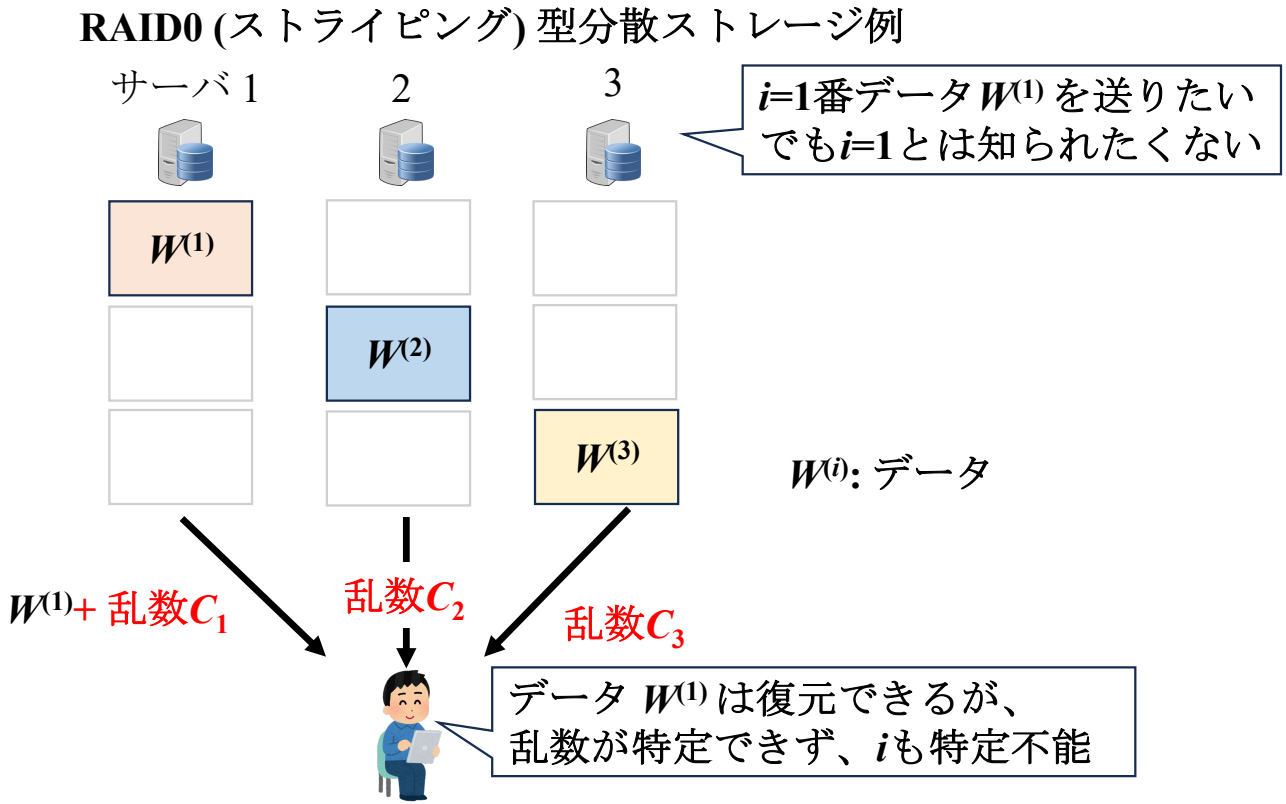
分散ストレージサーバからユーザへのプッシュ型データ配信において，
サーバのプライバシをユーザから保護する手法

[設定] 分散ストレージからデータ配信を行う
このとき、ユーザは分散ストレージ内のデータ配置を知っている



ユーザはどのサーバからデータ配信されたかわかれば，i が特定可能

[目的] 配信データ $W^{(i)}$ のインデックス i の秘匿
[手段] 乱数を各サーバへ配置．データ $W^{(i)}$ を乱数でシールし，どのサーバが $W^{(i)}$ を送るのかを秘匿



このとき，i によらず，ユーザが常に同じ計算手法でデータを復元可能なように，事前配置する乱数を設計

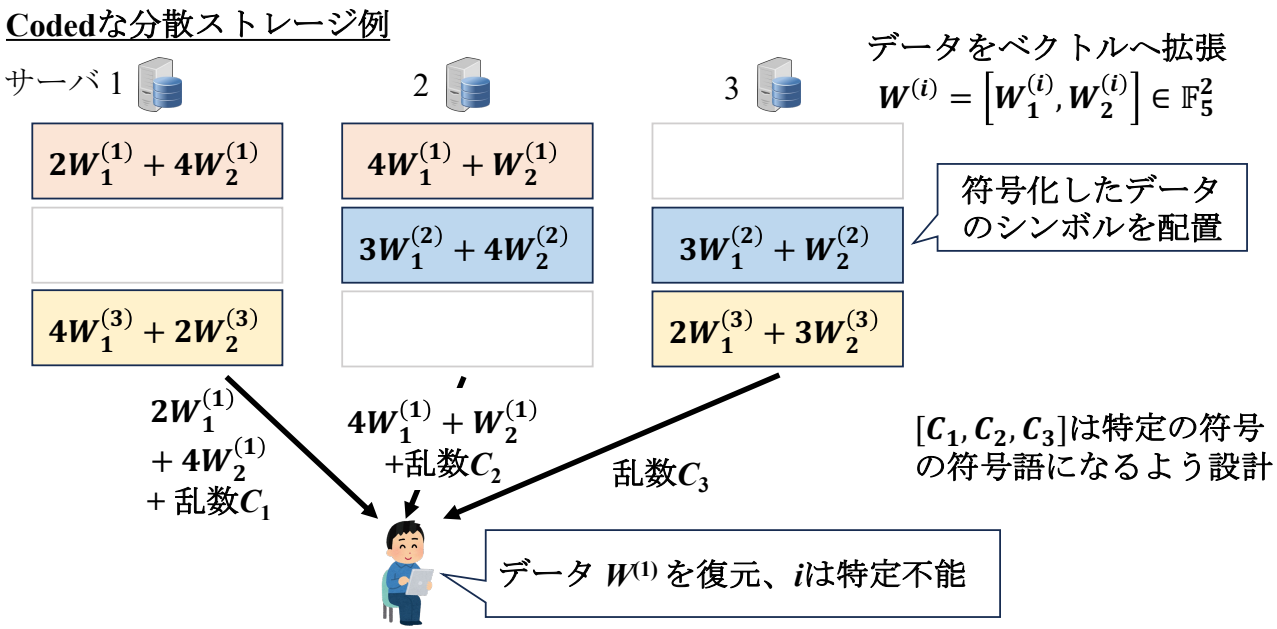
PID with Coded Storage [VR22]

オリジナル PID [Sun20]: データ $W^{(i)}$ をそのまま保存 (non-coded)

PID with Coded Storage [VR22]

$W^{(i)}$ を符号化して保存するように，PID のストレージモデルを拡張．

- PID の伝送効率 [2] を改善
- 特定のデータ配置の下での伝送率の理論限界を達成する，最大距離分離 (MDS) 符号ベースの手法を提案



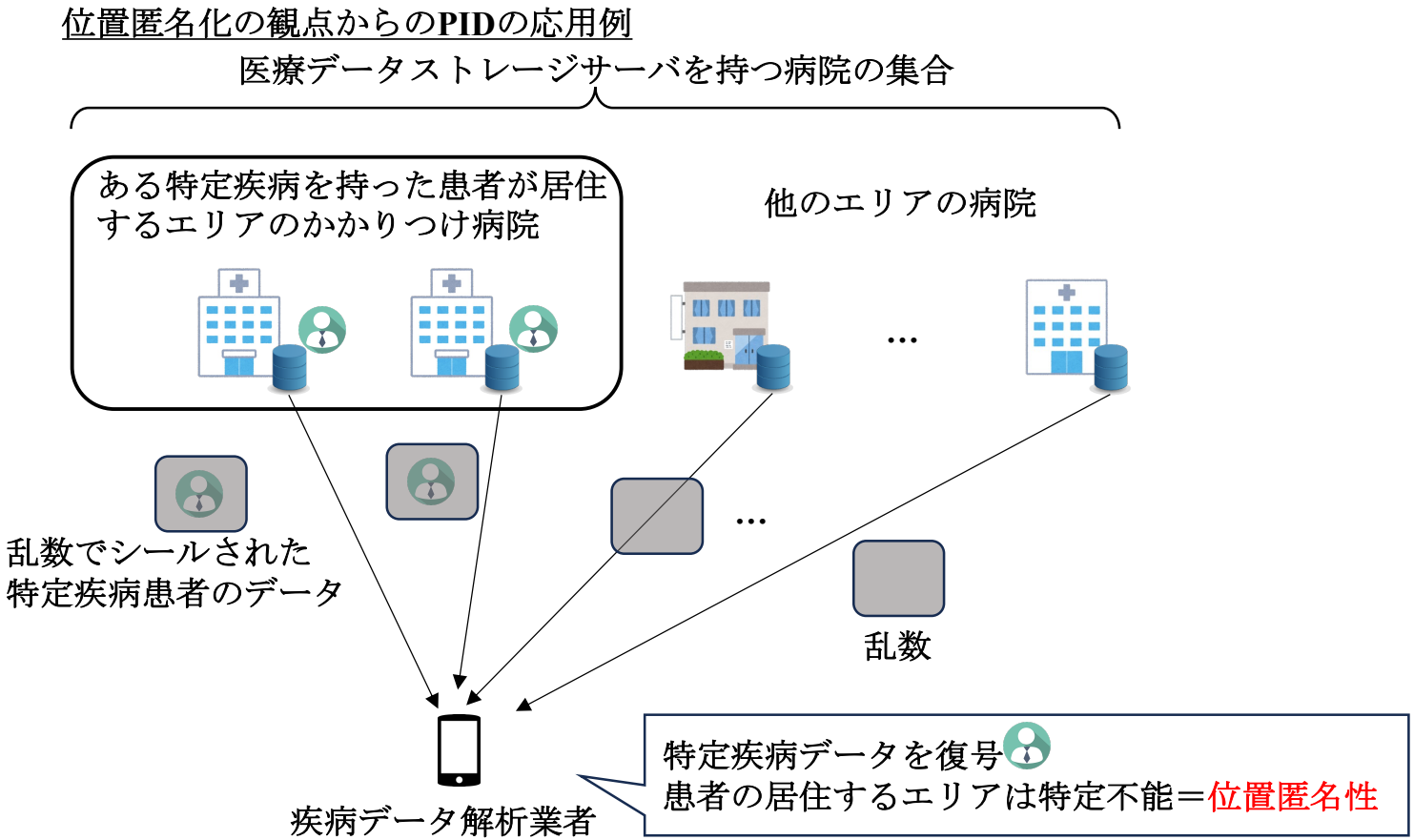
[2] $W^{(i)}$ のサイズ/伝送データサイズの合計

PID によって実現される位置匿名性

PID (with Coded Storage) におけるデータ配置の特徴

インデックス $i \Leftrightarrow W^{(i)}$ (の符号化データ) を保存するサーバの集合

$\Rightarrow$ インデックス i の秘匿は，$W^{(i)}$ を保存するサーバの位置匿名化と同義



PID with Coded Storage の課題

課題 1: データ配置の偏り・不均一性の考慮不足

RAID のような均一なデータ配置[3] のみを考慮したモデル化
⇒ 例のような匿名化では，データ配置の偏り・不均一性の考慮が必要
(※特定疾病患者は特定地域に偏在する等)

課題 2: MDS 符号の制限と，任意の符号ベースの手法の性能の未知性

既存手法は，MDS 符号の制約[4] により，大規模システム等で計算量的制限の可能性
⇒ 任意の符号ベースへの一般化が必要だが，その際の性能・性質が未知

[3] Bi-regular setting

[4] 符号長＝サーバ数が，有限体サイズに制約される [MS77]

本研究の貢献・主成果

課題 1 に対して: データ配置モデルを一般化した再定式化

- データ配置を一般化し，**位置匿名性の観点から PID with Coded Storage の問題を再定式化**
- 再定式化されたモデルにおいて，PID の成立条件を定義
  Correctness: ユーザが常にデータを復元可能
  Privacy: ユーザがデータを配信するサーバ集合を特定不能

課題 2 に対して: 任意の符号ベースでの成立条件の導出

再定式化されたモデルにおいて，**任意の線形符号 $\mathscr{C}$ をベースとした手法が，PID として成立するための $\mathscr{C}$ の十分条件を導出**

PID with Coded Storage の再定式化

再定式化: PID with Coded Storage の 2 段階のフェーズ

1. 配置フェーズ:

$\mathbb{F}$: 有限体，n: サーバ数，$W \triangleq [W_1, \ldots, W_l] \in \mathbb{F}^l$: データ (サイズ $l < n$)

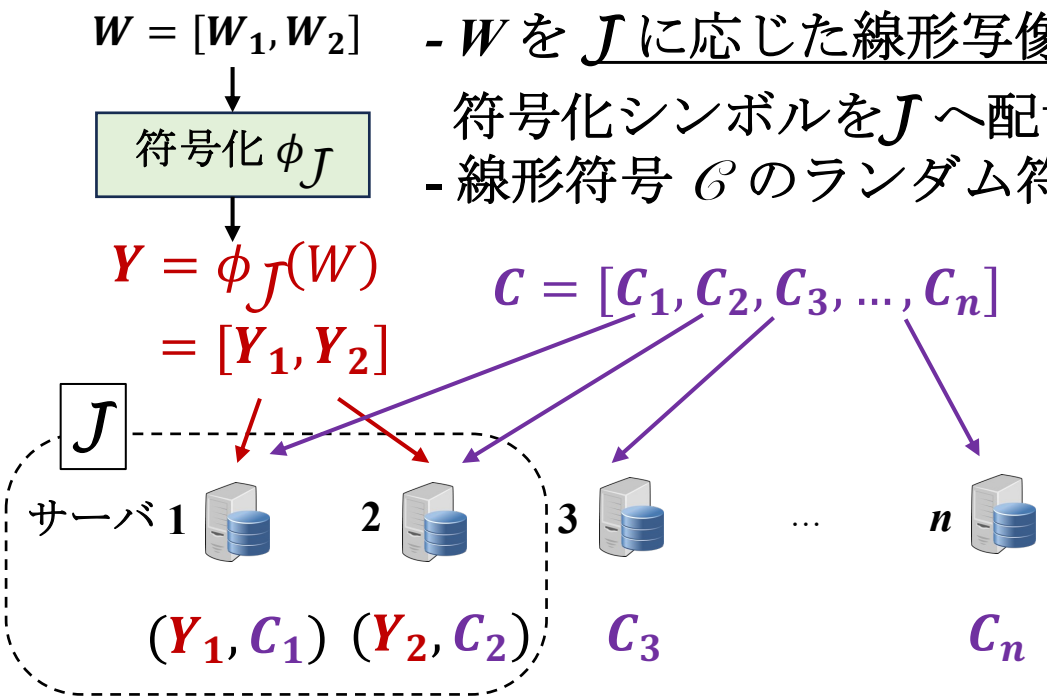
$\mathscr{C} \subseteq \mathbb{F}^n$: $n - l$ 次元の線形符号

$\mathcal{F} \subset \{\mathcal{J} \subset \{1, \ldots, n\} : |\mathcal{J}| \geq l\}$: 配置可能なサーバ集合の族

$\mathcal{J} \in \mathcal{F}$ より任意に選択.

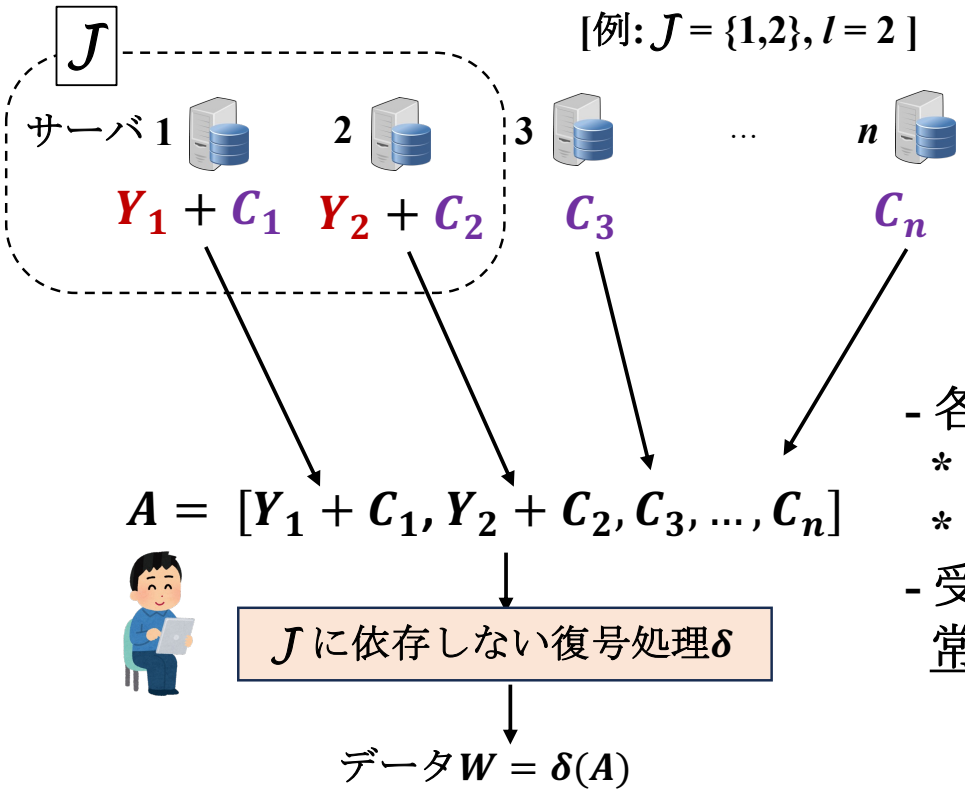
$C \in \mathscr{C}$ は W および $\mathcal{J}$ と独立ランダムに選択.

[例: $\mathcal{J} = \{1,2\}, l = 2$]



- W を $\mathcal{J}$ に応じた線形写像 $\phi_{\mathcal{J}}$ で符号化
  符号化シンボルを $\mathcal{J}$ へ配置
- 線形符号 $\mathscr{C}$ のランダム符号語 C の要素を各サーバへ配置

2. 配信フェーズ:



- 各サーバは、配信シンボルを生成・送信
  * 符号化シンボル + 符号語シンボル; or
  * 符号語シンボル
- 受信者は、配信シンボルを $\mathcal{J}$ に関係なく常に同じ線形写像δで処理

データのインデックス等を考慮せず，サーバの部分集合を直接考慮

$\mathscr{C}$ が MDS 符号，また $\mathcal{J}$ の族 $\mathcal{F}$ を制限した場合，既存手法 [VR22] に対応

再定式化されたモデルでは，以下の 2 つの関数の設計が必要:

- 配置フェーズにおける，$\mathcal{J}$ に応じた関数 $\phi_{\mathcal{J}}, \forall \mathcal{J} \in \mathcal{F}$
- 配信フェーズにおける，$\mathcal{J}$ に非依存な関数 δ

$\Rightarrow$ 主成果では，PID として満たすべき性質を保った上で，
これらの関数が設計可能な条件を導出

PID として成立するために満たさなければならない性質

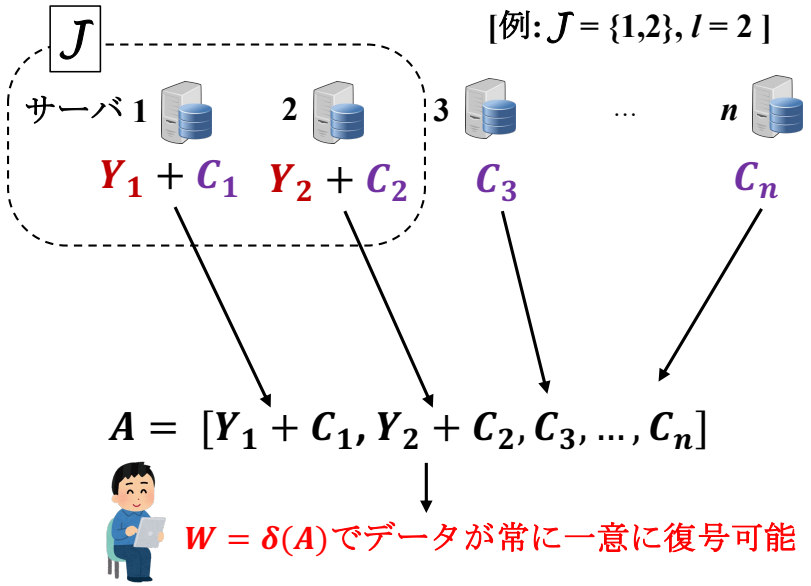
サーバ部分集合 $\mathcal{J}$ について再定式化した手法が，PID with Coded Storage として成立するために必要な 2 つの性質を定義

$H(\cdot)$: Shannon エントロピー

定義 1. Correctness

配置可能なサーバ集合 $\mathcal{J}$ の族 $\mathcal{F}$ について，**配信シンボル A から常にデータ W を一意に復元可能**．すなわち

$$H(W|A) = 0, \forall \mathcal{J} \in \mathcal{F}.$$

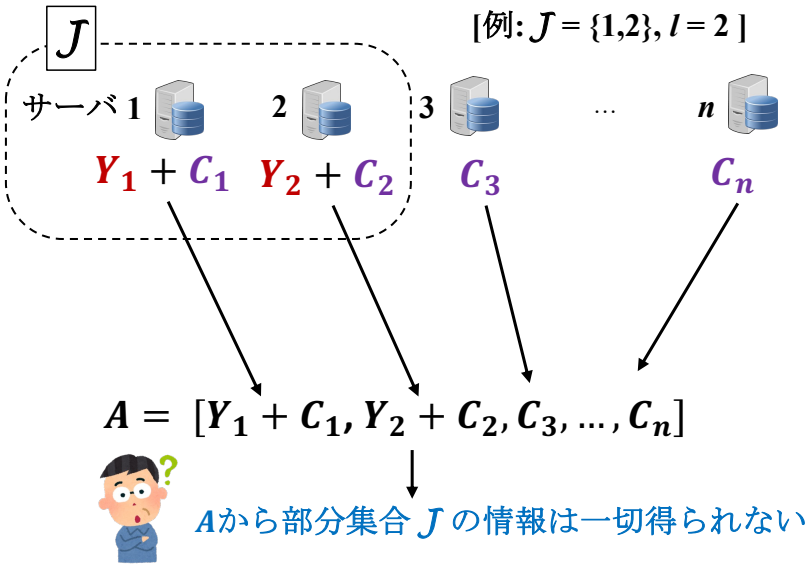


$I(\cdot;\cdot)$: 相互情報量

定義 2. Privacy

配置可能なサーバ集合 $\mathcal{J}$ の族 $\mathcal{F}$ について，**配信シンボル A からサーバ集合 $\mathcal{J}$ の情報を一切得られない**．すなわち

$$I(\mathcal{J};A)=0, \forall \mathcal{J}\in\mathcal{F}.$$



Correctness と Privacy を満たす $\phi_{\mathcal{J}}$ と δ が設計可能となる，$\mathcal{J}$ の族 $\mathcal{F}$ の十分条件を，主成果で導出

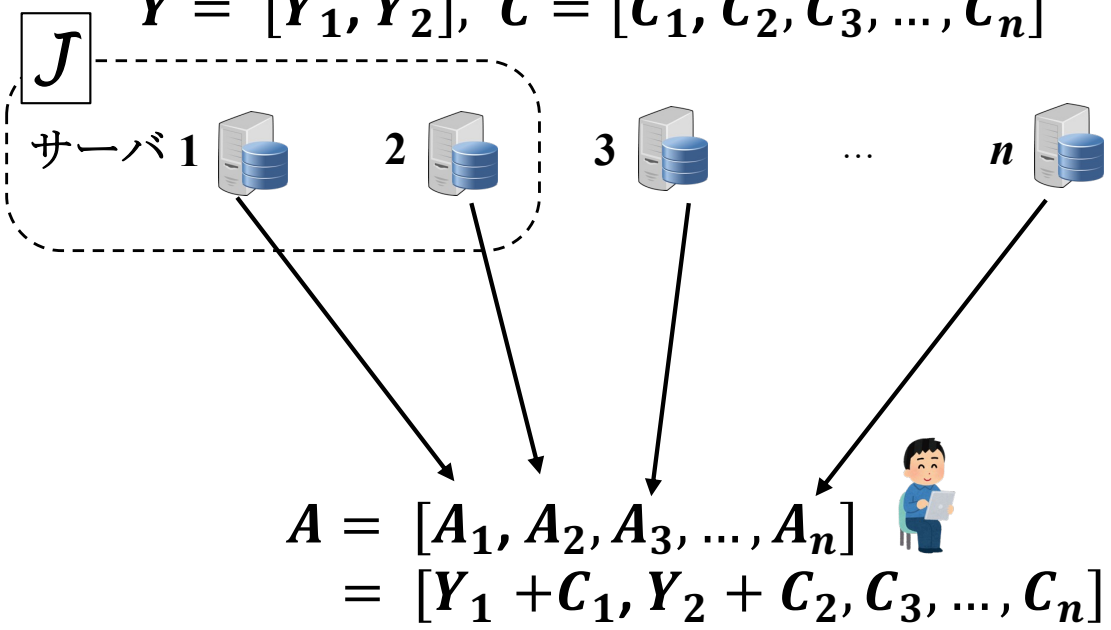
主成果: Correctness と Privacy を満たす条件

Correctness: コセット符号化の観点での Observation

- 符号化データ: $Y = \phi_{\mathcal{J}}(W) = [Y_j : j \in \mathcal{J}] \in \mathbb{F}^{|\mathcal{J}|}$.
- ランダムに選択した符号語: $C = [C_j : j = 1, \ldots, n] \in \mathscr{C} \subset \mathbb{F}^n$.
- 配信シンボル: $A = [A_j : j = 1, \ldots, n] \in \mathbb{F}^n$,

$$A_j = \begin{cases} Y_j + C_j & j \in \mathcal{J} \\ C_j & j \notin \mathcal{J} \end{cases}.$$

[例: $\mathcal{J} = \{1,2\}, l = 2$]



Y の $\mathcal{J}$ 成分以外を 0 にしたベクトル:
$\hat{Y} = [\hat{Y}_j : j = 1, \ldots, n] \in \mathbb{F}^n$, $\hat{Y}_j = Y_j$ if $j \in \mathcal{J}$, else 0

Observation

$A = \hat{Y} + C$. すなわち，A はコセット $\hat{Y} + \mathscr{C}$ よりランダムに選択されたとみなせる．
⇒ **コセット符号化** [OW84] の枠組み

⇓

コセット符号化の観点からの Correctness の解釈

復号関数 δ は，A からコセット $\hat{Y} + \mathscr{C}$ を一意に特定可能
⇔ シンドローム $AH^{\mathrm{T}} = (\hat{Y} + C)H^{\mathrm{T}} = \hat{Y}H \in \mathbb{F}^l$ が $W \in \mathbb{F}^l$ と同定可能
($H \in \mathbb{F}^{l \times n}$: $\mathscr{C}$ のパリティ検査行列, $CH^{\mathrm{T}} = 0$.)

すなわち，Correctness の成立条件は，$\hat{Y}H^{\mathrm{T}}$ と W が一意に対応する $\phi_{\mathcal{J}} : W \mapsto Y$ の設計可能条件に帰着

Correctness: 成立条件

$\hat{Y}H^{\mathrm{T}}$ と W が一意に対応する $\phi_{\mathcal{J}} : W \mapsto Y$ の設計可能条件の解析結果:

$d(\mathscr{C}^{\perp})$: 符号 $\mathscr{C}$ の双対符号 $\mathscr{C}^{\perp}$ の最小ハミング距離

主成果: 定理 1

符号化データ Y を配置しうるサーバ集合 $\mathcal{J} \subset \{1, \ldots, n\}$ の族 $\mathcal{F}$ について，

$$\mathcal{F} \subseteq \{\mathcal{J} \subset \{1, \ldots, n\} : |\mathcal{J}| \geq n - d(\mathscr{C}^{\perp}) + 1\}$$

のとき，$\mathcal{F}$ について Correctness を満たす手法が設計可能．

$$\Downarrow$$

符号化データ Y のシンボルを配置するサーバ数が $n - d(\mathscr{C}^{\perp}) + 1$ 以上[5] であれば，Correctness を満たす手法が設計可能．

[5] $d(\mathscr{C}^{\perp})$ の Singleton 限界 ($d(\mathscr{C}^{\perp}) \leq n - l + 1$) より，$n - d(\mathscr{C}^{\perp}) + 1 \geq l$

Privacy: A と $\mathcal{J}$ の相互情報量を解析

[Recall] $A = \hat{Y} + C$

$C \in \mathscr{C}$ は W および $\mathcal{J}$ と独立ランダムに選択

$\dim(\mathscr{C}) = n - l$，$W \in \mathbb{F}^l$ は一様分布

A と $\mathcal{J}$ の相互情報量を解析 (Proof Sketch):

$$I(A;\mathcal{J}) = I(A,\hat{Y};\mathcal{J}) - I(\hat{Y};\mathcal{J}|A)$$

$$\cdots$$

$$= H(A) - \underbrace{H(A,\hat{Y},W|\mathcal{J})}_{=H(C,\hat{Y},W|\mathcal{J})} + \underbrace{H(W|A,\hat{Y},\mathcal{J})}_{\substack{=0 \\ \text{by correctness}}} + \underbrace{H(\hat{Y}|A,\mathcal{J})}_{\substack{=H(\hat{Y}|A,W,\mathcal{J})=0 \\ \text{by correctness}}}$$

$$\cdots$$

$$= \underbrace{H(A)}_{\leq n} - \underbrace{H(C|\mathcal{J})}_{=H(C)=n-l} - \underbrace{H(W|\mathcal{J})}_{=l} \leq 0$$

相互情報量の非負性より $I(A;\mathcal{J}) = 0$

Privacy: 成立条件

前述の解析から，Correctness を満たす手法ならば，自動的に Privacy も満たされることが分かる．

主成果: 定理 2

符号化データ Y を配置しうるサーバ集合 $\mathcal{J} \subset \{1, \ldots, n\}$ の族 $\mathcal{F}$ について，手法が Correctness を満たすならば，$\mathcal{F}$ について Privacy も満たす．

まとめ

まとめ

まとめ:

- Private Information Delivery (PID) with Coded Storage を，位置匿名性の観点から再定式化
- 再定式化されたモデルにおいて，PID の性質に求められる Correctness と Privacy を定義
- 任意の線形符号 $\mathscr{C}$ をベースとした PID with Coded Storage の手法が Correctness と Privacy を満たすための十分条件を導出

今後の課題: 具体的な符号 $\mathscr{C}$ に基づく手法設計，実アプリケーション適用など

参考文献 I

[CGKS95] B. Chor, O. Goldreich, E. Kushilevitz, and M. Sudan, “Private information retrieval,” in *Proc. FOCS 1995*, Oct. 1995, pp. 41–50.

[MS77] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes*. North-Holland Mathematical Library, 1977.

[NKT26] K. Nakano, J. Kurihara, and T. Tanaka, “Extensive study on the security of private information delivery from coded storage,” to appear in *IEICE Trans. Fundamentals*, vol. E109-A, no. 3, Mar. 2026.

[OW84] L. H. Ozarow and A. D. Wyner, “Wire-tap channel II,” *AT&T Bell Lab. Tech. J.*, vol. 63, no. 10, pp. 2135–2157, Dec. 1984.

[Sun20] H. Sun, “Private information delivery,” *IEEE Trans. Inf. Theory*, vol. 66, no. 12, pp. 7672–7683, Dec. 2020.

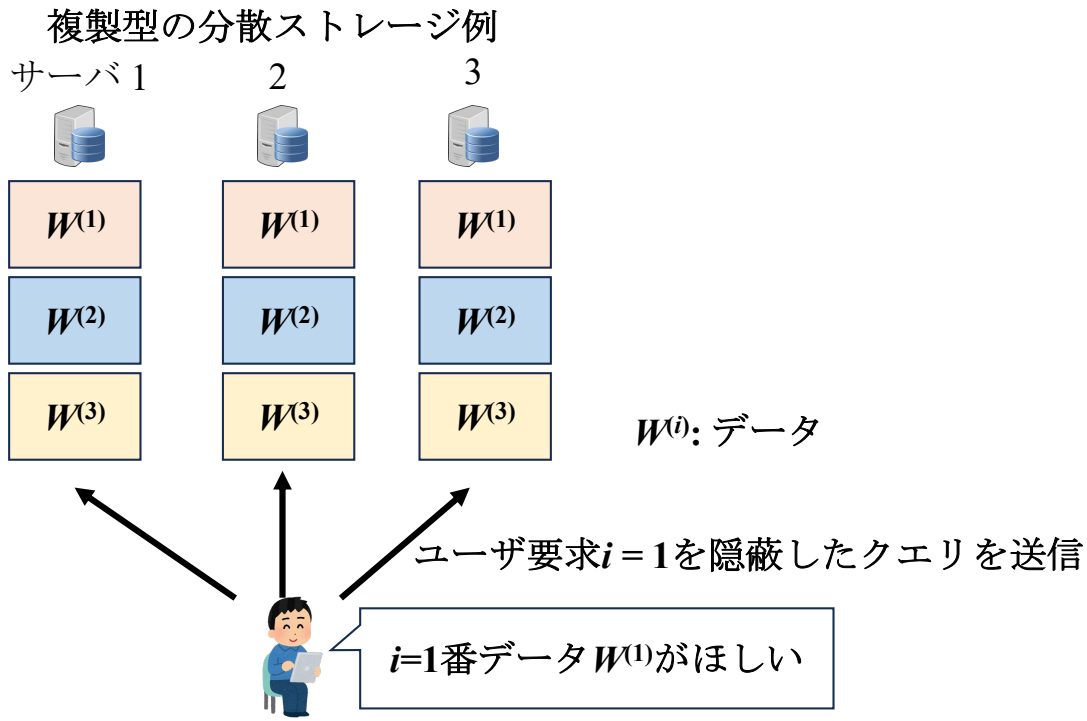
[VR22] K. Vaidya and B. S. Rajan, “Private information delivery with coded storage,” in *Proc. IEEE ISIT 2022*, Aug. 2022, pp. 2011–2015.

Appendix

情報理論・符号理論的な「プライバシ」保護手法

代表例: **Private Information Retrieval (PIR)** [CGKS95]

分散ストレージサーバからのプル型データ取得における，サーバに対してユーザのプライバシを保護する手法

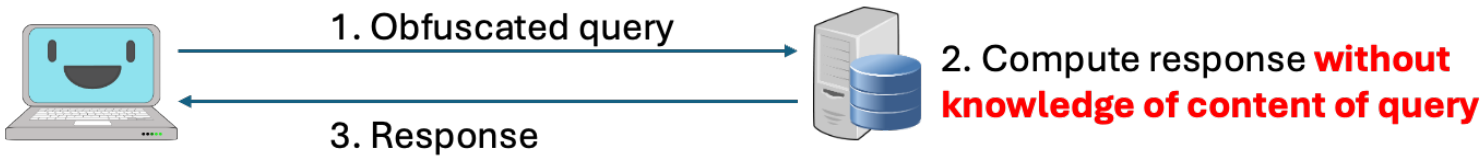


サーバ群はクエリに基づいて応答を生成．応答よりユーザは W_i を復元．

クエリからユーザ要求 (i.e, i) は特定不能 = プライバシ保護

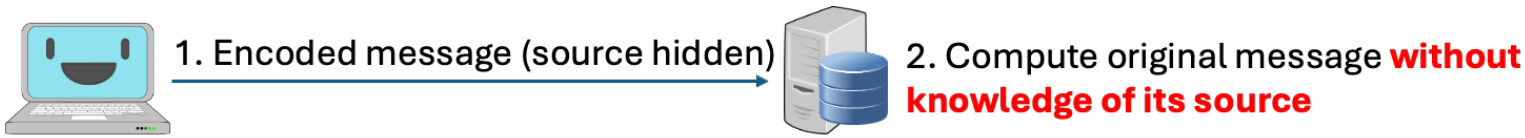
Private Information Retrieval (PIR) と Private Information Delivery (PID)

- **PIR**: 「**何を**」取得するのかを秘匿



Pull 型プロトコル．低遅延を実現可能な手法は未知．

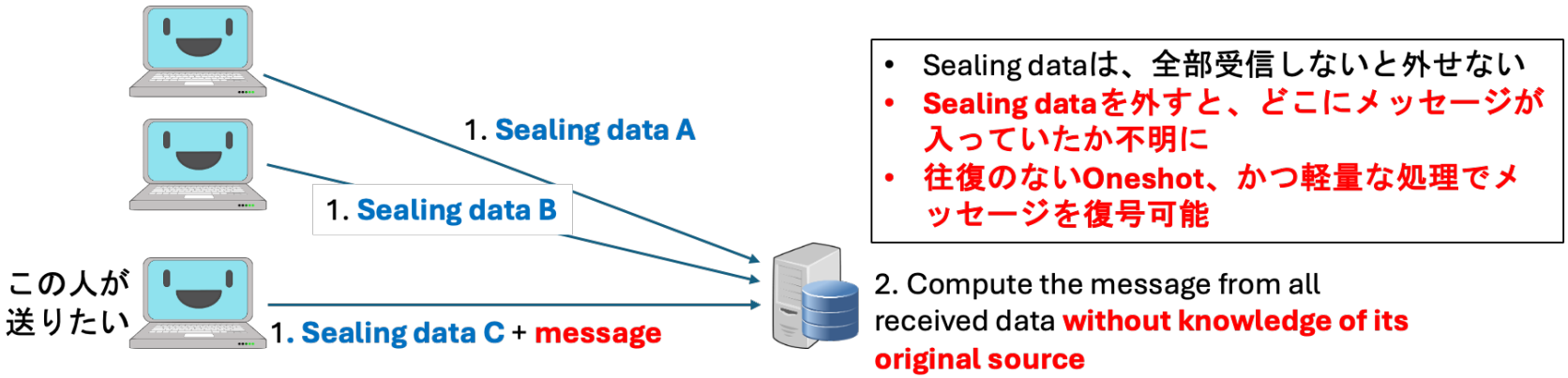
- **PID**: 「**誰が**」データを送るのかを秘匿



往復のない Push 型プロトコル．

既存の匿名化プロトコル (Tor 等) との大きな違い

既存手法の多くは「マルチホップ・マルチリレー型匿名化手法」だが，
PID は「**マルチソース型匿名化手法**」とみなせる．



受信者は，データを受け取れるものの，その本当の送信者を判別不能．

PID の特徴・特長:

- データ送信側では **PIR より小さい計算量**
- **匿名化のためのリレー不要**，直接送信