統計数理研究所オープンハウス　公開講演会「プライバシー保護とデータサイエンス」（2024/5/24）

差分プライバシーと統計解析への応用

村上 隆夫（統計数理研究所　学際統計数理研究系）

本講演の内容

- 差分プライバシー（Differential Privacy）
  - 個人のプライバシーを保護しながら統計解析を行う際の安全性指標のデファクト標準
  - Google、Apple、Microsoftなどが導入。2020年に米国国勢調査においても導入
  - 強固な安全性を保証すると言われているが、**安全性の概念が分かりづらい**とよく言われる。。。

- 本講演
  - 1) 差分プライバシーとはどのようなものか？
  - 2) どのような安全性を保証するのか？
  - 3) 差分プライバシーを満たす技術としてどのようなものがあるか？
  - といった差分プライバシーの概要について、なるべく分かりやすく説明

  - その後、具体的な統計解析への応用例をご紹介

本講演の内容

背景

（プライバシー問題、具体例、プライバシー保護研究の歴史）

差分プライバシー

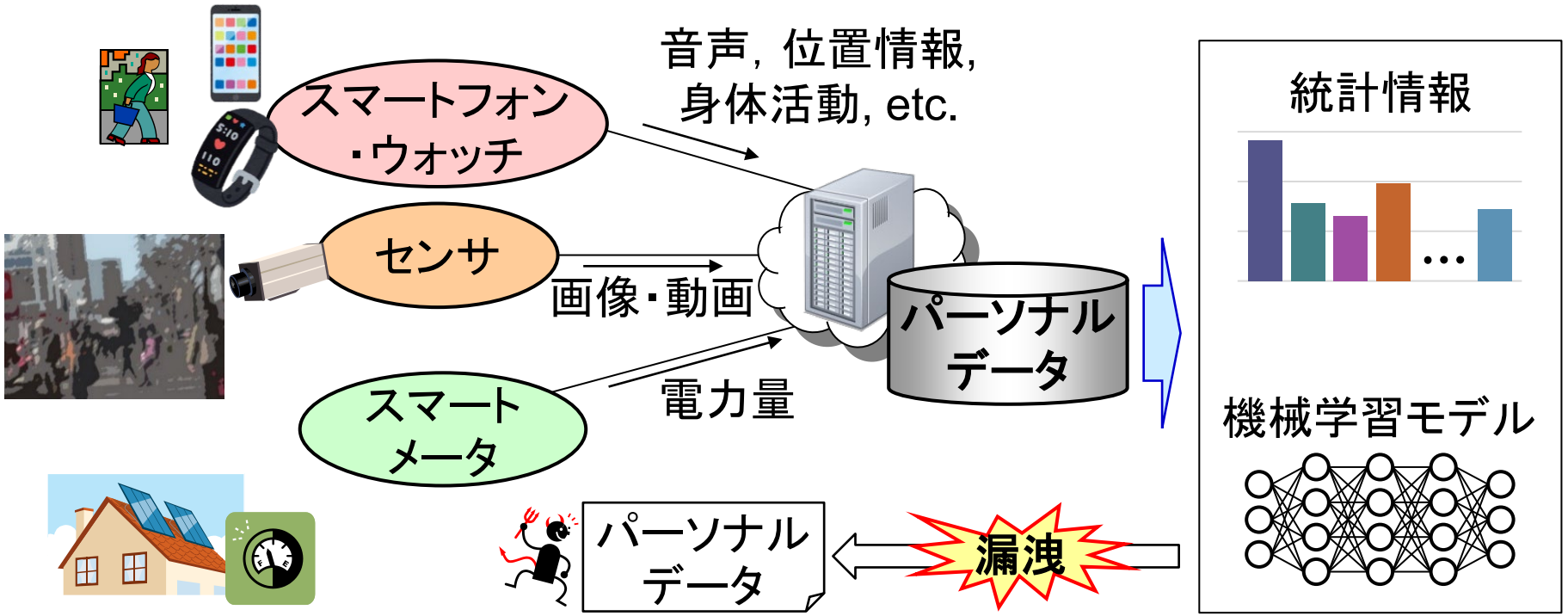
（差分プライバシーとは、どのような安全性を保証するか、差分プライバシーを満たす技術）

統計解析への応用例

（Yes/Noの分布推定、Googleの事例、グラフ統計解析）

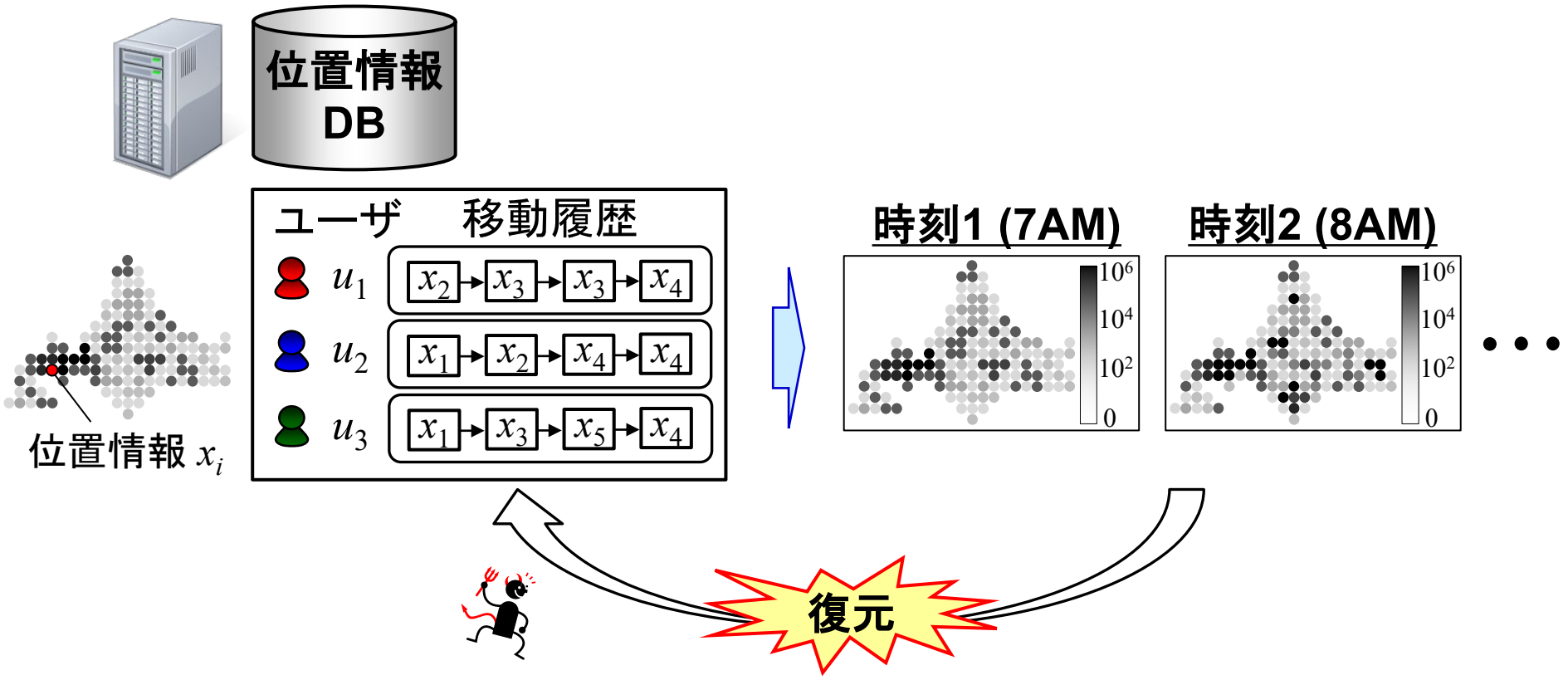
背景

- プライバシー問題
  - IoT、人工知能の普及に伴い、パーソナルデータの利活用への期待が高まっている
  - 一方、プライバシーの問題が懸念されている



背景

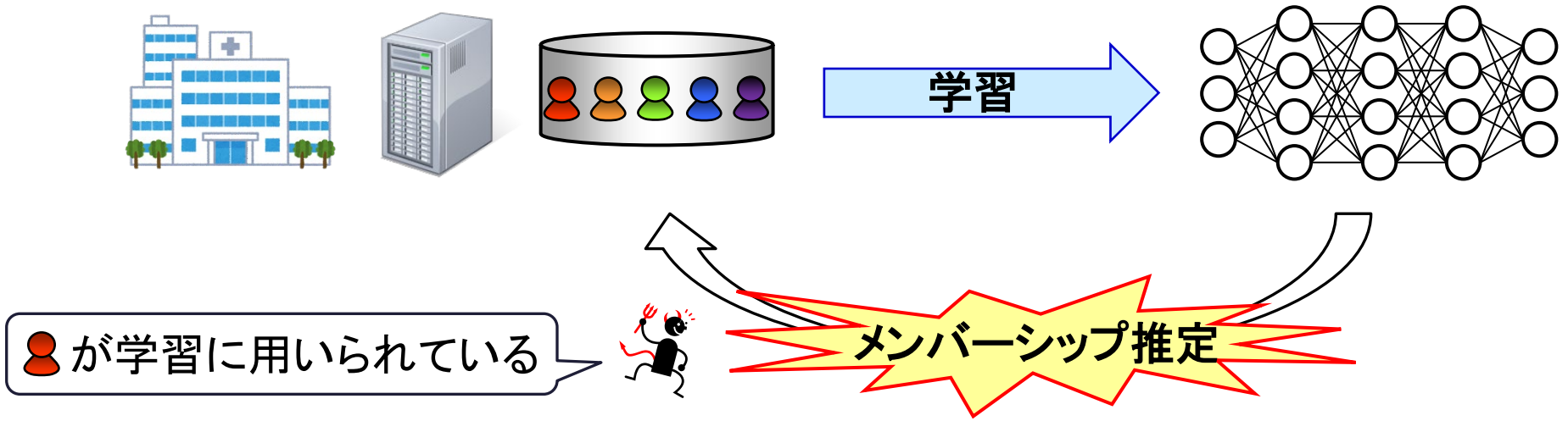
- 例1: 統計情報からのデータの漏洩
  - 時間帯ごとの人口分布から、各個人の位置情報が復元される恐れがある[1]



[1] A. Pyrgelis et al., “What Does The Crowd Say About You? Evaluating Aggregation-based Location Privacy,” PoPETs, 2017.

背景

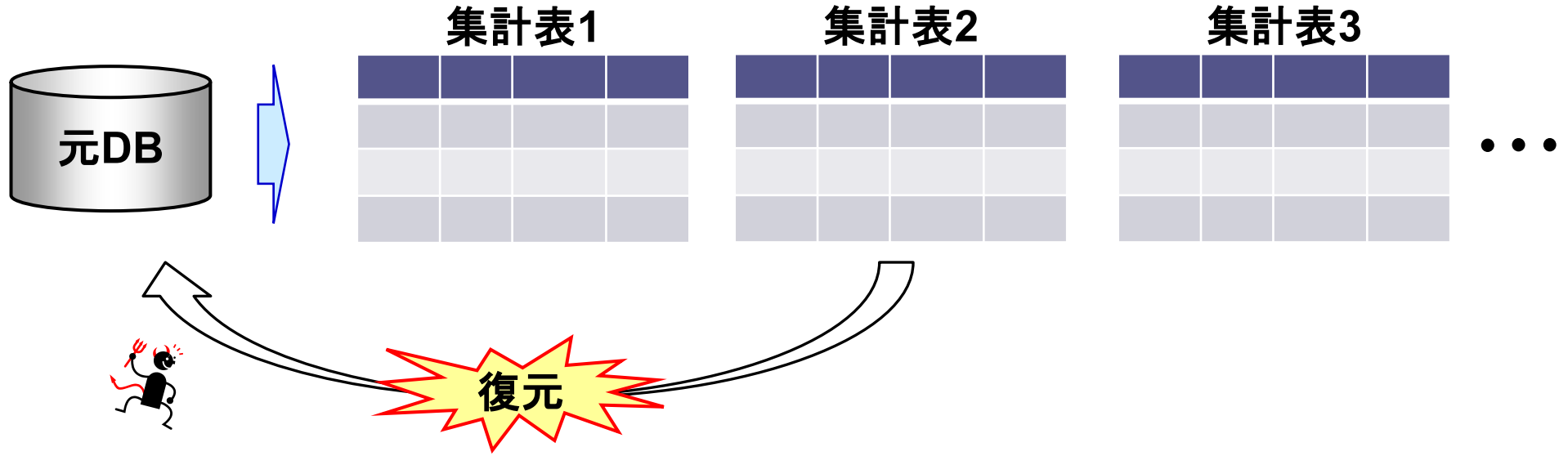
- 例2: 機械学習モデルからのデータの漏洩
  - 深層学習モデルから、どのユーザが学習に用いられたかが推定される恐れがある [2]



[2] R. Shokri et al., “Membership Inference Attacks Against Machine Learning Models,” S&P’17.

背景

- 例3：2010年米国国勢調査に対する再構築攻撃[3]
  - 2010年米国国勢調査の集計表から、元のデータベースを部分的に復元できることが示された

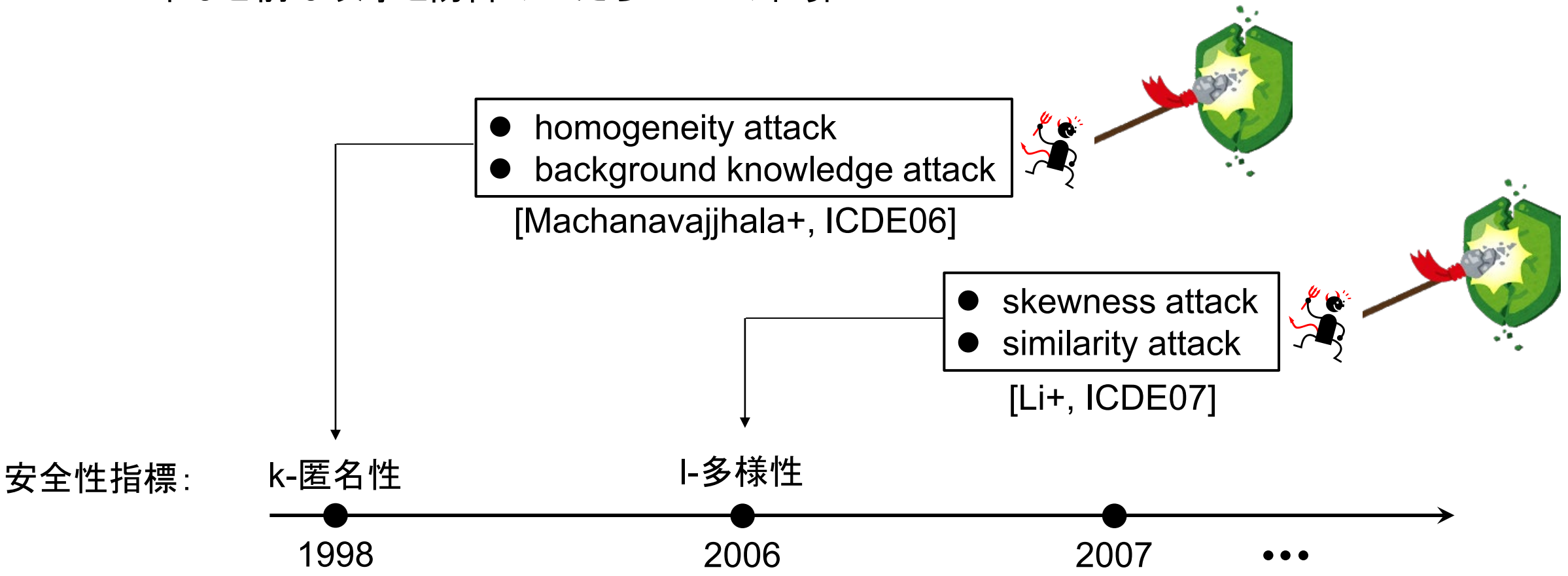


→ 2020年米国国勢調査では、差分プライバシーが導入[4]

[3] J. Abowd, Declaration of John Abowd, State of Alabama v. United States Department of Commerce. Case No. 3:21-CV-211-RAH-ECM-KCN, 2021.
[4] 寺田，“2020年米国国勢調査への差分プライバシーの導入に関する考察”，月刊統計（特集：プライバシー保護技術の新展開），2022.8

背景

- プライバシー保護研究の歴史
  - 20年ほど前は攻撃と防御のいたちごっこの世界



このような中、どのような攻撃に対しても安全性を理論的に保証する「差分プライバシー」が登場

本講演の内容

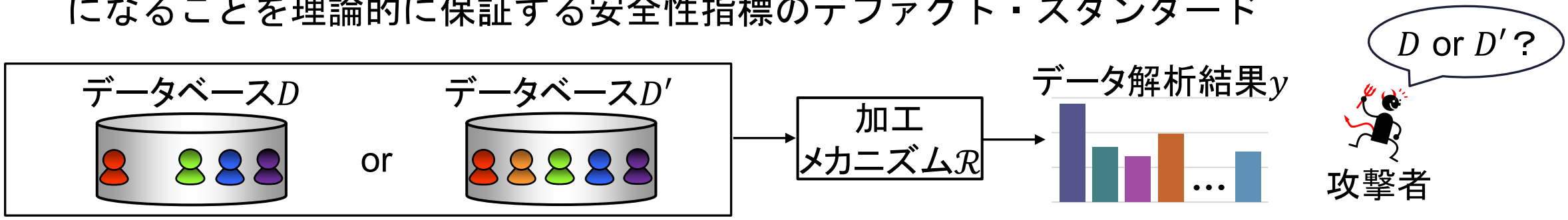
背景
（プライバシー問題、具体例、プライバシー保護研究の歴史）

差分プライバシー
（差分プライバシーとは、どのような安全性を保証するか、差分プライバシーを満たす技術）

統計解析への応用例
（Yes/Noの分布推定、Googleの事例、グラフ統計解析）

差分プライバシーとは？

- 差分プライバシー（DP: Differential Privacy）[5]
  - 「あるユーザのデータがDBにいてもいなくても、データ解析結果がほぼ同じ」になることを理論的に保証する安全性指標のデファクト・スタンダード



定義（ε-DP）

任意の隣接する（あるユーザが一方にのみ含まれる）データベースD, D'と任意の出力データyに対して

$$\Pr(y|D) \leq e^{\varepsilon}\Pr(y|D')$$

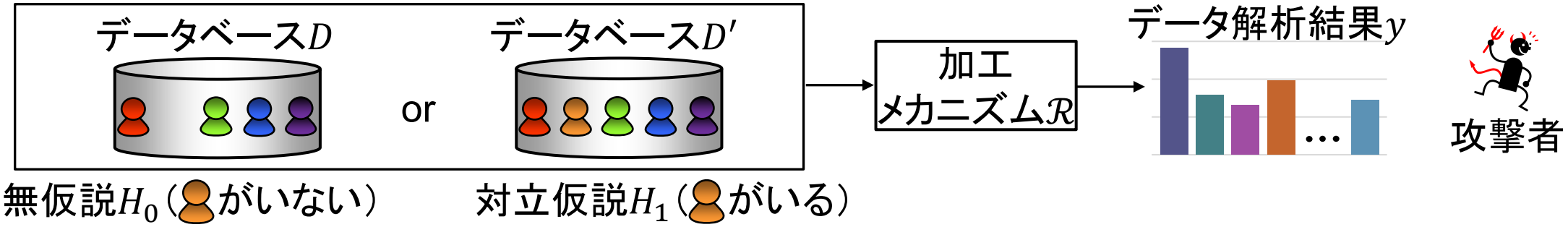
が成立するとき、加工メカニズム$\mathcal{R}$は**ε-DP**を満たすという（ε: privacy budget）

εが小さいとき（例：0～1）、$\Pr(y|D)$と$\Pr(y|D')$が近い値となる → 👤 がいるのか判別しづらい

[5] C. Dwork, A. Roth, “The Algorithmic Foundations of Differential Privacy,” Now Publishers Inc., 2014.

どのような安全性を保証するか？

- 仮説検定としての解釈[6]
  - 攻撃者が、yを基にH_0とH_1に関する仮説検定を行うことを考える
  - DPとtype I error（H_0を誤ってH_1と判定）とtype II error（H_1を誤ってH_0と判定）の関係は？



定理（[6] Theorem 2.4）

加工メカニズム$\mathcal{R}$が**ε-DP**を満たすならば、以下が成立する.

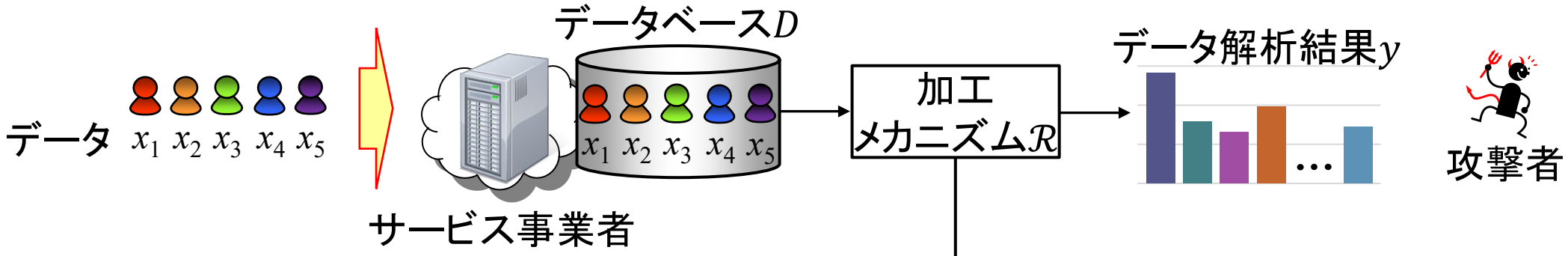
$$1 - (\text{type II error}) \leq e^{\varepsilon}(\text{type I error})$$

例：$\varepsilon = 0.1$ ⇒ **type II error $= 0.47$のとき、type I error ≥ 0.48**
（ほぼrandom guessの精度しか出ないので安全）

[6] L. Wasserman and S. Zhou, “A Statistical Framework for Differential Privacy,” Journal of the American Statistical Association, 2010.

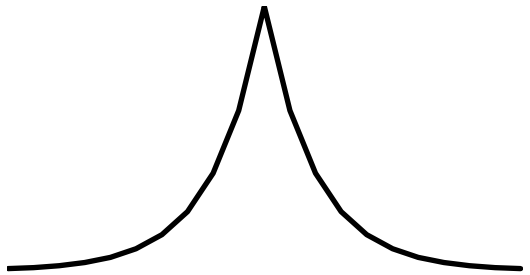
差分プライバシーを満たすメカニズム

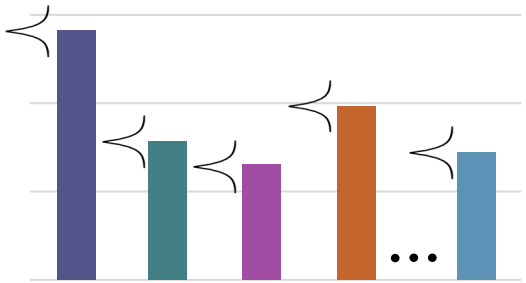
- 元々の差分プライバシーが仮定するモデル ＝ 中央集権型モデル（Central Model）
  - サービス事業者がユーザからデータを集める
  - データ解析結果y（例：ヒストグラム）にランダムなノイズを付与し、第三者提供 or 公開



例：ラプラスメカニズム

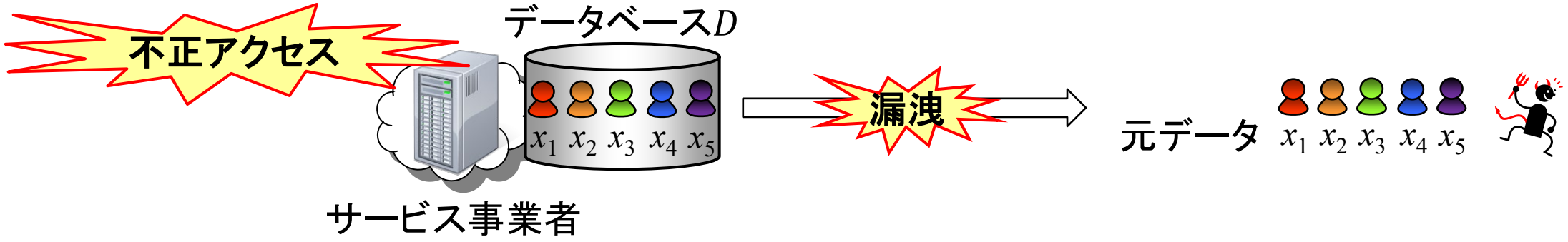
ヒストグラムの各binにラプラス分布に従ったランダムなノイズを付与する → **ε-DP**



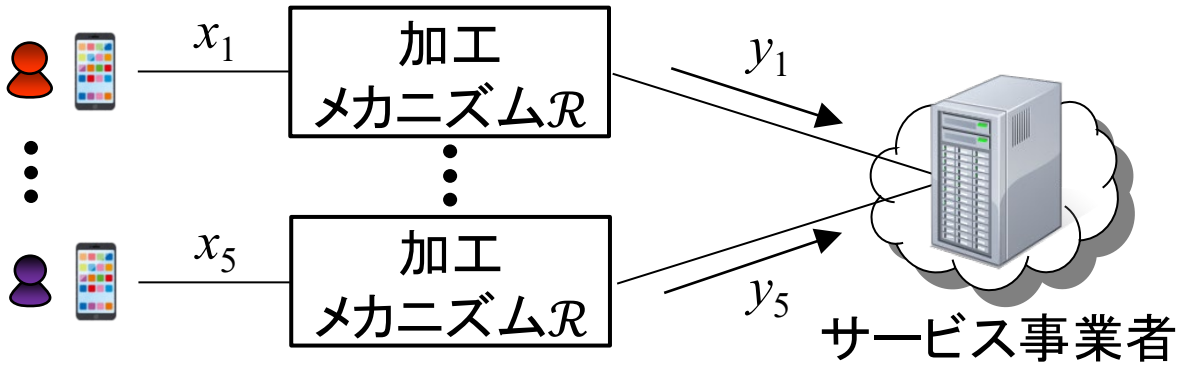


差分プライバシーを満たすメカニズム

- 中央集権型モデルの課題
  - 不正アクセスなどにより、全ユーザの元データが漏洩する恐れあり



- 局所型モデル（Local Model）
  - ユーザがサービス事業者を信用せず、自身のデータにノイズを加える
  - 出力データy_iは差分プライバシーを満たし、元データx_iに関する情報がほとんど漏洩しない



差分プライバシーを満たすメカニズム

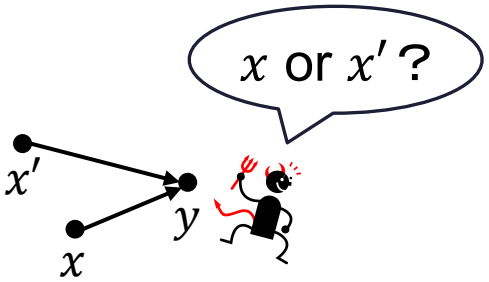
- 局所型差分プライバシー（LDP: Local Differential Privacy）

定義（ε-LDP）

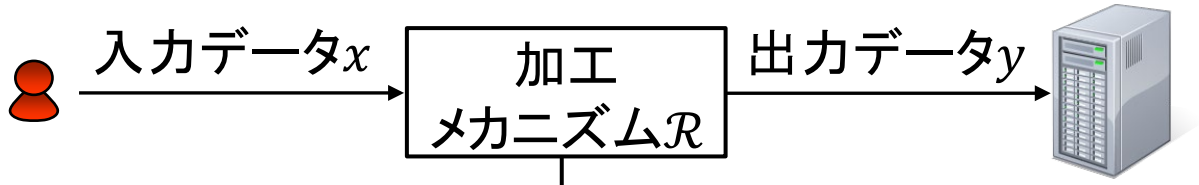
任意の元データ$x, x' \in \mathcal{X}$（$\mathcal{X}$：定義域）と出力データ$y \in \mathcal{Y}$（$\mathcal{Y}$：値域）に対して

$$\Pr(y|x) \leq e^{\varepsilon}\Pr(y|x')$$

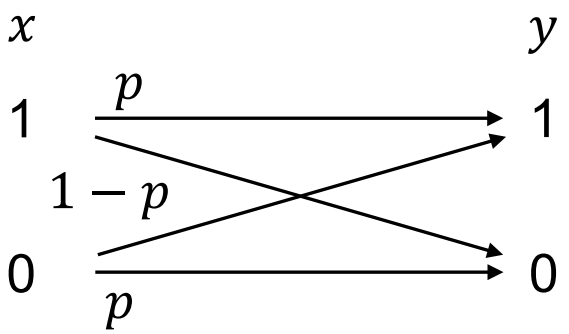
が成立するとき、加工メカニズム$\mathcal{R}$は**ε-LDP**を満たすという



εが小さいとき（例：0～1）、出力データyを見ても入力データxの値が分からない



例：RR (Randomized Response)



$$\left(p = \frac{e^{\varepsilon}}{e^{\varepsilon}+1}\right)$$

本講演の内容

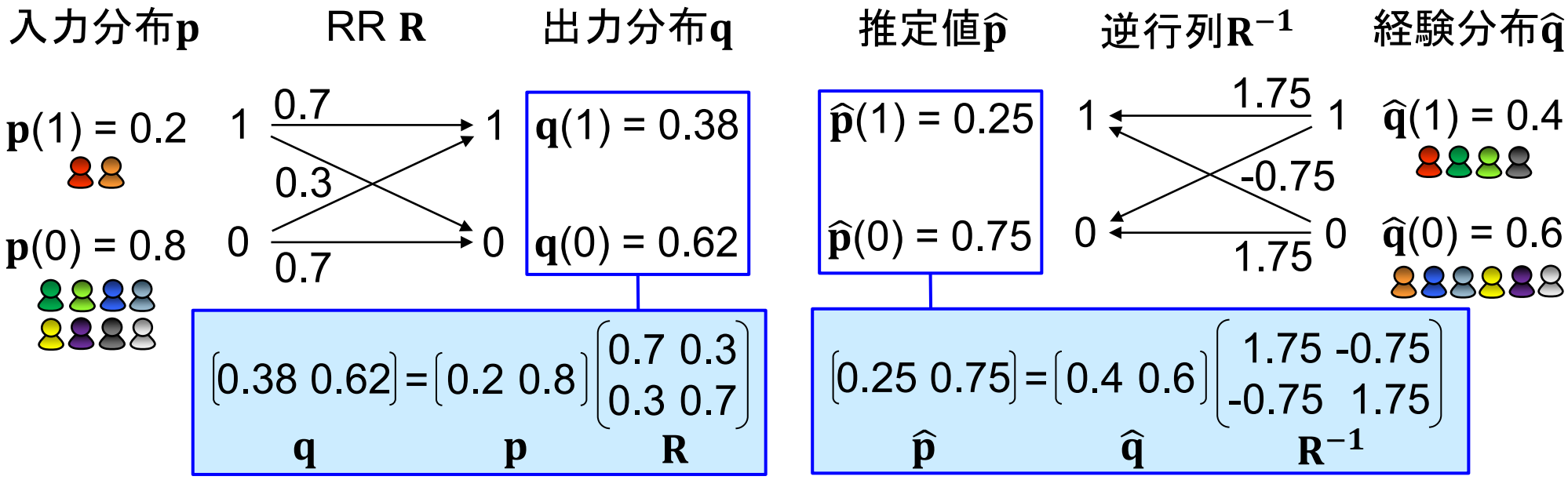
背景
（プライバシー問題、具体例、プライバシー保護研究の歴史）

差分プライバシー
（差分プライバシーとは、どのような安全性を保証するか、差分プライバシーを満たす技術）

統計解析への応用例
（Yes/Noの分布推定、Googleの事例、グラフ統計解析）

LDPを満たす分布推定

- 例1：Yes/Noの分布推定
  - 「今までにテストでカンニングをしたことがありますか？」（1: yes, 0: no） [7]
  - 加工メカニズム ⇒ RR（Randomized Response） を使用
  - 入力分布$\mathbf{p}$の推定法 ⇒ 逆行列法[8]を使用

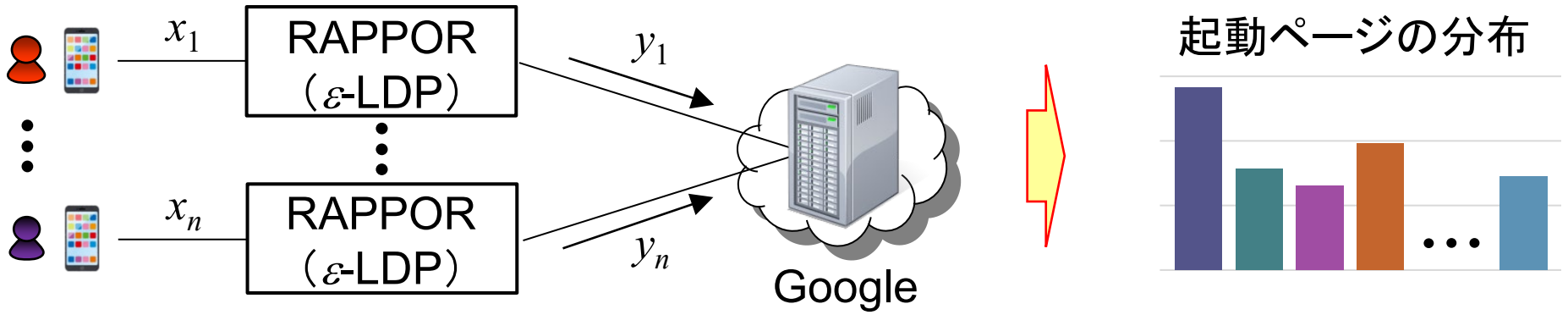


RRはε-LDPを満たす（$\varepsilon = 0.85$） ⇒ プライバシーを保護したまま、入力分布$\mathbf{p}$が推定できる

[7] B. Bullek et al., “Towards Understanding Differential Privacy: When Do People Trust Randomized Response Technique?,” CHI’17.
[8] R. Agrawal et al., “Privacy Preserving OLAP,” SIGMOD’05.

LDPを満たす分布推定

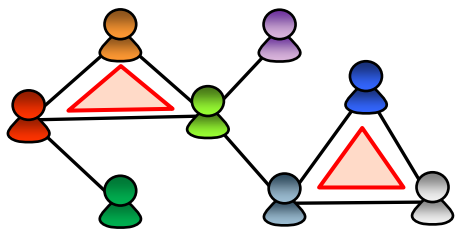
- 例2：Googleの事例[9]
  - Google独自のLDPメカニズムRAPPORを用いて、Chromeの起動ページの分布を推定

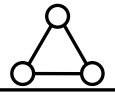
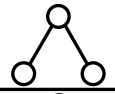
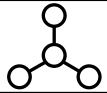


[9] U. Erlingsson et al., “RAPPOR: Randomized Aggregatable Privacy-Preserving Ordinal Response,” CCS’14.

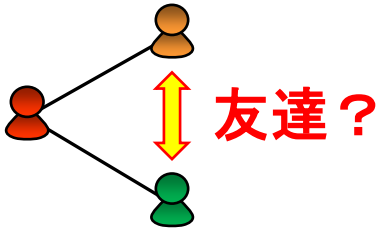
LDPを満たすグラフ統計解析[10]

- グラフ統計（部分グラフ数）
  - #triangles = 三角形の数、#k−stars = k個のnodeが繋がっている星の数



Shape	Name	Count
	Triangle	2
	2-star	15
	3-star	6

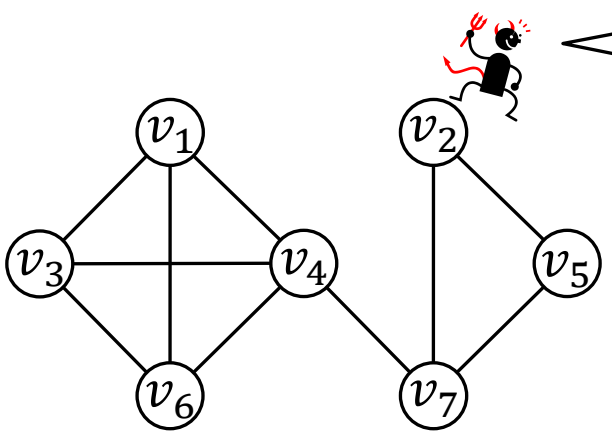
- 何に使えるか？
  - 「友達の友達が友達である確率」= $3 \times \frac{\#\text{triangles}}{\#2-\text{stars}}$ → 友達推薦の有効性が分かる



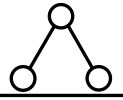
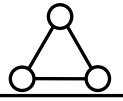
[10] J. Imola, T. Murakami, K. Chaudhuri, “Locally Differentially Private Analysis of Graph Statistics,” USENIX Security’21.

LDPを満たすグラフ統計解析[10]

- プライバシー問題
  - #triangles/#k−starsから、秘密にしたい友達情報（edge）が漏洩する可能性がある
  - 例：v_2が攻撃者の場合



$v_3, \ldots, v_7$同士の友達情報は全部知っている。
v_1は誰と友達なのだろう？

Shape	Name	Count
	2-star	20
	Triangle	5

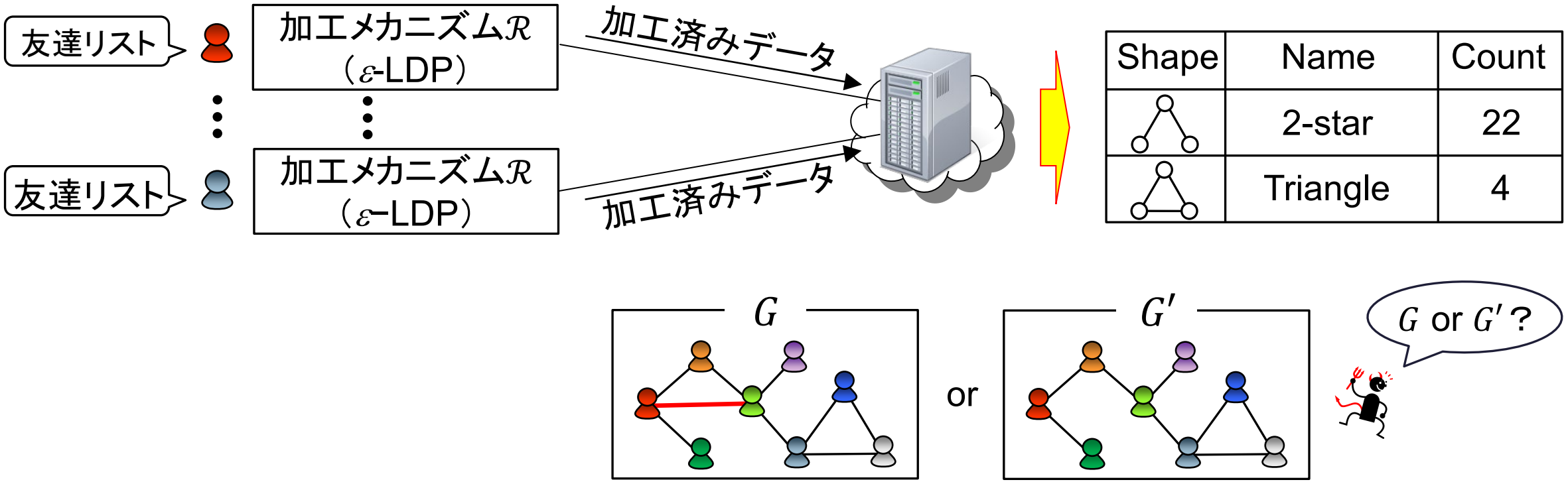
漏洩 v_1の友達はv_3, v_4, v_6

プライバシーを強固に保護するには、部分グラフ数（#triangles/#k−stars）にノイズを加える必要があり

[10] J. Imola, T. Murakami, K. Chaudhuri, “Locally Differentially Private Analysis of Graph Statistics,” USENIX Security’21.

LDPを満たすグラフ統計解析[10]

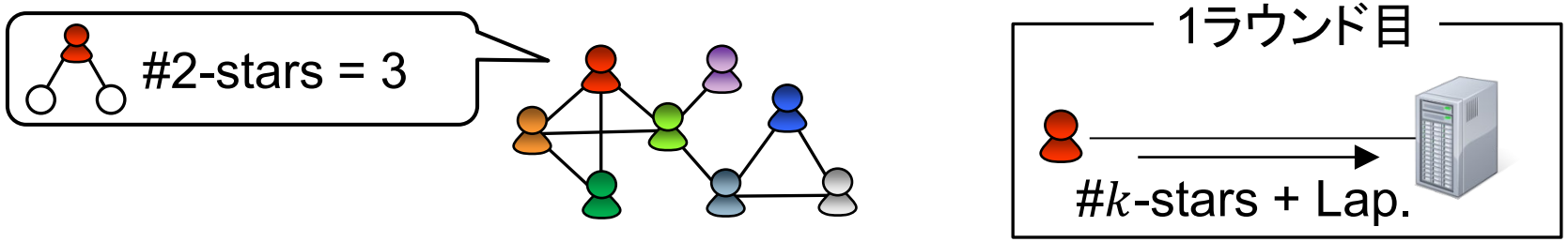
- LDP
  - 各ユーザが友達リスト（誰と繋がっているかの情報）に対してLDPを満たすノイズを加える
  - 攻撃者は、隣接グラフ（あるedgeが一方にのみ含まれる2つのグラフ）G, G'のどちらから来たのか区別できない → 友達関係を秘匿



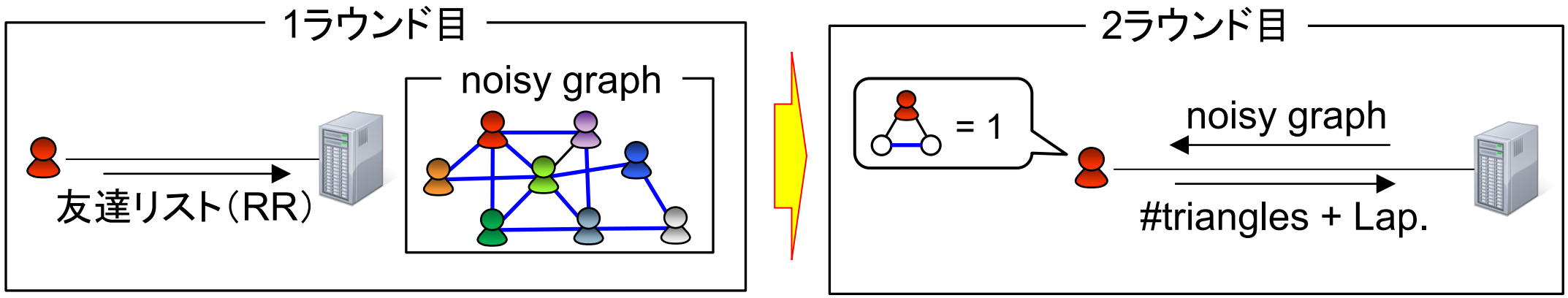
[10] J. Imola, T. Murakami, K. Chaudhuri, “Locally Differentially Private Analysis of Graph Statistics,” USENIX Security’21.

LDPを満たすグラフ統計解析[10]

- LDPを満たすk-starsの数え上げ[10]
  - 各ユーザは自身の#k-starsを知っている → 1ラウンドで高精度に推定できる（Lap: ラプラスノイズ）



- LDPを満たすtrianglesの数え上げ[10]
  - 各ユーザは友達同士が繋がっているか分からないため、自身の#trianglesを知らない
  - → 2ラウンド導入すれば、1辺だけnoisyなtriangleが数えられるようになり、高精度に推定できる

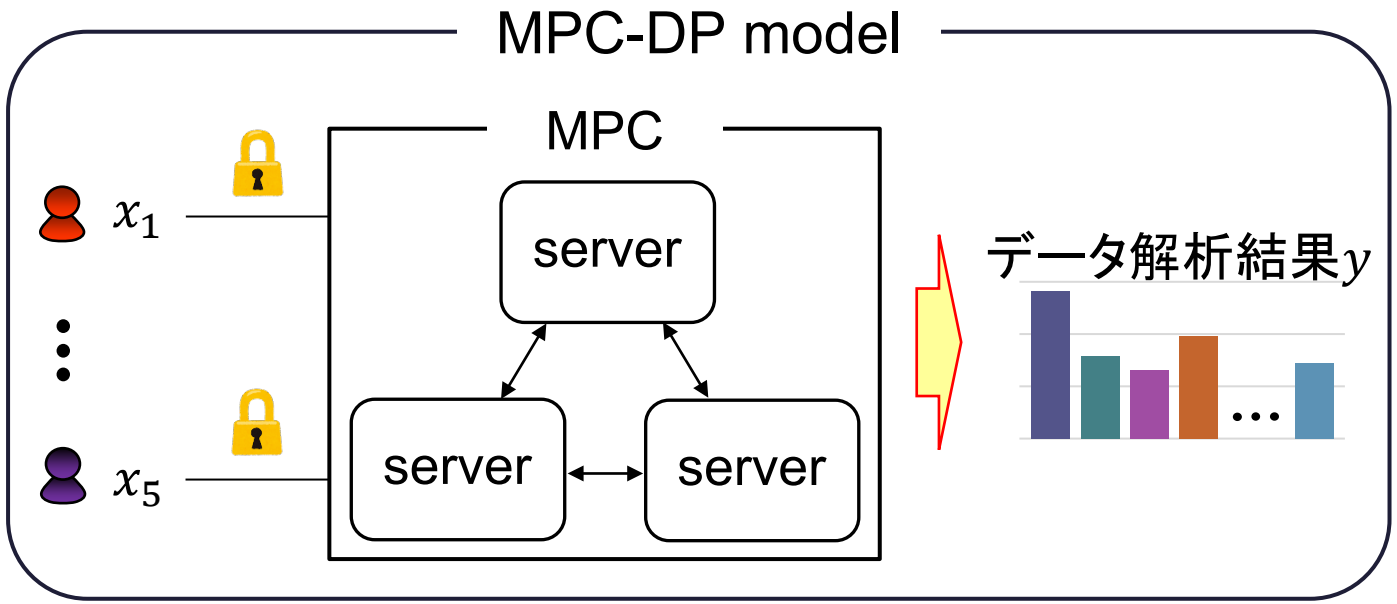


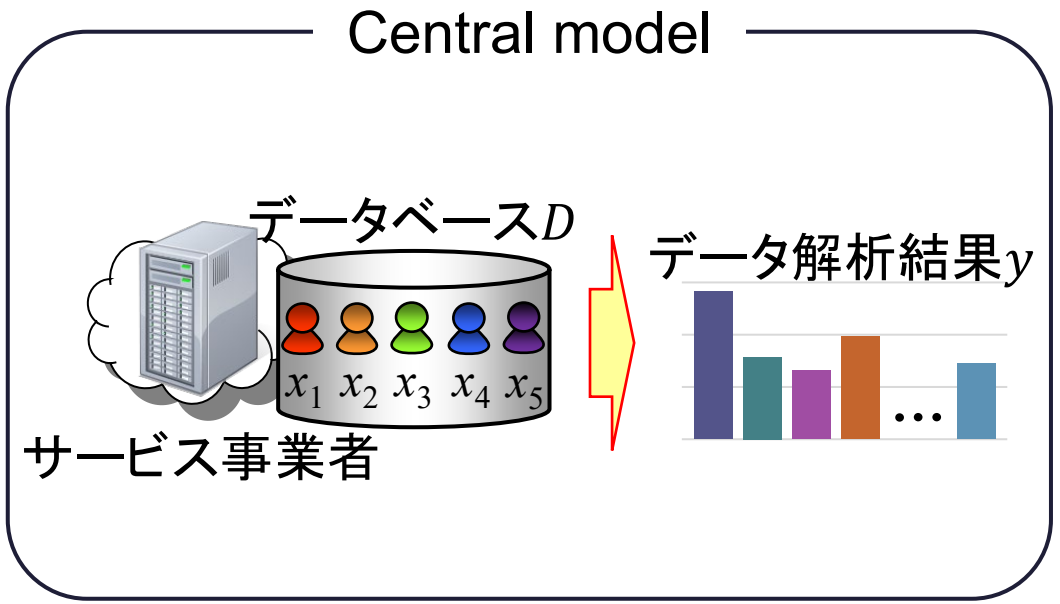
[10] J. Imola, T. Murakami, K. Chaudhuri, “Locally Differentially Private Analysis of Graph Statistics,” USENIX Security’21.

今後の展望

- LDPの課題
  - 各ユーザがノイズを加える分、ノイズ量が大きい → ユースケースによっては精度が全然でない
  - 例：ユーザ数が少ないと推定誤差が大きい

- 今後の展望：MPC-DP
  - 秘密計算（MPC: Multi-Party Computation）を用いて、データが秘匿された状態でノイズを付与
  - Local modelのように自身のデータは誰にも見られず、Central modelと同じ精度を達成できる





ご清聴ありがとうございました