実践セキュリティ特論 – プライバシー –
プライバシ強化技術（PETs）
講義スライド第３回、第４回

KDDI総合研究所　三本知明

講義の内容

プライバシに関する法規制について

・国内外のプライバシに関する法律の概要

・パーソナルデータに潜むリスク

プライバシ強化技術（１）：各プライバシ強化技術の基本

・匿名化

・差分プライバシ

・局所差分プライバシ

プライバシ強化技術（２）：各プライバシ強化技術の基本・応用例

・匿名化アルゴリズム

・サンプリング

・差分プライバシメカニズム（一部再掲）

・その他のプライバシ強化技術

・差分プライバシ・局所差分プライバシの応用例

演習：各チームによる匿名化・攻撃の説明

・各チームによる発表

・総評

匿名化アルゴリズム

データ提供におけるリスク評価指標：k-匿名性（再掲）

パーソナルデータを提供する際、直接識別情報を削除するだけでは特定や連結は防げない

・加工したパーソナルデータがどの程度プライバシ侵害のリスクがあるかを定量評価する必要がある

代表的なプライバシリスク評価指標：k-匿名性

・n人の個人からd個の属性を収集することを想定する

・個人 i から収集したデータを$\boldsymbol{x}_i = (x_{i1}, x_{i2}, \cdots, x_{id})$とし、全員分のデータ集合を$D = \{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n\}$とする

・ d_{QI}個の属性$X_1, \cdots, X_{d_{QI}}$を間接識別情報とする

　−k-匿名性における攻撃者モデルでは、攻撃者は間接識別情報を背景知識として持つと想定する

・個人 i のデータを間接識別情報の組み合わせとその他の情報の組み合わせとして$\boldsymbol{x}_i = \left(\boldsymbol{x}_i^{QI}, \boldsymbol{x}_i^{other}\right)$とする

・このときk-匿名性は以下のように定義される

定義：k-匿名性

Dを n 人の個人から集めたレコードの集合とする。またDに含まれる間接識別情報の値の組み合わせをAとする。

すべての$x^{QI} \in A$について、x^{QI}を含むレコードがDに含まれない、あるいは少なくともk個存在するとき、Dはk-匿名性を有する。

データ提供におけるリスク評価指標：k-匿名性（再掲）

k-匿名性を持つデータの具体例

直接識別情報　　間接識別情報

名前	年齢	性別	住所	年収	既往歴
青木	33	M	東京	1,100	うつ病
井上	34	M	埼玉	900	うつ病
	⋮	…	…	⋮	⋮
山田	47	M	京都	200	風邪
湯川	49	F	奈良	300	糖尿病
吉田	88	M	沖縄	10,000	肺がん

2-匿名化

10歳区切りで丸める（一般化）

地域単位で丸める（一般化）

年齢	性別	住所	年収	既往歴
30	M	関東	1,000	うつ病
30	M	関東	1,000	うつ病
⋮	…	…	⋮	⋮
50	-	近畿	250	風邪
50	-	近畿	250	糖尿病
88	-	沖縄	10,000	肺がん

外れ値の削除（レコード抑制）

属性値の削除（抑制）

グループ毎に平均化（マイクロアグリゲーション）

データ匿名化手法：匿名化

***k*-匿名化アルゴリズム**

・機械的にk-匿名化を行うアルゴリズムが多数提案されている

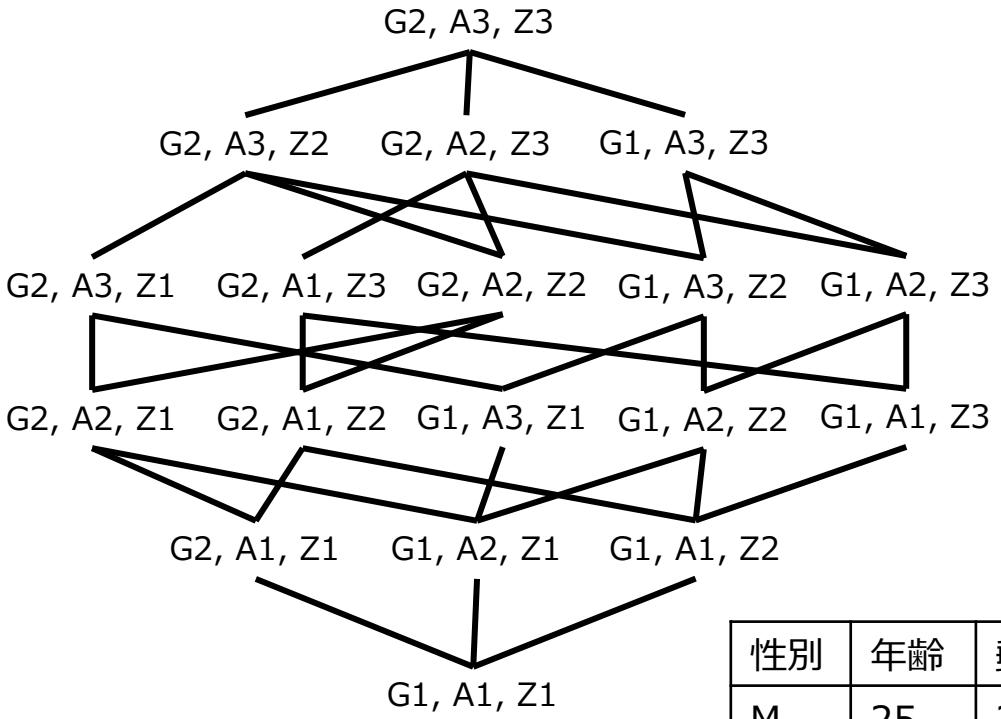
Incognito

・大局的再符号化の組み合わせからk-匿名化を行う

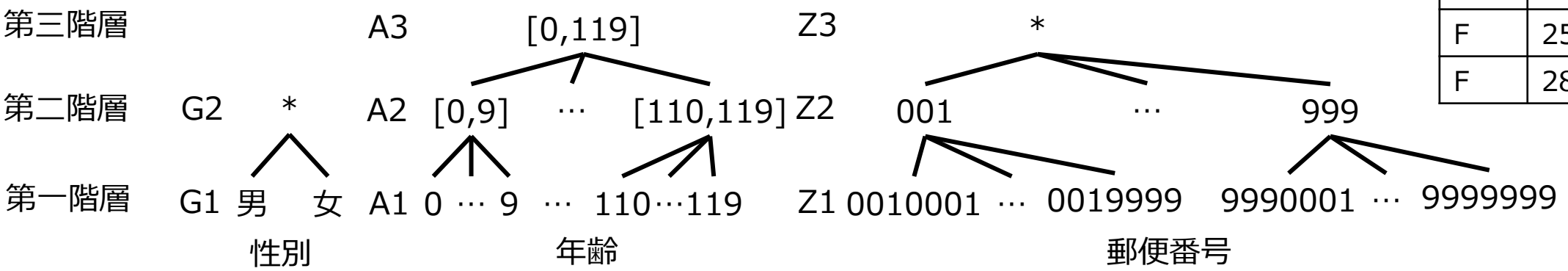
・特定の属性に着目したときにk個のデータが存在しない場合、その属性を加工しないとk-匿名性は満たさないという特徴（単調性）を利用する

ー2-匿名性を満たすことを考える

ー年齢だけ見ると2-匿名性を満たさない→A1は削除



性別	年齢	郵便番号
M	25	354-0025
M	29	354-0025
M	38	354-0038
M	31	354-0019
F	25	354-0045
F	28	354-0031



データ匿名化手法：匿名化

***k*-匿名化アルゴリズム**

・機械的にk-匿名化を行うアルゴリズムが多数提案されている

Incognito

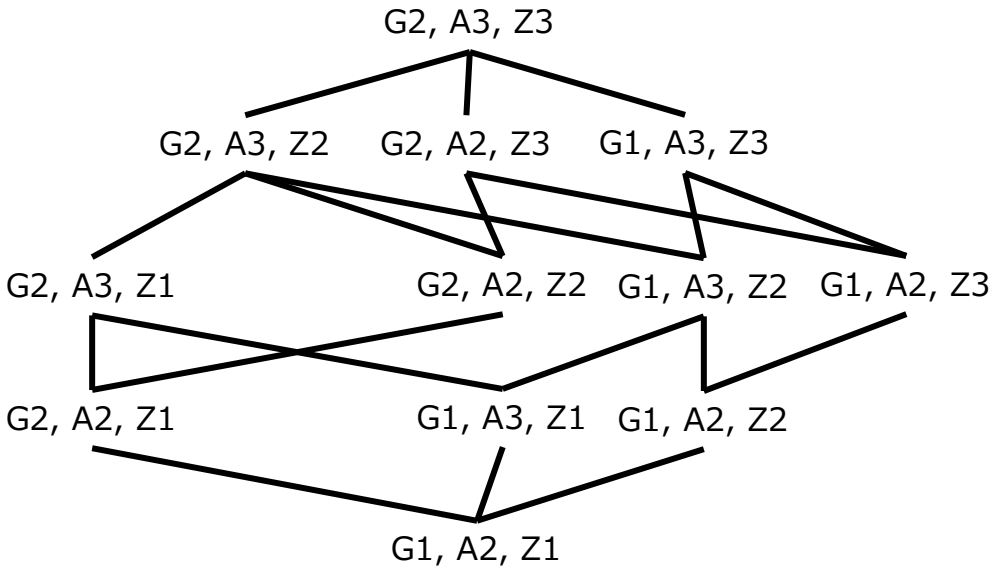
・大局的再符号化の組み合わせからk-匿名化を行う

・特定の属性に着目したときにk個のデータが存在しない場合、その属性を加工しないとk-匿名性は満たさないという特徴（単調性）を利用する

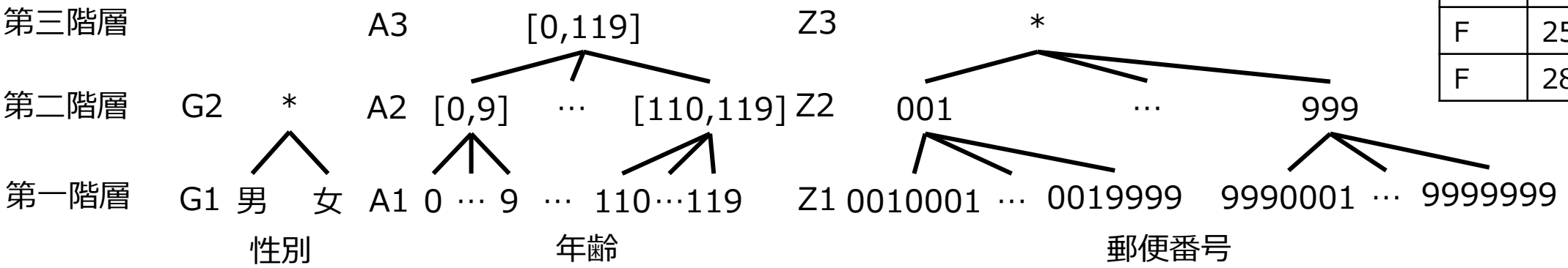
－2-匿名性を満たすことを考える

－年齢だけ見ると2-匿名性を満たさない→A1は削除

－郵便番号だけ見ると2-匿名性を満たさない→Z1は削除



性別	年齢	郵便番号
M	25	354-0025
M	29	354-0025
M	38	354-0038
M	31	354-0019
F	25	354-0045
F	28	354-0031



データ匿名化手法：匿名化

***k*-匿名化アルゴリズム**

・機械的に*k*-匿名化を行うアルゴリズムが多数提案されている

Incognito

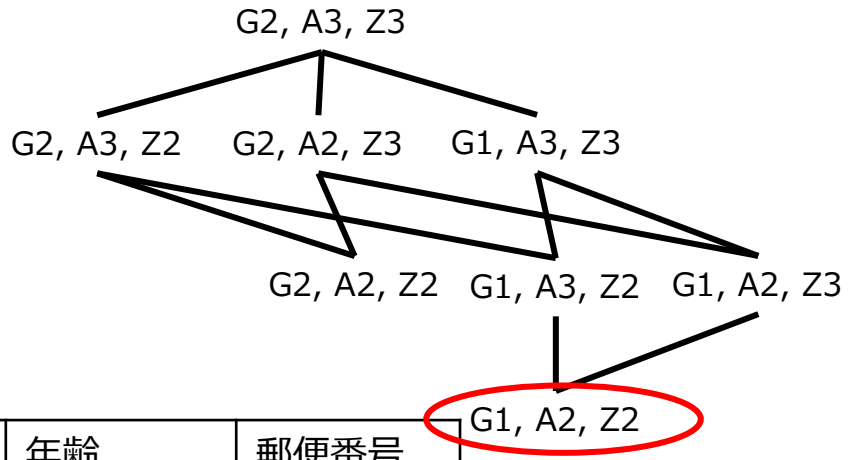
・大局的再符号化の組み合わせから*k*-匿名化を行う

・特定の属性に着目したときに*k*個のデータが存在しない場合、その属性を加工しないと*k*-匿名性は満たさないという特徴（単調性）を利用する

－2-匿名性を満たすことを考える

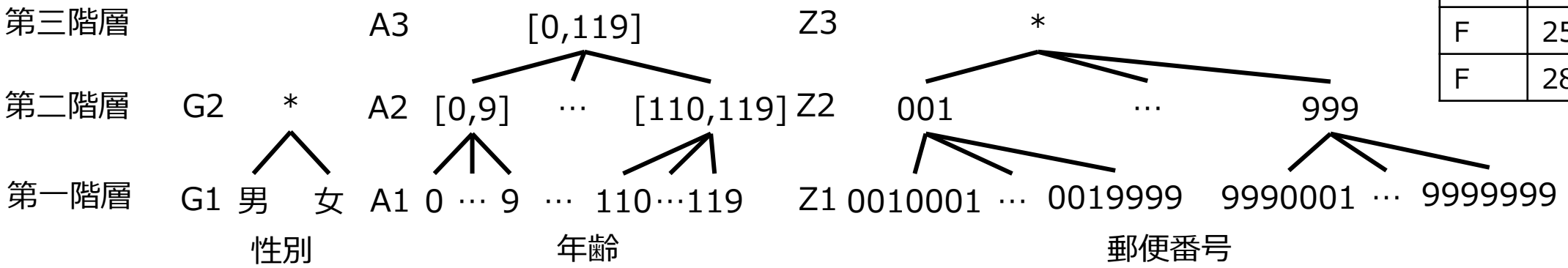
－年齢だけ見ると2-匿名性を満たさない→A1は削除

－郵便番号だけ見ると2-匿名性を満たさない→Z1は削除



性別	年齢	郵便番号
M	[20-29]	354
M	[20-29]	354
M	[30-39]	354
M	[30-39]	354
F	[20-29]	354
F	[20-29]	354

性別	年齢	郵便番号
M	25	354-0025
M	29	354-0025
M	38	354-0038
M	31	354-0019
F	25	354-0045
F	28	354-0031



データ匿名化手法：匿名化

k-匿名化アルゴリズム

・機械的にk-匿名化を行うアルゴリズムが多数提案されている

Mondrian

・マイクロアグリゲーションを再帰的に利用してk-匿名化を行う

ー入力データセットを一つのグループとして、グループを二分割してマイクロアグリゲーションを実行

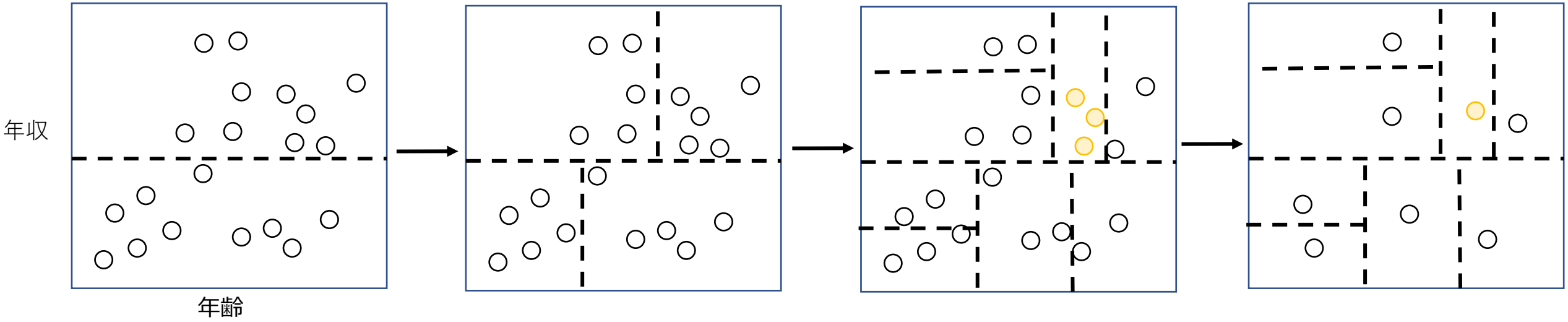
ー上記をグループ内のレコードが$2k$未満になるまで再帰的に繰り返す

ー以下は2-匿名化の例（グループ内のレコードが4個未満になるまで二分割を繰り返す）

年齢	年収
…	…
…	…
45	500
47	520
52	480
…	…
…	…

→

年齢	年収
…	…
…	…
48	500
48	500
48	500
…	…
…	…



k-匿名化アルゴリズムまとめ

k-匿名化アルゴリズム

・自動的にk-匿名化データを構築するアルゴリズムが提案されている

・実際のデータセットとユースケースを考えたうえで、細かく匿名化手法を考える必要がある

サンプリング

データ提供におけるリスク評価指標：標本一意性、母集団一意性

全数調査と標本調査

・母集団を対象とした調査を全数調査、標本を対象とした調査を標本調査という

　ー標本：母集団からランダムに抽出したレコードの集合

・調査対象のデータベースが母集団か標本かによってプライバシリスクの考え方は異なる

　ー標本において一意のレコードであっても、母集団において一意であるとは限らない

　ーk-匿名性の議論では標本一意であれば特定リスクが存在すると考えていた

・母集団一意性と標本一意性は個票開示問題（官庁統計における個人属性データの公開問題）において古くから議論されてきた

データ提供におけるリスク評価指標：標本一意性、母集団一意性

母集団一意となるレコード数の推定

・間接識別情報の組み合わせがK個あるとする

ー例えば年齢（0-119歳）、性別（男女）、住所（47都道府県）のとき、$K = 120 \times 2 \times 47$

性別	年齢	郵便番号
M	[20-29]	354
M	[20-29]	354
M	[30-39]	354
M	[30-39]	354
F	[20-29]	354
F	[20-29]	354

・これらの組み合わせをそれぞれ**セル**とよび、その間接識別情報の組み合わせを持つレコード数を**度数**とよぶ

ーi番目のセルの度数をF_iとし、度数がjとなるセルの数をS_jとする

・ポアソン分布を用いると、i番目のセルの度数がF_iとなる確率は

$$\Pr(F_i) = \frac{(N_0\pi_i)^{F_i} \cdot e^{-N_0\pi_i}}{F_i!}$$

ーポアソン分布：ある期間に平均λ回発生する事象がk回起こる確率を表す確率分布で、$\Pr(k) = \frac{\lambda^k e^{-\lambda}}{k!}$で表される

ーN_0：母集団の想定レコード数、π_i：i番目のセルへのデータの入りやすさを表すパラメータ

・さらにパラメータπ_iがガンマ分布（α, β：$\alpha\beta = 1/K$を満たす形状母数と尺度母数）に従うと想定すると

$$\pi_i = \mathrm{gamma}(\alpha, \beta)$$

ーガンマ分布：ある期間λに1回起こることが期待される事象が実際に起こるまでの時間を表す分布で、

$\mathbb{E}(\pi_i) = \alpha\beta = \frac{1}{K}, Var(\pi_i) = \alpha\beta^2 = \frac{\beta}{K}$

・このとき母集団一意となるレコード数の期待値は（式中略）

$$\mathbb{E}(S_1) = \frac{N_0}{(1 + N_0\beta)^{1+\frac{1}{K\beta}}}$$

データ提供におけるリスク評価指標：標本一意性、母集団一意性

母集団一意かつ標本一意となるレコード数の推定

・母集団からサンプリング確率λで抽出された標本を考える

・標本一意となるセル期待値は

$$\mathbb{E}(s_1) = \frac{\lambda N_0}{(1+\lambda N_0 \beta)^{1+\frac{1}{K\beta}}}$$

・標本一意かつ母集団一意となるセル（＝レコード）数の推定値Uは

$$U = s_1 \cdot \frac{\mathbb{E}(S_1)}{\mathbb{E}(s_1)} \cdot \lambda = s_1 \cdot \left(\frac{1+\lambda N_0 \beta}{1+N_0 \beta}\right)^{1+\frac{1}{K\beta}}$$

・βは直接求めることができず、標本から推定する

　$-s_f^2$は標本におけるセルの度数の不偏分散

$$\beta = \frac{1}{\lambda N_0}\left(\frac{K}{\lambda N_0} s_f^2 - 1\right)$$

サンプリングまとめ

母集団と標本

・データセットが母集団か標本かによってリスクの考え方が異なる

・k-匿名化では母集団、標本に関わらずすべてのセルに対して一意性を持たせないようにする

・データセットの分布を想定すると、標本において一意であるレコードの母集団一意性の推測が可能

差分プライバシメカニズム

統計分析におけるリスク評価指標：差分プライバシ（再掲）

定義：ϵ-差分プライバシ

クエリ$q \in Q$において、$d(D, D') = 1$であるような任意のデータベース$D, D' \in \mathcal{D}$、および任意の出力の部分集合$S \subseteq Y$について

$$\Pr(M(q, D) \in S) \leq \exp(\epsilon) \cdot \Pr(M(q, D') \in S)$$

$$\Leftrightarrow \frac{\Pr\big(M(q, D)\big)}{\Pr\big(M(q, D')\big)} \leq \exp(\epsilon)$$

であるとき、確率的メカニズムMはϵ-差分プライバシを満たす。ただし$\epsilon \geq 0$とする。

定理：差分プライバシが保証する秘匿性

差分プライバシは弱秘匿性を保証する

証明

・$d(D_0, D_m) = m$となるD_0, D_mを想定する

・このとき$d(D_0, D_1) = d(D_1, D_2) = \cdots = d(D_{m-1}, D_m) = 1$となるデータベース$D_1, D_2, \cdots D_{m-1}$が存在する

・Mがϵ-差分プライバシを満たす時、任意の$i \in \{0, \cdots, m\}$に対して以下が成り立つ

$$\frac{\Pr\big(M(q, D_i)\big)}{\Pr\big(M(q, D_{i+1})\big)} \leq \exp(\epsilon)$$

・したがって

$$\frac{\Pr\big(M(q, D_0)\big)}{\Pr\big(M(q, D_1)\big)} \cdot \frac{\Pr\big(M(q, D_1)\big)}{\Pr\big(M(q, D_2)\big)} \cdot \cdots \cdot \frac{\Pr\big(M(q, D_{m-1})\big)}{\Pr\big(M(q, D_m)\big)} = \frac{\Pr\big(M(q, D_0)\big)}{\Pr\big(M(q, D_m)\big)} \leq \exp(\epsilon \cdot m) = \exp\big(\epsilon \cdot d(D_0, D_m)\big) = c$$

統計分析におけるリスク評価指標：差分プライバシ（再掲）

定義：(ϵ, δ)-差分プライバシ

クエリ$q \in Q$において、$d(D, D') = 1$であるような任意のデータベース$D, D' \in \mathcal{D}$、および任意の出力の部分集合$S \subseteq Y$について、

$$\Pr(M(q, D) \in S) \leq \exp(\epsilon) \cdot \Pr(M(q, D') \in S) + \delta$$

であるとき、確率的メカニズムMは(ϵ, δ)-差分プライバシを満たす。ただし$\epsilon, \delta \geq 0$とする。

δの持つ意味

・(ϵ, δ)-差分プライバシはϵ-差分プライバシの緩和版であるといえる

－$\delta = 0$のときϵ-差分プライバシと等価であり、δが大きくなるほどMは確率的にϵ-差分プライバシを満たさなくなる

・(ϵ, δ)-差分プライバシを満たすメカニズムは$1 - \frac{2\delta}{e^{\epsilon}\epsilon}$の確率で$2\epsilon$-差分プライバシを保証する

－$\frac{2\delta}{e^{\epsilon}\epsilon}$の確率で珍しい値を持つレコード（外れ値）に関する情報が漏れていると考えられる

定理：(ϵ, δ)-差分プライバシを満たすメカニズムのϵ-差分プライバシにおけるプライバシ保証

(ϵ, δ)-差分プライバシを満たすメカニズムMについて、少なくとも確率$1 - \delta'$で以下が成り立つ

$$\Pr(M(D) = y) \leq e^{\epsilon'} \Pr(M(D') = y)$$

ここで$\epsilon' = 2\epsilon, \delta' = \frac{2\delta}{e^{\epsilon}\epsilon}$である。

差分プライバシを満たす確率的メカニズム（再掲）

敏感度（sensitivity）の導入

・数値属性$x_i \in \mathbb{R}$からなるデータベース$D = \{x_1, \cdots, x_n\}$とクエリ$q: \mathbb{R}^n \to \mathbb{R}$を考える

・$d(D, D') = 1$であるようなデータベースの組$D = \{x_1, \cdots, x_n\}, D' = \{x_1, \cdots, x_{n-1}, x'_n\}$を隣接データベースとよぶ

・1レコード異なるデータベースにクエリ出力が与える影響の大きさを敏感度（sensitivity）とよぶ

ー敏感度は以下で定義される。ただし$\|\cdot\|_p$はl_pノルムである。

$$\Delta_{p,q} = \max_{\forall D, D' : d(D,D')=1} \|q(D) - q(D')\|_p$$

敏感度の具体例

・敏感度はクエリに応じて大幅に異なる

・レコードの定義域を$x \in [0, m]$とする

ー頻度関数 $q_{1,frequency} = \max_{k \in [0,m]} |k - (k-1)| = 1$（$k$は$x_n$の属性値の頻度）

ー平均関数 $q_{1,average} = \frac{1}{n} \max_{x_i, x'_i \in [0,m]} |x_n - x'_n| = \frac{m}{n}$

ー最大値関数 $q_{1,max} = \max_{x_i, x'_i \in [0,m]} |x_n - x'_n| = m$

差分プライバシを満たす確率的メカニズム（再掲）

ラプラスメカニズム

・ラプラスメカニズム：真のクエリの出力にラプラス分布から生成した乱数を加えるメカニズム

ーラプラス分布：$Lap(R) = \frac{1}{2R} e^{-\frac{|x|}{R}}$

ラプラスメカニズムのアルゴリズム

1．データベースD、プライバシパラメータϵ、クエリqの敏感度$\Delta_{1,q}$を入力する

2．ラプラス分布$Lap\left(\frac{\Delta_{1,q}}{\epsilon}\right)$に従う確率変数$r$を計算する

3．$y = q(D) + r$を出力する

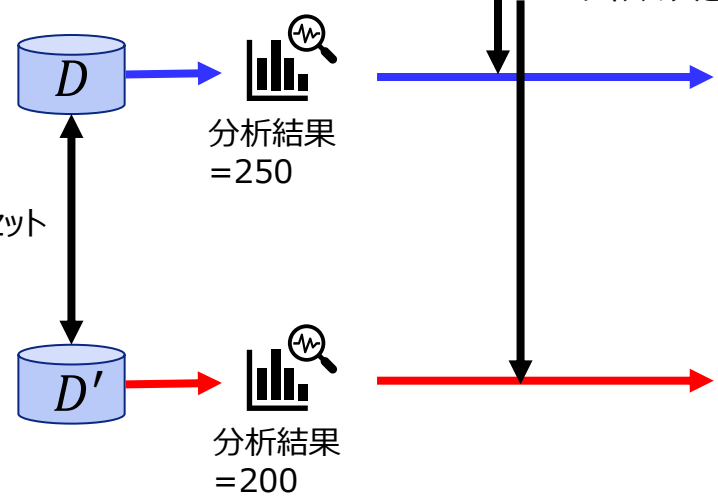
ラプラスメカニズムのイメージ

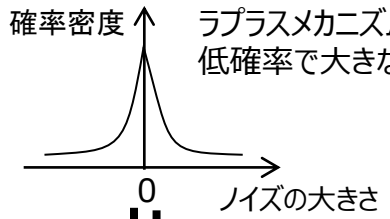
クエリ：40歳以上の貯金の平均値

名前	年齢	貯金
青木	33	1,000
⋮	⋮	⋮
山田	47	200
湯川	49	300

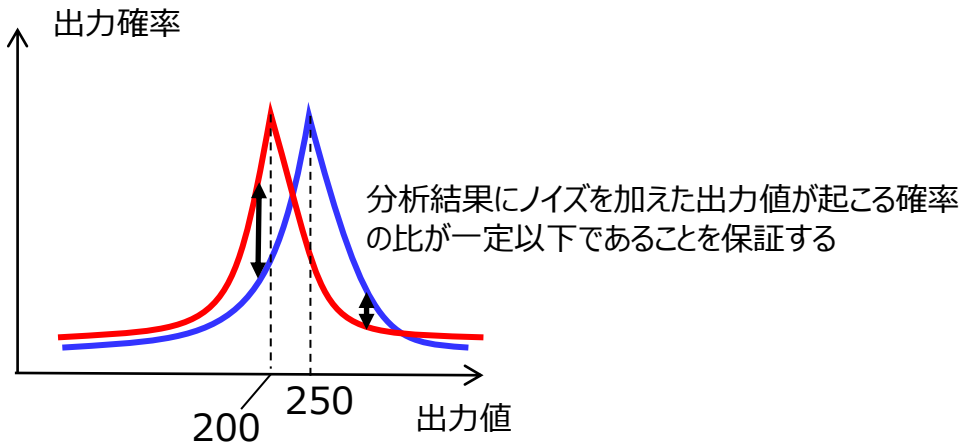
1レコード異なるデータセット

名前	年齢	貯金
青木	33	1,000
⋮	⋮	⋮
山田	47	200





ラプラスメカニズム：高確率で小さな値（0付近）、
低確率で大きな値となるノイズを分析結果に加える



分析結果にノイズを加えた出力値が起こる確率
の比が一定以下であることを保証する

差分プライバシを満たす確率的メカニズム（再掲）

定理：ラプラスメカニズムのプライバシ保証

ラプラスメカニズムはϵ-差分プライバシを満たす。

証明

・ラプラスメカニズムの応答値の確率密度分布は以下のように与えられる

$$p(M(D,q)=y)=\frac{1}{2R}\exp\left(-\frac{|r|}{R}\right)=\frac{\epsilon}{2\Delta_{1,q}}\cdot\exp\left(-\frac{\epsilon|y-q(D)|}{\Delta_{1,q}}\right)$$

・したがってラプラス分布における確率密度の比は以下の通り

$$\begin{aligned}\left|\frac{p(M(D,q)=y)}{p(M(D',q)=y)}\right| &= \left|\frac{\exp\left(-\frac{\epsilon|y-q(D)|}{\Delta_{1,q}}\right)}{\exp\left(-\frac{\epsilon|y-q(D')|}{\Delta_{1,q}}\right)}\right| \\ &= \left|\exp\left(\epsilon\cdot\frac{|y-q(D)|-|y-q(D')|}{\Delta_{1,q}}\right)\right| \\ &\leq \exp\left(\epsilon\cdot\frac{|q(D)-q(D')|}{\Delta_{1,q}}\right)(\because \text{三角不等式}) \\ &\leq \exp\left(\epsilon\cdot\frac{\Delta_{1,q}}{\Delta_{1,q}}\right)=e^{\epsilon}\end{aligned}$$

・ラプラスメカニズムはyを出力する確率の比を常にe^{ϵ}以下に抑えられるため、ϵ-差分プライバシを満たす

差分プライバシを満たす確率的メカニズム（再掲）

定理：ラプラスメカニズムの理論的有用性

ラプラスメカニズムは任意の$\delta \in (0,1]$について、以下が成り立つ。

$$\Pr\left(\|y - q(D)\|_1 > \frac{\Delta_{1,q}}{\epsilon} \cdot \ln\frac{1}{\delta}\right) \leq \delta$$

証明

・ラプラス分布に従う確率変数をrとすると、ラプラス分布の裾の確率は以下のように与えられる

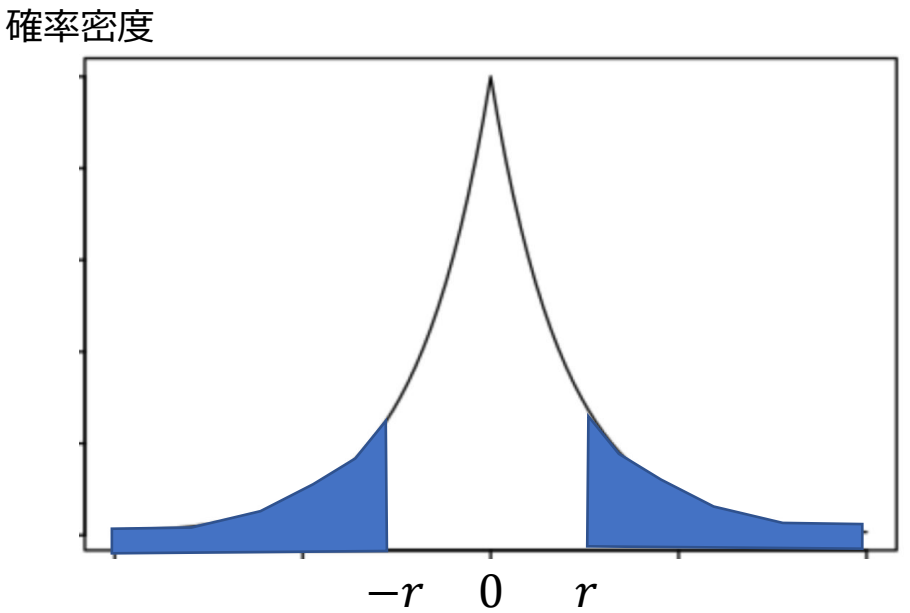
$$\Pr(|r| > t \cdot R) = \exp(-t)$$

$$\Leftrightarrow \Pr\left(\|y - q(D)\|_1 > \frac{\Delta_{1,q}}{\epsilon} \cdot \ln\frac{1}{\delta}\right) = \delta$$

プライバシと有用性のトレードオフ

・定理より、$\frac{\Delta_{1,q}}{\epsilon}$が大きいほど、誤差が大きくなる確率が高まることが分かる

・プライバシの強さと有用性はトレードオフの関係にある



差分プライバシを満たす確率的メカニズム（再掲）

ラプラスメカニズムの具体例：年齢

・年齢の定義域を[0,100]、データベースのサイズを$n = 100$、プライバシパラメータを$\epsilon = 0.1$とする

ークエリが平均値の場合、$\Delta_{1,q}^{ave} = \frac{100-0}{n} = \frac{100}{n} = O\left(\frac{1}{n}\right)$

ークエリが最大値の場合、$\Delta_{1,q}^{max} = 100 - 0 = 100 = O(1)$

$$\Pr\left(\|y - q(D)\|_1 > \frac{\Delta_{1,q}}{\epsilon} \cdot \ln\frac{1}{\beta}\right) \leq \beta$$

・平均値を出力とするラプラスメカニズムは95%の確率で真の値との誤差が$\frac{100}{\epsilon n}\ln\frac{1}{\delta} = \frac{1}{0.1}\ln\frac{1}{0.05} \approx 3$以下であることが保証される

ー$n = 100{,}000$まで増加させると、95%の確率で$\frac{100}{\epsilon n}\ln\frac{1}{\delta} = \frac{1}{0.1\times100}\ln\frac{1}{0.05} \approx 0.03$以下であることが保証される

・同様に最大値を出力とするラプラスメカニズムは95%の確率で$\frac{100}{\epsilon}\ln\frac{1}{\delta} = \frac{100}{0.1}\ln\frac{1}{0.05} \approx 2995.7$以下であることが保証される

ー$O(1)$なので、nをいくら増やしても精度の向上は見込めない

クエリと敏感度

・クエリによっては統計解析として問題のある結果となる

・差分プライバシを満たすために、敏感度を下げるための様々な研究がある

差分プライバシを満たす確率的メカニズム

指数メカニズム

・指数メカニズム：離散値をとるような出力空間における確率的メカニズム

　ーラプラスメカニズムは連続値を取る出力空間を対象としたメカニズム

・スコア関数$U: \mathcal{D} \times Y \to \mathbb{R}$を定義する

　ー$U(D, y)$は$q(D)$とyが近いほど高いスコアを取り、$q(D) = y$の時最大値を取るように設計する

・スコア関数の敏感度を以下のように定義する

$$\Delta_{U,q} = \max_{y \in Y} \max_{\forall D, D' : d(D,D')=1} |U(D, y) - U(D', y)|$$

指数メカニズムのアルゴリズム

１．データベースD、プライバシパラメータϵ、スコア関数U、スコア関数の敏感度$\Delta_{U,q}$を入力する

２．確率$\Pr(y|D) = \dfrac{\exp\left(\frac{\epsilon}{2\Delta_{U,q}} U(D,y)\right)}{\sum_{z \in Y} \exp\left(\frac{\epsilon}{2\Delta_{U,q}} U(D,z)\right)}$で$y \in Y$を選択し、出力する

定理：指数メカニズムのプライバシ保証

指数メカニズムはϵ-差分プライバシを満たす。

差分プライバシを満たす確率的メカニズム

定理：指数メカニズムのプライバシ保証

指数メカニズムはϵ-差分プライバシを満たす。

証明

・指数メカニズムの出力を$y = M(D, U)$とすると

$$\frac{\Pr(y|D)}{\Pr(y|D')} = \frac{\exp\left(\frac{\epsilon}{2\Delta_{U,q}} U(D,y)\right)}{\sum_{z\in Y} \exp\left(\frac{\epsilon}{2\Delta_{U,q}} U(D,z)\right)} \cdot \frac{\sum_{z\in Y} \exp\left(\frac{\epsilon}{2\Delta_{U,q}} U(D',z)\right)}{\exp\left(\frac{\epsilon}{2\Delta_{U,q}} U(D',y)\right)}$$

$$= \exp\left(\frac{\epsilon\big(U(D,y) - U(D',y)\big)}{2\Delta_{U,q}}\right) \cdot \frac{\sum_{z\in Y} \exp\left(\frac{\epsilon}{2\Delta_{U,q}} U(D',z)\right)}{\sum_{z\in Y} \exp\left(\frac{\epsilon}{2\Delta_{U,q}} U(D,z)\right)}$$

$$\leq \exp\left(\frac{\epsilon\Delta_{U,q}}{2\Delta_{U,q}}\right) \cdot \frac{\sum_{z\in Y} \exp\left(\frac{\epsilon}{2\Delta_{U,q}} \big(U(D,z) + \Delta_{U,q}\big)\right)}{\sum_{z\in Y} \exp\left(\frac{\epsilon}{2\Delta_{U,q}} U(D,z)\right)} \left(\because \Delta_{U,q} \geq U(D,y) - U(D',y)\right)$$

$$= \exp(\epsilon) \cdot$$

差分プライバシを満たす確率的メカニズム

定理：指数メカニズムの理論的有用性

指数メカニズムは任意の$t > 0$について、以下が成り立つ。

$$\Pr\left(M\big(H(D)\big) \leq H(D) - \frac{2\Delta_{U,q}}{\epsilon}\left(\ln\left(\frac{|Y|}{|Y_H|}\right) + t\right)\right) \leq e^{-t}$$

ここで$H(D) = \max_{y \in Y} U(D, y)$, $Y_H = \{y \in Y | U(D, y) = H(D)\}$であり、$M\big(H(D)\big)$は指数メカニズムを適用した際の$H(D)$。

証明

・略

差分プライバシを満たす確率的メカニズム

指数メカニズムの具体例：最頻値

・データベースのサイズを$n = 100{,}000$、プライバシパラメータを$\epsilon = 0.1$とする

・データベースにおいて、4つのカテゴリ$C_1, \cdots, C_4$が存在する属性の最頻値をクエリとする

・スコア関数を「指定したカテゴリに属するレコード数」として定義する

ーレコード数が多いカテゴリほど高いスコア値をとる

$$\Pr\left(M\big(H(D)\big) \leq H(D) - \frac{2\Delta_{U,q}}{\epsilon}\left(\ln\left(\frac{|Y|}{|Y_H|}\right) + t\right)\right) \leq e^{-t}$$

・このとき敏感度は

$$\Delta_{U,q} = \max_{C \in \{C_1,\cdots,C_4\}} \max_{\forall D,D' : d(D,D')=1} |U(D,C) - U(D',C)| = 1$$

・応答を期待するカテゴリのスコア関数の値は95%の確率で$H(D) - \frac{2}{0.1}\left(\ln\left(\frac{4}{1}\right) + 3\right) \approx H(D) - 87.7$以上が保証される

ー$t = 3$のとき、$e^{-3} \approx 0.05$

ーカテゴリC_1が最頻値で$|C_1| = 40{,}000$のとき、95%の確率で$|C_1| \geq 39{,}912.3$であることが保証される

ーもし$|C_2| = 39{,}912$であれば、5%の確率でC_2が出力される

複数回に渡るクエリに対する差分プライバシ

モジュラーアプローチ：差分プライバシメカニズムの出力値を組み合わせた統計解析

・差分プライバシを満たすメカニズムの出力を組み合わせた場合も差分プライバシは保証される

定理：直列合成定理

Dをデータベースとする。$i = 1, \dots, N$についてM_iをϵ_i-差分プライバシメカニズムとする。$M_i(q_i, D) = y_i$とすると、$M(q, D) = \{y_1, \cdots, y_N\}$を出力するメカニズムは$\sum_{i=1}^{N} \epsilon_i$-差分プライバシを満たす

証明

$$\frac{\Pr\big(M(q,D) = (y_1, \cdots, y_N)\big)}{\Pr\big(M(q,D') = (y_1, \cdots, y_N)\big)} = \frac{\Pr(M_1(q_1, D) = y_1)}{\Pr(M_1(q_1, D') = y_1)} \cdot \dots \cdot \frac{\Pr(M_N(q_N, D) = y_N)}{\Pr(M_N(q_N, D') = y_N)}$$
$$\leq e^{\epsilon_1} \cdot \dots \cdot e^{\epsilon_N} = e^{\sum_{i=1}^{N} \epsilon_i}$$

定理：並列合成定理

$D = \{D_1, \cdots, D_N\}$をデータベースとする。またmをϵ-差分プライバシメカニズムとする。$m(q, D_i) = y_i$とすると、$M(q, D) = \{y_1, \cdots, y_N\}$を出力するメカニズムは$\epsilon$-差分プライバシを満たす

証明

・略

プライバシパラメータの管理

・差分プライバシを満たすメカニズムの出力を組み合わせるとプライバシの保証が弱くなるため、プライバシパラメータの管理が必要

ーラプラスメカニズムは平均0のノイズを加えるため、繰り返しラプラスメカニズムを利用し、応答値の平均を取ると真の値が分かる

複数回に渡るクエリに対する差分プライバシ

合成定理の応用例

・相関係数をクエリとした差分プライバシメカニズムを考える

・相関係数Cは2つの属性A, Bの標準偏差と共分散から計算する

ーデータ数をn、属性A, Bのデータをそれぞれa_i, b_i、平均値をμ_A, μ_Bとすると、

$$C = \frac{s_{AB}}{s_A \cdot s_B} = \frac{\frac{1}{n}\sum_{i=1}^{n}(a_i - \mu_A)(b_i - \mu_B)}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(a_i - \mu_A)^2}\sqrt{\frac{1}{n}\sum_{i=1}^{n}(b_i - \mu_B)^2}}$$

・相関係数をクエリとした場合、標準偏差が0となるデータセットが存在するため、敏感度の定義ができない

ー例えば$a_1 = \cdots = a_n$となるデータベース

・そこで合成定理を利用して、相関係数をクエリとした差分プライバシメカニズムMの構築が可能

ーA, Bそれぞれの標準偏差、および共分散をクエリとした差分プライバシメカニズム$M_{s_A}, M_{s_B}, M_{s_{AB}}$を設計する

ー$M_{s_A}(q_{s_A}, D) = \tilde{s}_A, M_{s_B}(q_{s_B}, D) = \tilde{s}_B, M_{s_{AB}}(q_{s_{AB}}, D) = \tilde{s}_{AB}$を計算する

ー$\tilde{C} = \frac{\tilde{s}_{AB}}{\tilde{s}_A \cdot \tilde{s}_B}$を計算することと差分プライバシを満たす相関係数をクエリとしたメカニズムMは同義

・このときMがϵ-差分プライバシを満たすには、$M_{s_A}, M_{s_B}, M_{s_{AB}}$をそれぞれ$\epsilon/3$-差分プライバシを満たすようにすればよい

その他PETs

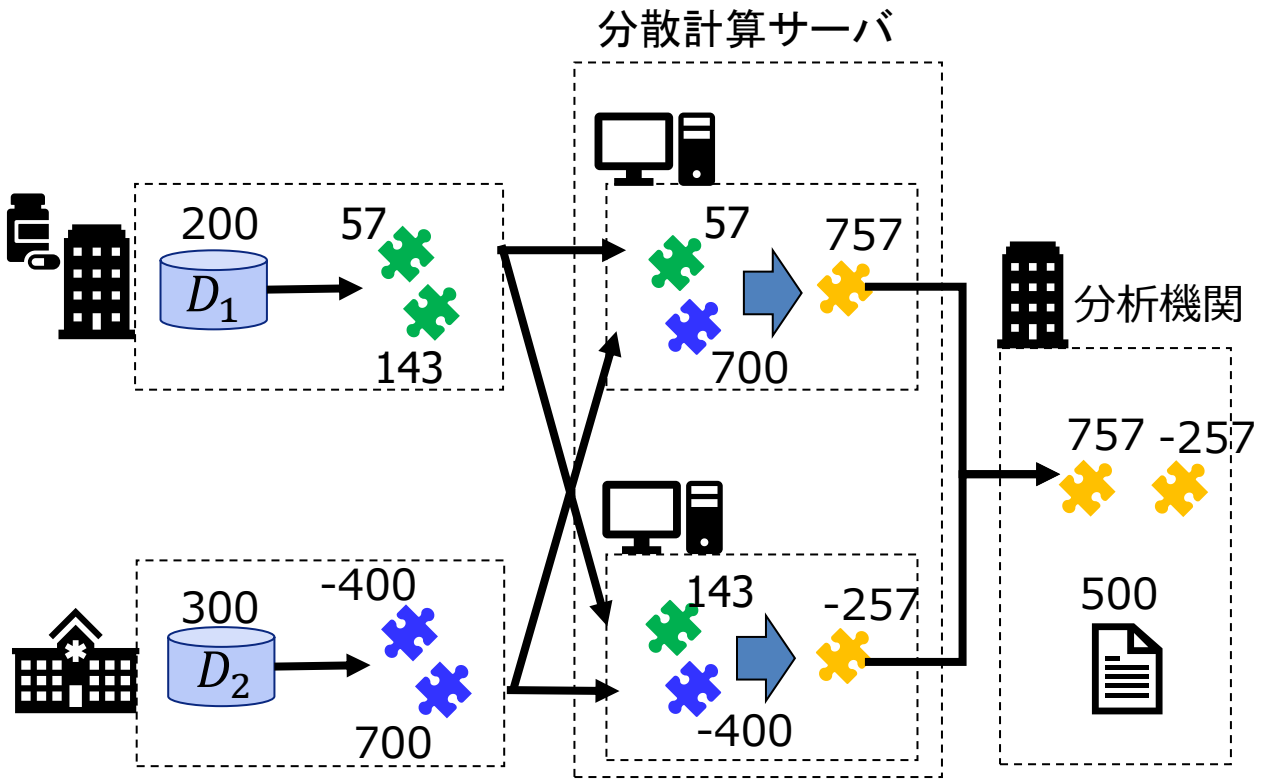
その他のプライバシ強化技術（PETs）：秘密計算

秘密計算：データを分散あるいは暗号化した状態で分析する手法

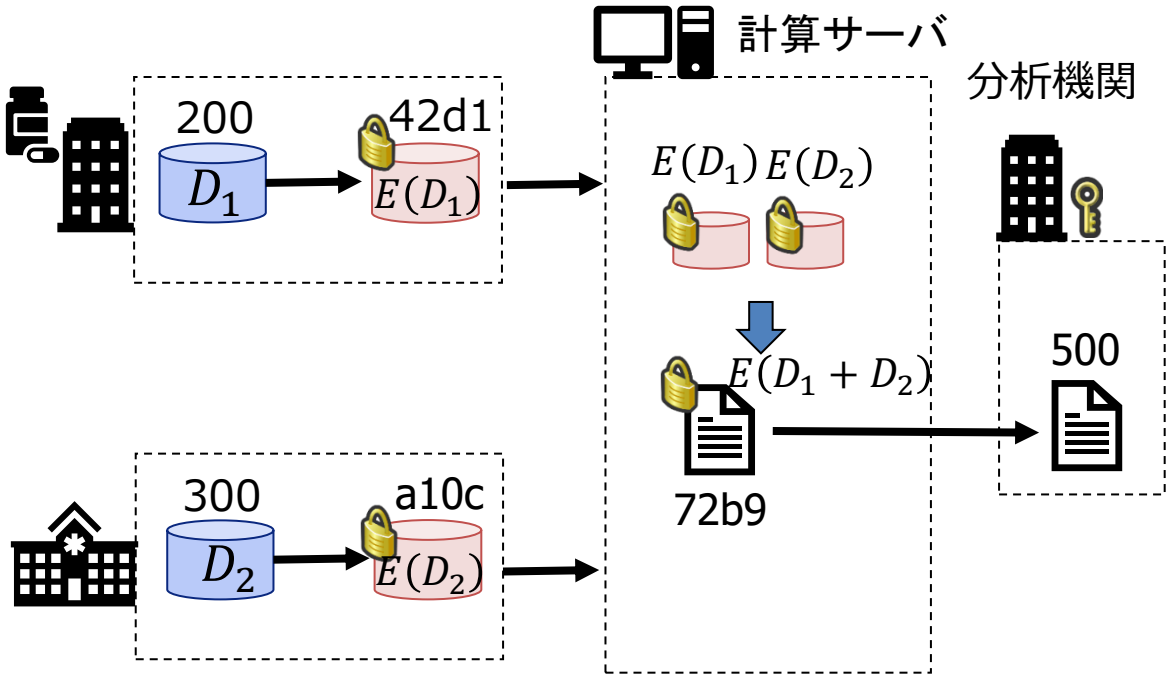
・秘密分散（左図）：データを分散化した状態で処理を行い、特定の計算を実現

・準同型暗号（右図）：データを暗号化した状態で加算・乗算が可能な暗号を利用して特定の計算を実現

・Garbled Circuit：暗号技術を利用してAND、ORの回路を構築し、これらを組み合わせて特定の計算を実現



秘密分散法



準同型暗号法

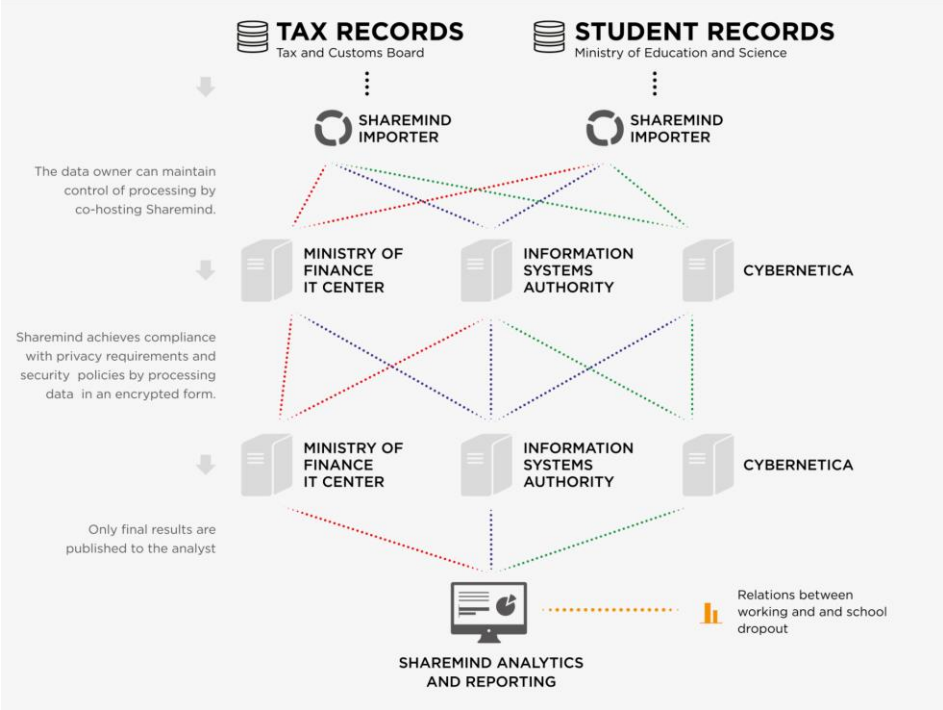
秘密計算の実用例

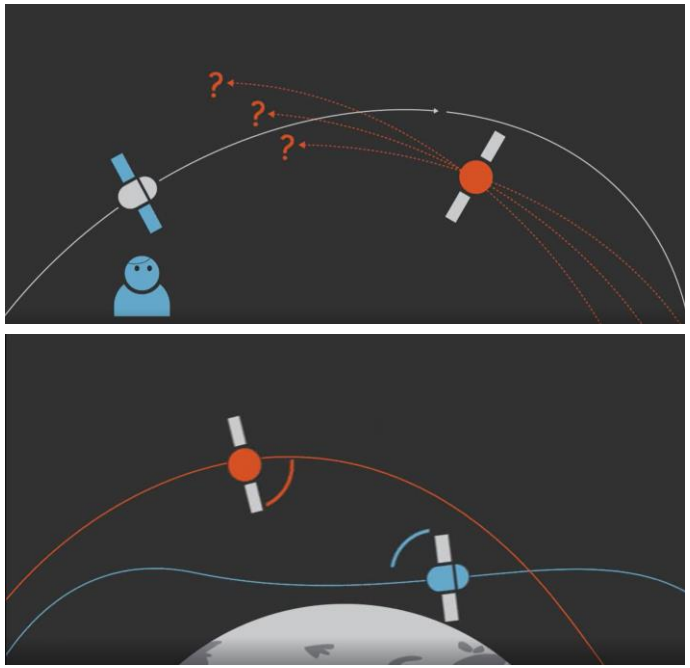
学生のアルバイト量と落第率の相関

・学生のアルバイト量と落第率の相関は、教育科学省が保持する学生データと税務・関税局が保持する労働者・納税データを掛け合わせれば簡単に割り出すことができるが、個人情報保護の観点からそれぞれのデータを共有することができなかった

人工衛星の軌道予測と衝突防止対策

・人工衛星の互いの位置や飛行ルートを共有すれば衝突は避けられるが、これらは各国・各事業者の機密情報であり、共有することができなかった





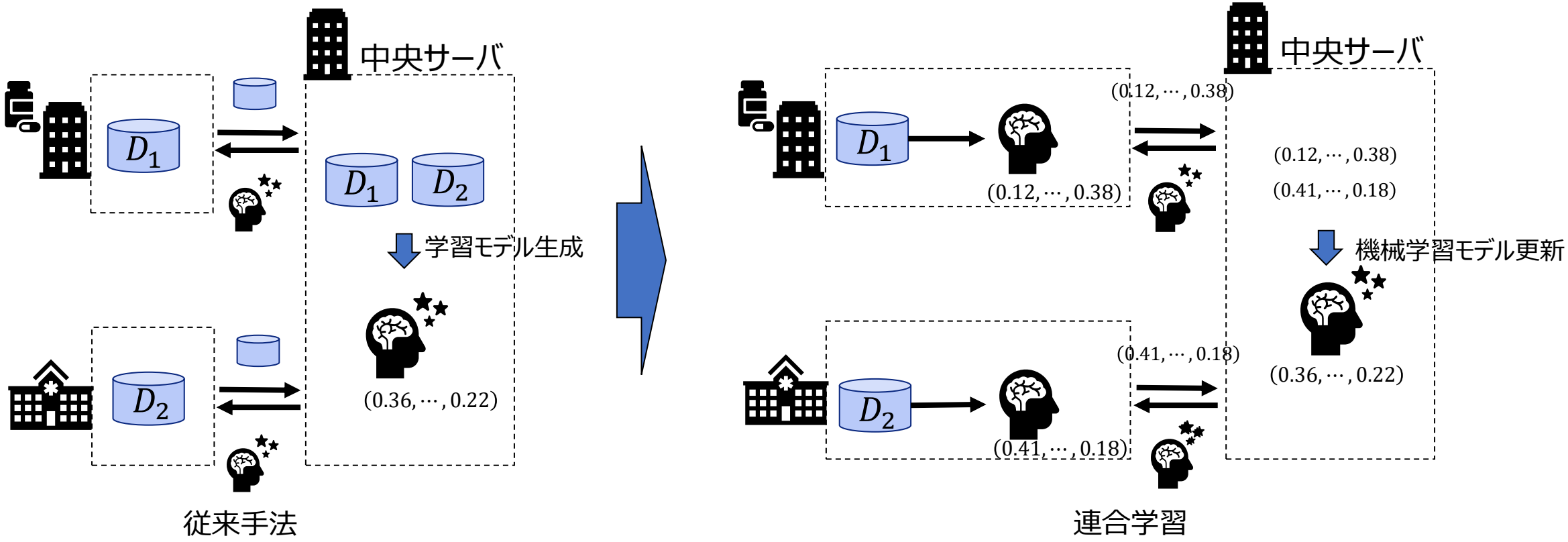
https://sharemind.cyber.ee/big-data-analytics-protection/
https://sharemind.cyber.ee/satellite-collision-security/

その他のプライバシ強化技術（PETs）：連合学習

<u>連合学習：学習データを分散した状態で機械学習モデルを生成・更新する手法</u>

・各企業が持つデータを他社に渡すことなく、共同で機械学習モデルを作成する

・秘密計算に比べて計算サーバの計算量が小さい



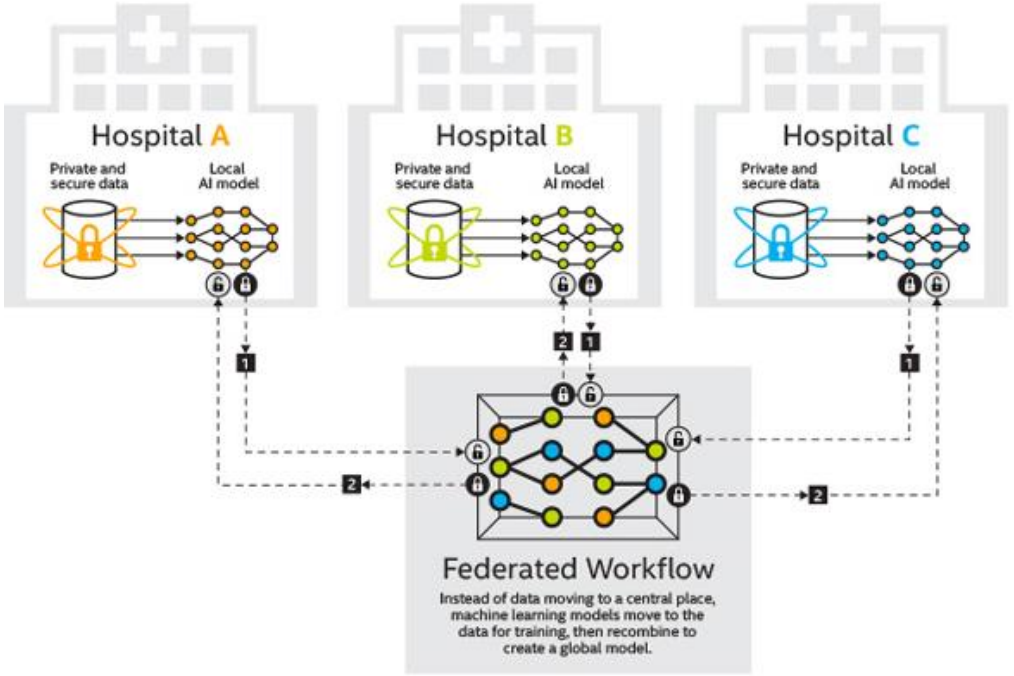
連合学習の実用例

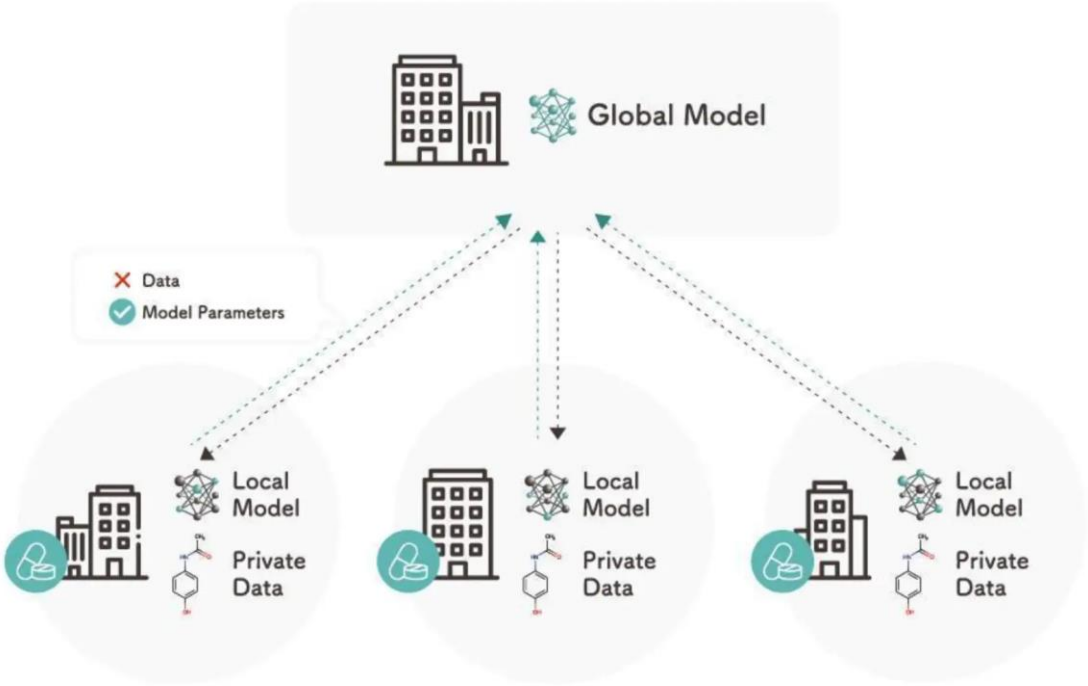
脳腫瘍を特定するAIモデルの開発

・インテルとペンシルバニア大学は連合学習により脳腫瘍を識別するAIモデルの学習を可能とする技術を共同開発し、プライバシ保護を行わない場合と比較して99%以上の精度が出ることを実証

AI創薬に向けた連合学習機能を有する機械学習ライブラリの開発

・京都大学とAI創薬企業Elixにより、化合物データを対象とした連合学習による機械学習モデルの構築が可能なオープンソースのライブラリkMoLがリリースされた





https://newsroom.intel.com/news/intel-works-university-pennsylvania-using-privacy-preserving-ai-identify-brain-tumors/#gs.vhqho5

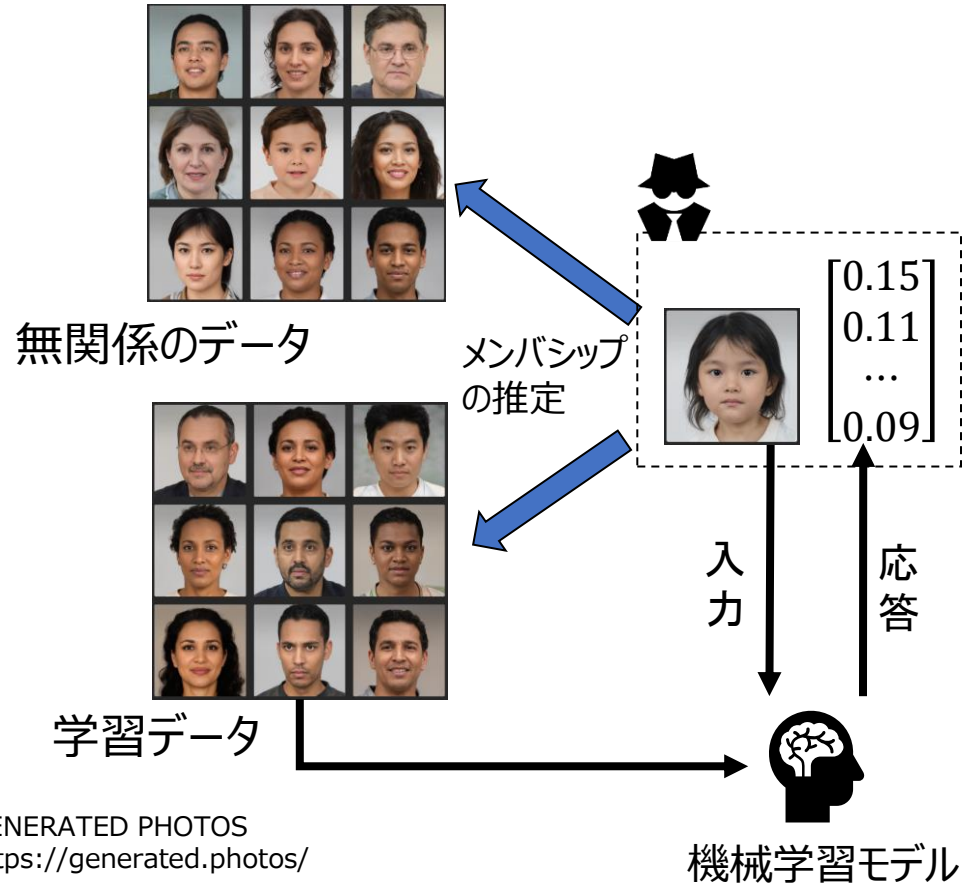
https://prtimes.jp/main/html/rd/p/000000007.000027687.html

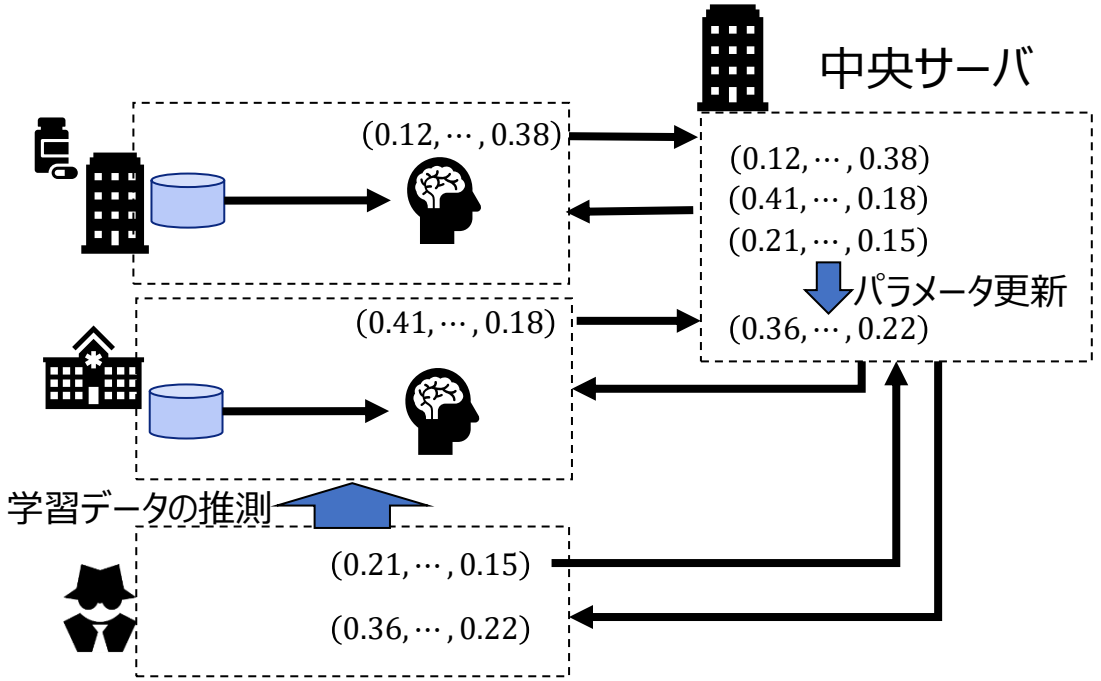
複雑化するプライバシ情報に対する攻撃

機械学習におけるプライバシへの攻撃も存在する

・メンバシップ推定（左図）：あるデータが機械学習モデルの学習データとして使われたかどうかを推測する

・学習データの推測（右図）：連合学習において学習モデル構築時のパラメータから学習データを推測する





差分プライバシメカニズム・局所差分プライバシメカニズム応用例

差分プライベート連合学習

Federated Learning

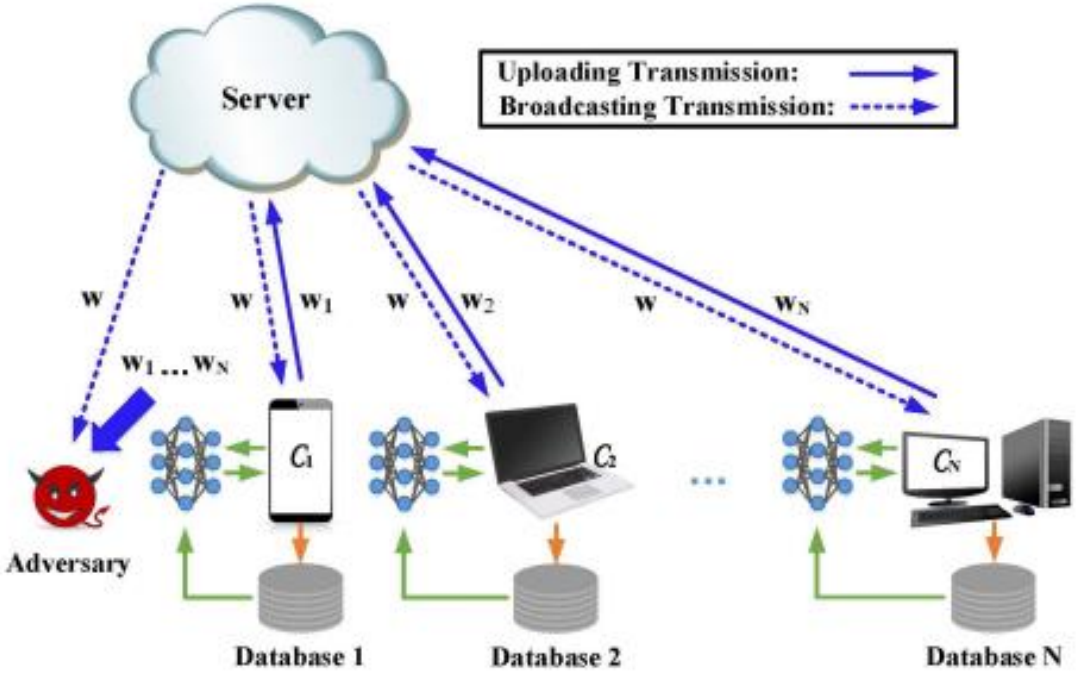
・連合学習のトレーニングプロセス

１．各クライアントがローカルデータベースに基づき
パラメータを更新する

２．サーバはパラメータの統合を行う

３．サーバは統合したパラメータを各クライアントに送信する

４．各クライアントはパラメータをもとにモデルを更新する



・サーバSとN個のクライアント$C_i(i = 1, \dots, N)$から構成される連合学習システムを考える

ークライアントC_iはローカルデータベースD_iを保有する

・連合学習ではサーバで集約する機械学習モデルの損失関数を最小化するようなベクトル$\boldsymbol{w}$を見つける以下の最適化問題を解く

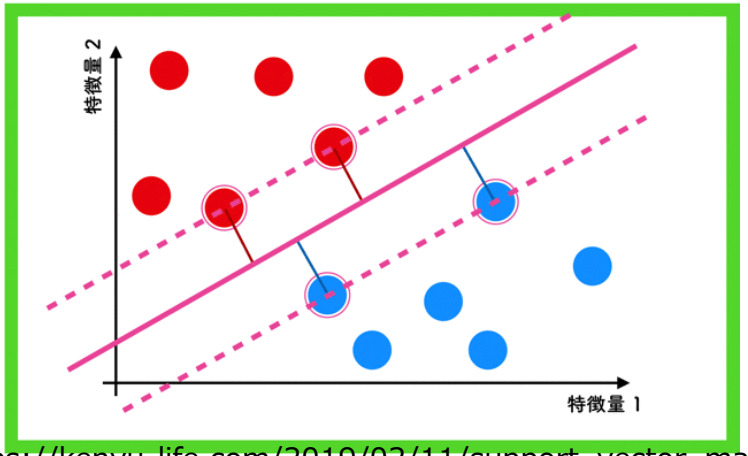
ーw_i：C_iで計算したパラメータベクトル

ーw：サーバで結合したパラメータベクトル

ー$p_i = |D_i| / \sum^N |D_j|$：クライアント$C_i$の重み

ー$F_i(\cdot)$：C_iの局所損失関数

$$\boldsymbol{w} = \sum_{i=1}^{N} p_i \cdot \boldsymbol{w}_i \cdots (1),$$
$$\boldsymbol{w}^* = \arg\min_{\boldsymbol{w}} \sum_{i=1}^{N} p_i \cdot F_i(\boldsymbol{w}, D_i) \cdots (2).$$



https://kenyu-life.com/2019/02/11/support_vector_machine/

差分プライベート連合学習

連合学習における敏感度（uplink）

・各クライアントは保有するデータベースD_iの各データD_{ij}に対して、損失関数が最小となるベクトルを見つける

ーバッチサイズはトレーニングサンプル数と等しいと想定すると、

$$\boldsymbol{w}_i = \arg\min_{\boldsymbol{w}} F_i(\boldsymbol{w}, D_i) = \frac{1}{|D_i|} \cdot \sum_{j=1}^{|D_i|} \arg\min_{\boldsymbol{w}} F_i\left(\boldsymbol{w}, D_{ij}\right)$$

・D_{im}とD'_{im}が異なるレコードとすると、クライアントC_iのパラメータベクトルを出力としたときの敏感度は

$$\begin{aligned}
\Delta_{UP}^{i} &= \max_{\forall D_i, D_i'} \|\boldsymbol{w}_i - \boldsymbol{w}_i'\| \\
&= \max_{\forall D_i, D_i'} \left\| \frac{1}{|D_i|} \cdot \sum_{j=1}^{|D_i|} \arg\min_{\boldsymbol{w}} F_i\left(\boldsymbol{w}, D_{ij}\right) - \frac{1}{|D_i'|} \cdot \sum_{j=1}^{|D_i'|} \arg\min_{\boldsymbol{w}} F_i\left(\boldsymbol{w}, D'_{ij}\right) \right\| \\
&= \max_{\forall D_i, D_i'} \left\| \frac{1}{|D_i|} \cdot \left(\arg\min_{\boldsymbol{w}} F_i(\boldsymbol{w}, D_{im}) - \arg\min_{\boldsymbol{w}} F_i(\boldsymbol{w}, D'_{im}) \right) \right\|
\end{aligned}$$

・モデルのトレーニングでは、極端なパラメータの変更が行われないように、clipping threshold C というものが設定される

ーこのとき常に$|\boldsymbol{w}_i| \leq C$

・したがって、$\Delta_{UP}^{i} = \frac{2C}{|D_i|}$

・連合学習のuplink全体を考えた敏感度は$\Delta_{UP} = \max_i \{\Delta_{UP}^{i}\}$

差分プライベート連合学習

連合学習における敏感度（downlink）

・サーバは各クライアントから送られる$\boldsymbol{w}_i$の加重平均$\boldsymbol{w} = p_1\boldsymbol{w}_1 + \cdots + p_N\boldsymbol{w}_N$を求める

・クライアントC_iから送られるパラメータが異なると想定すると
　ーローカルデータベースの最小サイズをmとする

$$\Delta_{DOWN}^{i} = \max_{\forall D_i, D_i'} \|\boldsymbol{w} - \boldsymbol{w}'\|$$
$$= \max_{\forall D_i, D_i'} \|(p_1\boldsymbol{w}_1 + \cdots + p_i\boldsymbol{w}_i + p_N\boldsymbol{w}_N) - (p_1\boldsymbol{w}_1 + \cdots + p_i\boldsymbol{w}_i' + \cdots + p_N\boldsymbol{w}_N)\|$$
$$= \max_{\forall D_i, D_i'} \|p_i(\boldsymbol{w}_i - \boldsymbol{w}_i')\|$$
$$= p_i \cdot \Delta_{UP}$$
$$\leq \frac{2p_iC}{|D_i|}$$
$$\leq \frac{2p_iC}{m}$$

・連合学習のdownlink全体を考えた敏感度は$\Delta_{DOWN} = \max_i\{\Delta_{DOWN}^{i}\} = \max_i\left\{\frac{2p_iC}{m}\right\}$

差分プライベート連合学習

<u>差分プライバシを満たす連合学習のプライバシ</u>

・既存研究においては(ϵ, δ)-差分プライバシを満たすメカニズムとしてガウスメカニズムを利用している

ー$\sigma \geq c \cdot \frac{\Delta_{2,f}}{\epsilon}\left(c \geq \sqrt{2\ln(1.25/\delta)}\right)$のとき、$M(D) = f(D) + n$（ガウシアンノイズ$n \sim N(0, \sigma^2)$）が$(\epsilon, \delta)$-差分プライバシを満たすことが知られている

・連合学習全体を通して差分プライバシを満たすには、uplinkとdownlinkで差分プライバシを満たす必要がある

ーt回目のパラメータ更新において、各クライアントC_iはローカルデータベースD_iを用いてパラメータ$\boldsymbol{w}_i^t$を更新後、ノイズを加えたパラメータ$M_{UP}\left(\boldsymbol{w}_i^t\right) = \boldsymbol{w}_i^t + n_i$をサーバに送信する

ーサーバは各クライアントから送られたパラメータを利用して$\boldsymbol{w}^{t+1} = \sum_{i=1}^{N} p_i \cdot M\left(\boldsymbol{w}_i^t\right)$を更新、ノイズを加えたパラメータ$M_{DOWN}(\boldsymbol{w}^{t+1}) = \boldsymbol{w}^{t+1} + n = \tilde{\boldsymbol{w}}^{t+1}$を各クライアントに送信する

・uplink、downlinkの各パラメータ更新において、プライバシパラメータϵ, δで差分プライバシを適用し、パラメータ更新をT回行うと、連合学習全体としては$(2T\epsilon, 2T\sigma)$-差分プライバシを満たす

ー差分プライバシの直列合成定理と並列合成定理による

・実際のパラメータ更新回数は100回を超えることもあり、各回のノイズがかなり大きくなる

ー最近の研究ではベクトルの圧縮やスパース性を用いてより効率的な手法が提案されている

差分プライベート連合学習

差分プライバシを満たす連合学習の有用性

・プライバシ強度ϵを満たすとき、T回目のパラメータ更新後の最適な損失関数$F(\boldsymbol{w}^*)$と起こりうる損失関数$F(\widetilde{\boldsymbol{w}}^T)$の差の期待値は

$$\mathbb{E}[F(\widetilde{\boldsymbol{w}}^T) - F(\boldsymbol{w}^*)] \leq P^T\Theta + \left(\frac{\kappa_1 CT}{m\sqrt{N}\cdot\epsilon} + \frac{\kappa_0 C^2T^2}{m^2N\cdot\epsilon^2}\right)(1-P^T)$$

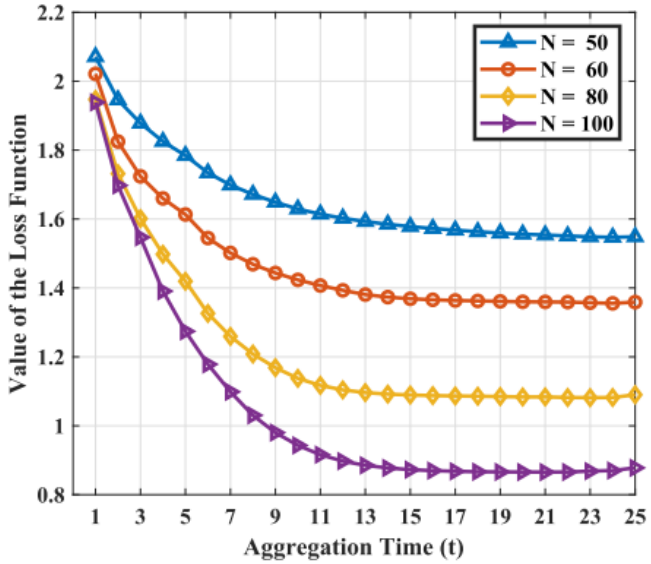
・この理論値をもとに各パラメータと有用性の分析が可能

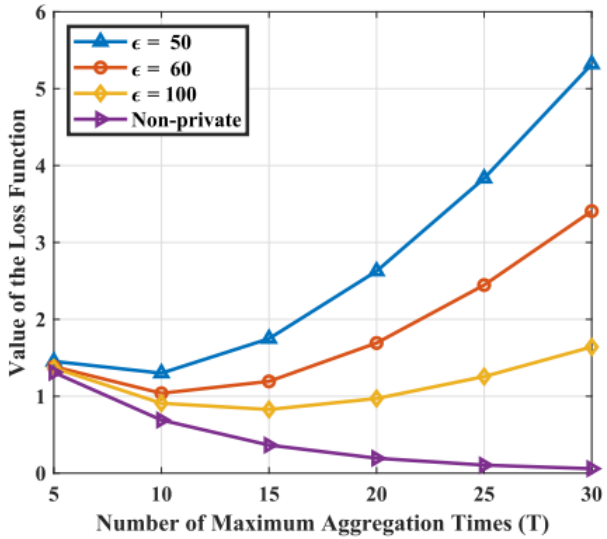
ークライアント数Nについて、$\mathbb{E}[F(\widetilde{\boldsymbol{w}}^T) - F(\boldsymbol{w}^*)]$は単調減少

→クライアント数が多いほど有用性が向上する

ー上記式をTで微分すると、N, ϵが十分大きい場合、Tについて下に凸な関数となる

→有用性が最大となるパラメータ更新回数が存在する





局所差分プライバシメカニズムの実用

Learning with Privacy at Scale

・Apple社による局所差分プライバシの実用

・ユーザのキーボード入力を知ることはUXの向上に役立つ

　ーユーザのキーボード入力を直接知ることはプライバシ侵害につながる

　ー局所差分プライバシメカニズムを利用して、統計情報のみ取得する

・全体像

　ーPrivatization：ユーザによる局所差分プライバシメカニズムを用いた処理

　ーIngestor：不要な間接識別情報の削除処理

　ーAggregator：ユースケースに応じた統計分析処理

局所差分プライバシメカニズムの実用

Privatization：　（スマホ内で自動で）ユーザが実施する処理

・デバイスでイベントが発生するたびに、局所差分プライバシメカニズムを実施し、一時的に保存する

・一定時間経過後、デバイスに保存されているデータをランダムサンプリングし、サーバに送信する

・サーバに送信する際、デバイスID等のユーザ識別子を削除する

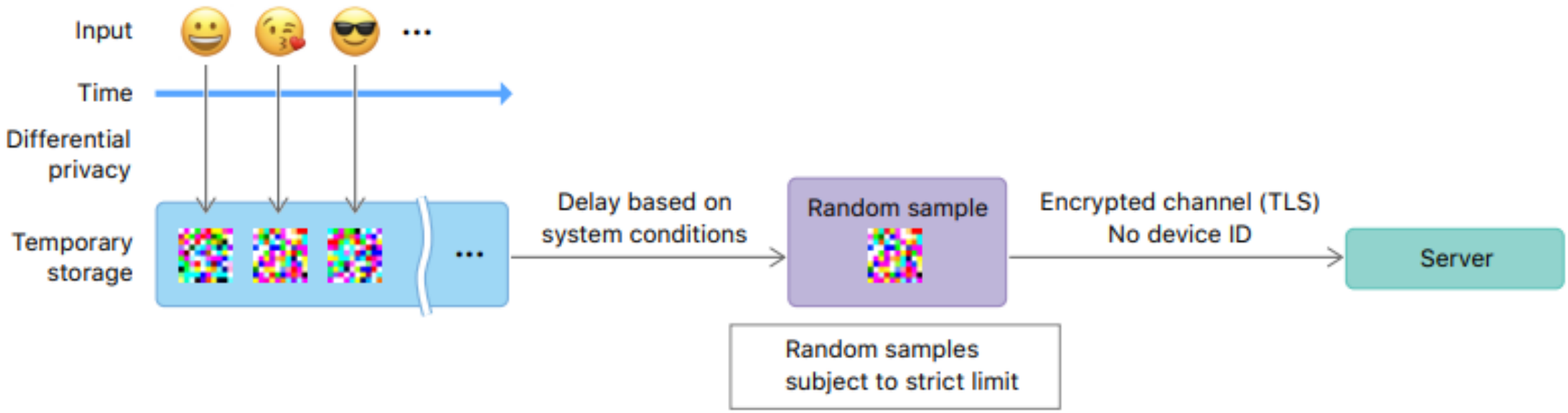


Figure 2: Privatization Stage

局所差分プライバシメカニズムの実用

Ingestion and Aggregation：個人の特定を防ぐための前処理

・各ユーザから送られたデータはユーザ識別子とIPアドレスが削除される

・Ingestor

ーデータのタイムスタンプ等の情報を削除し、ユースケースごとにレコードを分離するバッチ処理を行う

・Aggregator

ーユースケースごとにヒストグラムを生成する

ー複数のユースケースからデータが結合されることはない

ー所定のしきい値を超える要素のみがヒストグラムに反映される

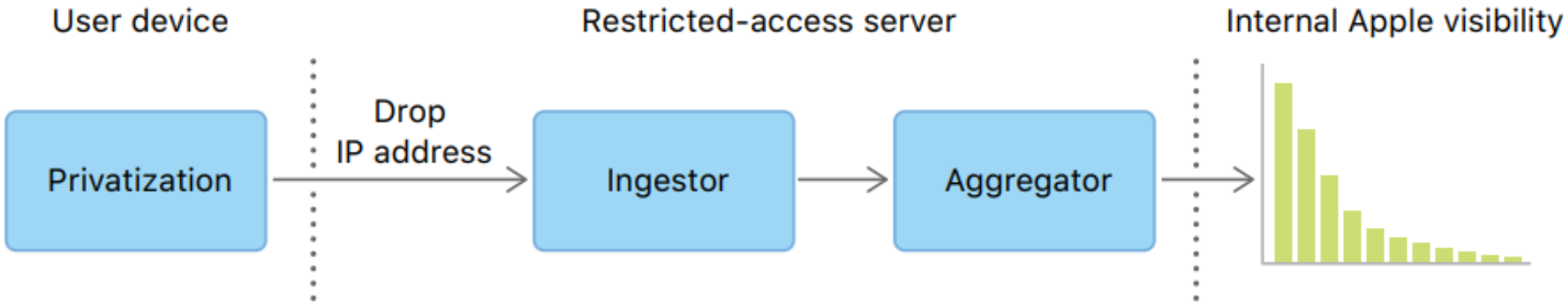


Figure 1: System Overview

局所差分プライバシメカニズムの実用

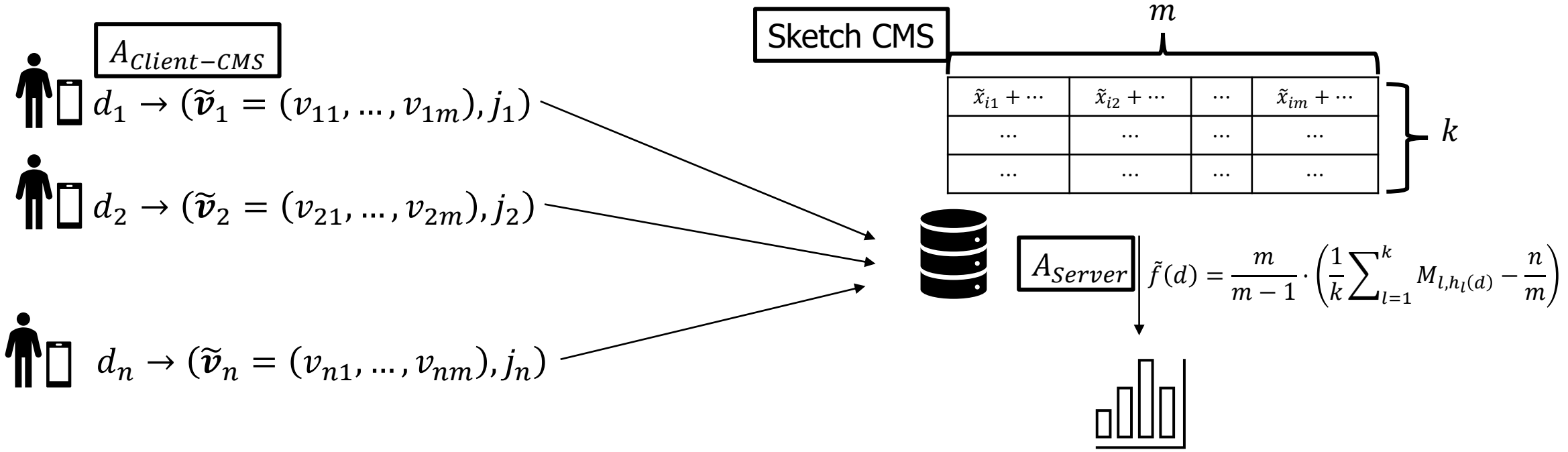
頻度分析メカニズム

・ユーザは事前に辞書登録されたm種類のデータについて、以下のステップを通じてApple社にデータを提供する

－$A_{\text{client-CMS}}$：ユーザは局所差分プライバシを利用し、辞書登録されたデータをm次元ベクトル$\boldsymbol{v}$に変換して提供する

－$\text{Sketch} - \text{CMS}$：サーバは提供されたデータを集約し、行列$M \in \mathbb{R}^{k \times m}$を生成する

－A_{Server}：サーバは行列Mをもとにユーザの入力したデータの分布を推測する



局所差分プライバシメカニズムの実用

クライアントの処理： $A_{\text{client-CMS}}(d, \epsilon, H) \rightarrow (\tilde{\boldsymbol{v}}, j)$

1．ユーザは事前にk個のハッシュ関数$H = \{h_1, \cdots, h_k\}(h_i(\cdot) \in [1, m])$を共有している

2．各ユーザはランダムにハッシュ関数h_jを選択、$\boldsymbol{v} = \left(v_1, \cdots v_{h_i(d)-1}, v_{h_i(d)}, v_{h_i(d)+1} \cdots, v_m\right) = (-1, \cdots, -1, 1, -1 \cdots, -1)$とする

3．$\boldsymbol{v}$の各要素について、$\frac{1}{e^{\epsilon/2}+1}$の確率でビット反転させ、$\tilde{\boldsymbol{v}}$とする（ランダム化応答を利用、$\epsilon$-LDPを満たす）

4．$\tilde{\boldsymbol{v}}, j$を出力する

サーバの処理１： $\text{Sketch} - \text{CMS}(D, \epsilon, k, m) \rightarrow M$

1．データセット$D = \{(\tilde{\boldsymbol{v}}_1, j_1), \cdots, (\tilde{\boldsymbol{v}}_n, j_n)\}$を入力とする

2．各ユーザからのデータに対して、$\tilde{\boldsymbol{x}}_i = k \cdot \left(\frac{1}{2} \cdot \frac{e^{\epsilon/2}+1}{e^{\epsilon/2}-1} \cdot \tilde{\boldsymbol{v}}_i + \frac{1}{2} \cdot \mathbf{1}\right)$を計算する

3．空の行列$M \in \{0\}^{k \times m}$を用意する

4．各ユーザからのデータに対して、$M_{j_i, l} = M_{j_i, l} + \tilde{x}_{i,l}$を計算、$M$を出力する

サーバ側の処理２： $A_{\text{Server}}(d, M, \epsilon, H) \rightarrow \tilde{f}(d)$

5．データdのユーザの合計入力回数$f(d)$を$\tilde{f}(d) = \frac{m}{m-1} \cdot \left(\frac{1}{k} \sum_{l=1}^{k} M_{l, h_l(d)} - \frac{n}{m}\right)$として予測する

局所差分プライバシメカニズムの実用

ユーザの処理の具体例（$\epsilon = \infty$）：$A_{\text{client-CMS}}(d, \epsilon, H) \rightarrow (\tilde{\boldsymbol{v}}, j)$

・ハッシュの数が3個、データの種類が4種類、すなわち$k = 3, m = 4$の場合

・ユーザが$j = 2$を選択し、$h_2(d) = 1$だったとすると、$\boldsymbol{v} = (1, -1, -1, -1)$

・ $\boldsymbol{v}$に局所差分プライバシメカニズムを適用し、$\tilde{\boldsymbol{v}} = (1, -1, -1, -1)$とする

ーここでは説明のため$\epsilon = \infty$（ノイズなし）

ー$h_2(d)$の箇所のみ1、ほかはすべて-1となり、ϵが小さくなるにしたがって確率的に値が反転する

・$(\tilde{\boldsymbol{v}}, j)$を返す

局所差分プライバシメカニズムの実用

サーバの処理１の具体例（$\epsilon = \infty$）:$\mathrm{Sketch-CMS}(D, \epsilon, k, m) \rightarrow M$

・ハッシュの数が3個、データの種類が4種類、ユーザが4人、すなわち$k = 3, m = 4, n = 4$の場合

・$D = \{(\tilde{\boldsymbol{v}}_1, j_1), \cdots, (\tilde{\boldsymbol{v}}_n, j_n)\} = \{((-1,1,-1,-1), 2), ((1,-1,-1,-1), 1), ((-1,1,-1,-1), 2), ((-1,-1,1,-1), 2)\}$とする

　$-h_2(d_1) = 2, h_1(d_2) = 1, h_2(d_3) = 2, h_2(d_4) = 3$

・$\tilde{\boldsymbol{x}}_1 = k \cdot \left(\frac{1}{2} \cdot \frac{e^{\epsilon/2}+1}{e^{\epsilon/2}-1} \cdot \tilde{\boldsymbol{v}}_1 + \frac{1}{2} \cdot \mathbf{1}\right) = 3 \cdot \left(\frac{1}{2} \cdot 1 \cdot (-1,1,-1,-1) + \frac{1}{2}(1,1,1,1)\right) = 3 \cdot (0,1,0,0) = (0,3,0,0)$

　$-\epsilon = \infty$では正しい$h_{j_i}(d_i)$の箇所のみk、その他は0となる

・$M_{j_i,l} = M_{j_i,l} + \tilde{x}_{i,l}$を計算

3	0	0	0
0+0+0	3+3+0	0+0+3	0+0+0

・$M = \begin{bmatrix} 3 & 0 & 0 & 0 \\ 0 & 6 & 3 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$を返す

局所差分プライバシメカニズムの実用

サーバの処理２の具体例（$\epsilon = \infty$）　： $A_{\mathrm{Server}}(d, M, \epsilon, H) \to \tilde{f}(d)$

・$D = \{(\tilde{\boldsymbol{v}}_1, j_1), \cdots, (\tilde{\boldsymbol{v}}_n, j_n)\} = \{((-1,1,-1,-1), 2), ((1,-1,-1,-1), 1), ((-1,1,-1,-1), 2), ((-1,-1,1,-1), 2)\}$とする

$$-M = \begin{bmatrix} 3 & 0 & 0 & 0 \\ 0 & 6 & 3 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

・$\tilde{f}(d) = \frac{m}{m-1}\left(\frac{1}{k}\sum M_{l,h_l(d)} - \frac{n}{m}\right)$

$$-\tilde{f}(d_3) = \frac{4}{3} \cdot \left(\frac{1}{3} \cdot \left(M_{1,h_1(d_3)} + M_{2,h_2(d_3)} + M_{3,h_3(d_3)}\right) - \frac{4}{4}\right) = \frac{4}{3}\left(\frac{1}{3} \cdot (0 + 6 + 3 + 0) - 1\right) = \frac{8}{3}$$

正しいd_3の出現回数× k

d_3が出現した回数× k
今回は$h_2(d_1) = h_2(d_3)$となるためd_1も含む

確率的に他のデータとも被る
今回は$h_3(d_3) = h_2(d_4)$となるためd_4を含む

確率的にd_3を入力したときのハッシュ値と異なるデータを入力としたときのハッシュ値が一致する

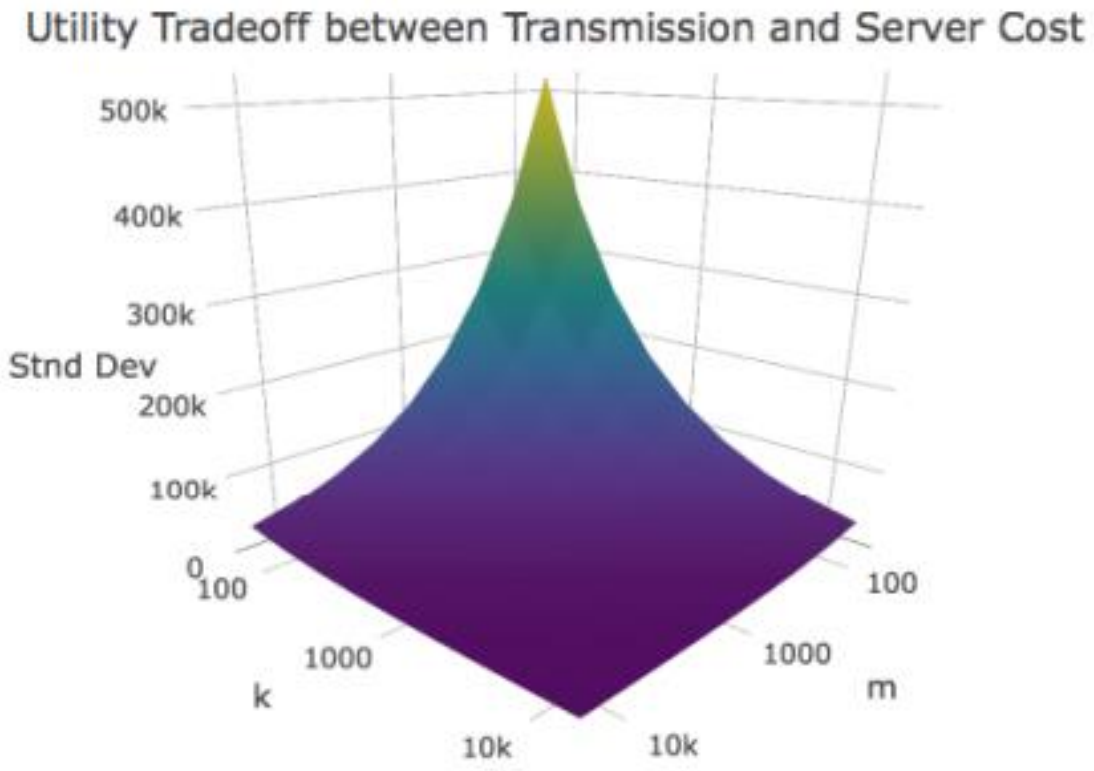
局所差分プライバシメカニズムの実用

頻度分析メカニズムの有用性１

・証明は省略するが、以下が成り立つ

$-\mathbb{E}\left[\tilde{f}(d)\right] = f(d)$

$-Var\left[\tilde{f}(d)\right] \leq \left(\frac{m}{m-1}\right)^2 \cdot \left(\left(\frac{e^{\epsilon/2}+1}{e^{\epsilon/2}-1}\right)^2 + \frac{1}{m} + \frac{\sum_{d' \in \mathcal{D}} f(d')^2}{n \cdot k \cdot m}\right) \cdot n$



頻度分析メカニズムの有用性２

・mはユーザの帯域幅、kはサーバの計算量に関係する

　−m, kが大きいほど頻度分析の精度は高くなる

　−mが大きいとユーザの通信量が大きくなり、kが大きいとサーバでの計算時間が長くなる

・$\epsilon = 2, n = 100{,}000{,}000$において、頻度の標準偏差と$k, m$の関係を実験的に評価

局所差分プライバシメカニズムの実用

課題点

・顔文字などの事前登録されたデータにしか対応できない

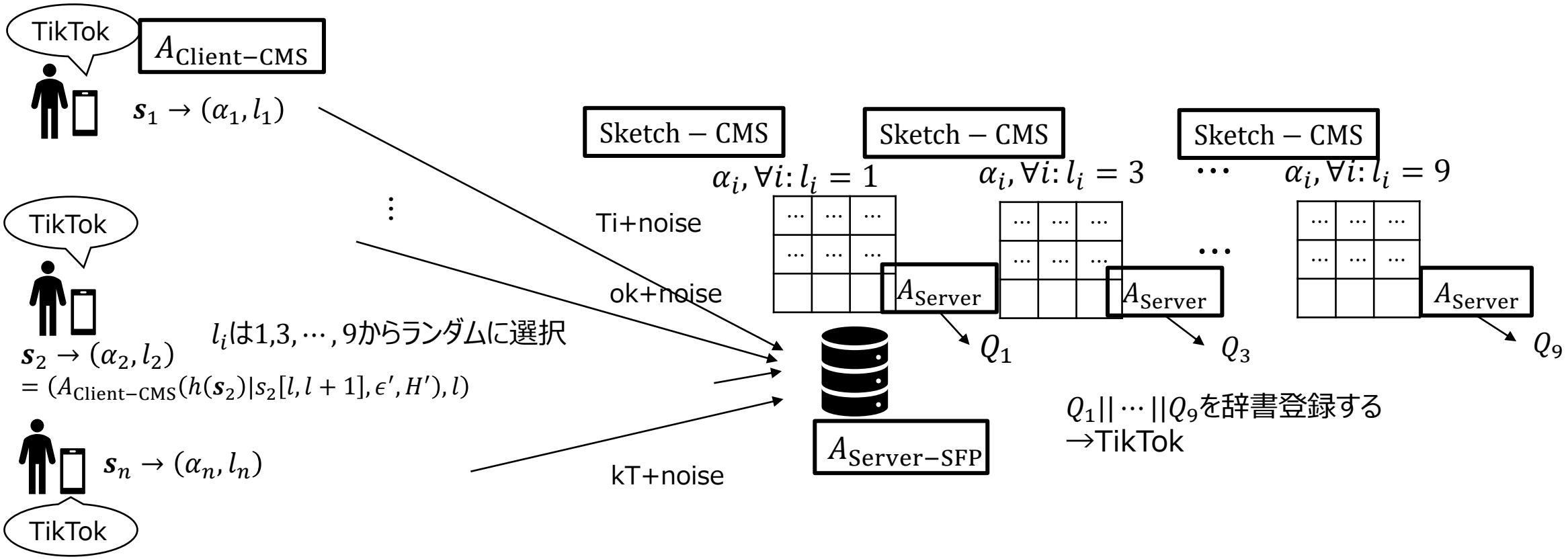
ーアルファベットからなるk字の文字列の場合、サーバ側は26^k回ループが発生し、またかなりの誤検知が発生する

局所差分プライバシメカニズムの実用

<u>改良メカニズム</u>

・事前登録されたデータだけでなく、任意の文字列に対しても頻度分析が可能なメカニズムを提案

　ー入力文字列に対して辞書登録を行うことが可能



局所差分プライバシメカニズムの実用

メカニズムの適用先

・Webサイト上に表示される動画の音声のon, offの判定（$m = 1024, k = 65536, \epsilon = 8$）

ーユーザが音声の自動再生がoffの動画をonにした、あるいは音声の自動再生がonの動画をそのままにしていた場合、
そのサイトでは音声の自動再生をonにすべきと判定

ー（ハッシュ関数を用いて1024bitにマッピングした）ドメイン単位でアクセスURLの収集を行う

・新語の発見と単語の自動修正の改善（$\mathrm{m} = 1024, k = 2048, \epsilon = 6$）

ー最新の単語トレンド辞書を作成し、予測変換候補にする

ーwyd (What you doing?)などの略語やMayweatherなどのトレンド単語の辞書作成が可能となり、予測変換の候補とする

ーlovやknoなど、よく起こる単語の誤字のカテゴリ化が可能となり、自動修正の候補とする

・地域ごとの利用回数の多い絵文字の発見（$m = 1024, k = 65536, \epsilon = 4$）

ー地域ごとに使われる絵文字を分析し、
QuickTypeの改善が可能となる

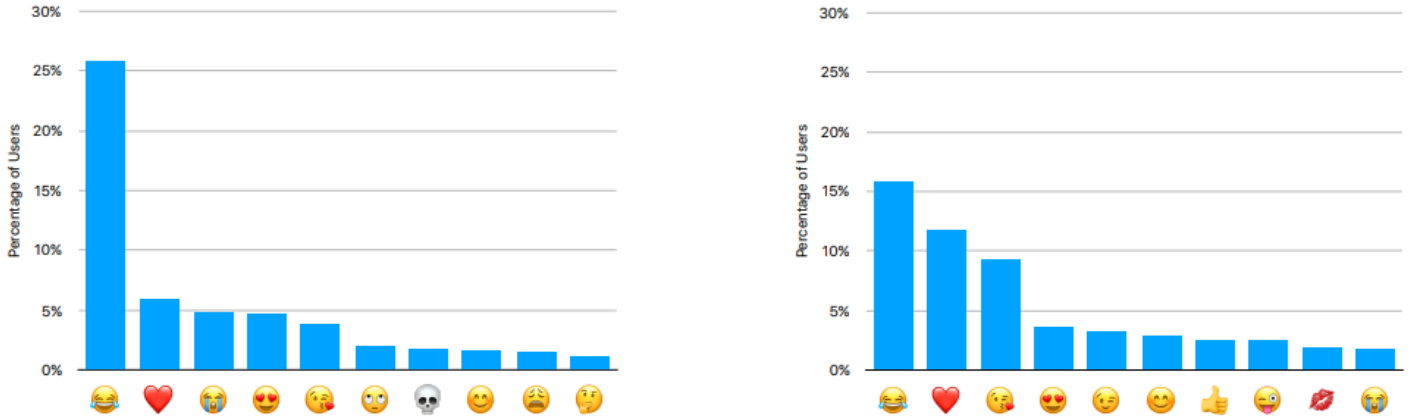
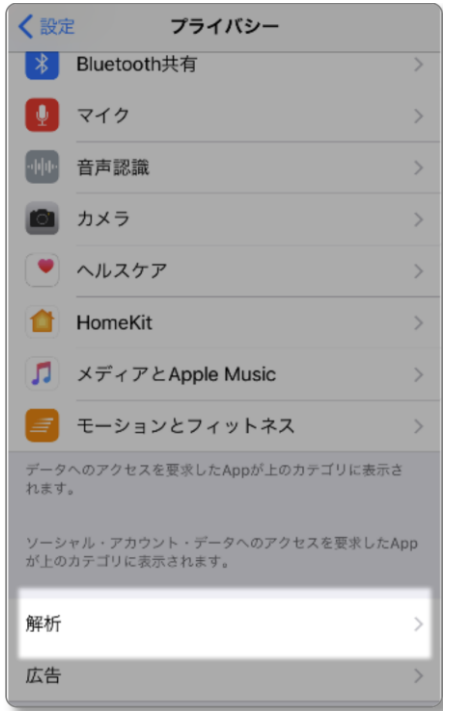


Figure 5: Emojis in Different Keyboard Locales

局所差分プライバシメカニズムの実用



演習

演習課題

設定

・A銀行は保有する顧客情報を匿名加工情報に加工し、B社に販売したいと考えています

ーA銀行は顧客の定期預金のキャンペーン施策のために作成したデータをB社に提供予定です

ーB社は手広く様々なターゲティング広告を行う企業で、適切な広告配信のためにパーソナルデータを必要としています

・匿名化のノウハウがないA銀行はプライバシの専門家（講義参加者）に匿名加工メカニズム開発の依頼を検討しています

・A銀行は皆さんに顧客情報のサンプルデータを提供し、1ヶ月後（6/17）にコンペを開催予定です

ー想定されるリスクや有用性維持の方法、実際の匿名化メカニズムなどの提案を受けて、どのチームに開発を依頼するかを決定する予定です

演習課題

<u>演習の流れ</u>

・本日～5/31：匿名化フェーズ

－各チームXにそれぞれ異なる100レコードのデータセット（teamX_original.csv）を配布

－各チームはデータセットに対して匿名化処理を行って提出

・提出するデータは、**匿名化データセット（teamX_anonymized.csv）**

・6/1～6/11：攻撃フェーズ

－各チームXにチームYの匿名化データセット（teamY_anonymized.csv）と、それぞれ異なる200レコードのデータセット（teamY_candidate.csv）（うち100レコードはteamY_original.csvのレコード）を配布

－各チームはデータセットに対してレコードの再識別攻撃を行って提出

・提出するデータは、**対応表（teamX_attack_teamY.csv）**

・6/17：発表

－各チームは15分間の発表（発表10分、質疑5分）を行う

・匿名化の方針（実装できなくてもよい）：想定リスク、有用性担保 etc.

・攻撃の方針（実装できなくてもよい）：どういった情報が取れるか、他チームの弱点 etc.

・実際の匿名化、攻撃のアルゴリズムの説明

・アピールポイントなど

演習課題

6/1より攻撃フェーズがオープンとなります

・攻撃フェーズに先立って、各チームの元レコード（teamX_original.csv）を含む200レコードのデータセット（teamX_candidate.csv）を公開します

<u>設定</u>

・A銀行との打ち合わせ時に、A銀行のデータ管理者から次のような質問を受けました

「実は当行では、実際の顧客データ（teamX_original.csv）にダミーデータを追加して保管しているデータセット（teamX_candidate.csv）があります。最悪ケースを考えて、情報漏洩によりこのデータセット（teamX_candidate.csv）がB社に渡ったとしても、顧客のプライバシ情報を守りたいです。他の専門家チームからも匿名化データのサンプルを受け取ったのですが、どう考えたらいいものでしょうか」

・少なくとも加工データから元データが分かること（連結）はリスクであるため、匿名化データの連結に対する耐性を評価する必要があります

ー各チームで他の専門家チームの匿名化データへの連結（再識別）攻撃を行い、対応表（teamX_attack_teamY.csv）を提出してください

・またA銀行のデータ管理者の質問を受け、匿名化の方針や匿名化アルゴリズムを追加検討しても構いません

演習課題

各チーム発表