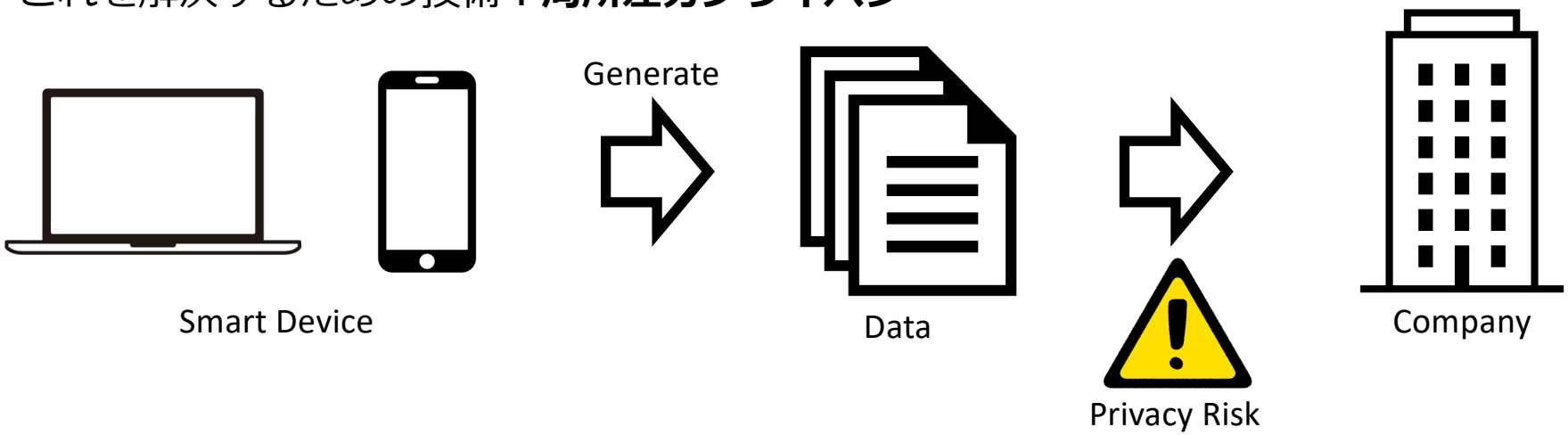
Key-ValueデータのLDPプロトコルPCKVの推定値操作攻撃の提案と評価

谷口輝海, 菊池浩明

明治大学

背景

- 企業は、ユーザデバイス上で生じるデータをサービス向上のために活用したい。
- データをそのまま収集するとプライバシが侵害される恐れ。
- ユーザはプライバシを守りたい。
- これを解決するための技術：**局所差分プライバシ**



Key-Valueデータの LDP方式

PrivKV [Ye, et al. 2019]

ユーザの Key-Value 集合 S_i

key k	value v
ワイルドスピード	*
007	4
タイタニック	*
変態仮面	5
君の名は。	*

*は評価なし

$\frac{1}{5}$ kから一様にサンプリング

k	v
007	4

PrivKV

送信データ

j	k_j^*	v_j^+
007	1	-1

データが疎だと評価したデータが送信されない

$j = 007$

v_j : vを正規化 → v_jを2値化 → v_j'をランダマイズ $RR(v_j')$ → k_jをランダマイズ $RR(k_j)$

$< 007, 4 > \rightarrow < 1, 0.5 > \rightarrow < 1, 1 > \rightarrow < 1, -1 > \rightarrow < 1, -1 >$

PCKV [Gu, et al. 2020]

ユーザの Key-Value 集合 S_i

key k	value v
ワイルドスピード	*
007	4
タイタニック	*
変態仮面	5
君の名は。	*

Padding data

key k	value v
ワイルドスピード	*
007	4
タイタニック	*
変態仮面	5
君の名は。	*
DummyKey1	0
DummyKey2	0
DummyKey3	0

padding length ℓ

S_iからサンプリングする確率 $\frac{|S_i|}{\max(|S_i|, \ell)}$

$\frac{2}{3}$

k	v
007	4

$\frac{1}{3}$

PCKV

送信データ

PCKV-UE	$(0, -1, 0, 0, 0, 1, 0, 0)$
PCKV-GRR	(君の名は。, -1)

vを正規化＆2値化 →

PCKV-UE

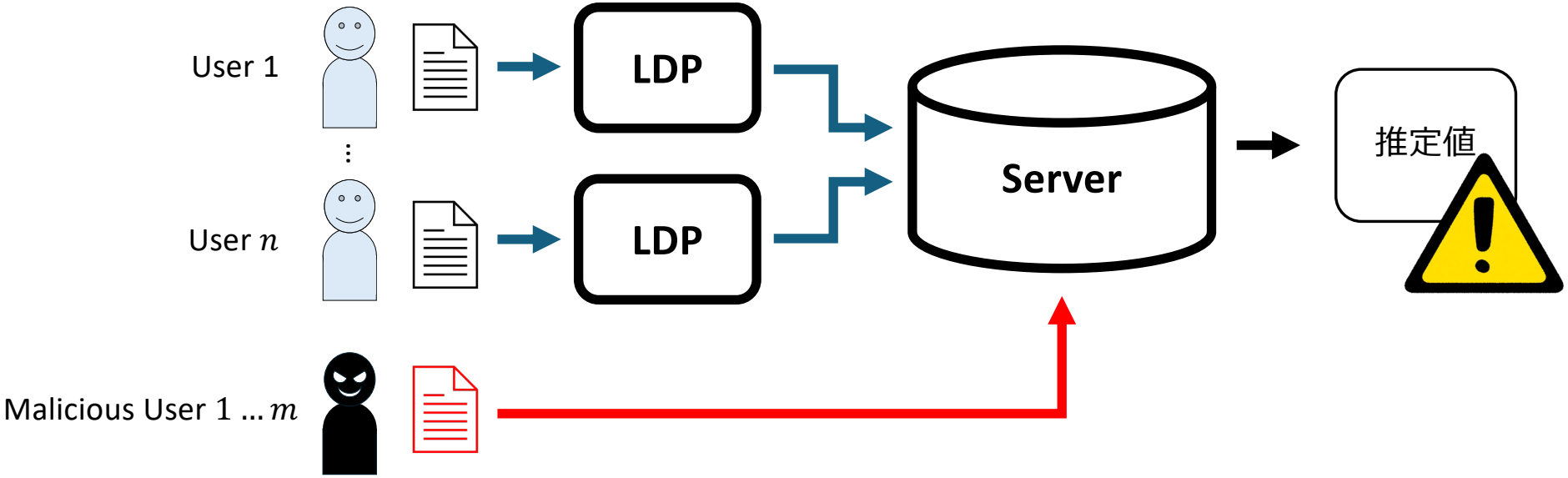
$\boldsymbol{y} = (0, 0, 0, 0, 0, 0, 0, 0)$ → 007に対応する箇所をv → $\boldsymbol{y} = (0, 1, 0, 0, 0, 0, 0, 0)$ → 各値をランダム化 → $\boldsymbol{y} = (0, -1, 0, 0, 0, 1, 0, 0)$

PCKV-GRR

$< k, v > = < 007, 1 >$ → keyとvalueを同時にランダム化 → $< k, v > = <$ 君の名は。, $-1 >$

ポイズニング攻撃

- 送信データを細工して推定値を操作する



代表的な攻撃　特定のカテゴリの頻度を増加させる。[Cao, et al. 2021]

特定のkeyの頻度と平均を増加させる。[Wu, et al. 2022]

平均値を攻撃者が意図した値に近づける。[Li, et al. 2023]

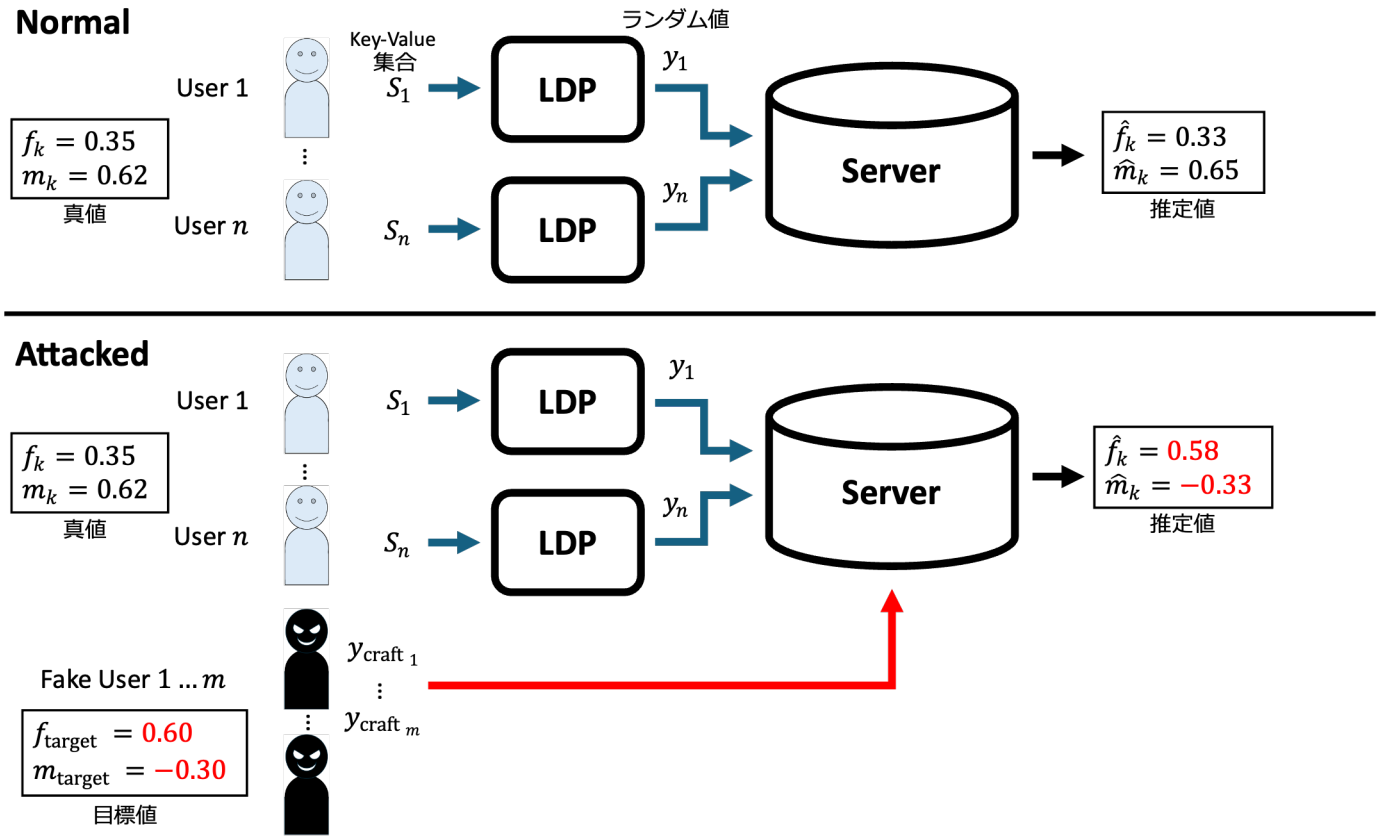
本研究の貢献

	攻撃目的＼対象統計量	分散	平均値	頻度
利得最大化攻撃	真値から増加させる		Wu et al 2022	Cao et al 2021
推定値操作攻撃	攻撃者の目標値にする	Li et al 2023	本研究	

研究目的

- Key-Valueデータのための**LDP**方式に対し、keyの推定頻度と推定平均を攻撃者の意図した値に近づけるポイズニング攻撃を提案する。
- 提案攻撃を**PrivKV, PCKV-UE, PCKV-GRR**に適用し、攻撃精度を比較評価する。

研究概要　推定値操作攻撃



問題設定

- 正規ユーザn人の頻度f_kと平均μ_kとする。ここで、標的となる頻度$f_{k,t}$と平均$\mu_{k,t}$にするために、m人の偽ユーザを挿入する。このとき、m人の偽ユーザの送信データを決定したい。
- 最終的に達成したい式

$$\begin{cases} E[\hat{f}_k] = f_{k,t} \\ E[\hat{\mu}_k] = \mu_{k,t} \end{cases}$$

提案方式

PCKV-UEに対する攻撃の適用例

PCKV-UEにおける頻度推定

$$E\left[\hat{f}_k\right] = E\left[\frac{\frac{n_1^k + n_{-1}^k}{n} - b}{a-b}\ell\right]$$

PCKV-UEにおける平均値推定

$$E\left[\hat{\mu}_k\right] = E\left[\frac{\ell\left(n_1^k - n_{-1}^k\right)}{a(2p-1)n\hat{f}_k}\right]$$

m人の偽ユーザを挿入(m_1^k, m_{-1}^k, m_0^k)

$$E\left[\hat{f}_k\right] = E\left[\frac{\frac{n_1^k + n_{-1}^k + m_1^k + m_{-1}^k}{n+m} - b}{a-b}\ell\right] = f_{k,t}$$

$$E\left[\hat{\mu}_k\right] = E\left[\frac{\ell\left(n_1^k - n_{-1}^k + m_1^k - m_{-1}^k\right)}{a(2p-1)(n+m)\hat{f}_k}\right] = m_{k,t}$$

攻撃者の目標値

攻撃者が推定する
真の頻度と平均

$$\begin{cases} m_1^k + m_{-1}^k = \dfrac{a-b}{\ell}\left((m+n)f_{k,t} - f_k\right) + mb \\ m_1^k - m_{-1}^k = \dfrac{a(2p-1)}{\ell}\left((m+n)f_{k,t}\mu_{k,t} - nf_k\mu_k\right) \\ m_1^k + m_{-1}^k + m_0^k = m \\ 0 \le m_1^k, m_{-1}^k, m_0^k \le m \end{cases}$$

Research Question

RQ 1. 与えられた目標値を達成するための不正者数mはどれくらいか？

RQ 2. 3つのLDP方式で、推定値操作攻撃に対して最も安全な方式はどれか？

実験・評価指標

- 推定精度評価：推定値$\hat{f}_k, \hat{m}_k$と真値f_k, m_kの平均二乗誤差を計算

$$MSE_f = \frac{1}{|K|}\sum_{k\in K}\left(\hat{f}_k - f_k\right)^2, MSE_\mu = \frac{1}{|K|}\sum_{k\in K}(\hat{\mu}_k - \mu_k)^2$$

- 攻撃精度評価：推定値$\hat{f}_k, \hat{m}_k$と目標値$f_{k,t}, m_{k,t}$の平均二乗誤差を計算

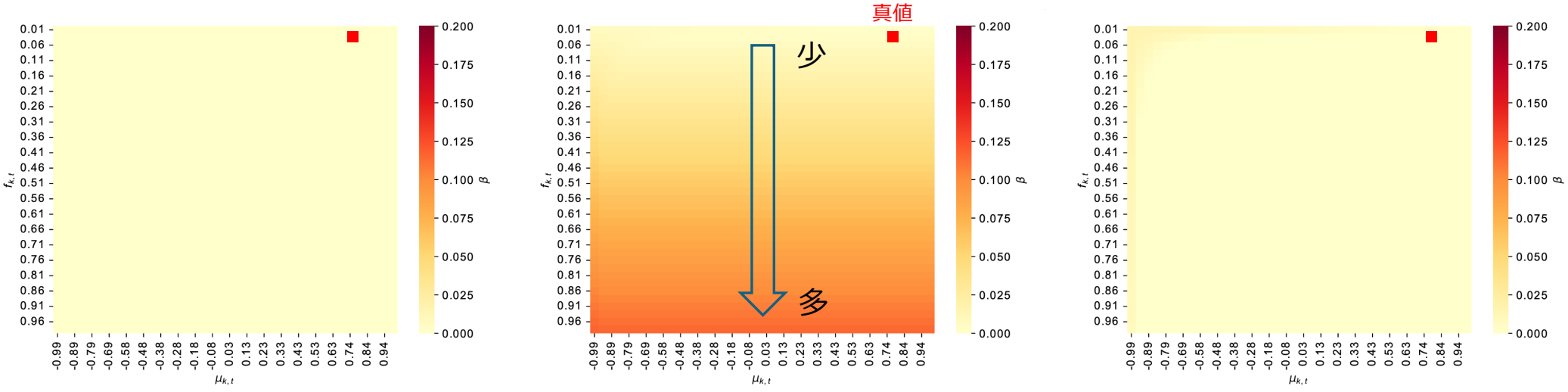
$$MSE_{f_k} = \frac{1}{N}\sum_{i\in[N]}\left(\hat{f}_k^{(i)} - f_{k,t}\right)^2, MSE_{\mu_k} = \frac{1}{N}\sum_{i\in[N]}\left(\hat{\mu}_k^{(i)} - \mu_{k,t}\right)^2$$

- 使用データセット：Clothing Dataset

User 数	key 数	record 数
105,508	5850	192,198

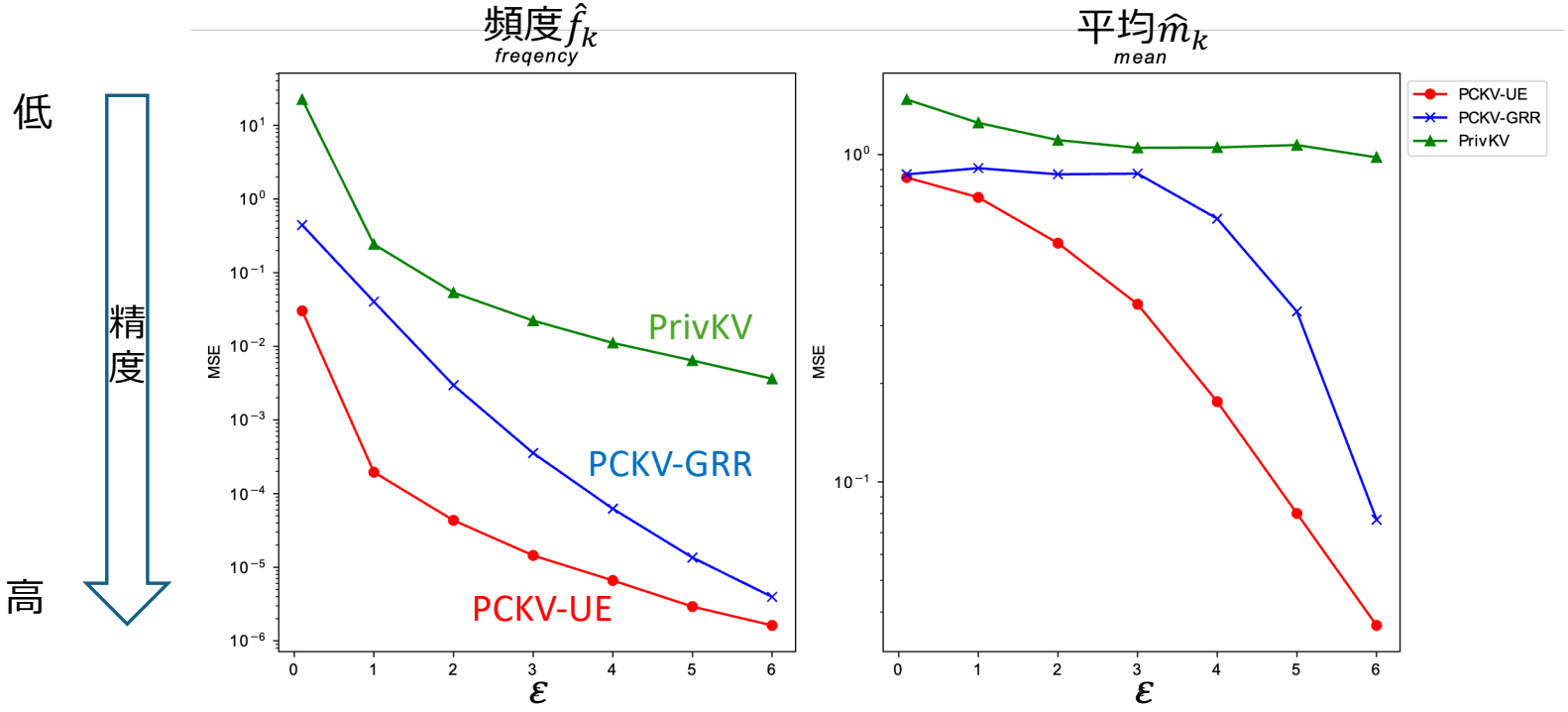
方程式が解を持つためのmの必要条件

- 前述の連立式を満たす必要十分なmの範囲を、目標値（$f_{k,t}, m_{k,t}$）を変更して計算
- mの下限を用いてβ（偽ユーザの割合）を計算したときのヒートマップ



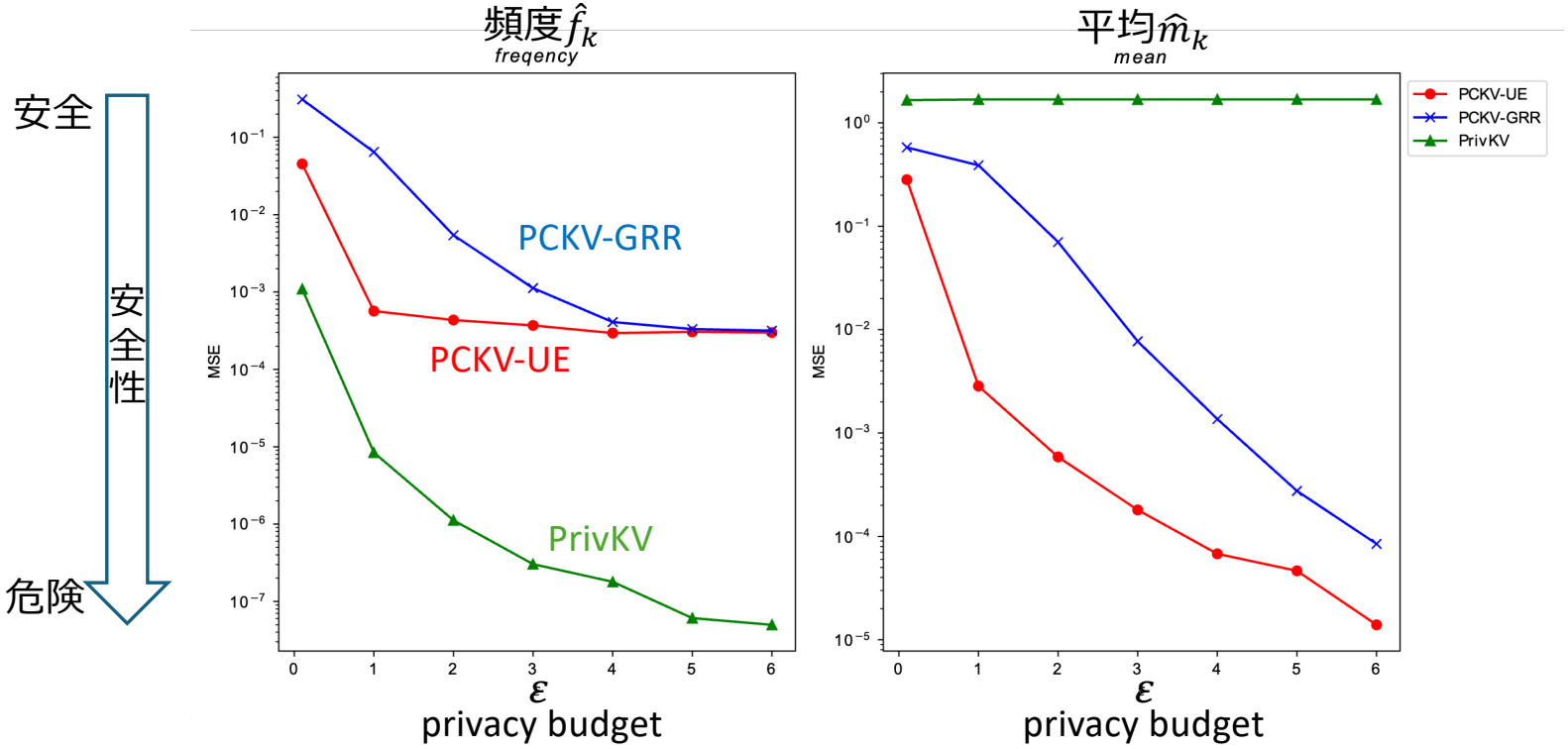
実験結果① 推定精度

- εによる3方式の推定精度推移



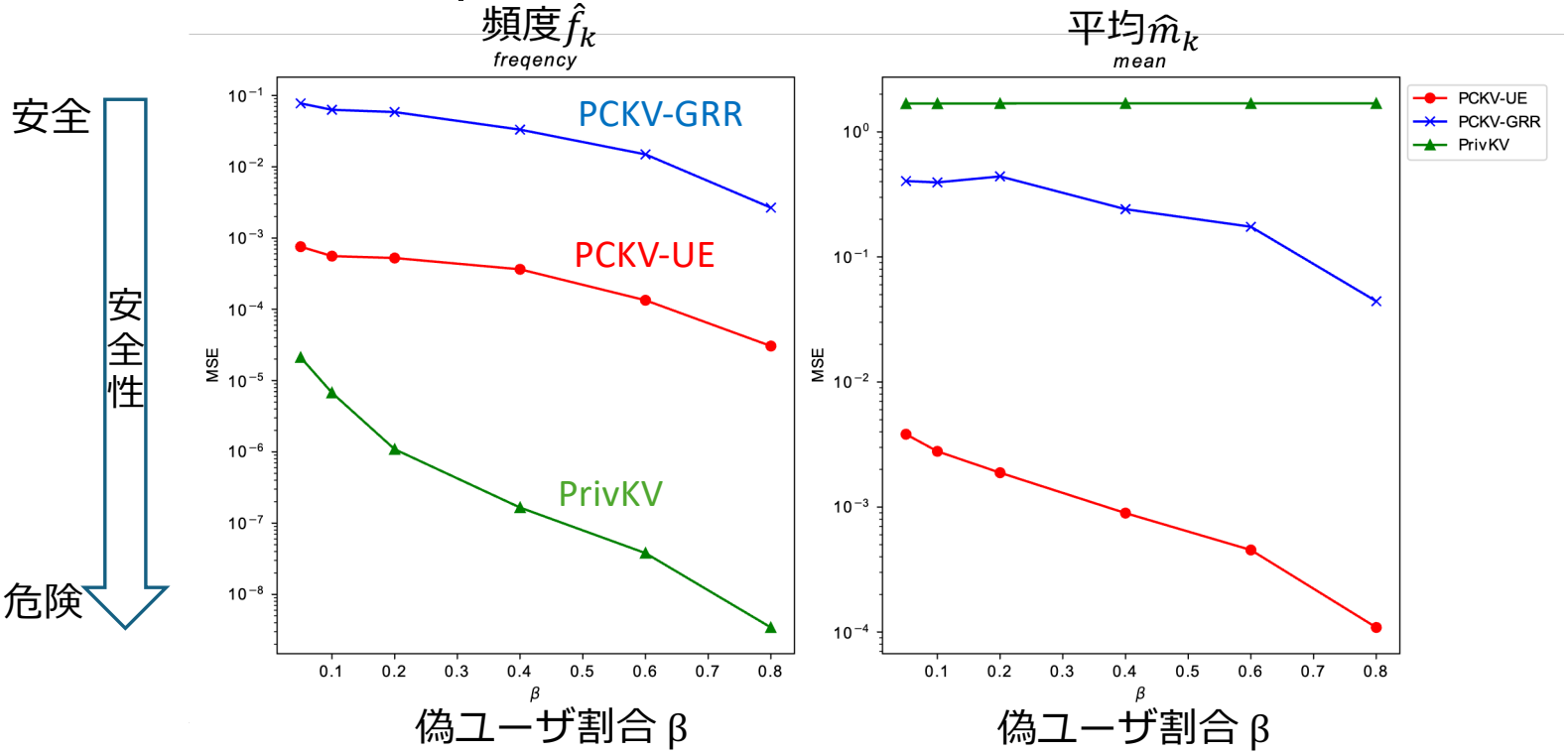
実験結果② 攻撃精度（ε）

- ε による3方式の攻撃精度推移



実験結果③ 攻撃精度（β）

- 偽ユーザの割合 β による攻撃精度推移



考察

- PrivKVが平均値への攻撃に対して最も安全な理由
  - εを大きくしても（攻撃をしなかった場合の）平均値の推定精度は大きく改善せず、ランダム化のノイズにより推定値の操作が困難になるため
- 頻度の操作結果
  - keyの摂動方式における値域の違いが影響している
  - PrivKV : 2値のRR
  - PCKV-GRR : K値のRR
  - PCKV-UE : K値のUE

まとめ

- Key-Valueデータ収集のためのLDP方式PrivKV, PCKV-UE, PCKV-GRRに対して、keyの頻度とvalueの平均を攻撃者の目標値に操作する推定値操作攻撃を提案した
- RQ 1. 与えられた目標値を達成するための不正者数mはどれくらいか？
  - PrivKV, PCKV-GRRでは少数(全体の**1%**未満)で攻撃可、PCKV-UEでは目標値によって最大**10%**程度必要
- RQ 2. 3つのLDP方式で、推定値操作攻撃に対して最も安全な方式はどれか？
  - 総合的に見て**PCKV-GRR**が最も安全。
  - 推定精度ではPCKV-UEに劣り、平均値への攻撃ではPrivKVが最も安全であるが、PrivKVは平均値の推定精度が低い。
- 今後
  - 攻撃誤差の理論的分析
  - 防御手法についての検討