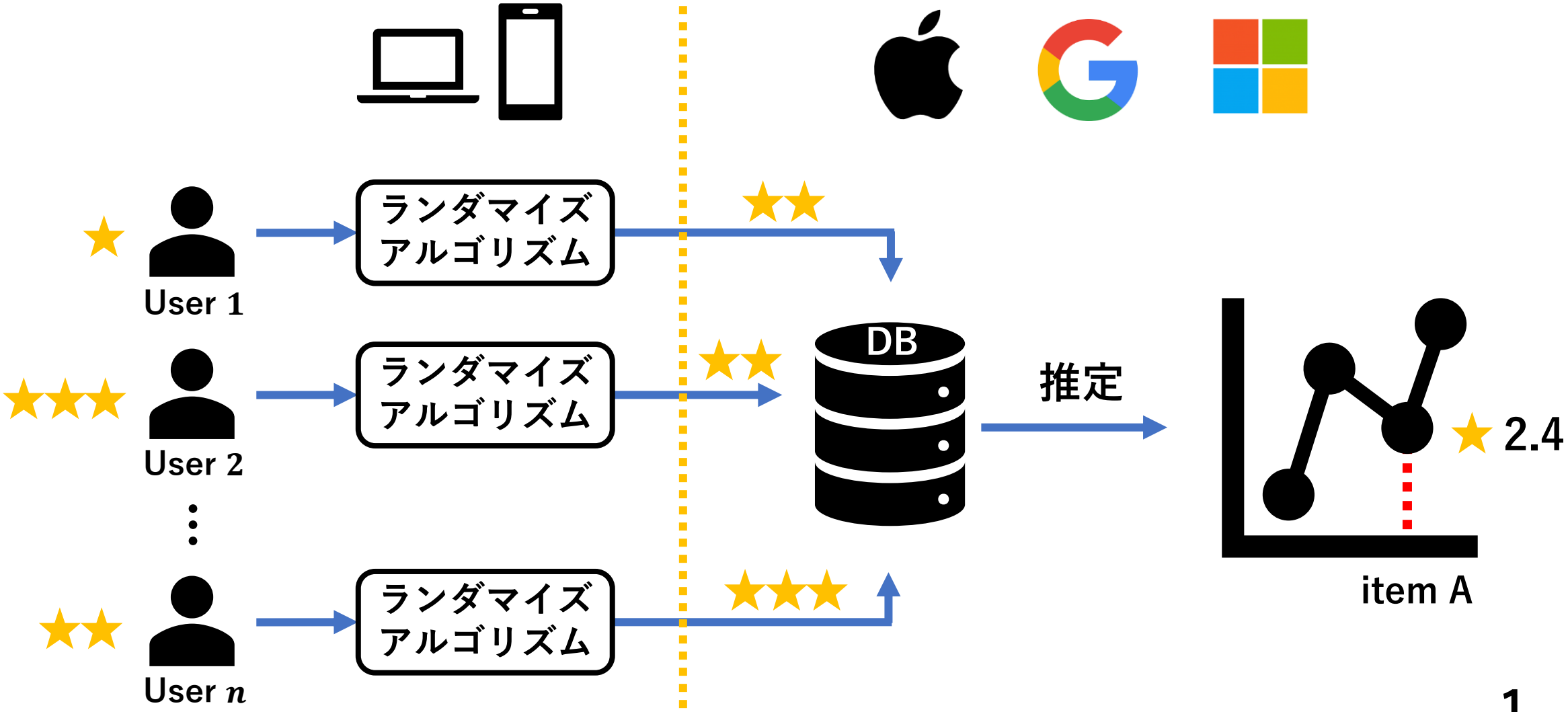
コンピュータセキュリティシンポジウム2022

ポイズニング攻撃に対してロバストなEMアルゴリズムを用いたkey-valueデータにおけるLDPプロトコル

堀込 光，菊池浩明（明治大学），
Chia-Mu Yu（National Yang Ming Chiao Tung Univercity）

局所差分プライバシー(LDP)



PrivKV [Q.Ye, et al., 2019]

key-valeデータ

k	v
千と千尋の神隠し	2
タイタニック	*
アナと雪の女王	*
君の名は。	4
もののけ姫	3

サンプリング $< k, v >$

k	v
君の名は。	4

PrivKV

出力

a	k^* (フラグ)	v^+
君の名は。	1	-1

a = “君の名は。”

VPP (valueのランダマイズ)

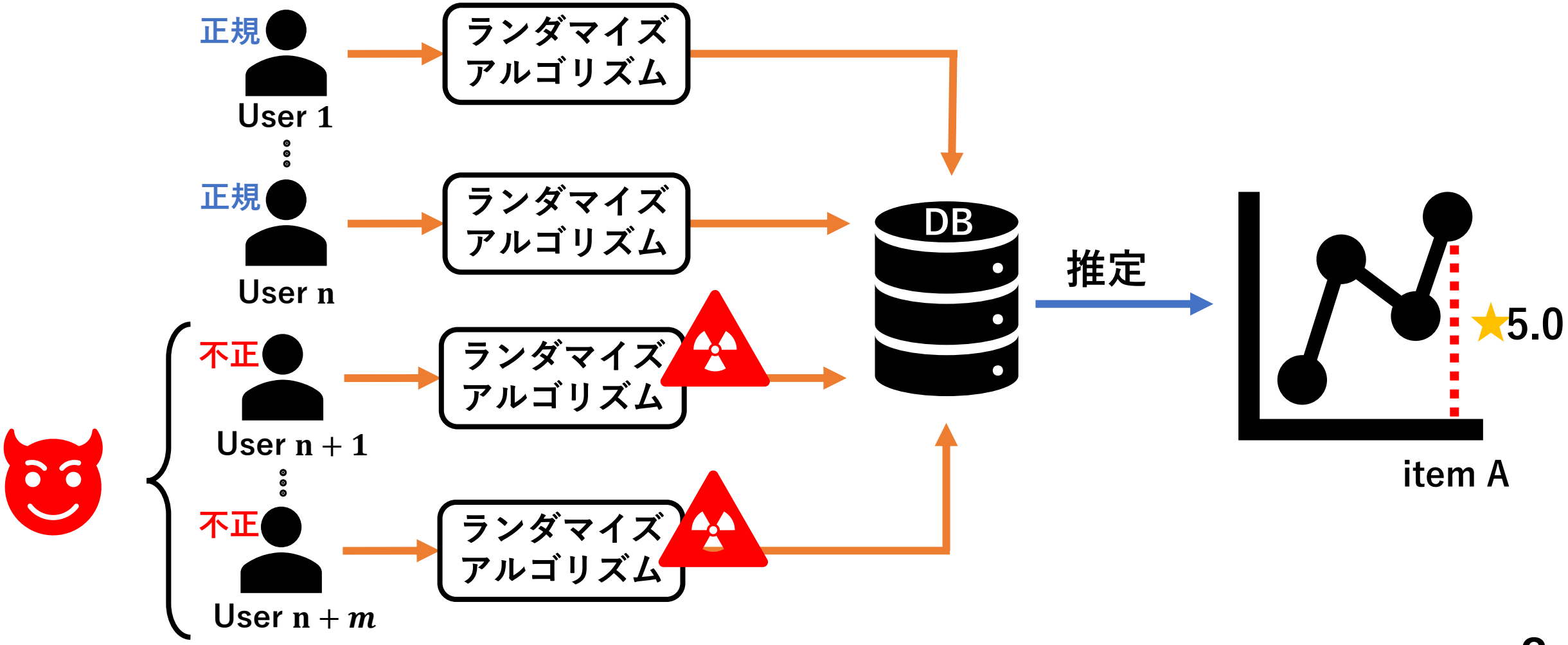
RR(keyのランダマイズ)

$< k'_a, v'_a > \rightarrow v'_a$の2値化 $\rightarrow < k'_a, v^*_a > \rightarrow RR(v^*_a) \rightarrow < k'_a, v^+_a > \rightarrow RR(k'_a) \rightarrow < k^*_a, v^+_a >$

$< 1, 0.5 > \longrightarrow < 1, 1 > \longrightarrow < 1, -1 > \longrightarrow < 1, -1 >$

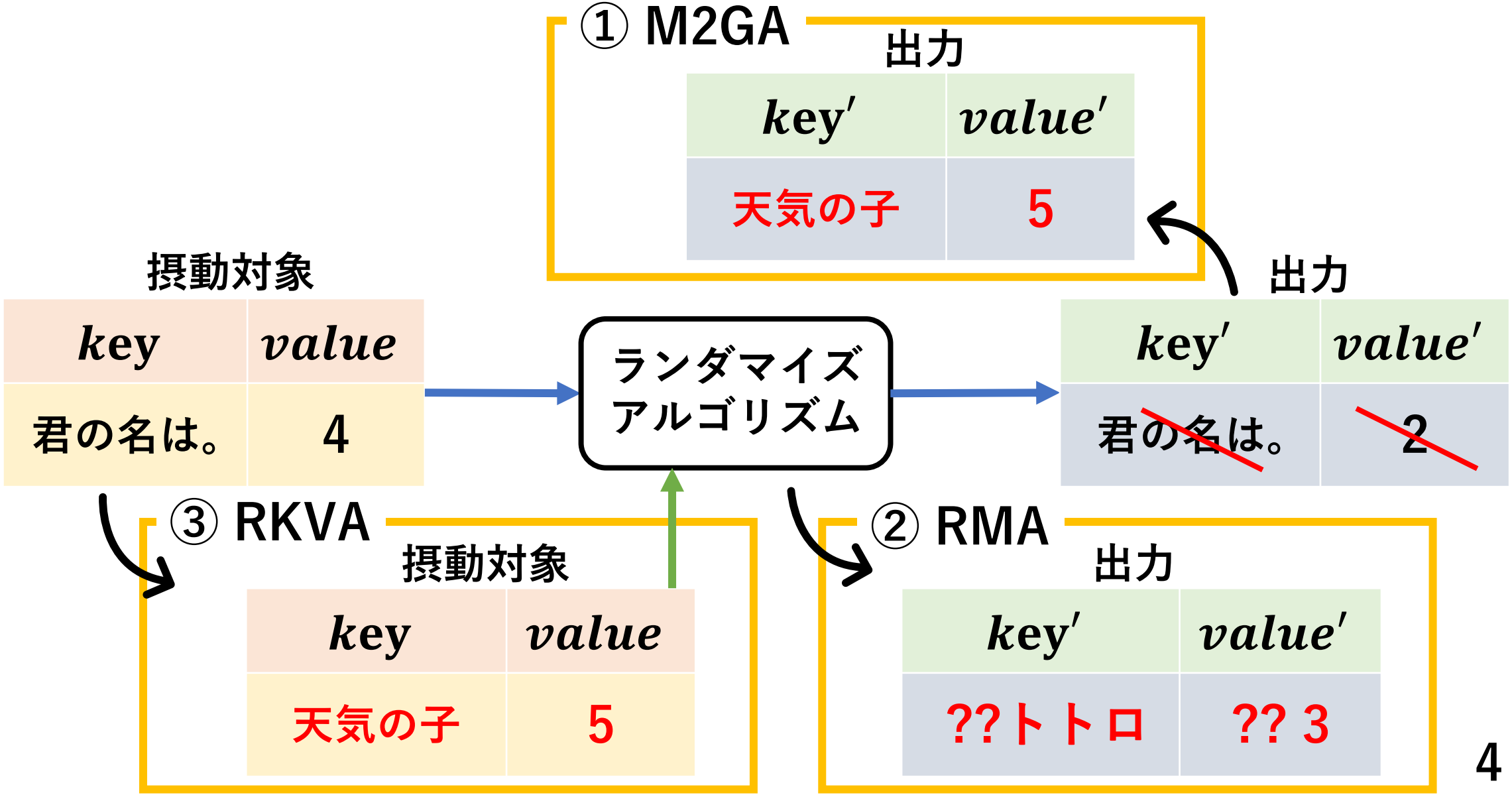
ポイズニング攻撃：攻撃者による推定値操作

攻撃者の目的：特定のアイテムに対する推定値を操作すること



X.Cao, J.Jia, N.Z. Gong, "Data poisoning attacks to local differential privacy protocols", USENIX Security Symposium, pp.947-964, 2021.

3つのポイズニング攻撃手法 [Y.Wu, et al., 2022]



ポイズニング攻撃による強度 [X.cao, et al., 2021]

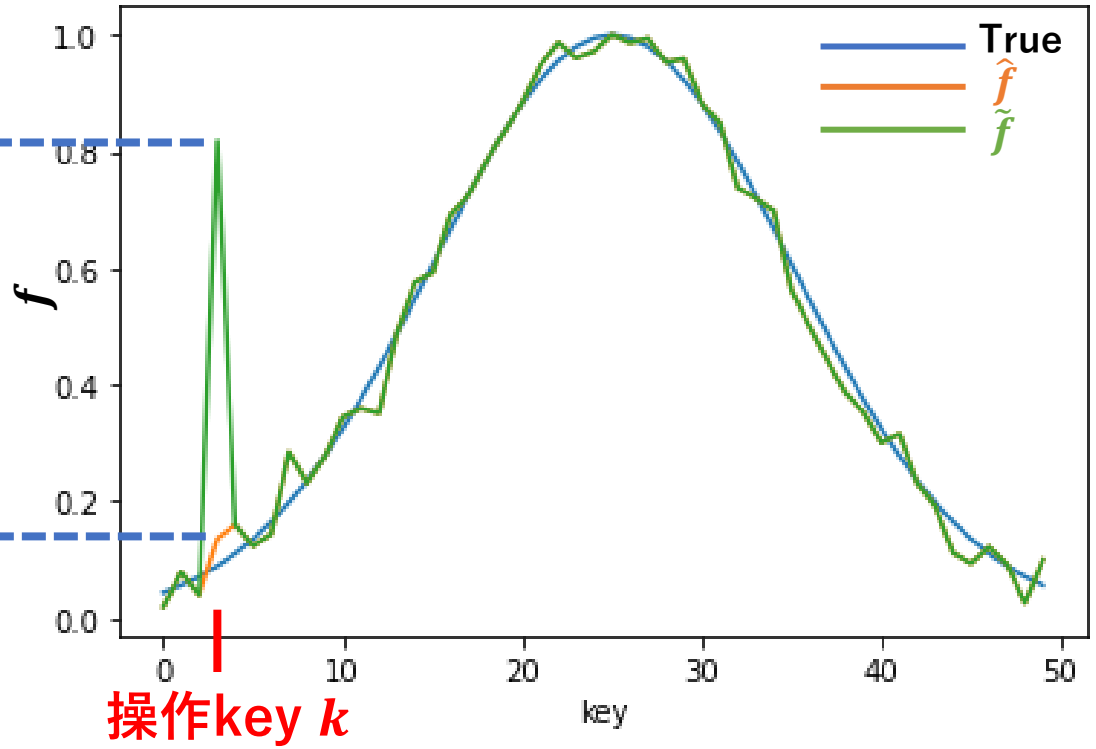
ポイズニング攻撃による推定度数の変化量 $\Delta\hat{f}_k = \tilde{f}_k - \hat{f}_k$

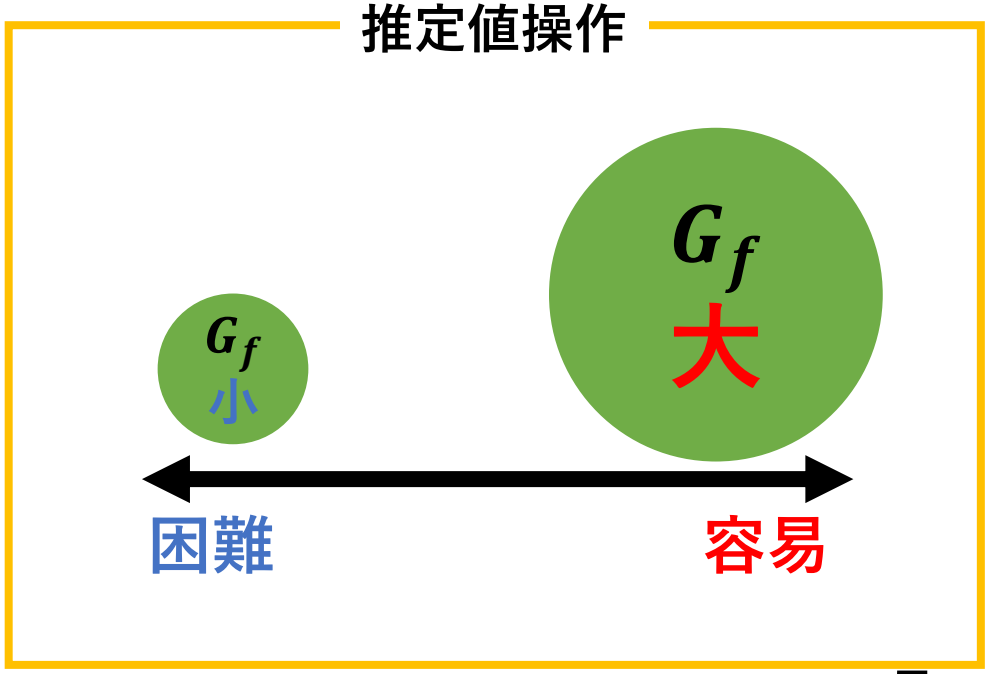
操作対象keyの集合 $T = \{k_1, k_2, \ldots k_r\}$

操作利得 $G_f = \sum_{k \in T} \mathbb{E}[\Delta\hat{f}_k]$

攻撃あり $\tilde{f}_k$

攻撃なし $\hat{f}_k$



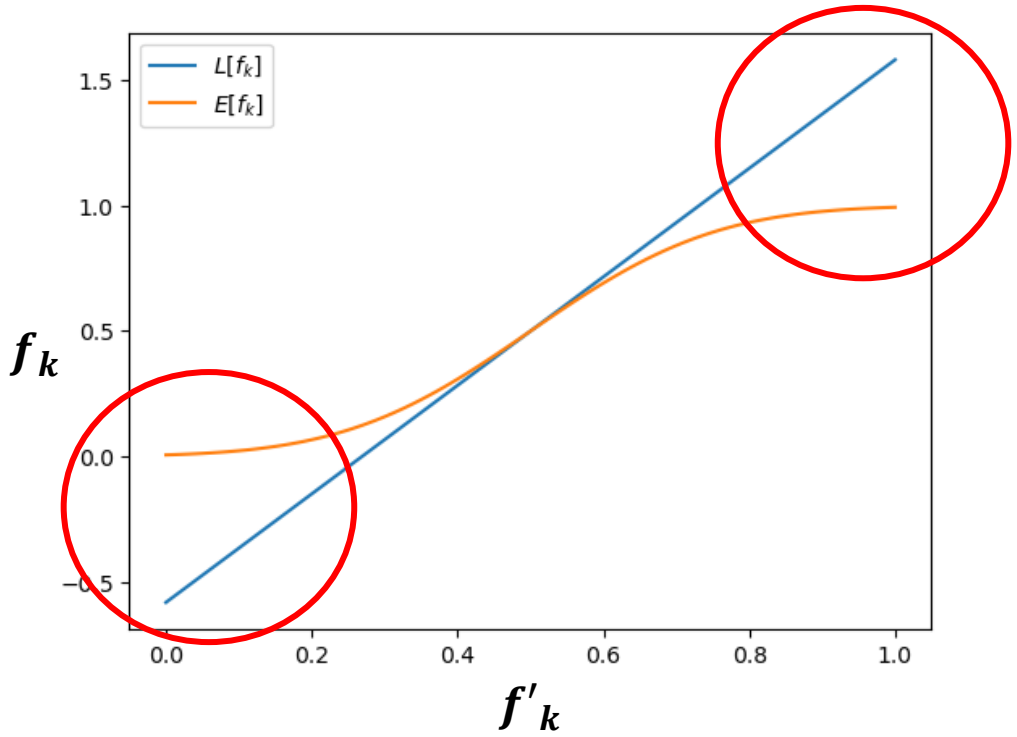


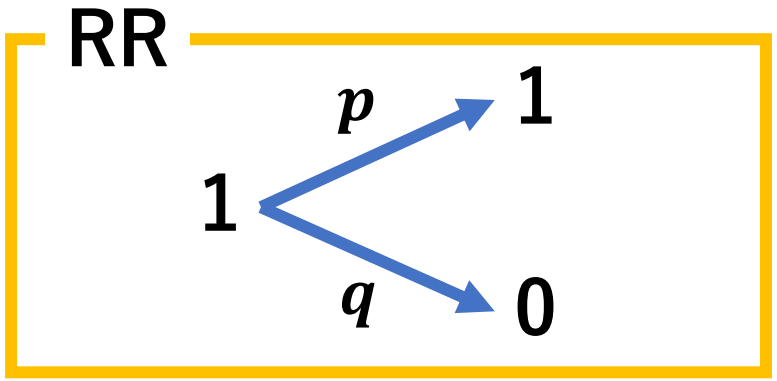
ポイズニング攻撃に対するPrivKVの問題点

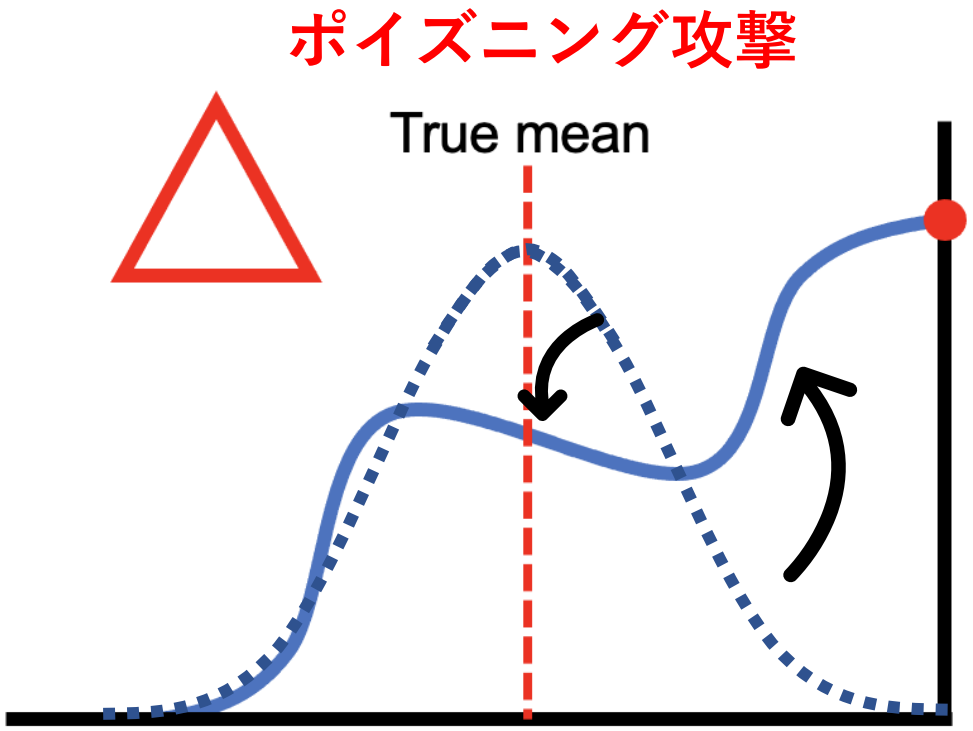
- **PrivKVの推定手法：最尤推定法**

$$f'_k = f_k p + (1 - f_k) q$$

$$L[f_k] = \frac{f'_k - q}{p - q}$$





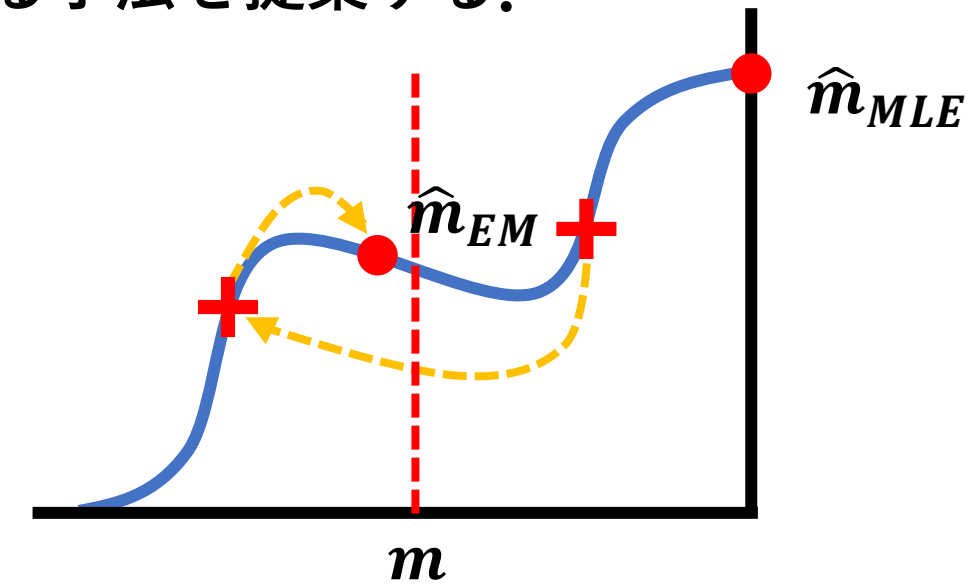


研究概要

	摂動化	推定
PrivKV	VPP+RR	MLE
本提案		EM

- 解決手法

PrivKVにEM(Expectation Maxmization)アルゴリズム[宮川雅巳,1987]を適用し，推定する手法を提案する．



提案手法
EM(Expectation Maximization)アルゴリズムの適用

摂動の流れ

a = “君の名は。”　　VPP (valueのランダマイズ)　　keyのランダマイズ

$< k'_a, v'_a > \rightarrow v'_a$の2値化 $\rightarrow < k'_a, v^*_a > \rightarrow RR(v^*_a) \rightarrow < k'_a, v^+_a > \rightarrow RR(k'_a) \rightarrow < k^*_a, v^+_a >$

$v'_a = [-1, +1]$　　$v^*_a = \{-1, 1\}$　　$v^+_a = \{-1, 0, 1\}$

事前確率を求める対象の設定

案1　$< k'_a, v'_a > \longrightarrow < k^*_a, v^+_a >$　×

案2　$< k'_a, v^*_a > \longrightarrow < k^*_a, v^+_a >$　○

$< k^*_a, v^+_a > = \{< 1, 1 >, < 1, -1 >, < 0, 0 >\}$

$< k'_a, v^*_a > = \{< 1, 1 >, < 1, -1 >, < 0, 1 >, < 0, -1 >\}$

・出力 $< k^*_a, v^+_a >$ を用いて $< k'_a, v^*_a >$ の事前確率を推定する。

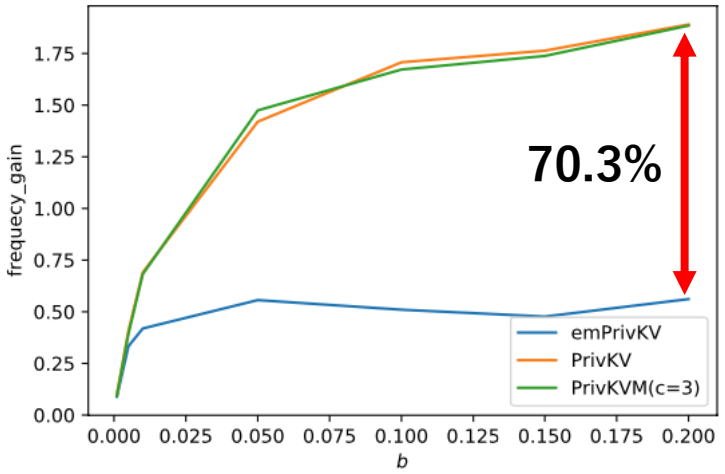
実験

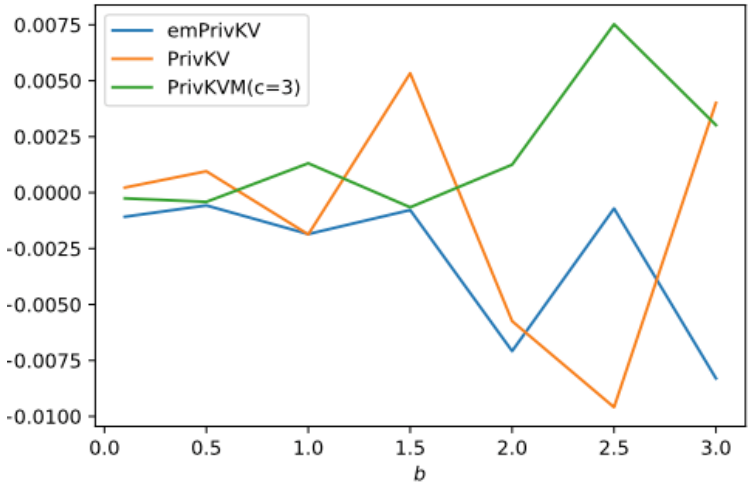
- 目的：提案手法，PrivKV，PrivKVの対話型プロトコルPrivKVM(対話回数=3)について3つのポイズニング攻撃，M2GA, RMA，RKVAに対する頻度利得と値利得を比較する．
- データ：keyとvalueがガウス分布に従う合成データ
- 評価実験
  - 偽ユーザ割合β，安全性指標ε，操作するアイテム数r，ユーザ数nを変化させ，3つのポイズニング攻撃を行い$\Delta\hat{f}$と$\Delta\hat{m}$を算出する．
  - それぞれの攻撃とパラメータで50回の試行を行い$\Delta\hat{f}$と$\Delta\hat{m}$の平均値を頻度利得，値利得とする．
  - 初期値を$\beta = 0.05$，$\varepsilon = 1$，r=1，$n = 10^4$とする．

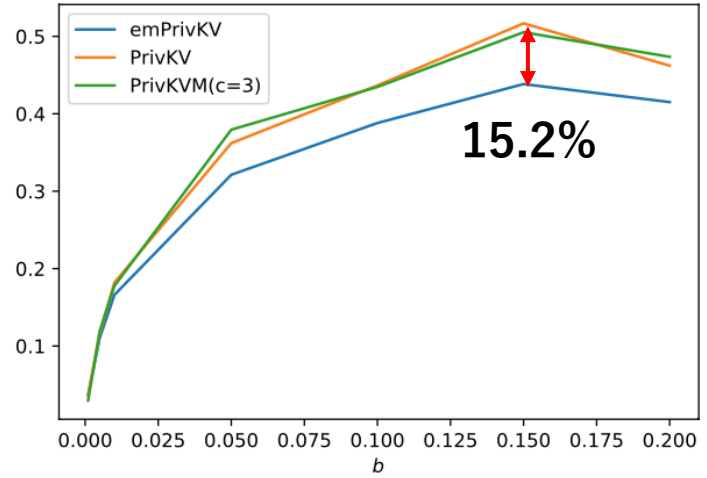
結果：頻度利得

M2GA　RMA　RKVA

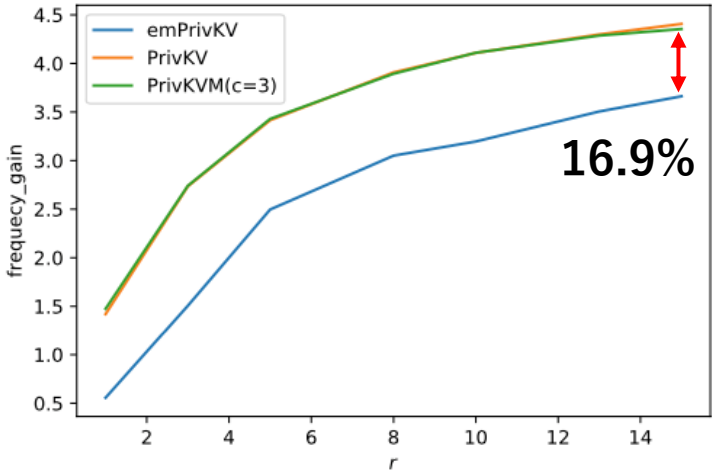
β

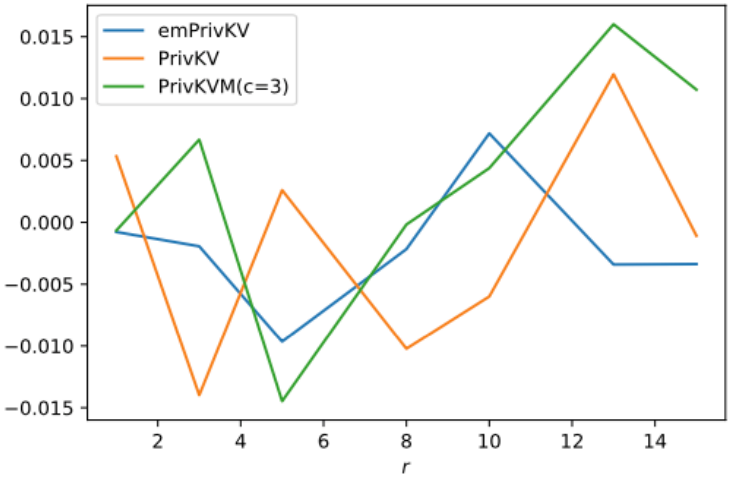


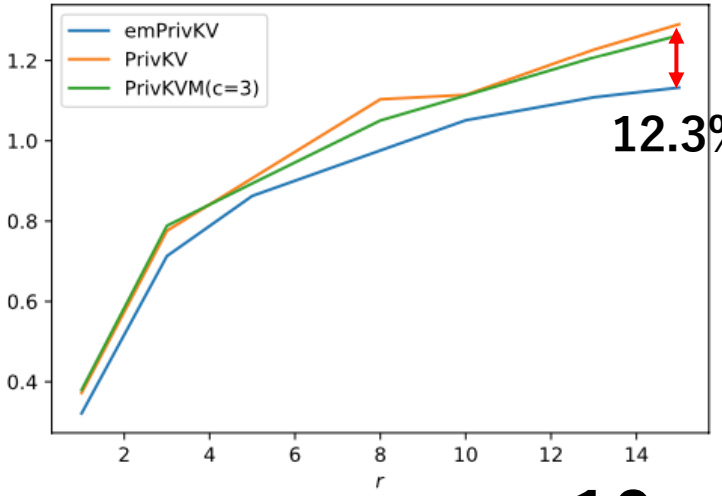




r



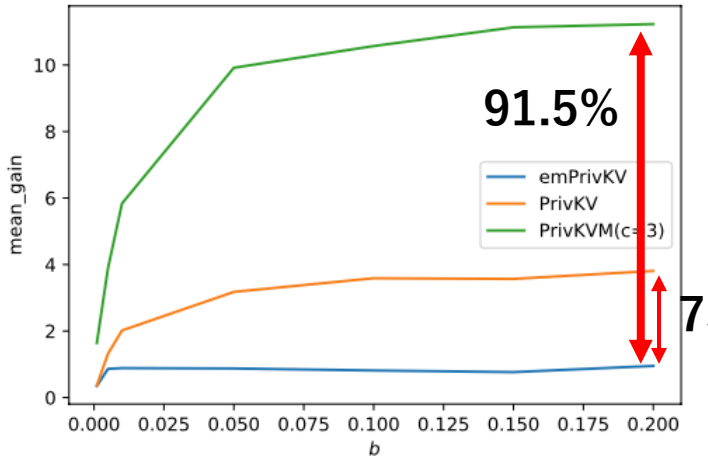


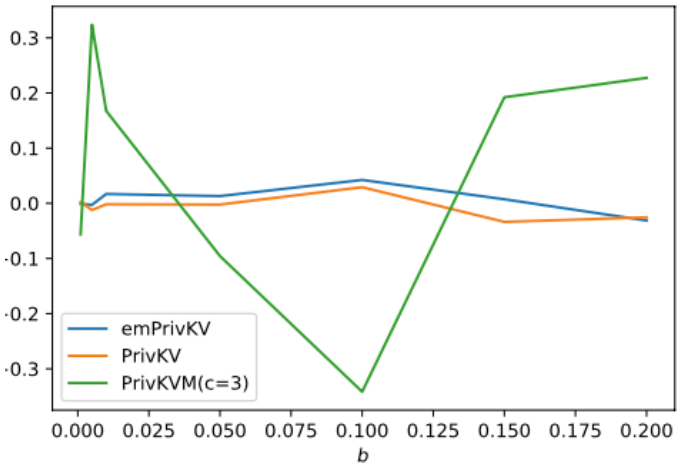


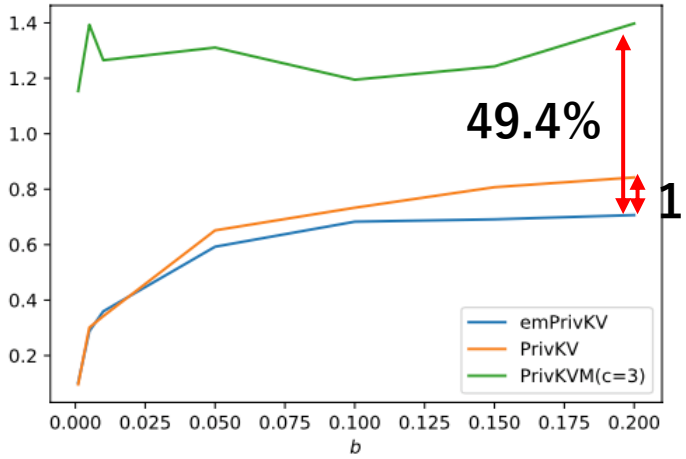
結果 mean gain

M2GA | **RMA** | **RKVA**

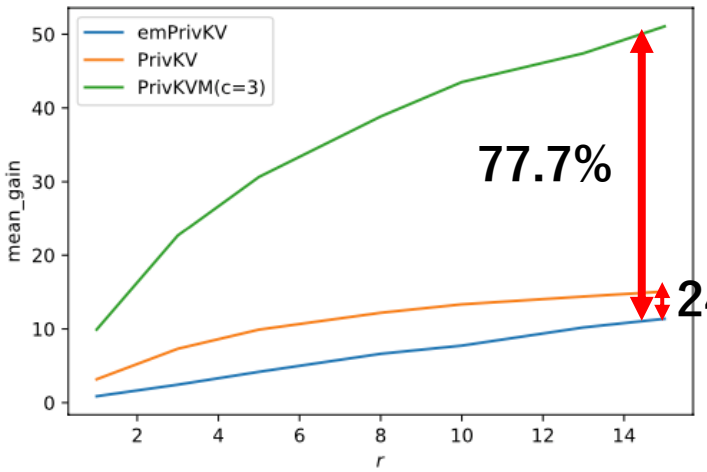
β

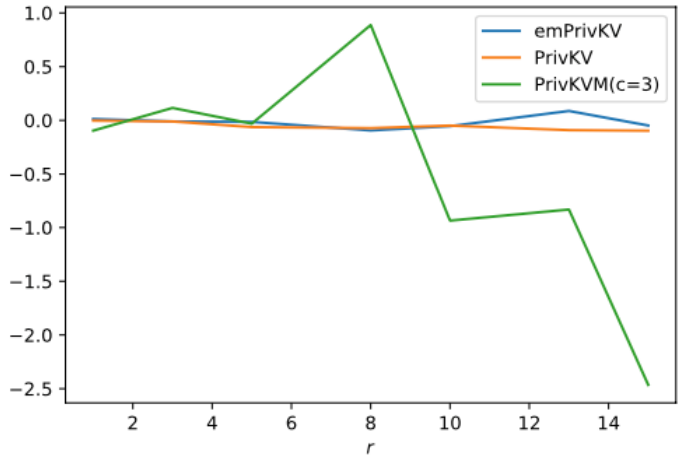


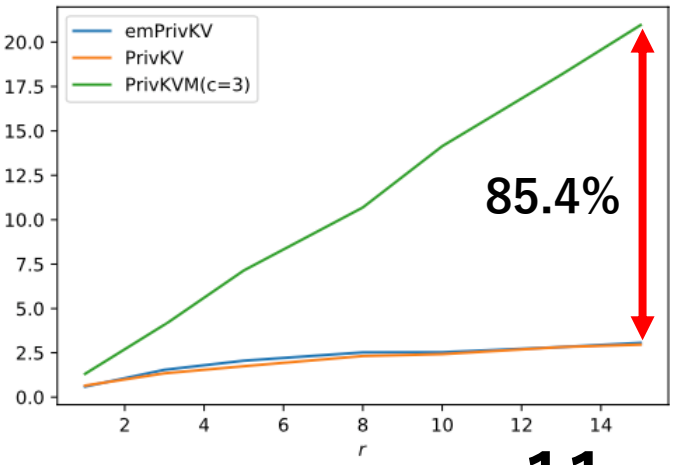




r







まとめ

- 本研究では，LDPプロトコル PrivKVがポイズニング攻撃に対して脆弱であることを示し，EMアルゴリズムを用いた，よりロバストなLDPプロトコルを提案した.

- 3つのポイズニング攻撃に，提案手法が効果的であることを示した．特にM2GAの頻度利得と値利得を小さくする.

- 作成した合成データと比較して，より疎であるビックデータに対しても提案手法は有効であると考えられる.