Session 8
Privacy and Security

- Practical Differential Privacy via Grouping and Smoothing
  - Georgios Kellaris, Stavros Papadopoulos (HKUST) :
- CorrectDB: SQL Engine with Practical Query Authentication
  - Summet Bajaj, Radu Sion (Stony Brook Univ) :
  - こちらは超概要だけ

Practical Differential Privacy via Grouping and Smoothing

- *one-time* publishing of *non-overlapping* countsを対象に差分プライバシを適用する際，前処理としてグルーピングを行うことでutilityを上げる
- *one-time* and *non-overlapping* count

	p1	p2	p3	p4
u1	4	2	0	6
u2	0	0	2	3
u3	3	5	7	2

place	count
p1	7
p2	7
p3	9
p4	11

one-time:1回しかpublishしない
(更新がないDB, 上記の問合せは1回だけ実行）

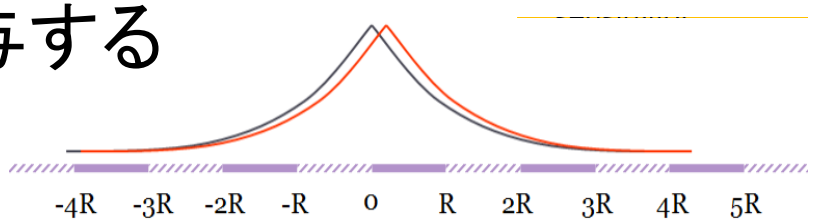
non-overlapping:
各セルの値はただ1つのカウントに影響

Laplace Perturbation Algorithm (LPA):
各集約演算の結果にラプラスノイズを追加

既存のDPアルゴリズムの問題点

- Laplace Perturbation Algorithm (LPA)：集計結果にラプラスノイズを付与する

$$t = c + < Lap\left(\frac{\Delta_1(Q)}{\epsilon}\right) >^d$$



DEFINITION 3. *The L_p* **sensitivity** *of* $\mathbf{Q}$ *w.r.t.* $\mathcal{D}$ *is*

$$\Delta_p(\mathbf{Q}, \mathcal{D}) = \max_{D, D' \in \mathcal{D}} \left\| \mathbf{Q}(D) - \mathbf{Q}(D') \right\|_p$$

for all neighboring $D, D' \in \mathcal{D}$. *When there is no ambiguity on* $\mathcal{D}$, *we simply use symbol* $\Delta_p(\mathbf{Q})$.

あるタプルがあるかないかで結果がどの程度変わるか

- $1 \leq \Delta_1(Q) \leq d$

ヒストグラムの場合：
1ユーザが$\Delta_1(Q)$ 個のバケットカウントに影響

- non-overlapping countの場合
  - 1ユーザが任意のバケットカウントに影響

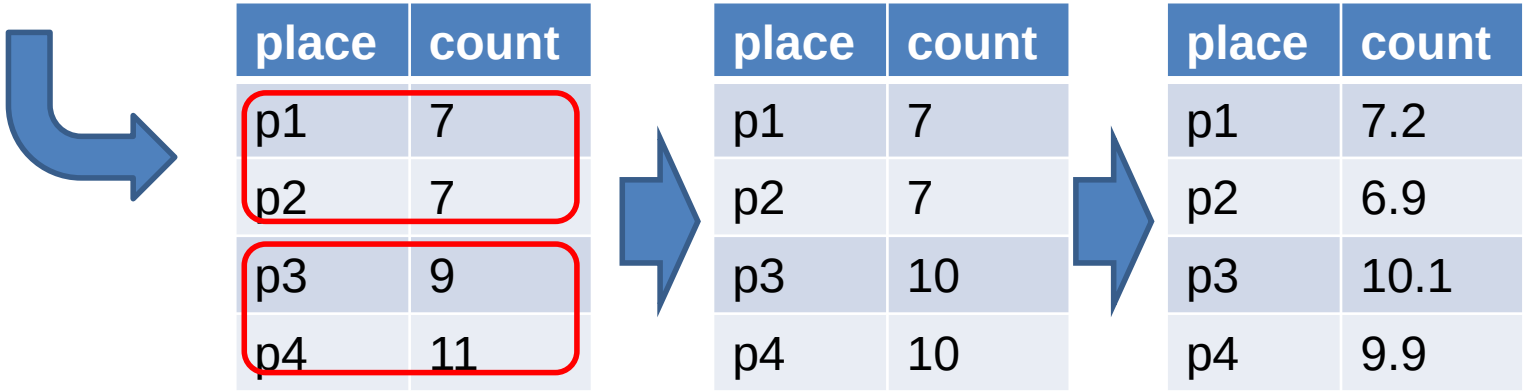
	p1	p2	p3	p4
u1	4	2	0	1
u2	0	0	2	3
u3	3	5	7	1

Grouping and Smoothing (GS)

- columnをグループ化 (Grouping)
- グループごとにcountの平均をとる (Smoothing)
- 平均値に対してラプラスノイズを付与する

	p1	p2	p3	p4
u1	4	2	0	6
u2	0	0	2	3
u3	3	5	7	2

一人のユーザの影響を平均化する



place	count
p1	7
p2	7
p3	9
p4	11

place	count
p1	7
p2	7
p3	10
p4	10

place	count
p1	7.2
p2	6.9
p3	10.1
p4	9.9

グルーピングの戦略

- Random Grouping (GS-R)
  - カラム数がdの時
    $$d_g = \lfloor d/\Delta_1(Q) \rfloor$$
    のグループに分ける
  - 各グループのsensitivityは
    1以下となる (LEMMA 1)

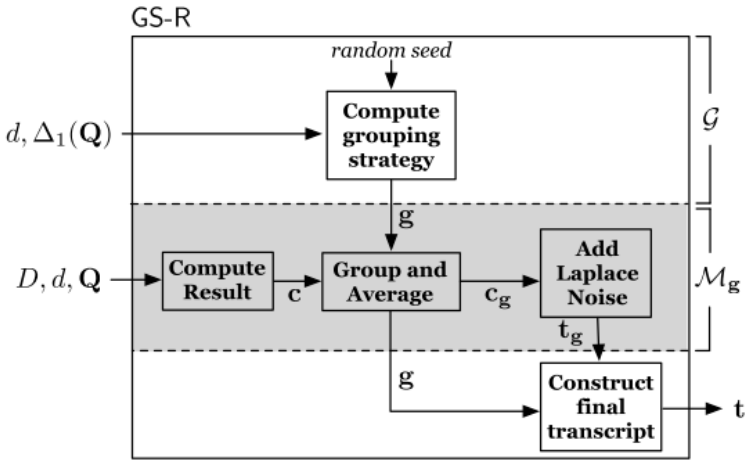


Figure 1: Outline of GS-R

- Effective Grouping via Sampling(GS-S)
  - ソートしてグルーピングだと逆に
    1ユーザの有無でグルーピング
    そのものが変わってしまい
    ϵ-差分プライバシを満たさなくなる
  - サンプルとってラプラスノイズくわえた
    結果を使ってグループ化する

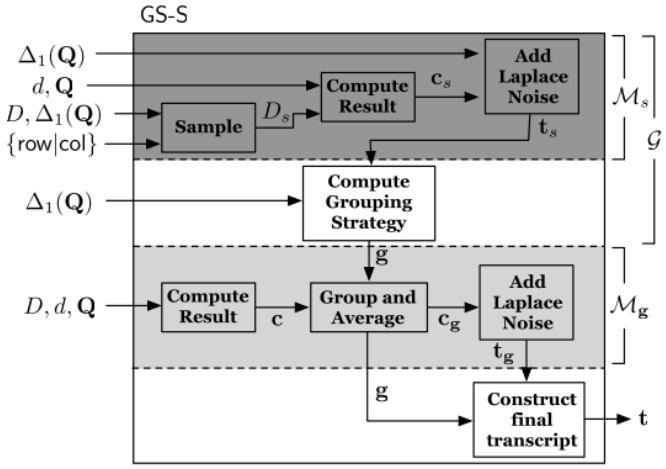


Figure 3: Outline of GS-S

Fine-tuning Step (GS)

- グルーピング戦略を
  いくつか用意して
  ベストな戦略を選ぶ

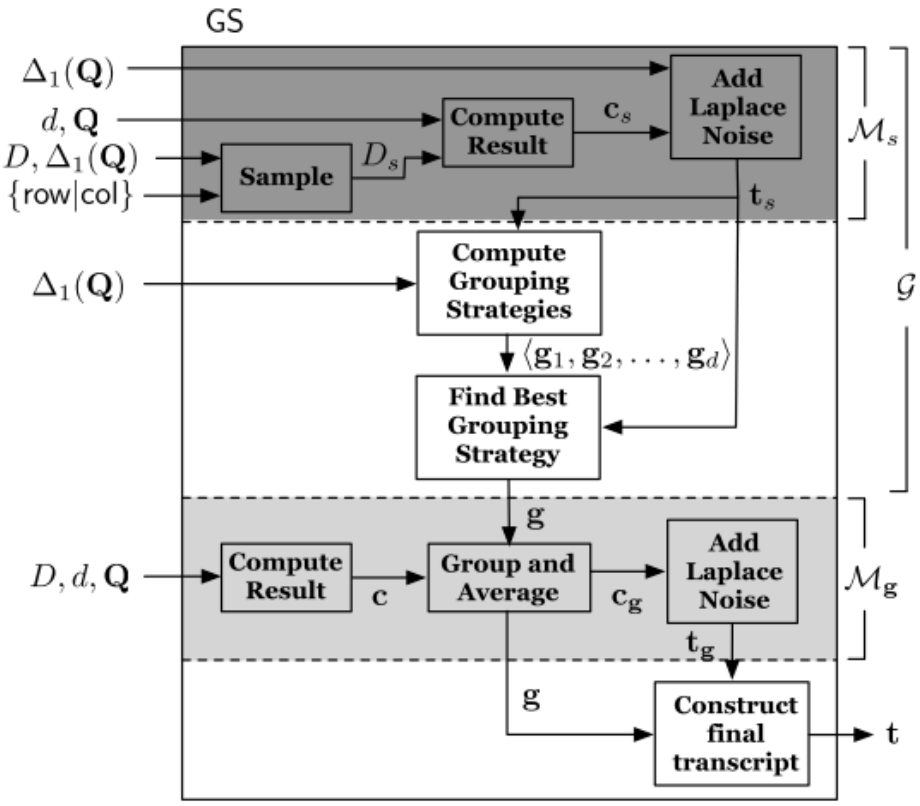
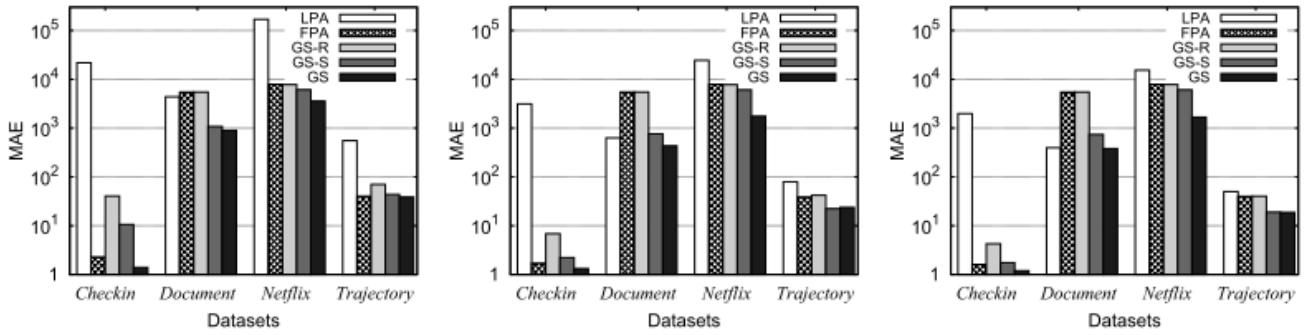


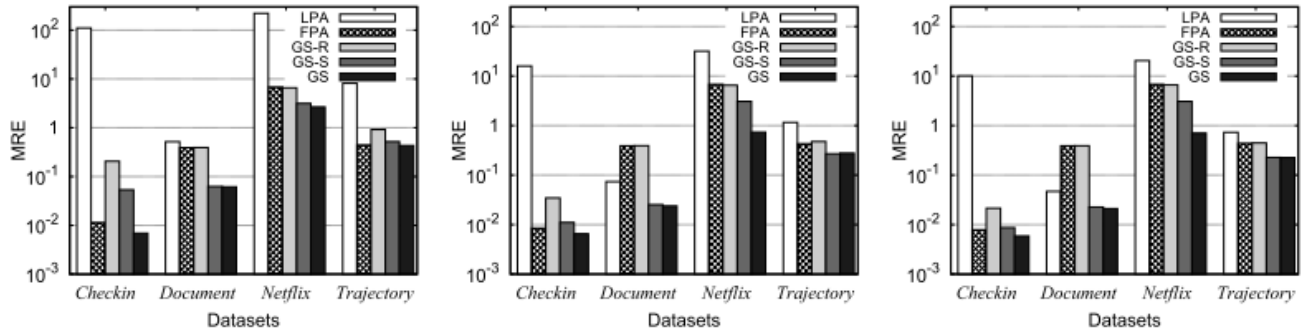
Figure 5: Outline of GS

Experimental Evaluation



(a) $\epsilon = 0.1$ (b) $\epsilon = \ln 2$ (c) $\epsilon = \ln 3$

Figure 9: MAE vs. datasets



(a) $\epsilon = 0.1$ (b) $\epsilon = \ln 2$ (c) $\epsilon = \ln 3$

Figure 10: MRE vs. datasets

CorrectDB: SQL Engine with Practical Query Authentication

- アウトソーシングデータベースを対象とした
  問合せ認証を完備したデータベースエンジンの開発
  - 問合せ認証アルゴリズムを
    サーベイして各演算に対して
    最もよいアルゴリズムを選定
  - trusted hardwareを使って
    認証することで処理速度を保証

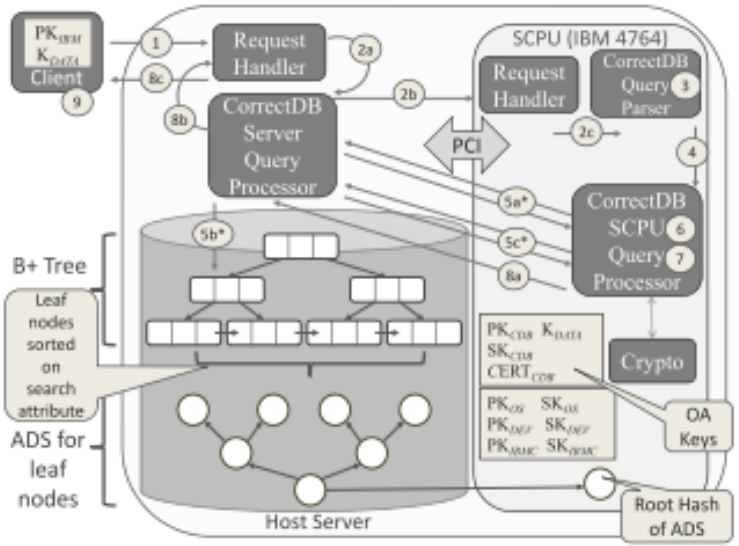


Figure 2: CorrectDB Architecture

Name	Year	QCT	QCP	Ops	Updates
Tree Based					
MHT [8]	2003	✓	✓	S,R	×
VBT [24]	2004	✓	✓	S,R	✓
EMBT [15]	2006	✓	✓	S,R,J	✓
Singh et al [28]	2008	✓	✓	S	✓
XBT [26]	2009	✓	✓	R	✓
AIM [32]	2009	✓	✓	S,R,J	×
MRT [33]	2009	✓	✓	S_p	✓
AABT [16]	2010	✓	✓	A	✓
MR-SKY [17]	2011	✓	✓	S_p	×
Yang et al [34]	2011	✓	✓	R	×
Signature Based					
AGS [21, 20]	2004	✓	×	S,R	×
DSAC [22]	2005	✓	✓	S,R	✓
Pang et al [23]	2005	✓	✓	S,R,J	✓
VKDT,VRT [7]	2006	✓	✓	S_p	×
Pang et al [25]	2009	✓	✓	S,R,J	✓
VNA [12]	2010	✓	✓	S_p	✓
Others					
DICT [10]	2001	✓	✓	K	✓
AUDIT [31]	2007	✓	×	S,J	✓
HLT [35]	2012	✓	✓	S,R,J,A	✓

Table 1: Summary of existing approaches (QCT - Query Correctness, QCP - Query Completeness, S - Select, R -