

# A Unified View of Frequency Estimation and their Attacks on Local Differential Privacy

Al Mehdi Saadat Chowdhury, Dhaval Pankaj Tanna, Deepak Vellanki, Chirag Manjeshwar  
School of Computing and Augmented Intelligence  
Arizona State University

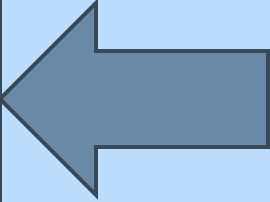
# Outline



## Introduction:

- Differential Privacy
- Local Differential Privacy
- Pure LDP and Frequency Estimation
- Attack Problem

## Frequency Estimation Techniques:

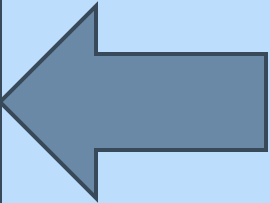
- RAPPOR
  - K Randomized Response (kRR)
  - Optimized Unary Encoding (OUE)
  - Optimized Local Hashing (OLH)
- 



## Attacks:

- Attack Types
- Attacking kRR
- Attacking OUE
- Attacking OLH

## Evaluation:

- Comparison between Estimators
  - Gain from Attacks
  - Impacts of Parameters on Attacks
- 

Conclusion

# Introductory Concepts

# Differential Privacy

.

$$\Pr[\mathcal{M}(x) \in \mathcal{S}] \leq e^\epsilon \Pr[\mathcal{M}(y) \in \mathcal{S}] + \delta$$

Equation from report

TODO: describe terms

# Local Differential Privacy

$$\forall y \in \text{Range}(\mathcal{M}) : \Pr[\mathcal{M}(v_1) = y] \leq e^\epsilon \Pr[\mathcal{M}(v_2) = y]$$

Equation from report  
(from section 2.1 in OUE & OLH paper)

$$\Pr(PE(v_1) = y) \leq e^\epsilon \Pr(PE(v_2) = y)$$

Alternative Equation from attacks paper (section 2.1 definition 1)  
[this looks more similar to differential privacy definition than the one used in report]

TODO: select eqn & describe terms

# Pure Local Differential Privacy

$$\Pr[PE(v_1) \in \{y | v_1 \in \text{Support}(y)\}] = p^*, \quad \forall_{v_2 \neq v_1} \Pr[PE(v_2) \in \{y | v_1 \in \text{Support}(y)\}] = q^*$$

Equation from report

(from oue & olh paper, section 3 definition 3)

$$\Pr(PE(v_1) \in \{y | v_1 \in S(y)\}) = p$$

$$\Pr(PE(v_2) \in \{y | v_1 \in S(y)\}) = q$$

Alternative equation

(from attacks section 2.1)

TODO: select eqn & describe terms

# Frequency Estimation Problem

$$\tilde{f}_v = \frac{\frac{1}{n} \sum_{i=1}^n \mathbf{1}_{S(y_i)}(v) - q}{p - q}$$

$$\sum_{i=1}^n E[\mathbf{1}_{S(y_i)}(v)] = n(f_v(p - q) + q)$$

Equations from section 2.1 (attacks paper)

TODO: describe terms

# Frequency Estimation Techniques



# RAPPOR

$$B'_i = \begin{cases} 1 & \text{with probability } \frac{1}{2}f \\ 0 & \text{with probability } \frac{1}{2}f \\ B_i & \text{with probability } 1 - f \end{cases} \quad P(S_i = 1) = \begin{cases} q & \text{if } B'_i = 1 \\ p & \text{if } B'_i = 0 \end{cases}$$

Equations from section 2  
(Rappor algorithm)

$$t_{ij} = \frac{c_{ij} - \left(p + \frac{1}{2}fq - \frac{1}{2}fp\right)N_j}{(1-f)(q-p)}$$

Equation from section 4  
(Decoding)

# K Randomized Response

$$Pr[PE(v) = i] = \begin{cases} p = \frac{e^\epsilon}{e^\epsilon + d - 1}, & \text{if } i = v \\ q = \frac{1-p}{d-1}, & \text{otherwise} \end{cases}$$

Equation from report (attacks paper section 2.1)

$$\text{Var}^*[\tilde{c}_{DE}(i)] = n \cdot \frac{d-2+e^\epsilon}{(e^\epsilon-1)^2}$$

Equation from OUE & OLH paper (section 4.1)

TODO: describe terms

# Optimized Unary Encoding

$$Pr[PE(v) = i] = \begin{cases} p = \frac{1}{2}, & \text{if } i = v \\ q = \frac{1}{e^\epsilon + 1}, & \text{otherwise} \end{cases}$$

Equation from report (attacks paper section 2.1)

$$\text{Var}^*[\tilde{c}_{OUE}(i)] = n \cdot \frac{4e^\epsilon}{(e^\epsilon - 1)^2}$$

Equation from OUE & OLH paper (section 4.3)

TODO: describe terms

# Optimized Local Hashing

$$\forall_{i \in [d]} Pr[y = \langle H, x \rangle] = \begin{cases} p = \frac{e^\epsilon}{e^\epsilon + d - 1}, & \text{if } x = i \\ q = \frac{1}{e^\epsilon + d - 1}, & \text{otherwise} \end{cases}$$

Equation from report (attacks paper section 2.1)

$$\text{Var}^*[\tilde{c}_{OLH}(i)] = n \cdot \frac{4e^\epsilon}{(e^\epsilon - 1)^2}$$

Equation from OUE & OLH paper (section 4.4)

TODO: describe terms

# Attack – Problem Formulation

$$G(Y) = \sum_{t \in T} E[\Delta \tilde{f}_t]$$

Equation from attacks paper (section 3.1)

$$\max_{\mathbf{Y}} G(\mathbf{Y})$$

Equation from report (section 3.1 in attacks paper)

TODO: describe terms

# Attacks on Frequency Estimators

# Attack Types

- Describe RPA, RIA, MGA
- Equation related to MGA:

$$y^* = \arg \max_{y \in \mathcal{D}} \sum_{t \in T} \mathbb{1}_{S(y)}(t).$$

Todo: convert to latex then svg & describe terms

# Attacking kRR

$$G_{\text{RPA}}^{\text{kRR}} = \frac{rm}{d(n+m)(p-q)} - c$$

$$G_{\text{RIA}}^{\text{kRR}} = \frac{(p+(r-1)q)m}{(n+m)(p-q)} - c$$

$$G_{\text{MGA}}^{\text{kRR}} = \frac{m}{(n+m)(p-q)} - c$$

Section 3.3 attacks paper

TODO: describe terms



# Attacking OUE

$$G_{\text{RPA}}^{\text{OUE}} = \frac{rm}{2(n+m)(p-q)} - c$$

$$G_{\text{RIA}}^{\text{OUE}} = \frac{(p+(r-1)q)m}{(n+m)(p-q)} - c$$

$$G_{\text{MGA}}^{\text{OUE}} = \frac{rm}{(n+m)(p-q)} - c$$

Section 3.4 attacks paper

TODO: describe terms

# Attacking OLH

$$G_{\text{RPA}}^{\text{OLH}} = \frac{rm}{d'(n+m)(p-q)} - c$$

$$G_{\text{RIA}}^{\text{OLH}} = \frac{[p+(r-1)q]m}{(n+m)(p-q)} - c$$

$$G_{\text{MGA}}^{\text{OLH}} = \frac{rm}{(n+m)(p-q)} - c$$

Section 3.4 attacks paper

TODO: describe terms

# Evaluation

# Comparison between Estimators

- TODO: add table

# Gain from Attacks

- TODO: convert img to ppt table

	kRR	OUE	OLH
Random perturbed-value attack (RPA)	$\beta(\frac{r}{d} - f_T)$	$\beta(r - f_T)$	$-\beta f_T$
Random item attack (RIA)	$\beta(1 - f_T)$	$\beta(1 - f_T)$	$\beta(1 - f_T)$
Maximal gain attack (MGA)	$\beta(1 - f_T) + \frac{\beta(d-r)}{e^\epsilon - 1}$	$\beta(2r - f_T) + \frac{2\beta r}{e^\epsilon - 1}$	$\beta(2r - f_T) + \frac{2\beta r}{e^\epsilon - 1}$
Standard deviation of estimation	$\frac{r\sqrt{d-2+e^\epsilon}}{(e^\epsilon - 1)\sqrt{n}}$	$\frac{2re^{\epsilon/2}}{(e^\epsilon - 1)\sqrt{n}}$	$\frac{2re^{\epsilon/2}}{(e^\epsilon - 1)\sqrt{n}}$

$$\beta = \frac{m}{n+m}$$

# Impacts of Parameters on Attacks

$$\text{Frequency Gain} \propto \frac{1}{e^\varepsilon - 1}$$

To do: Add graphs

# Conclusion

# Summary

- TODO



# Open Problems

- TODO

Thank You