Sublinear Space Private Algorithms Under the Sliding Window Model

Jalaj Upadhyay

JOHNS HOPKINS
WHITING SCHOOL OF ENGINEERING
Differential Privacy

$x_1$
$x_2$
$\vdots$
$x_n$

$x_1 :$
$x'_2 :$
$\vdots$
$x_n :$

A
Differential Privacy

A

queries/tasks

private random coin

A(x)

⋯

queries/tasks

private random coin

A(x')
Differential Privacy

Output distribution is close

private random coin

\( A \)

queries/tasks

\( A(x) \)

queries/tasks

\( A(x') \)

private random coin
Differential Privacy

\[ x \text{ and } x' \text{ are } \text{neighbor} \text{ if they differ in one data point} \]

Output distribution is close
Differential Privacy

$x$ and $x'$ are neighbor if they differ in one data point

Differential Privacy [DMNS06]
Algorithm $A$ is $\alpha$-differentially private if
- for all neighboring data sets $x$ and $x'$
- for all possible outputs $S$, $\Pr[A(x) \in S] \leq e^\alpha \cdot \Pr[A(x') \in S]$
\( x \) and \( x' \) are *neighbor* if they differ in one data point.

Differential Privacy [DMNS06]
Algorithm \( A \) is \( \alpha \)-differentially private if
- for all neighboring data sets \( x \) and \( x' \)
- for all possible outputs \( S \),
  \[
  \Pr[A(x) \in S] \leq e^\alpha \cdot \Pr[A(x') \in S] 
  \]

\( \alpha = 0 \): perfect privacy
  no utility
As \( \alpha \) increases, less privacy
  more utility
Differential Privacy

$x$ and $x'$ are neighbor if they differ in one data point

Algorithm $A$ is $\alpha$-differentially private if
- for all neighboring data sets $x$ and $x'$
- for all possible outputs $S$,
  \[ \Pr[A(x) \in S] \leq e^{\alpha} \cdot \Pr[A(x') \in S] \]

\[ \alpha = 0 \]: perfect privacy
  no utility

As $\alpha$ increases, less privacy
  more utility

Output distribution is close

Allows utility-privacy trade-off
Differential Privacy Under Sliding Window

• Differential privacy overview of Apple
  “Apple retains the collected data for a maximum of three months”
Differential Privacy Under Sliding Window

• Differential privacy overview of Apple

“Apple retains the collected data for a maximum of three months”
Differential Privacy Under Sliding Window

• Differential privacy overview of Apple
  “Apple retains the collected data for a maximum of three months”

Goal of this paper

• Formalize privacy under sliding window model
• Design sublinear space private algorithms in the sliding window model
Problem Studied: Private $\ell_1$ heavy hitters

- $x$ be an $n$-dimensional vector
- Output all indices $i \in [n], \ x_i \geq \phi \ |x|_1$ and estimate of $x_i$
- Allowed to accept $i \in [n]$ if $x_i \geq (\phi - \rho) \ |x|_1$
Problem Studied: Private $\ell_1$ heavy hitters

- $x$ be an $n$-dimensional vector
- Output all indices $i \in [n], \ x_i \geq \phi \| x \|_1$ and estimate of $x_i$
- Allowed to accept $i \in [n]$ if $x_i \geq (\phi - \rho) \| x \|_1$

Main Theorem

There is an efficient $o(w)$ space $(\epsilon, \delta)$-DP algorithm that returns a set of indices, $I$, and estimates $\hat{x}_i$ for $i \in I$,

- If $x_i \geq \phi \| x \|_1$, then $|x_i - \hat{x}_i| \leq \rho \| x \|_1 + O \left( \frac{1}{\epsilon} \log w \right)$
- Does not include any $i$ if $x_i < (\phi - 3 \rho) \| x \|_1 + O \left( \frac{\phi}{\epsilon} \log w \right)$
Problem Studied: Private $\ell_1$ heavy hitters

- $x$ be an $n$-dimensional vector
- Output all indices $i \in [n]$, $x_i \geq \phi \|x\|_1$ and estimate of $x_i$
- Allowed to accept $i \in [n]$ if $x_i \geq (\phi - \rho) \|x\|_1$

Main Theorem
There is an efficient $o(w)$ space $(\epsilon, \delta)$-DP algorithm that returns a set of indices, $I$, and estimates $\hat{x}_i$ for $i \in I$,

- If $x_i \geq \phi \|x\|_1$, then $|x_i - \hat{x}_i| \leq \rho \|x\|_1 + \mathcal{O}\left(\frac{1}{\epsilon} \log \frac{w}{\delta}\right)$
- Does not include any $i$ if $x_i < (\phi - 3 \rho) \|x\|_1 + \mathcal{O}\left(\frac{\phi}{\epsilon} \log \frac{w}{\delta}\right)$
Other Results and Open Problems

• Algorithm extends to continual observation under sliding window

• Current non-private framework do not extend to privacy
  • Lower bound using standard packing argument

• Space lower bound on estimating $\ell_1$-heavy hitters
  • Reduction to communication complexity problem
Other Results and Open Problems

• Algorithm extends to continual observation under sliding window

• Current non-private framework do not extend to privacy
  • Lower bound using standard packing argument

• Space lower bound on estimating $\ell_1$-heavy hitters
  • Reduction to communication complexity problem

Characterize what is possible to compute privately under the sliding window model