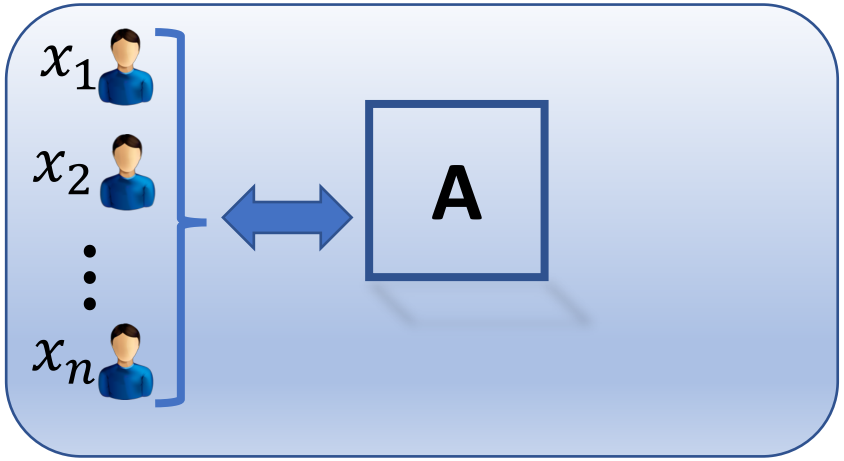
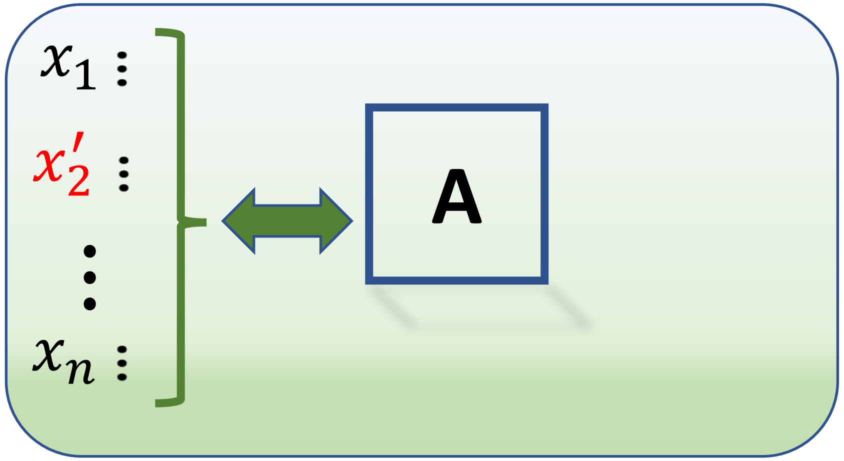
Sublinear Space Private Algorithms Under the Sliding Window Model

Jalaj Upadhyay

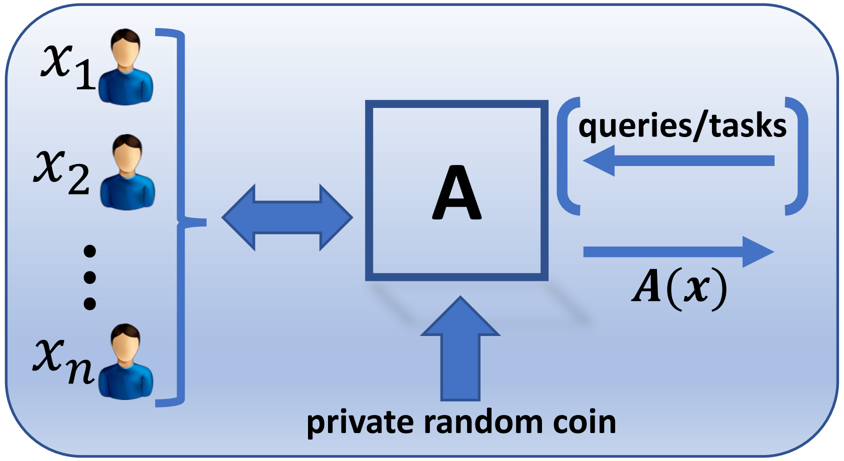
JOHNS HOPKINS
WHITING SCHOOL
of ENGINEERING

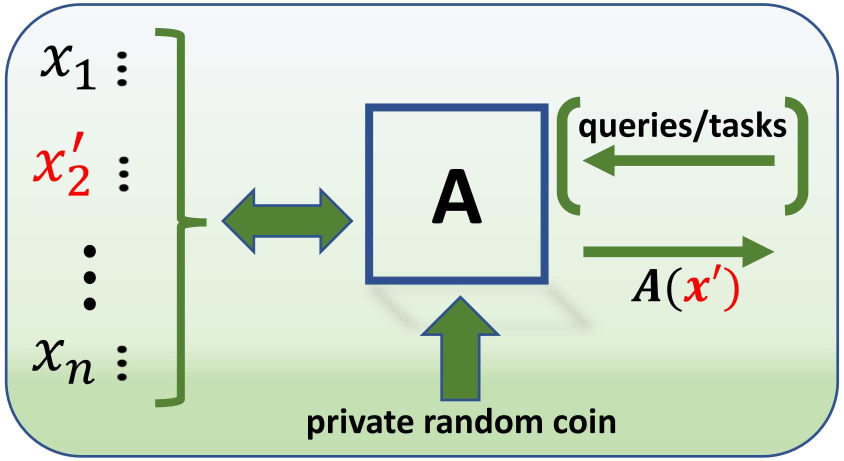
Differential Privacy



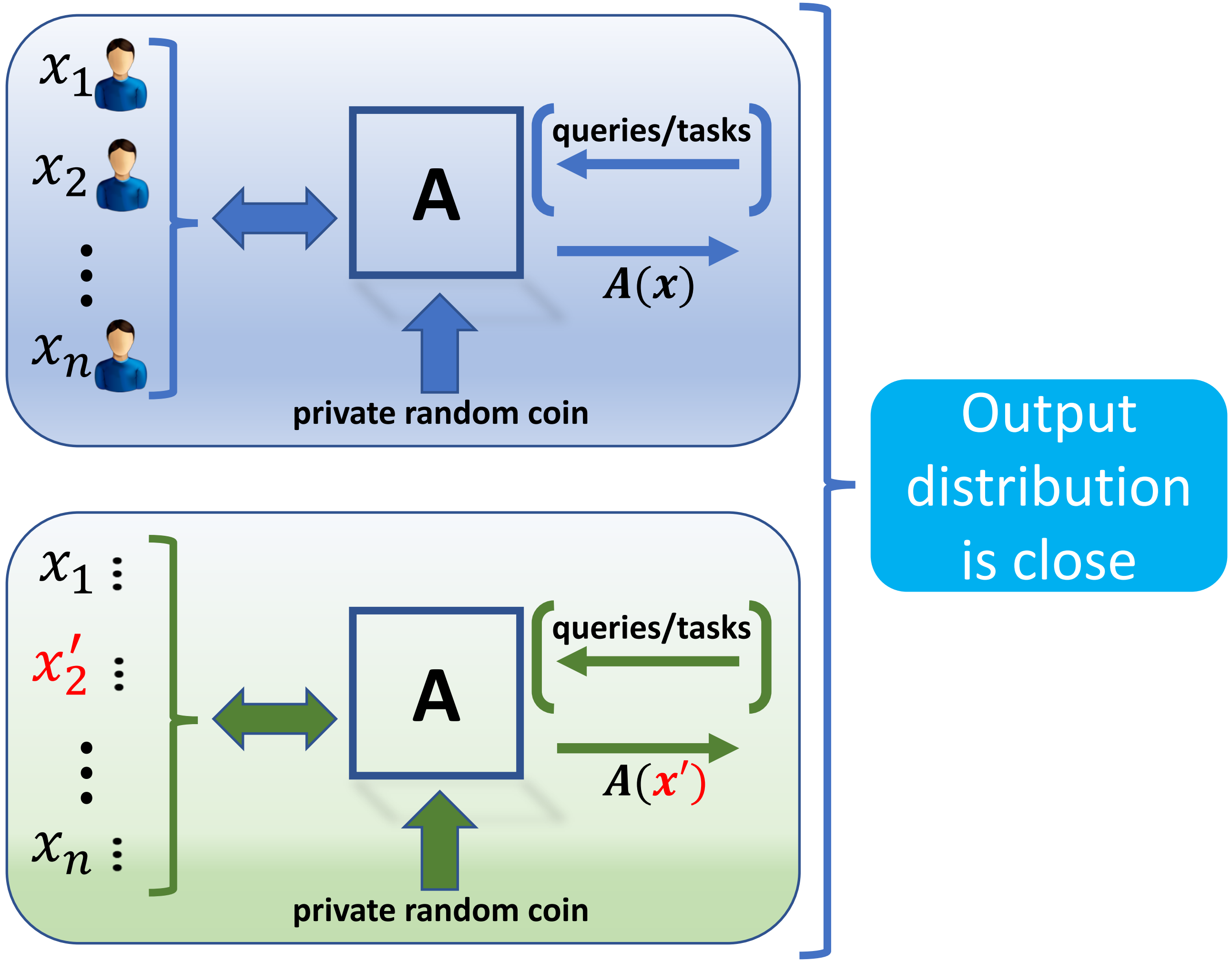


Differential Privacy

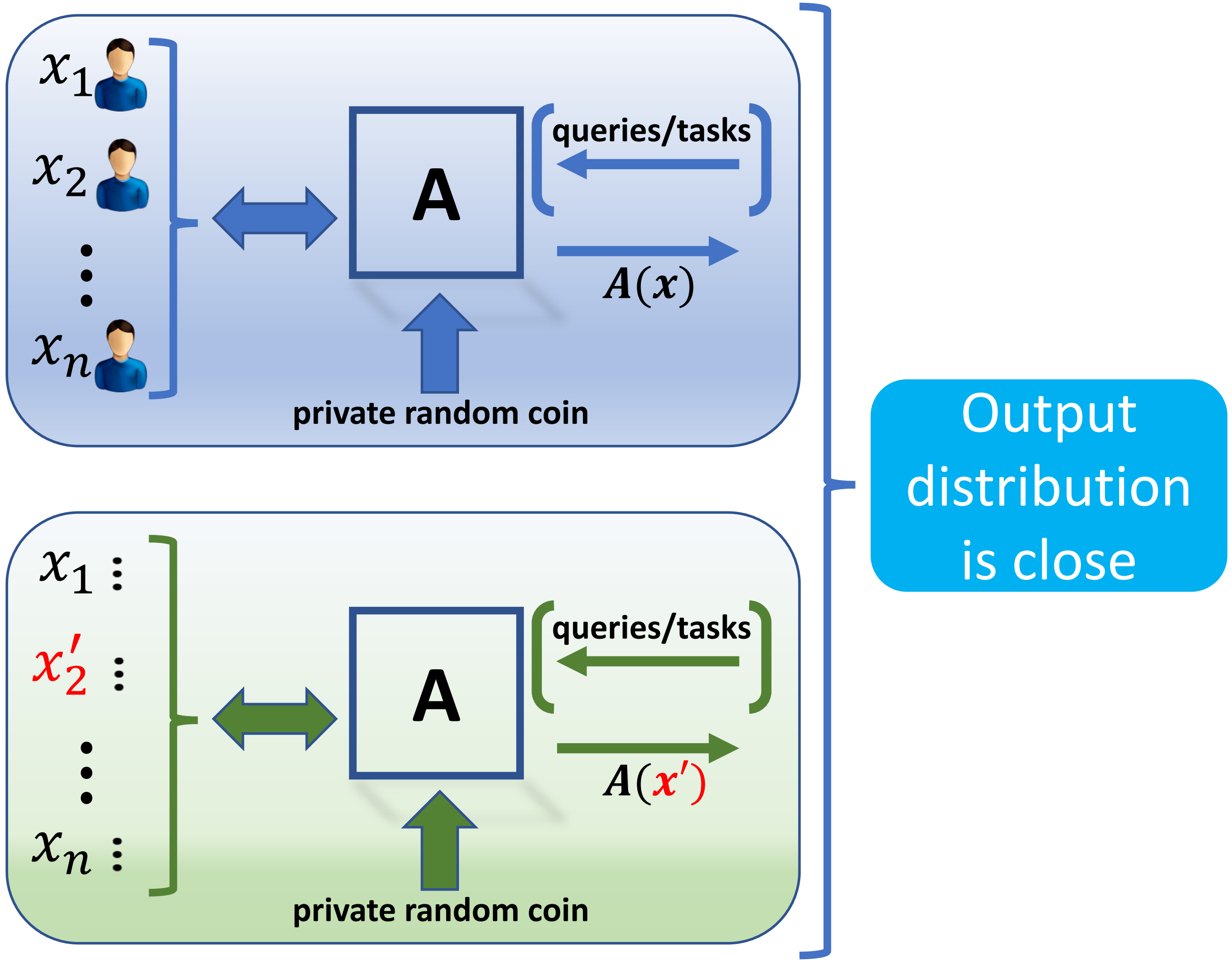




Differential Privacy

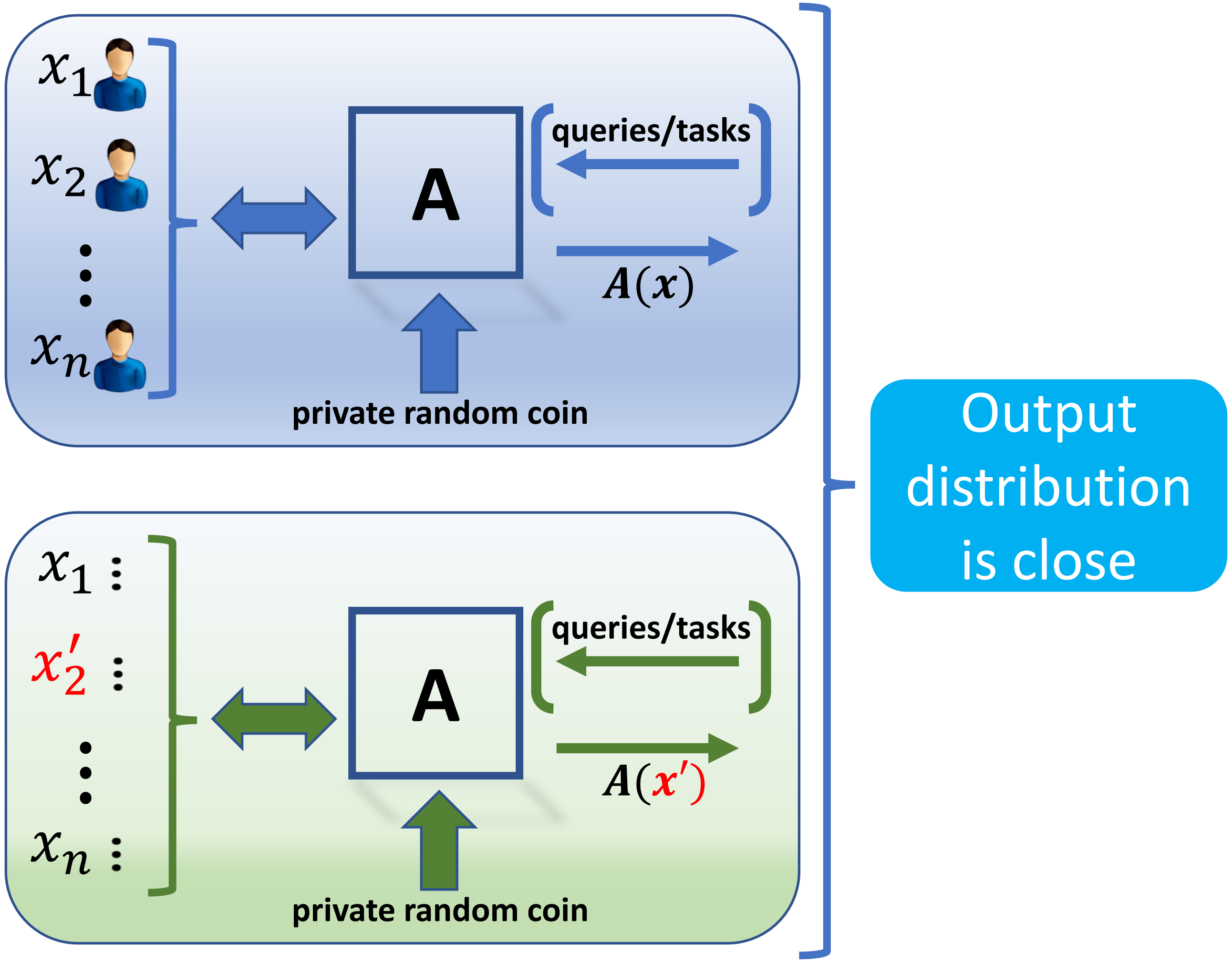


Differential Privacy



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

Differential Privacy



$\boldsymbol{x}$ and $\boldsymbol{x'}$ are *neighbor* if
they differ in one data point

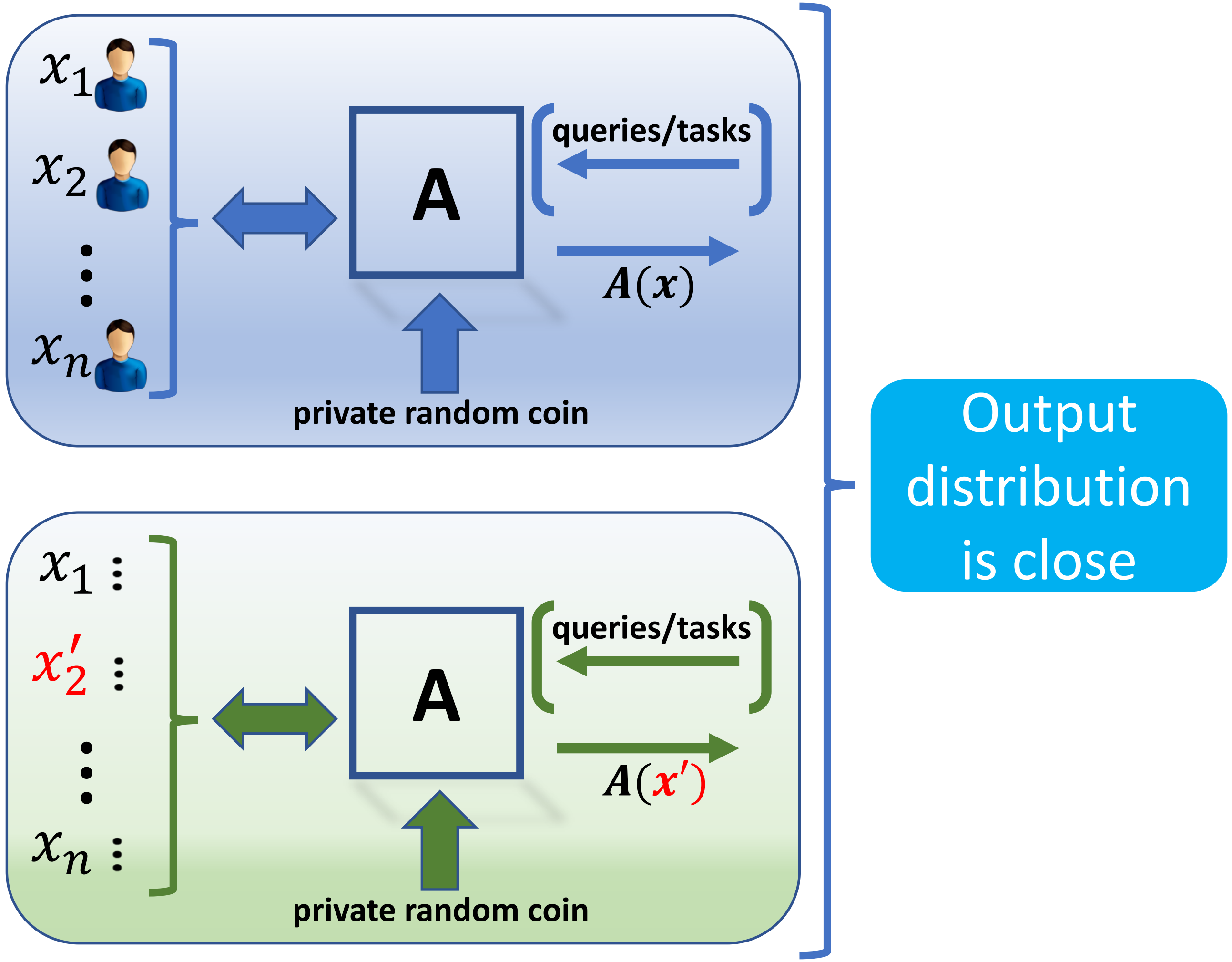
Differential Privacy [DMNS06]

Algorithm A is α-differentially private if

- for all neighboring data sets $\boldsymbol{x}$ and $\boldsymbol{x}'$
- for all possible outputs S,

$$\Pr[A\,(\boldsymbol{x}) \in S] \leq e^{\alpha} \cdot \Pr[\,A(\boldsymbol{x}') \in S]$$

Differential Privacy



$\boldsymbol{x}$ and $\boldsymbol{x'}$ are *neighbor* if
they differ in one data point

Differential Privacy [DMNS06]

Algorithm A is α-differentially private if

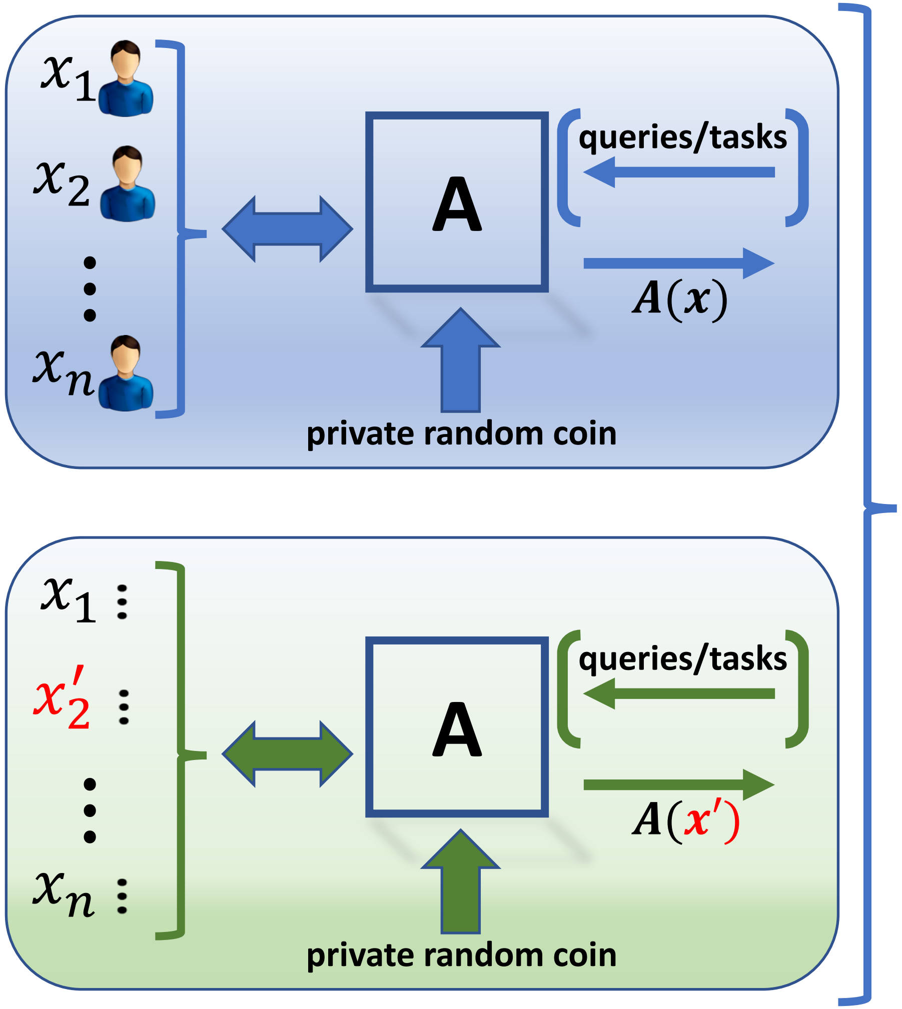
- for all neighboring data sets $\boldsymbol{x}$ and $\boldsymbol{x}'$
- for all possible outputs S,

$$\Pr[A\ (\boldsymbol{x}) \in S] \le e^{\alpha} \cdot \Pr[\ A(\boldsymbol{x}') \in S]$$

$\alpha = 0$: perfect privacy
no utility

As α increases, less privacy
more utility

Differential Privacy



Output distribution is close

$\boldsymbol{x}$ and $\boldsymbol{x'}$ are *neighbor* if
they differ in one data point

Differential Privacy [DMNS06]

Algorithm A is α-differentially private if

- for all neighboring data sets $\boldsymbol{x}$ and $\boldsymbol{x'}$
- for all possible outputs S,

$$\Pr[A\ (\boldsymbol{x}) \in \mathrm{S}] \leq e^{\alpha} \cdot \Pr[\ A(\boldsymbol{x'}) \in S]$$

$\alpha = 0$: perfect privacy
no utility
As α increases, less privacy
more utility

Allows utility-privacy trade-off

Differential Privacy Under Sliding Window

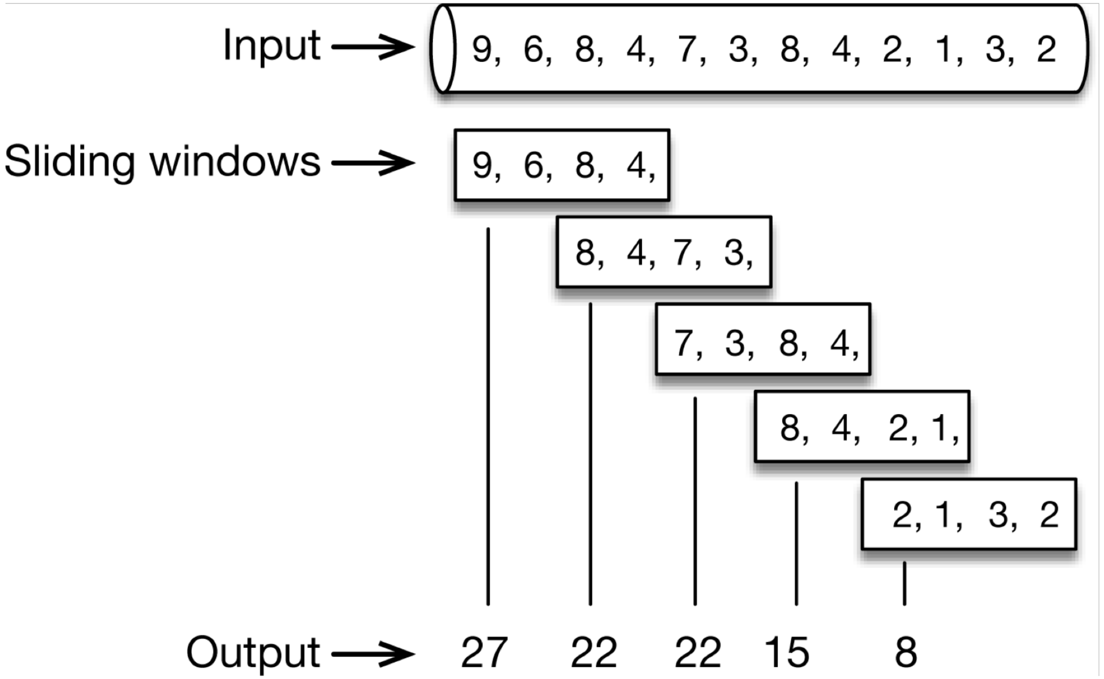
- Differential privacy overview of Apple

  "*Apple retains the collected data for a maximum of three months*"

Differential Privacy Under Sliding Window

- Differential privacy overview of Apple

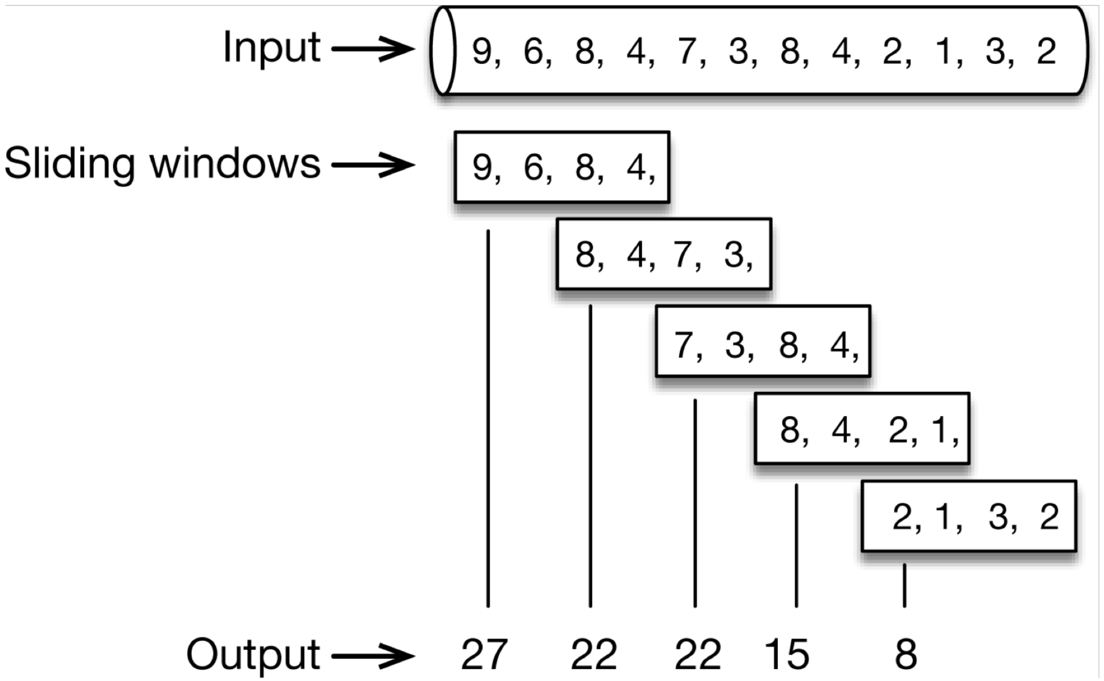
  "Apple retains the collected data for a maximum of three months"



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 ℓ_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 ℓ_1 heavy hitters

- x be an n-dimensional vector
- Output all indices $i \in [n]$, $x_i \geq \phi \parallel x \parallel_1$ and estimate of x_i
- Allowed to accept $i \in [n]$ if $x_i \geq (\phi - \rho) \parallel x \parallel_1$

Main Theorem

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

- If $x_i \geq \phi \parallel x \parallel_1$, then $|x_i - \hat{x}_i| \leq \rho \parallel x \parallel_1 + O\left(\frac{1}{\epsilon} \log w\right)$
- Does not include any i if $x_i < (\phi - 3\rho) \parallel x \parallel_1 + O\left(\frac{\phi}{\epsilon} \log w\right)$

Problem Studied: Private ℓ_1 heavy hitters

- x be an n-dimensional vector
- Output all indices $i \in [n]$, $x_i \geq \phi \parallel x \parallel_1$ and estimate of x_i
- Allowed to accept $i \in [n]$ if $x_i \geq (\phi - \rho) \parallel x \parallel_1$

Main Theorem

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

- If $x_i \geq \phi \parallel x \parallel_1$, then $|x_i - \hat{x}_i| \leq \rho \parallel x \parallel_1 + O\left(\frac{1}{\epsilon} \log w\right)$
- Does not include any i if $x_i < (\phi - 3\rho) \parallel x \parallel_1 + O\left(\frac{\phi}{\epsilon} \log w\right)$

Price of privacy

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 ℓ_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 ℓ_1-heavy hitters
  - Reduction to communication complexity problem

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