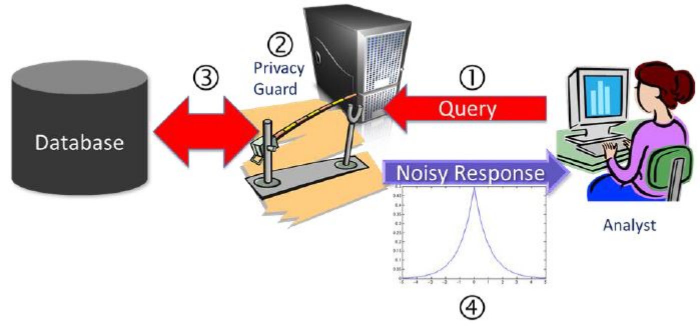
Differential Privacy in the Streaming Model

Jalaj Upadhyay

Pennsylvania State University

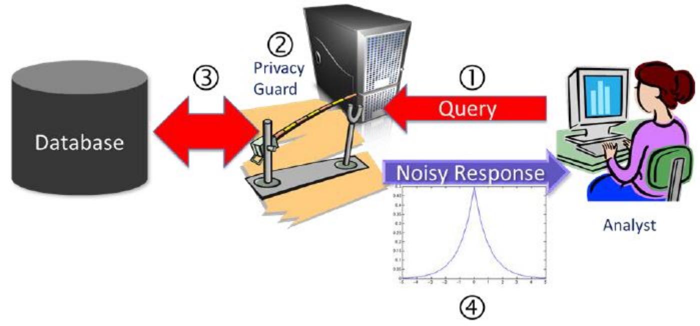
January 07, 2016

Differential Privacy: The Framework



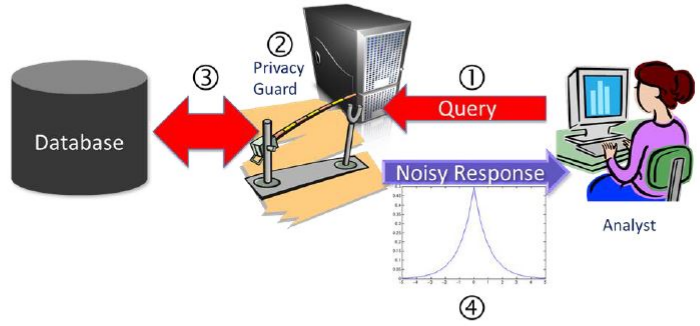
Analyst wishes to get some task done on the $\mathrm{Database}$

Differential Privacy: The Framework



Privacy guard provides privacy of individuals in the $\mathrm{Database}$

Differential Privacy: The Framework



The privacy guard performs the task on the $\mathrm{Database}$

Differential Privacy: The Mathematical Formulation

The idea is that absence or presence of an individual entry should not change the output “by much"

Differential Privacy: The Mathematical Formulation

The idea is that absence or presence of an individual entry should not change the output “by much"

Definition. A randomized algorithm, $\mathfrak{M}$, gives (ε, δ)-differential privacy if, for all “neighboring data," $\mathbf{D}$ and $\widetilde{\mathbf{D}}$, and for all $S \subseteq \mathsf{Range}(\mathfrak{M})$,

$$\mathsf{Pr}\left[\mathfrak{M}(\mathbf{D}) \in S\right] \leq \exp(\varepsilon)\mathsf{Pr}\left[\mathfrak{M}(\widetilde{\mathbf{D}}) \in S\right] + \delta$$

Differential Privacy: The Mathematical Formulation

The idea is that absence or presence of an individual entry should not change the output “by much"

Definition. A randomized algorithm, $\mathfrak{M}$, gives (ε, δ)-differential privacy if, for all “neighboring data," $\mathbf{D}$ and $\widetilde{\mathbf{D}}$, and for all $S \subseteq \mathrm{Range}(\mathfrak{M})$,

$$\Pr\left[\mathfrak{M}(\mathbf{D}) \in S\right] \leq \exp(\varepsilon) \Pr\left[\mathfrak{M}(\widetilde{\mathbf{D}}) \in S\right] + \delta$$

We restrict how the privacy guard can access the database

Differentially Private Streaming Model of Computation

Privacy Guard

Private Matrix

8	1	6
3	5	7
4	9	2
5	3	7
6	2	1
2	6	7
2	1	9
1	1	6
1	3	1

Differentially Private Streaming Model of Computation

Privacy Guard

$$\begin{array}{|c|}\hline 8\\ \hline 3\\ \hline 4\\ \hline 5\\ \hline 6\\ \hline 2\\ \hline 2\\ \hline 1\\ \hline 1\\ \hline \end{array} \quad \Leftarrow \quad \begin{array}{|c|c|}\hline 1 & 6\\ \hline 5 & 7\\ \hline 9 & 2\\ \hline 3 & 7\\ \hline 2 & 1\\ \hline 6 & 7\\ \hline 1 & 9\\ \hline 1 & 6\\ \hline 3 & 1\\ \hline \end{array}$$

Private Matrix

- Operates on the stream
- Update the data structure $\begin{pmatrix} \color{red}{4} \\ \color{red}{6} \\ \color{red}{1} \end{pmatrix}$

Differentially Private Streaming Model of Computation

Privacy Guard			Private Matrix
	1 5 9 3 2 6 1 1 3	$\Leftarrow$	6 7 2 7 1 7 9 6 1

- Operates on the stream
- Update the data structure $\begin{pmatrix} 4 & 3 \\ 6 & 2 \\ 1 & 8 \end{pmatrix}$

Differentially Private Streaming Model of Computation

Privacy Guard

$$\begin{array}{|c|}\hline 6\\ \hline 7\\ \hline 2\\ \hline 7\\ \hline 1\\ \hline 7\\ \hline 9\\ \hline 6\\ \hline 1\\ \hline \end{array}$$

- Operates on the stream
- Update the data structure $\begin{pmatrix} 4 & 3 & 4 \\ 6 & 2 & 8 \\ 1 & 8 & 9 \end{pmatrix}$

Differentially Private Streaming Model of Computation

An analyst comes along

Differentially Private Streaming Model of Computation

An analyst comes along

$\xrightarrow{\text{request to do a task}}$

Differentially Private Streaming Model of Computation

An analyst comes along

$$\xrightarrow{\text{request to do a task}}$$

- uses

$$\begin{pmatrix} 4 & 3 & 4 \\ 6 & 2 & 8 \\ 1 & 8 & 9 \end{pmatrix}$$

Differentially Private Streaming Model of Computation

An analyst comes along

$$\xrightarrow{\text{request to do a task}}$$

$$\xleftarrow{\text{performs the task}}$$

- uses

$$\begin{pmatrix} 4 & 3 & 4 \\ 6 & 2 & 8 \\ 1 & 8 & 9 \end{pmatrix}$$

Differentially Private Streaming Model of Computation

cannot figure out individual information

Differentially Private Streaming Model of Computation

cannot figure out individual information

Privacy goal achieved

Differentially Private Streaming Model of Computation

Following are the extra parameters

1. number of passes over the matrix
2. space requirement of the data structures
3. time required to update the data structures

Differentially Private Streaming Model of Computation

Following are the extra parameters

1. number of passes over the matrix
2. space requirement of the data structures
3. time required to update the data structures

Differentially Private Streaming Model of Computation

Following are the extra parameters

1. number of passes over the matrix
2. space requirement of the data structures
3. time required to update the data structures

The Main Idea

Non-private Setting

The Main Idea

Non-private Setting

Data-structure is a sketch generated using random matrix

The Main Idea

Non-private Setting

Data-structure is a sketch generated using random matrix

$\Downarrow$

Efficient one-pass streaming algorithms

The Main Idea

Private Setting

The Main Idea

Private Setting

Special distribution of random matrices

The Main Idea

Private Setting

Special distribution of random matrices

$+$

Sketch generated using a random matrix picked from this distribution

The Main Idea

Private Setting

Special distribution of random matrices

$+$

Sketch generated using a random matrix picked from this distribution

$\Downarrow$

Differentially private one-pass streaming algorithms

The Main Idea

Private Setting

Special distribution of random matrices

$$+$$

Sketch generated using a random matrix picked from this distribution

$$\Downarrow$$

Differentially private one-pass streaming algorithms

First Approach

Streaming Private Sketch Generator (PSG_1)

Pick a random Gaussian matrix $\mathbf{\Phi}$
Multiply $\mathbf{\Phi}$ to the streamed column

First Approach

Streaming Private Sketch Generator (PSG_1)

Pick a random Gaussian matrix $\mathbf{\Phi}$
Multiply $\mathbf{\Phi}$ to the streamed column

Theorem. If the singular values of the streamed matrix to PSG_1 algorithm are at least $\sigma_1 := \left(4\sqrt{r\log(2/\delta)}\log(r/\delta)\right)/\varepsilon$, then PSG_1 preserves (ε, δ)-differential privacy

First Approach

Streaming Private Sketch Generator (PSG_1)

Pick a random Gaussian matrix Φ
Multiply Φ to the streamed column

Theorem. If the singular values of the streamed matrix to PSG_1 algorithm are at least $\sigma_1 := \left(4\sqrt{r\log(2/\delta)}\log(r/\delta)\right)/\varepsilon$, then PSG_1 preserves (ε, δ)-differential privacy

Similar result was shown by [BBDS12] for non-streaming algorithms

First Approach

Streaming Private Sketch Generator (PSG$_2$)

Pick a random Gaussian matrix $\mathbf{\Phi}$
Multiply $\mathbf{\Phi}^{\mathsf{T}}\mathbf{\Phi}$ to the streamed column

Theorem. If the singular values of the streamed matrix to the PSG$_2$ algorithm are at least $\sigma_2 := \left(4r\log(r/\delta)\right)/\varepsilon$, then PSG$_2$ preserves (ε, δ)-differential privacy.

A Meta Algorithm

- Get a stream in the form of column vector

A Meta Algorithm

- Get a stream in the form of column vector
- Perturb the vector to lift the singular values

A Meta Algorithm

- Get a stream in the form of column vector
- Perturb the vector to lift the singular values
- Feed it to PSG_1 or PSG_2

A Meta Algorithm

- Get a stream in the form of column vector
- Perturb the vector to lift the singular values
- Feed it to PSG_1 or PSG_2
- Perform any post-processing

Another Candidate for Φ: Update-time Efficiency

1. Pick $\{\mathbf{g}_1, \cdots, \mathbf{g}_n\} \sim \mathcal{N}(0,1)^n$
2. Divide it into r equal blocks of vectors $\mathbf{\Phi}_1, \cdots, \mathbf{\Phi}_r$.

$$\mathbf{P} := \begin{pmatrix} \mathbf{\Phi}_1 & \mathbf{0}^{n/r} & \cdots & \mathbf{0}^{n/r} \\ \mathbf{0}^{n/r} & \mathbf{\Phi}_2 & \cdots & \mathbf{0}^{n/r} \\ \vdots & \ddots & \ddots & \vdots \\ \mathbf{0}^{n/r} & \cdots & \mathbf{0}^{n/r} & \mathbf{\Phi}_r \end{pmatrix}$$

Compute $\mathbf{\Phi} = \sqrt{\frac{1}{r}}\mathbf{P}\mathbf{\Pi}\mathbf{W}$, where $\mathbf{W}$ is a randomized Hadamard matrix and $\mathbf{\Pi}$ is a permutation matrix

Thank you for your attention