UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

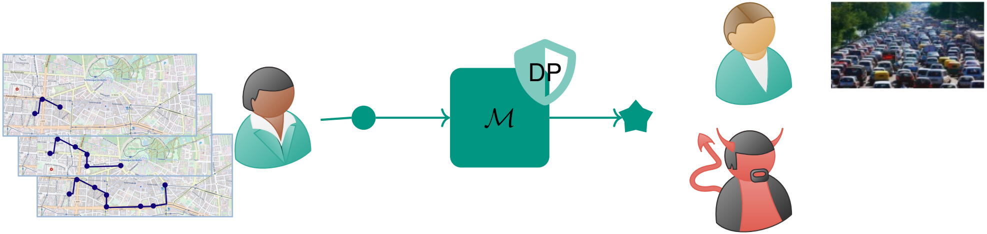
Karlsruhe Institute of Technology

Composability Properties of Differential Privacy for General Granularity Notions

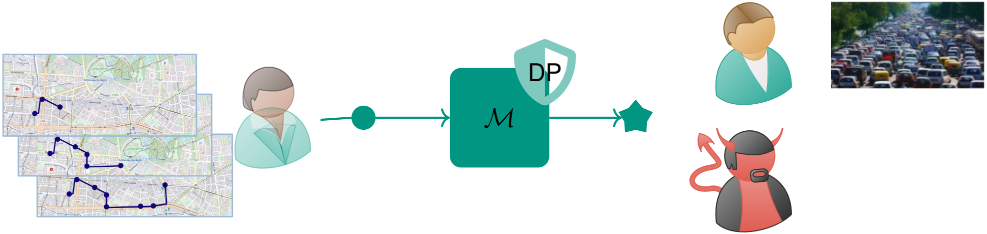
37th IEEE Computer Security Foundations Symposium

Patricia Guerra-Balboa, Àlex Miranda-Pascual, Javier Parra-Arnau, Thorsten Strufe | 12th July 2024

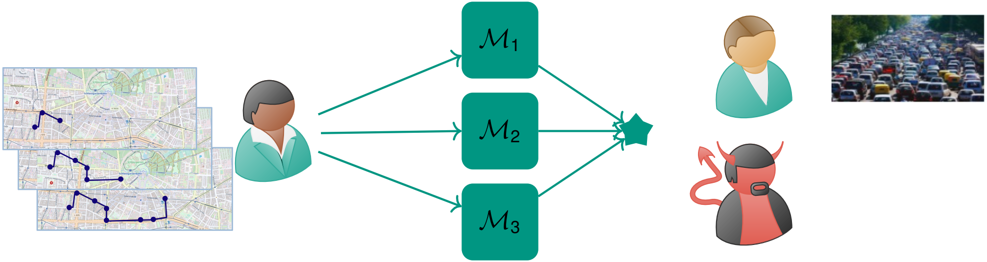
Differential Privacy



Differential Privacy

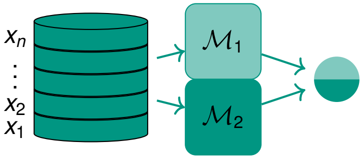


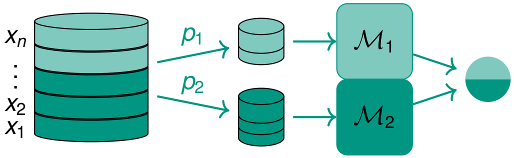
Differential Privacy



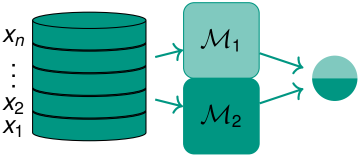
- **Composition:** We apply more than one mechanism to the database
  - To discretize a **complex** problem
  - To manage **continuous data releases**, for instance in **streaming**.

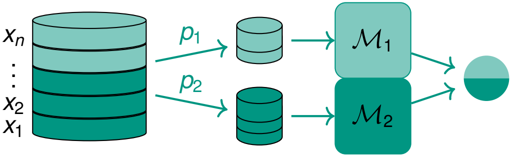
Related work: Composition until now





Related work: Composition until now

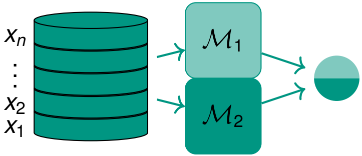


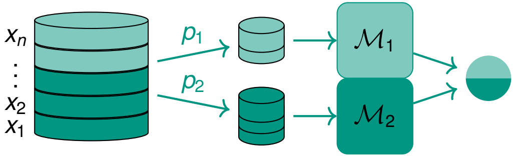


Sequential Composition Theorem

- Use-cases: streaming data, multiple query answering

Related work: Composition until now

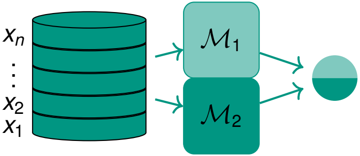


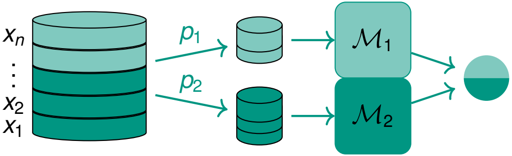


Sequential Composition Theorem

- Use-cases: streaming data, multiple query answering
- $x_i \in \mathcal{X}$

Related work: Composition until now

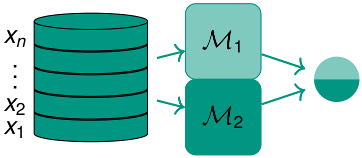


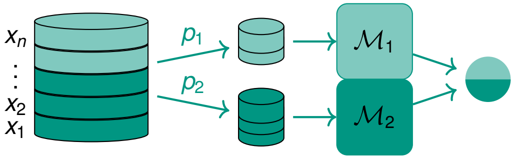


Sequential Composition Theorem

- Use-cases: streaming data, multiple query answering
- $x_i \in \mathcal{X}$
- $\mathcal{M}_i$ is unbounded DP in $\mathbb{D}_{\mathcal{X}}$
- $\mathcal{M} = (\mathcal{M}_1(D), \ldots, \mathcal{M}_k(D))$

Related work: Composition until now

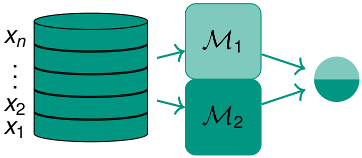




Sequential Composition Theorem

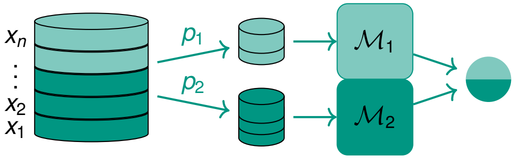
- Use-cases: streaming data, multiple query answering
- $x_i \in \mathcal{X}$
- $\mathcal{M}_i$ is unbounded DP in $\mathbb{D}_{\mathcal{X}}$
- $\mathcal{M} = (\mathcal{M}_1(D), \ldots, \mathcal{M}_k(D))$
- $\varepsilon = \sum_{i\in[k]} \varepsilon_i$

Related work: Composition until now



Sequential Composition Theorem

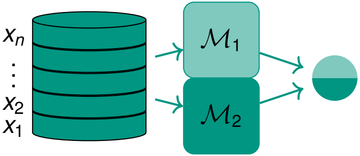
- Use-cases: streaming data, multiple query answering
- $x_i \in \mathcal{X}$
- $\mathcal{M}_i$ is unbounded DP in $\mathbb{D}_{\mathcal{X}}$
- $\mathcal{M} = (\mathcal{M}_1(D), \ldots, \mathcal{M}_k(D))$
- $\varepsilon = \sum_{i \in [k]} \varepsilon_i$



Parallel Composition Theorem

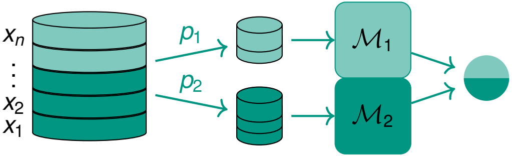
- Use-cases: federated learning, Mini-Batch training

Related work: Composition until now



Sequential Composition Theorem

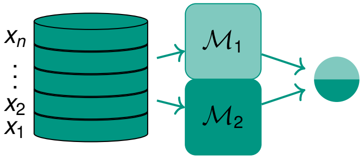
- Use-cases: streaming data, multiple query answering
- $x_i \in \mathcal{X}$
- $\mathcal{M}_i$ is unbounded DP in $\mathbb{D}_{\mathcal{X}}$
- $\mathcal{M} = (\mathcal{M}_1(D), \ldots, \mathcal{M}_k(D))$
- $\varepsilon = \sum_{i \in [k]} \varepsilon_i$



Parallel Composition Theorem

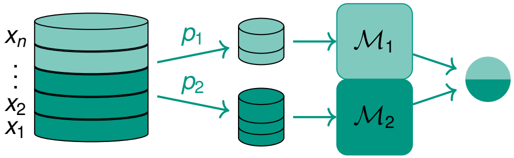
- Use-cases: federated learning, Mini-Batch training
- $x_i \in \mathcal{X}$ and $\{p_i\}$ partition of $\mathcal{X}$

Related work: Composition until now



Sequential Composition Theorem

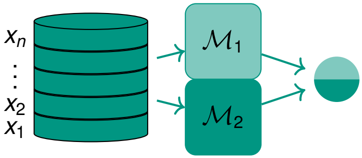
- Use-cases: streaming data, multiple query answering
- $x_i \in \mathcal{X}$
- $\mathcal{M}_i$ is unbounded DP in $\mathbb{D}_{\mathcal{X}}$
- $\mathcal{M} = (\mathcal{M}_1(D), \dots, \mathcal{M}_k(D))$
- $\varepsilon = \sum_{i \in [k]} \varepsilon_i$

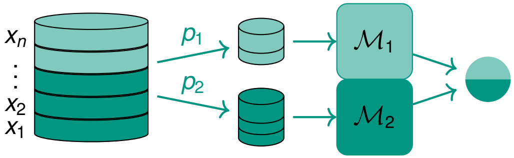


Parallel Composition Theorem

- Use-cases: federated learning, Mini-Batch training
- $x_i \in \mathcal{X}$ and $\{p_i\}$ partition of $\mathcal{X}$
- $\mathcal{M}_i$ unbounded DP in $\mathbb{D}_{\mathcal{X}_i}$
- $\mathcal{M} = (\mathcal{M}_1(p_1(D)), \dots, \mathcal{M}_k(p_k(D)))$

Related work: Composition until now





Sequential Composition Theorem

- Use-cases: streaming data, multiple query answering
- $x_i \in \mathcal{X}$
- $\mathcal{M}_i$ is unbounded DP in $\mathbb{D}_{\mathcal{X}}$
- $\mathcal{M} = (\mathcal{M}_1(D), \ldots, \mathcal{M}_k(D))$
- $\varepsilon = \sum_{i\in[k]} \varepsilon_i$

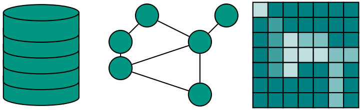
Parallel Composition Theorem

- Use-cases: federated learning, Mini-Batch training
- $x_i \in \mathcal{X}$ and $\{p_i\}$ partition of $\mathcal{X}$
- $\mathcal{M}_i$ unbounded DP in $\mathbb{D}_{\mathcal{X}_i}$
- $\mathcal{M} = (\mathcal{M}_1(p_1(D)), \ldots, \mathcal{M}_k(p_k(D)))$
- $\varepsilon = \max_{i\in[k]} \varepsilon_i$

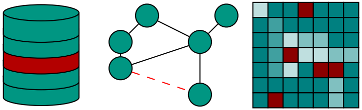
Problem Statement

Q1: Composition in general data domains and granularities

New data domains



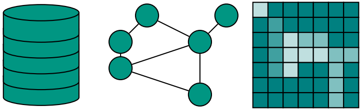
Other privacy requirements



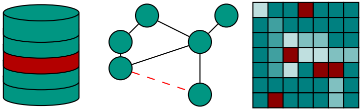
Problem Statement

Q1: Composition in general data domains and granularities

New data domains



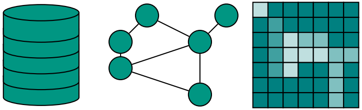
Other privacy requirements



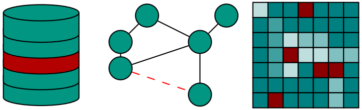
- ✔ New privacy requirements are modeled through the **neighborhood definition**, also called **granularity**

Q1: Composition in general data domains and granularities

New data domains



Other privacy requirements

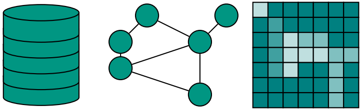


- ✓ New privacy requirements are modeled through the **neighborhood definition**, also called **granularity**
- $\Pr(\mathcal{M}(D) \in S) \leq e^{\varepsilon} \Pr(\mathcal{M}(D') \in S) \; \forall D \sim_{\mathcal{G}} D'$

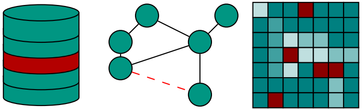
Problem Statement

Q1: Composition in general data domains and granularities

New data domains



Other privacy requirements

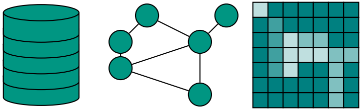


- ✔ New privacy requirements are modeled through the **neighborhood definition**, also called **granularity**
- $\Pr(\mathcal{M}(D) \in S) \leq e^{\varepsilon} \Pr(\mathcal{M}(D') \in S) \; \forall D \sim_{\mathcal{G}} D'$
- **?** Do composition theorems work with other granularities?

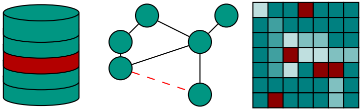
Problem Statement

Q1: Composition in general data domains and granularities

New data domains



Other privacy requirements



- ✔ New privacy requirements are modeled through the **neighborhood definition**, also called **granularity**
- $\Pr(\mathcal{M}(D) \in S) \leq e^{\varepsilon} \Pr(\mathcal{M}(D') \in S) \; \forall D \sim_{\mathcal{G}} D'$
- **?** Do composition theorems work with other granularities?
- ✘ NO $\rightarrow$ Parallel does not hold for bounded DP
  - ✘ In extreme cases leading to $\varepsilon = \infty$ privacy leakage

How can we compute the privacy leakage in general granularities?

Problem Statement

Q2: What happen when we use other composition strategies?

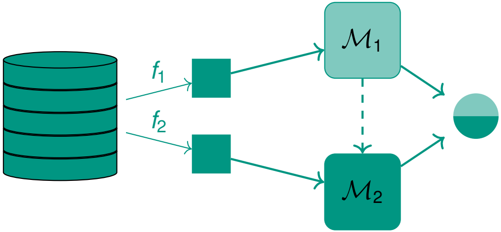
Parallel

Sequential

Problem Statement

Q2: What happen when we use other composition strategies?

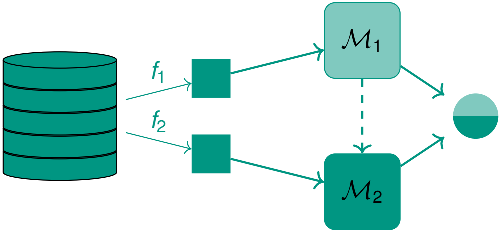
Parallel **Sequential**



Problem Statement

Q2: What happen when we use other composition strategies?

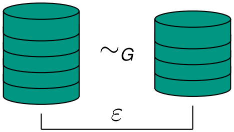
Parallel **Sequential**



Can we compute tighter bounds on the privacy leakage for arbitrary functions f?

d-privacy
Generalizing Differential Privacy by Chatzikokolakis at al.

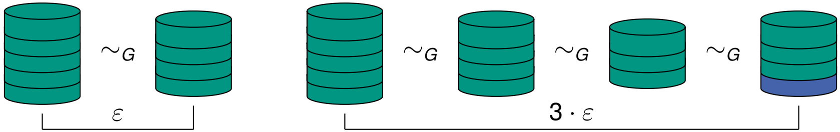
- Group Privacy: Given any granularity $\mathcal{G}$



d-privacy

Generalizing Differential Privacy by Chatzikokolakis at al.

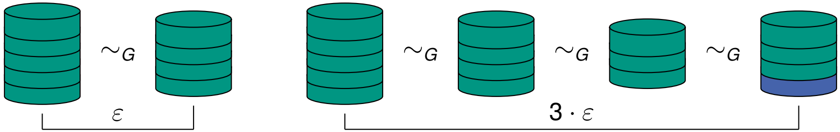
- Group Privacy: Given any granularity $\mathcal{G}$



d-privacy

Generalizing Differential Privacy by Chatzikokolakis at al.

- Group Privacy: Given any granularity $\mathcal{G}$



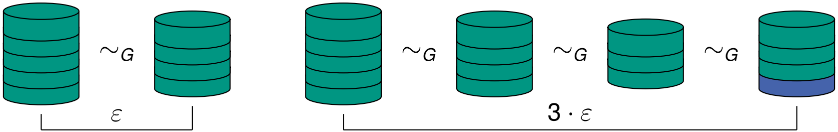
d-privacy

$\mathcal{M} : \mathbb{D} \to \mathrm{Range}(\mathcal{M})$ is d-private if for all $S \subseteq \mathrm{Range}(\mathcal{M})$

$$\mathrm{P}(\mathcal{M}(D) \in S) \leq \mathrm{e}^{d_{\mathbb{D}}(D,D')}\, \mathrm{P}(\mathcal{M}(D') \in S).$$

d-privacy
Generalizing Differential Privacy by Chatzikokolakis at al.

- Group Privacy: Given any granularity $\mathcal{G}$



d-privacy

$\mathcal{M} : \mathbb{D} \to \mathrm{Range}(\mathcal{M})$ is d-private if for all $S \subseteq \mathrm{Range}(\mathcal{M})$

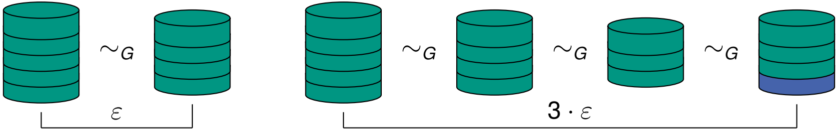
$$\mathrm{P}(\mathcal{M}(D) \in S) \leq \mathrm{e}^{d_{\mathbb{D}}(D,D')}\,\mathrm{P}(\mathcal{M}(D') \in S).$$

- $\mathbb{D}$ arbitrary data domain

d-privacy

Generalizing Differential Privacy by Chatzikokolakis at al.

- Group Privacy: Given any granularity $\mathcal{G}$



d-privacy

$\mathcal{M} : \mathbb{D} \to \mathrm{Range}(\mathcal{M})$ is *d*-private if for all $S \subseteq \mathrm{Range}(\mathcal{M})$

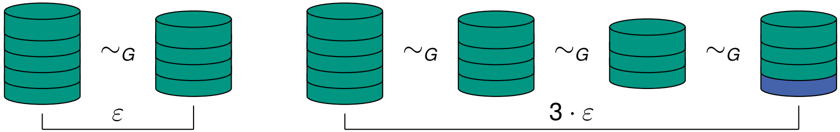
$$\mathrm{P}(\mathcal{M}(D) \in S) \leq \mathrm{e}^{d_{\mathbb{D}}(D,D')}\, \mathrm{P}(\mathcal{M}(D') \in S).$$

- $\mathbb{D}$ arbitrary data domain
- *d* sets the level of indistinguishability between two databases

d-privacy

Generalizing Differential Privacy by Chatzikokolakis at al.

- Group Privacy: Given any granularity $\mathcal{G}$



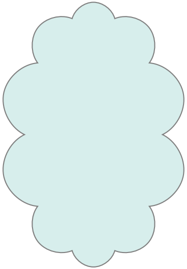
d-privacy

$\mathcal{M} : \mathbb{D} \to \mathrm{Range}(\mathcal{M})$ is d-private if for all $S \subseteq \mathrm{Range}(\mathcal{M})$

$$\mathrm{P}(\mathcal{M}(D) \in S) \leq \mathrm{e}^{d_{\mathbb{D}}(D,D')}\,\mathrm{P}(\mathcal{M}(D') \in S).$$

- $\mathbb{D}$ arbitrary data domain
- d sets the level of indistinguishability between two databases
- DP $\leftrightarrow$ d-privacy

General composition theorem

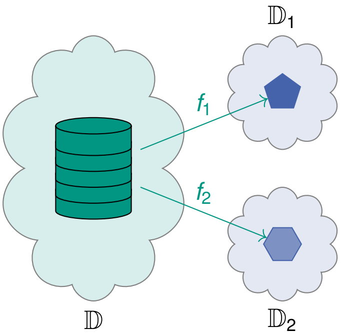


$\mathbb{D}$

General Composition Theorem

- $\mathbb{D}$ be a database class

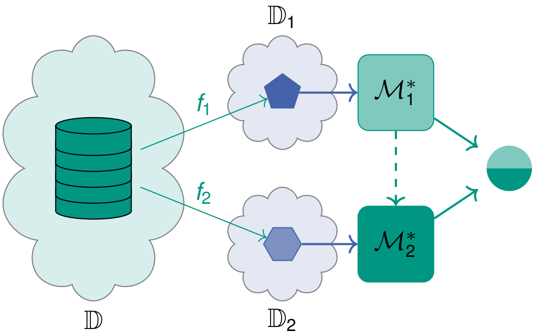
General composition theorem



General Composition Theorem

- $\mathbb{D}$ be a database class
- $f_i : \mathbb{D} \to \mathbb{D}_i$ be a deterministic map

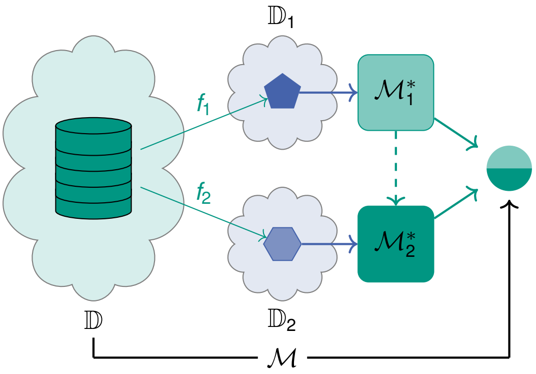
General composition theorem



General Composition Theorem

- $\mathbb{D}$ be a database class
- $f_i : \mathbb{D} \to \mathbb{D}_i$ be a deterministic map
- $\mathcal{M}_i^* : \mathbb{D}_i \to \mathcal{R}_i$ be d_i-private

General composition theorem



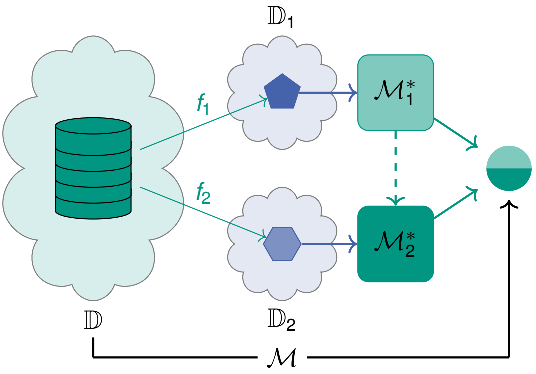
General Composition Theorem

- $\mathbb{D}$ be a database class
- $f_i : \mathbb{D} \to \mathbb{D}_i$ be a deterministic map
- $\mathcal{M}_i^* : \mathbb{D}_i \to \mathcal{R}_i$ be d_i-private

Then $\mathcal{M} = (\mathcal{M}_1^* \circ f_1, \ldots, \mathcal{M}_k^* \circ f_k)$ is $d_{\mathbb{D}}$-private with

$$d_{\mathbb{D}}(D, D') = \sum_{i=1}^{k} d_i(f_i(D), f_i(D')).$$

General composition theorem



- If $d(f_i(D), f_i(D')) = \infty \Rightarrow$ No privacy

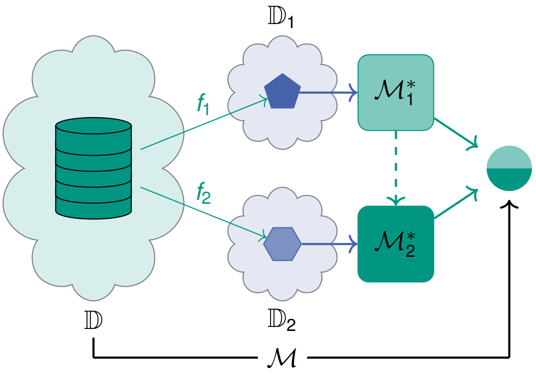
General Composition Theorem

- $\mathbb{D}$ be a database class
- $f_i : \mathbb{D} \to \mathbb{D}_i$ be a deterministic map
- $\mathcal{M}_i^* : \mathbb{D}_i \to \mathcal{R}_i$ be d_i-private

Then $\mathcal{M} = (\mathcal{M}_1^* \circ f_1, \ldots, \mathcal{M}_k^* \circ f_k)$ is $d_{\mathbb{D}}$-private with

$$d_{\mathbb{D}}(D, D') = \sum_{i=1}^{k} d_i(f_i(D), f_i(D')).$$

General composition theorem



General Composition Theorem

- $\mathbb{D}$ be a database class
- $f_i : \mathbb{D} \to \mathbb{D}_i$ be a deterministic map
- $\mathcal{M}_i^* : \mathbb{D}_i \to \mathcal{R}_i$ be d_i-private

Then $\mathcal{M} = (\mathcal{M}_1^* \circ f_1, \ldots, \mathcal{M}_k^* \circ f_k)$ is $d_{\mathbb{D}}$-private with

$$d_{\mathbb{D}}(D, D') = \sum_{i=1}^{k} d_i(f_i(D), f_i(D')).$$

- If $d(f_i(D), f_i(D')) = \infty \Rightarrow$ No privacy
- If $f_i(D) = f_i(D') \Rightarrow$ Tighter bound $\longrightarrow \sum_{i\,:\, f_i(D) \neq f_i(D')} r_i \varepsilon_i$

General composition theorem

General Composition Theorem

General composition theorem

General Composition Theorem

$$d_i(D, D') = \varepsilon_i |(D \cup D)' \backslash (D \cap D')|$$

&

p partition

$$\downarrow$$

$$\varepsilon = \max_i \varepsilon_i$$

General composition theorem

General Composition Theorem

$d_i(D, D') = \varepsilon_i |(D \cup D)' \setminus (D \cap D')|$
&
p partition
$\downarrow$
$\varepsilon = \max_i \varepsilon_i$

For all $d_i(D, D')$
&
$f = id$
$\downarrow$
$d = \sum_i d_i$

General composition theorem

General Composition Theorem

$d_i(D, D') = \varepsilon_i |(D \cup D)' \setminus (D \cap D')|$
&
p partition
$\downarrow$
$\varepsilon = \max_i \varepsilon_i$

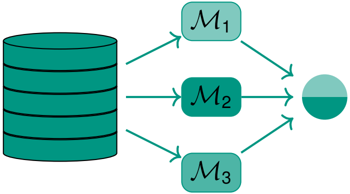
For all $d_i(D, D')$
&
$f = id$
$\downarrow$
$d = \sum_i d_i$

We derive the **conditions** needed to obtain $\max_i \varepsilon_i$

We give examples of intermediate bounds between sequential and parallel

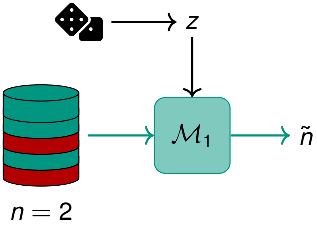
We derive a **privacy amplification** respect to sequential composition in the “common-domain setting”

The common domain setting



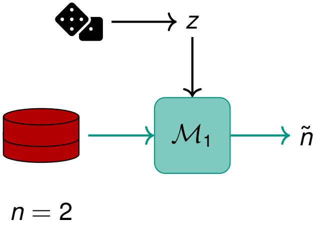
- Generalized Sequential: $d = \sum d_i \ (\sum \varepsilon_i)$

The common domain setting



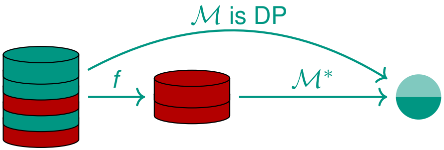
- Generalized Sequential: $d = \sum d_i \ (\sum \varepsilon_i)$
- $\mathcal{M}(D) = |D_{\leq 18}| + Z$ with $Z \sim Lap(\frac{\Delta}{\varepsilon})$

The common domain setting



- Generalized Sequential: $d = \sum d_i \ (\sum \varepsilon_i)$
- $\mathcal{M}(D) = |D_{\leq 18}| + Z$ with $Z \sim Lap(\frac{\Delta}{\varepsilon})$
- $\Pr(\mathcal{M}(D) \in S) = \Pr(\mathcal{M}(D_{\leq 18}) \in S)$

The common domain setting



- Generalized Sequential: $d = \sum d_i \; (\sum \varepsilon_i)$
- $\mathcal{M}(D) = |D_{\leq 18}| + Z$ with $Z \sim Lap(\frac{\Delta}{\varepsilon})$
- $\Pr(\mathcal{M}(D) \in S) = \Pr(\mathcal{M}(D_{\leq 18}) \in S)$
- We say that $\mathcal{M}$ is ***f*-dependent** if there exists $\mathcal{M}^*$ with domain $f(\mathbb{D})$ such that

$$\mathcal{M} = \mathcal{M}^* \circ f.$$

Tighter composition bound under f-dependency

Theorem

- $\mathbb{D}$ be a database class

Tighter composition bound under f-dependency

Theorem

- $\mathbb{D}$ be a database class
- $f_i : \mathbb{D} \to \mathbb{D}_i$ be a deterministic map

Tighter composition bound under f-dependency

Theorem

- $\mathbb{D}$ be a database class
- $f_i : \mathbb{D} \to \mathbb{D}_i$ be a deterministic map
- $\mathcal{M} : \mathbb{D} \to \mathcal{R}_i$ be d_i-private and f_i-dependent

Tighter composition bound under f-dependency

Theorem

- $\mathbb{D}$ be a database class
- $f_i : \mathbb{D} \to \mathbb{D}_i$ be a deterministic map
- $\mathcal{M} : \mathbb{D} \to \mathcal{R}_i$ be d_i-private and f_i-dependent

Then $\mathcal{M} = (\mathcal{M}_1, \dots, \mathcal{M}_k)$ is $d_{\mathbb{D}}$-private with

$$d_{\mathbb{D}} = \sum_{i=1}^{k} d_i^{f_i} \leq \sum_{i=1}^{k} d_i$$

Tighter composition bound under f-dependency

Theorem

- $\mathbb{D}$ be a database class
- $f_i : \mathbb{D} \to \mathbb{D}_i$ be a deterministic map
- $\mathcal{M} : \mathbb{D} \to \mathcal{R}_i$ be d_i-private and f_i-dependent

Then $\mathcal{M} = (\mathcal{M}_1, \ldots, \mathcal{M}_k)$ is $d_{\mathbb{D}}$-private with

$$d_{\mathbb{D}} = \sum_{i=1}^{k} d_i^{f_i} \leq \sum_{i=1}^{k} d_i$$

$$d_i^{f_i}(D, D') = \min_{\substack{\tilde{D}, \tilde{D}' \in \mathbb{D} \\ f_i(\tilde{D}) = f_i(D) \\ f_i(\tilde{D}') = f_i(D')}} d_i(\tilde{D}, \tilde{D}').$$

Tighter composition bound under f-dependency

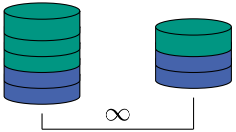
Theorem

- $\mathbb{D}$ be a database class
- $f_i : \mathbb{D} \to \mathbb{D}_i$ be a deterministic map
- $\mathcal{M} : \mathbb{D} \to \mathcal{R}_i$ be d_i-private and f_i-dependent

Then $\mathcal{M} = (\mathcal{M}_1, \ldots, \mathcal{M}_k)$ is $d_{\mathbb{D}}$-private with

$$d_{\mathbb{D}} = \sum_{i=1}^{k} d_i^{f_i} \leq \sum_{i=1}^{k} d_i$$

$$d_i^{f_i}(D, D') = \min_{\substack{\tilde{D}, \tilde{D}' \in \mathbb{D} \\ f_i(\tilde{D}) = f_i(D) \\ f_i(\tilde{D}') = f_i(D')}} d_i(\tilde{D}, \tilde{D}').$$



Tighter composition bound under f-dependency

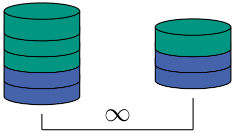
Theorem

- $\mathbb{D}$ be a database class
- $f_i : \mathbb{D} \to \mathbb{D}_i$ be a deterministic map
- $\mathcal{M} : \mathbb{D} \to \mathcal{R}_i$ be d_i-private and f_i-dependent

Then $\mathcal{M} = (\mathcal{M}_1, \ldots, \mathcal{M}_k)$ is $d_{\mathbb{D}}$-private with

$$d_{\mathbb{D}} = \sum_{i=1}^{k} d_i^{f_i} \leq \sum_{i=1}^{k} d_i$$

$$d_i^{f_i}(D, D') = \min_{\substack{\tilde{D}, \tilde{D}' \in \mathbb{D} \\ f_i(\tilde{D}) = f_i(D) \\ f_i(\tilde{D}') = f_i(D')}} d_i(\tilde{D}, \tilde{D}').$$

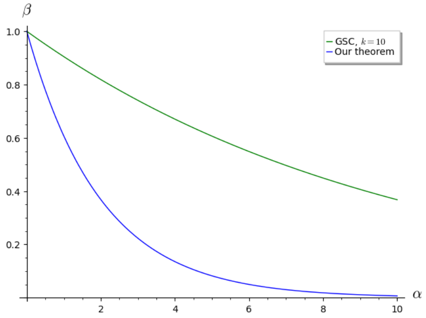


$$d^f(D, D') = 0$$

Improved privacy leakage bound for bounded DP

Corollary

Let p be a k-partitioning function. For all $i \in [k]$, let $\mathcal{M}_i : \mathbb{D} \to \mathcal{R}_i$ be mechanisms satisfying bounded ε_i-DP and p_i-dependent. Then mechanism $\mathcal{M} = (\mathcal{M}_1, \ldots, \mathcal{M}_k)$ with domain $\mathbb{D}$ is bounded ε-DP with $\varepsilon = \max_{i,j \in [k];\, i \neq j}(\varepsilon_i + \varepsilon_j)$.



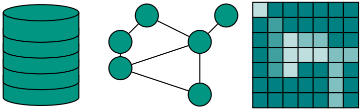
- ✔ Thanks to our theorem we have an improved bound

$$\varepsilon = \max_{i,j \in [k];\, i \neq j}(\varepsilon_i + \varepsilon_j) < \sum_{i=1}^{k} \varepsilon_i$$

Conclusions

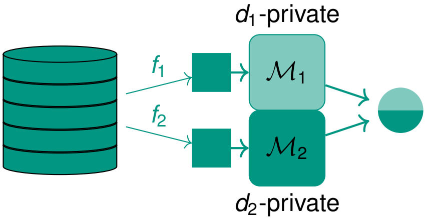
Conclusions

- ✓ Composability is inherent to DP



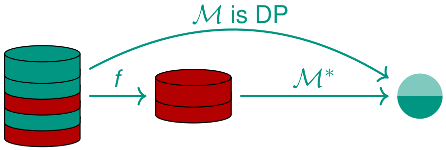
Conclusions

- ✓ Composability is inherent to DP
- ✓ We provide a tighter privacy leakage estimation under
  - ✓ any composition strategy & pre-processing functions
  - ✓ under mixed privacy requirements



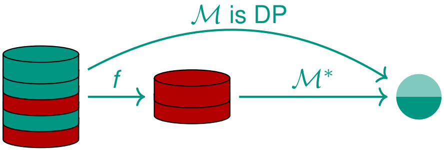
Conclusions

- ✓ Composability is inherent to DP
- ✓ We provide a tighter privacy leakage estimation under
  - ✓ any composition strategy & pre-processing functions
  - ✓ under mixed privacy requirements
- ✓ We provide tighter privacy bounds under mechanisms dependencies. Solving the problem of bounded parallel.



Conclusions

- ✓ Composability is inherent to DP
- ✓ We provide a tighter privacy leakage estimation under
  - ✓ any composition strategy & pre-processing functions
  - ✓ under mixed privacy requirements
- ✓ We provide tighter privacy bounds under mechanisms dependencies. Solving the problem of bounded parallel.
- ✓ All theorems have been extended to approximate, zero-concentrated and Gaussian DP.



Conclusions

- Composability is inherent to DP
- We provide a tighter privacy leakage estimation under
  - any composition strategy & pre-processing functions
  - under mixed privacy requirements
- We provide tighter privacy bounds under mechanisms dependencies. Solving the problem of bounded parallel.
- All theorems have been extended to approximate, zero-concentrated and Gaussian DP.

Thanks for your attention!

Figure: For more details check our paper