Unlocking the Potential of Composition for General Neighborhood Definitions

Àlex Miranda-Pascual*
Karlsruhe Institute of Technology, Germany
Universitat Politècnica de Catalunya, Spain
alex.pascual@kit.edu

Patricia Guerra-Balboa*
Karlsruhe Institute of Technology, Germany
patricia.balboa@kit.edu

Javier Parra-Arnau
Universitat Politècnica de Catalunya, Spain
javier.parra@upc.edu

Thorsten Strufe
Karlsruhe Institute of Technology, Germany
strufe@kit.edu

***Abstract*—This paper is an extended summary of previously published results [1].**

The composability properties of differential privacy (DP) are key to the construction of most DP algorithms. However, the new neighborhood definitions and data domains in the literature are not covered by the original composition theorems. For instance, the parallel composition theorem does not translate well to general neighborhoods. These limitations make it difficult to compute accurate estimates of the privacy loss when composing DP mechanisms in new settings.

To overcome this problem, we prove a general composition theorem in a general framework, defined for any kind of data domain or neighborhood definition. We also study the hypothesis needed to obtain the best composition bounds. Our theorems cover both parallel and sequential composition settings, as well as any effect of preprocessing, allowing us to compute the final privacy loss of a composition with greatly improved accuracy.

I. Introduction

Differential privacy (ε-DP) [2] is a well-known privacy notion in the field of data protection. One advantage of DP over other privacy notions is that DP possesses the key property of *composability*: It is possible to form a new DP mechanism by composing a finite number of given DP mechanisms and to directly compute the privacy provided as a result of this composition. Composability is key for the construction of most DP algorithms, in particular, the privacy protection of adaptive updates (e.g., in a streaming scenario or private learning) requires composition to be computed.

Currently, DP composition is represented by two results: *sequential composition* [2] and *parallel composition* [3]. Parallel composition is applied when all combined mechanisms access mutually disjoint databases, the maximum loss before combination corresponding to the total privacy loss after composition. Sequential composition covers any case when arbitrary DP mechanisms with access to the entire data are combined. The total privacy loss in sequential composition is computed as the sum of the losses of each composed mechanism.

The problem is that the sequential and parallel composition theorems were originally stated for tabular databases in the *unbounded* [4] scenario and do not easily adapt to the newer settings. Nowadays, the literature works both with different *database domains* (i.e., classes of the input databases of a privacy mechanism) and with different *neighborhood definitions* (also called *granularity notions* [2]), such as *bounded DP*, *edge-DP* and *free-lunch DP* [5] and the existing composition theorems may not extend directly in these cases. For instance, Li, Lyu, Su, *et al.* [6] show that the proof of the parallel composition theorem [3] does not hold if we change the original granularity to bounded DP. Since composition for new granularities is non-trivial or even impossible, the lack of a general composition framework risks misapplying DP composition, for example, by using parallel composition in a bounded scenario.

To provide a context where all granularities can be composed and where the final privacy loss can be systematically computed, we use a general mathematical framework based on the notion of d-privacy [7] to present the general composition theorem (III.1). Our results compute the privacy loss after composition in any domain and granularity notion, both existing and future, and even allow combining different domains and granularity notions. Besides, our result allow a more accurate calculation of the privacy loss upon any possible composition of DP mechanisms, including the effect of preprocessing, leading to better bounds than the sum obtained using sequential composition (see Example III.2).

Furthermore, we study the application of our theorem to the setting where we compose over disjoint databases (as in parallel composition). Here, we derive sufficient conditions to obtain the "$\max \varepsilon_i$" bound. Besides, for the cases where this bound cannot be achieved, we provide a new variation on composition (see Section III-A) that allows us to achieve better results. In particular, we provide a solution to the open problem of Li, Lyu, Su, *et al.* [6] by giving the lowest possible privacy loss for the composition of bounded DP mechanisms executed on mutually disjoint databases (Corollary III.5).

Our contributions are as follows:

- We prove the general composition theorem (III.1), a new result that allows us to reduce the estimated privacy loss and design improved DP mechanisms in general contexts. We propose corollaries for composition over disjoint inputs.
- We introduce a new setting that allows us to reduce the

*These authors contributed equally.

privacy loss in some cases. In particular, we provide the minimum privacy loss for the bounded case when the mechanism accesses disjoint parts of the database.

All proofs are in the long version of [1] (arXiv:2308.14649).

II. Background

Differential privacy (DP) [2], in its original definition, aims to hide the presence or absence of any private record in the database such that an analyst can extract statistics about the whole population while limiting the ability of an adversary to learn private information about individuals. Formally,

Definition II.1 (Differential privacy [2])**.** Let $\mathbb{D}_{\mathcal{X}}$ be the class of all databases drawn from $\mathcal{X}$ and $\varepsilon \geq 0$. We say a randomized mechanism $\mathcal{M}$ with domain $\mathbb{D}_{\mathcal{X}}$ is *ε-differentially private* (ε-DP) if for all *neighboring* $D, D' \in \mathbb{D}_{\mathcal{X}}$ and all measurable $S \subseteq \mathrm{Range}(\mathcal{M})$,

$$\mathrm{P}\{\mathcal{M}(D) \in S\} \leq \mathrm{e}^{\varepsilon}\, \mathrm{P}\{\mathcal{M}(D') \in S\}. \tag{II.1}$$

An important part of DP is the concept of *neighborhood*, also referred to as the *granularity notion* of DP [2]. In Definition II.1, two databases $D, D' \in \mathbb{D}_{\mathcal{X}}$ are *neighboring* if and only if they "differ" on at most one element, i.e., $|D \triangle D'| := |(D \cup D') \backslash (D \cap D')| \leq 1$. This granularity notion, called the *unbounded* granularity notion [4], protects the presence or absence of a record in a database. However, in many use cases, one aims to protect a different property about the dataset, e.g., the value of a specific sensitive attribute, or entire groups of individuals. To adapt DP to protect different sensitive properties one must change the granularity in the original definition. We can define any granularity over any class of databases $\mathbb{D}$ as follows:

Definition II.2 ($\mathcal{G}$-neighborhood)**.** Given a database class $\mathbb{D}$, we define the *$\mathcal{G}$-neighborhood relation* as a binary symmetric relation $\sim_{\mathcal{G}}$ between elements in $\mathbb{D}$. We say that $D, D' \in \mathbb{D}$ are *$\mathcal{G}$-neighboring* if $D \sim_{\mathcal{G}} D'$.

In this case, we say a mechanism $\mathcal{M}$ with domain $\mathbb{D}$ is *$\mathcal{G}$ ε-DP* ($\varepsilon \geq 0$) if it verifies $\mathrm{P}\{\mathcal{M}(D) \in S\} \leq \mathrm{e}^{\varepsilon}\, \mathrm{P}\{\mathcal{M}(D') \in S\}$ for all $D \sim_{\mathcal{G}} D'$ in $\mathbb{D}$ and all measurable $S \subseteq \mathrm{Range}(\mathcal{M})$.

We denote the unbounded neighborhood by $\mathcal{U}$. Another particularly popular granularity notion is the *bounded* ($\mathcal{B}$) neighborhood definition [4]. A pair of databases $D, D' \in \mathbb{D}$ are *bounded neighboring* if D can be obtained from D' by changing the value of exactly one record for another (i.e., the Hamming distance between D and D' is one). Therefore, *bounded DP* protects the values of the records [5], and it is broadly used in several contexts such as private learning.

Note that, given a data domain $\mathbb{D}$, we can construct a *canonical metric* $d_{\mathbb{D}}^{\mathcal{G}}$ for each granularity $\mathcal{G}$ over $\mathbb{D}$ by defining the distance $d_{\mathbb{D}}^{\mathcal{G}}(D, D')$ as the minimum number of neighboring databases in $\mathbb{D}$ you need to cross to obtain D' from D (with $d_{\mathbb{D}}^{\mathcal{G}}(D, D') = \infty$ if it is not possible) [7]. In particular, $d_{\mathbb{D}}^{\mathcal{G}}(D, D') = 1$ if and only if $D \sim_{\mathcal{G}} D'$ (and $D \neq D'$).

Then, from the group property of DP [3], $\mathcal{M}$ is $\mathcal{G}$ ε-DP if and only if for all $D, D' \in \mathbb{D}$ and all measurable $S \subseteq \mathrm{Range}(\mathcal{M})$,

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

This property motivates using metrics over $\mathbb{D}$ to measure privacy protection, which is formalized with $d_{\mathbb{D}}$-privacy [7]. Note that having an (extended pseudo-)metric $d_{\mathbb{D}}$ implies that $(\mathbb{D}, d_{\mathbb{D}})$ is an (extended pseudo-)metric space, which we will call *privacy space*. Thus, $d_{\mathbb{D}}$-privacy is defined as follows:

Definition II.3 ($d_{\mathbb{D}}$-privacy [7])**.** Let $(\mathbb{D}, d_{\mathbb{D}})$ be a privacy space. Then, a randomized mechanism $\mathcal{M}$ with domain $\mathbb{D}$ is *$d_{\mathbb{D}}$-private* if for all $D, D' \in \mathbb{D}$ and all measurable $S \subseteq \mathrm{Range}(\mathcal{M})$, $\mathrm{P}\{\mathcal{M}(D) \in S\} \leq \mathrm{e}^{d_{\mathbb{D}}(D,D')}\, \mathrm{P}\{\mathcal{M}(D') \in S\}$.

A direct relation exists between original DP and $d_{\mathbb{D}}$-privacy: For any granularity notion $\mathcal{G}$ over $\mathbb{D}$, a mechanism $\mathcal{M}$ with domain $\mathbb{D}$ is $\varepsilon d_{\mathbb{D}}^{\mathcal{G}}$-private if and only if it is $\mathcal{G}$ ε-DP [1].

A. Introduction to the Composition Theorems

In many cases, the information we need to extract can be obtained as a function of some query answers. This can be modeled as $s = h(s_1, \ldots, s_k)$, where h is an arbitrary deterministic function and $s_i = f_i(D)$ is the output of an arbitrary query (where f_i can even be the identity). Instead of protecting s directly, we can discretize the problem by obtaining $\tilde{s}_i$, the private output of s_i, and computing $\tilde{s} = h(\tilde{s}_1, \ldots, \tilde{s}_k)$. To do this, we take k ε-DP mechanisms $\mathcal{M}_i^*$ such that $\mathcal{M}_i^*(f_i(D)) = \tilde{s}_i$ and consider the composed mechanism $\mathcal{M}$ (i.e., such that $\mathcal{M}(D) = (\mathcal{M}_1^*(f_1(D)), \ldots, \mathcal{M}_k^*(f_k(D)))$ for all $D \in \mathbb{D}$). Being able to derive the privacy level of $\mathcal{M}$ from those of $\mathcal{M}_i^*$ becomes a useful tool: The question arises whether the composition of the mechanisms $\mathcal{M}$ is ε-DP and what the value of ε would be in this case.

The original DP answers this question thanks to the *composability property*: The composition $\mathcal{M}$ of k ε_i-DP mechanisms $\mathcal{M}_i$ is also ε-DP, where ε depends directly on $\varepsilon_1, \ldots, \varepsilon_k$. This property holds even if $\mathcal{M}_i^*(f_i(D))$ uses the output of $\mathcal{M}_j^*(f_j(D))$ for any or all $j < i$ in its computation (the so-called *adaptative* composition), which we will implicitly assume possible for the rest of the paper.

In the classic unbounded scenario of DP, there are two composition theorems: the *sequential* and *parallel composition theorems*. Sequential composition refers to the case where every $\mathcal{M}_i^*$ takes as input the whole database D (i.e., $f_i = \mathrm{id}$):

Theorem II.4 (Sequential composition (SC) [2])**.** *For $i \in [k]$, let $\mathcal{M}_i$ with domain $\mathbb{D}_{\mathcal{X}}$ be an [unbounded] ε_i-DP mechanism. Consider the mechanism $\mathcal{M}$ with domain $\mathbb{D}$ such that $\mathcal{M}(D) = (\mathcal{M}_1(D), \ldots, \mathcal{M}_k(D))$ for all $D \in \mathbb{D}_{\mathcal{X}}$. Then $\mathcal{M}$ is [unbounded] $(\sum_{i=1}^{k} \varepsilon_i)$-DP.*

Alternatively, the composition is *parallel* if each $\mathcal{M}_i^*$ uses only data from a subset $D_i \subseteq D$ that is not used by any other, i.e., f_i defines a partition. Formally,

Definition II.5 (Partitioning function)**.** A partition $\{\mathcal{X}_i\}_{i \in [k]}$ of $\mathcal{X}$ extends naturally as a partition function $p = \{p_i\}_{i \in [k]}$ of the elements $D \in \mathbb{D}_{\mathcal{X}}$, i.e., $p_i(D) \subseteq D$ is the multiset such that element $x \in D$ has multiplicity $m_{p_i(D)}(x) = \mathbf{1}_{\mathcal{X}_i}(x)\, m_D(x)$.

Theorem II.6 (Parallel composition [6])**.** *Let $\{p_i(D)\}_{i \in [k]}$ be a partition of D for all $D \in \mathbb{D}_{\mathcal{X}}$ as defined in Definition II.5. For $i \in [k]$, let $\mathcal{M}_i$ with domain $\{p_i(D) \mid D \in \mathbb{D}_{\mathcal{X}}\}$ be an [unbounded] ε_i-DP mechanism. Consider the mechanism $\mathcal{M}$ with domain $\mathbb{D}_{\mathcal{X}}$ such that $\mathcal{M}(D) = (\mathcal{M}_1(p_1(D)), \ldots, \mathcal{M}_k(p_k(D)))$ for all $D \in \mathbb{D}_{\mathcal{X}}$. Then $\mathcal{M}$ is [unbounded] $(\max_{i \in [k]} \varepsilon_i)$-DP.*

Parallel composition is highly desirable because it provides a lower bound than the one obtained through the more general sequential composition. However, unlike sequential composition that extends from unbounded DP to other granularity notions without problem and without altering the expression final privacy budget $\sum_{i=1}^{n} \varepsilon_i$ [8], parallel composition does not extend well for other granularity notions and general $\mathbb{D}$: For example, if we consider the same hypothesis of Theorem II.6 but we impose $\mathcal{M}_i$ to be bounded DP instead, then we can obtain that $\mathcal{M}$ is not bounded DP [1].

Similarly, the generalization of Theorem II.4 to $d_{\mathbb{D}}$-privacy is direct [7]. However, the extension of parallel composition remains unexplored for general granularities. This opens a new question about how to measure the privacy of composed mechanisms in general. To answer this question, we introduce general composition rules that allow composing DP mechanisms with different domains and granularities.

III. The General Composition Theorem

In this section, we introduce our generalized version of the composition results. Since the theorem does not impose any condition on the privacy metric of the initial $\mathcal{M}_i$, our results can be used for any privacy space and any possible independent composition strategy, generalizing both Theorems II.4 and II.6.

Theorem III.1 (Generalized composition theorem)**.** *Let $\mathbb{D}$ be a database class and, for all $i \in [k]$, let $(\mathbb{D}_i, d_i)$ be a privacy space, and let $f_i \colon \mathbb{D} \to \mathbb{D}_i$ be a deterministic map. For all $i \in [k]$, let $\mathcal{M}_i^* \colon \mathbb{D}_i \to \mathcal{R}_i$ be d_i-private mechanisms. Then mechanism $\mathcal{M} = (\mathcal{M}_1^* \circ f_1, \dots, \mathcal{M}_k^* \circ f_k)$ is $d_{\mathbb{D}}$-private with*

$$d_{\mathbb{D}}(D, D') \coloneqq \sum_{i=1}^{k} d_i(f_i(D), f_i(D')) \quad \text{for all } D, D' \in \mathbb{D}.$$

When we apply this theorem to $\mathcal{M}_i$ that satisfy $\mathcal{G}_i$ ε_i-DP mechanism we obtain that the composed mechanism $\mathcal{M}$ is $\mathcal{G}$ ε-DP with

$$\varepsilon = \max_{D \sim_{\mathcal{G}} D'} \sum_{i=1}^{k} \varepsilon_i d_{\mathbb{D}_i}^{\mathcal{G}_i}(f_i(D), f_i(D')) \leq \max_{D \sim_{\mathcal{G}} D'} \sum_{i \colon f_i(D) \neq f_i(D')} r_i \varepsilon_i$$

where $r_i \coloneqq \max_{D \sim_{\mathcal{G}} D'} d_{\mathbb{D}}^{\mathcal{G}_i}(f_i(D), f_i(D'))$ for any well-defined granularity $\mathcal{G}$ in the domain $\mathbb{D}$.

It is important to note that our composition theorem (III.1) provides the privacy level of the resulting mechanism by construction. This means that we cannot generally impose the privacy level of the composed mechanism $\mathcal{M}$, but we can compute it. In particular, sequential composition for $d_{\mathbb{D}}$-privacy can be obtained for all metrics: If we select $\mathbb{D}_i = \mathbb{D}$ and $f_i = \mathrm{id}$, we obtain that $\mathcal{M}$ is $(\sum_{i=1}^{k} d_i)$-private. We can also end up with extreme cases where $d_{\mathbb{D}}(D, D') = \infty$ for certain $D, D' \in \mathbb{D}$, which does not provide privacy between these databases. Theorem III.1 showcases the effect of preprocessing and allows considering interesting composition strategies that provide tighter, more precise bounds than sequential composition even if the inputs are not disjoint:

Example III.2. Consider a database class $\mathbb{D} = \mathbb{D}_{\mathcal{X}}$ where each record $x \in \mathcal{X}$ corresponds to an ambulance and includes its position and the labels of at least three hospitals associated with the ambulance. We want to know the number of available ambulances for each of the k hospitals, so we consider $\mathcal{M}$ such that $\mathcal{M}(D) = (\mathcal{M}^*(f_1(D)), \dots, \mathcal{M}^*(f_k(D)))$ where $\mathcal{M}^*$ outputs the noisy count of records in its input, and $f_i(D)$ is the subdatabase of $D \in \mathbb{D}$ of ambulances assigned to hospital i. Since each ambulance only collaborates with at most three hospitals, for all $D \sim_{\mathcal{U}} D'$ there are at most three indices $i \in [k]$ such that $f_i(D) \sim_{\mathcal{U}} f_i(D')$, and $f_i(D) = f_i(D')$ for all other indices. Applying then the composition theorem (III.1), we obtain that $\mathcal{M}$ is $d_{\mathbb{D}}$-private with $d_{\mathbb{D}}(D, D') = \sum_{i=1}^{k} d_{\mathbb{D}}^{\mathcal{U}}(f_i(D), f_i(D')) \leq 3 d_{\mathbb{D}}^{\mathcal{U}}(D, D') < k d_{\mathbb{D}}^{\mathcal{U}}(D, D')$.

Unlike the sequential setting, the case in which the mechanisms take as input disjoint subsets of the initial database (as in parallel composition) does not generally yield analogous results to Theorem II.6. If we take $\{f_i\}_{i \in [k]} = \{p_i\}_{i \in [k]}$, a *$k$-partitioning function* (as in Definition II.5), and we apply Theorem III.1, we obtain that $\mathcal{M}$ is $d_{\mathbb{D}}$-private with $d_{\mathbb{D}}(D, D') = \sum_{i=1}^{k} d_i(p_i(D), p_i(D')) \leq I_p(D, D')(\max_{i \in [k]} \Delta p_i \, d_i(D, D'))$ for all $D, D' \in \mathbb{D}$, where $I_p(D, D') \coloneqq \#\{i \mid p_i(D) \neq p_i(D')\}$ and Δp_i is the smallest value such that $d_i(p_i(D), p_i(D')) \leq \Delta p_i \, d_i(D, D')$. This fact is coherent with what we know: Assuming $\mathbb{D} = \mathbb{D}_{\mathcal{X}}$ and $\mathbb{D}_i = p_i(\mathbb{D})$, if $d_i = \varepsilon_i d_{\mathbb{D}_i}^{\mathcal{U}}$, then $I_p(D, D') = 1$ and $\Delta p_i \leq 1$ for all $D \sim_{\mathcal{U}} D'$, and therefore $d_{\mathbb{D}} = (\max_{i \in [k]} \varepsilon_i) d_{\mathbb{D}}^{\mathcal{U}}$. However, if we consider instead bounded DP mechanisms, then there exist $D, D' \in \mathbb{D}$ such that $d_i(D, D') = d_{\mathbb{D}_i}^{\mathcal{B}}(p_i(D), p_i(D')) = \infty$ for some i and thus $d_{\mathbb{D}}(D, D') = \infty$ [1]. Therefore, we have no better expression for $d_{\mathbb{D}}$ without imposing extra conditions.

For canonical metrics over a granularity $\mathcal{G}$, the best bound generally achievable is the maximum over the privacy bounds of $\mathcal{M}_i$, or formally: If $\mathcal{M}_i$ are $\varepsilon_i d_{\mathbb{D}_i}^{\mathcal{G}}$-private mechanisms, then the composed mechanism $\mathcal{M}$ is $(\max_{i \in [k]} \varepsilon_i) d_{\mathbb{D}}^{\mathcal{G}}$-private. This bound is achieved for unbounded DP (i.e., Theorem II.6), but not for bounded DP. To ensure we achieve the best bound, it is sufficient that $\sum_{i=1}^{k} d_{\mathbb{D}_i}^{\mathcal{G}}(p_i(D), p_i(D')) = d_{\mathbb{D}}^{\mathcal{G}}(D, D')$. This equation can be hard to check in general, but it holds if the partitioning function verifies:

(C1) $d_{\mathbb{D}}^{\mathcal{G}}$*-compatibility*: For all $\mathcal{G}$-neighboring $D, D' \in \mathbb{D}$, there exists at most one $j \in [k]$ such that $p_i(D) = p_i(D')$ for all $i \neq j$, i.e., $I_p(D, D') \leq 1$ for all $D \sim_{\mathcal{G}} D'$; and

(C2) the sensitivity Δp_i of p_i with respect to $d_{\mathbb{D}}^{\mathcal{G}}$ and $d_{\mathbb{D}_i}^{\mathcal{G}}$ is at most 1 (i.e., $d_{\mathbb{D}_i}^{\mathcal{G}}(p_i(D), p_i(D')) \leq 1$ if $d_{\mathbb{D}}^{\mathcal{G}}(D, D') = 1$).

However, these conditions may be difficult to achieve. In particular, for the bounded granularity notion we have that $I_p(D, D')$ is either 1 or 2 for all $D \sim_{\mathcal{B}} D'$, and $\Delta p_i = \infty$.

Addressing C1 for the bounded case is potentially simple, we just need to perform the maximum over the privacy budget of two mechanisms instead of one (since $I_p(D, D') \leq 2$). However, C2 is much harder to address. In the next section, we introduce a new setting, the *common-domain*, that in particular allows us to drop condition C2 and compute tight bounds for composition over disjoint databases in bounded DP.

A. Common-Domain Setting

In this section, we analyze the particular case where we have k $d_{\mathbb{D}}$-private mechanisms $\mathcal{M}_i$ with domain $\mathbb{D}$, protecting any database of $D \in \mathbb{D}$, but the computation of $\mathcal{M}$ depends exclusively on the information contained in $f_i(D)$ and not on the total information of D. For instance, when we use a Laplace mechanism [2] to compute the number of records belonging

to a certain subclass, the mechanism can take as input the whole database but the output only depends on the records in the subclass. In this case, we can provide new composition rules that allow us to obtain better privacy bounds.

We formalize this scenario with the *common-domain setting* that relates to the perspective in which $\mathcal{M}_i = \mathcal{M}_i^* \circ f_i$ are d_i-private instead of $\mathcal{M}_i^*$, i.e., $\mathcal{M}_i$ and $\mathcal{M}$ have the same "common" domain $\mathbb{D}$. Importantly, if $\mathcal{M}_i$ are d_i-private, we can bound the privacy loss by at least $d_{\mathbb{D}}(D, D') = \sum_{i=1}^k d_i(D, D') < \infty$ if all $d_i(D, D') < \infty$; unlike when we impose the privacy constraints in $\mathcal{M}_i^*$, where we can obtain $d_{\mathbb{D}}(D, D') = \infty$ even if all $d_i(D, D') < \infty$. In the common-domain setting, note that $\mathcal{M}_i$ depends exclusively on $f_i(D)$, which is formalized under the notion of *dependency*:

Definition III.3 (Dependency)**.** Let $\mathcal{M}$ be a randomized mechanism and let f be a deterministic map, both with domain $\mathbb{D}$. We say that $\mathcal{M}$ is *f-dependent* if there exists $\mathcal{M}^*$ with domain $f(\mathbb{D})$ such that $\mathcal{M} = \mathcal{M}^* \circ f$.

Under these conditions, we obtain the following result:

Theorem III.4 (Composition theorem for common domain)**.** *For all $i \in [k]$, let $(\mathbb{D}, d_i)$ be a privacy space, and let f_i be a deterministic map over $\mathbb{D}$. For all $i \in [k]$, let $\mathcal{M}_i \colon \mathbb{D} \to \mathcal{R}_i$ be a mechanism satisfying d_i-privacy and f_i-dependency. Then mechanism $\mathcal{M} = (\mathcal{M}_1, \dots, \mathcal{M}_k)$ is $d_{\mathbb{D}}$-private*[1] with $d_{\mathbb{D}} \coloneqq \sum_{i=1}^k d_i^{f_i}$, where*

$$d_i^{f_i}(D, D') \coloneqq \min_{\tilde{D} \in S_{f_i}(D),\, \tilde{D}' \in S_{f_i}(D')} d_i(\tilde{D}, \tilde{D}')$$

and $S_{f_i}(D) = \{\tilde{D} \in \mathbb{D} \mid f_i(\tilde{D}) = f_i(D)\}$.

Note that $\sum_{i=1}^k d_i^{f_i}(D, D') \leq \sum_{i \colon f_i(D) \neq f_i(D')} d_i(D, D')$ are better bounds than $\sum_{i=1}^k d_i$ given by the general composition (Theorem III.1). We can also easily translate the result for granularities: If we take $\mathcal{M}_i$ to be $\mathcal{G}$ ε_i-DP (i.e., $\varepsilon_i d_{\mathbb{D}}^{\mathcal{G}}$-private), we obtain that $\mathcal{M}$ is $\mathcal{G}$ ε-DP (i.e., $\varepsilon d_{\mathbb{D}}^{\mathcal{G}}$-private) with

$$\varepsilon = \max_{D \sim_{\mathcal{G}} D'} \sum_{i \colon f_i(D) \neq f_i(D')} \varepsilon_i.$$

Looking into disjoint inputs in this setting, we see condition C2 is trivially satisfied. Therefore, we only need to impose that the partitioning function is $d_{\mathbb{D}}^{\mathcal{G}}$-compatible (C1) to ensure that the composition $\mathcal{M}$ of k $\varepsilon_i d_{\mathbb{D}}^{\mathcal{G}}$-private, p_i-dependent mechanism $\mathcal{M}_i \colon \mathbb{D} \to \mathcal{R}_i$ is $\varepsilon d_{\mathbb{D}}^{\mathcal{G}}$-private with $\varepsilon = \max_{i \in [k]} \varepsilon_i$.

In general, Theorem III.4 is not affected by the sensitivities Δp_i, resulting in a smaller privacy loss than that obtained by sequential composition. This is especially useful in the bounded case, where dropping condition C2 allows us to compute a reasonable bound when considering a partition of the database:

Corollary III.5. *Let p be a k-partitioning function of Definition II.5. For all $i \in [k]$, let $\mathcal{M}_i \colon \mathbb{D} \to \mathcal{R}_i$ be mechanisms satisfying bounded ε_i-DP and p_i-dependent. Then mechanism $\mathcal{M} = (\mathcal{M}_1, \dots, \mathcal{M}_k)$ with domain $\mathbb{D}$ is bounded ε-DP with $\varepsilon = \max_{i,j \in [k];\, i \neq j} (\varepsilon_i + \varepsilon_j)$.*

We thus provide a solution to the problem posed in [6], obtaining a tight bound for composition over disjoint databases in bounded DP, which was previously missing.

[1]We refer to it as $d_{\mathbb{D}}$-privacy* since $d_{\mathbb{D}}$ is not a metric (it does not verify the triangle inequality).

IV. Conclusions

In this paper, we provide a general composition theorem (III.1) for $d_{\mathbb{D}}$-privacy. This result facilitates the computation of the final privacy guarantee of any composed mechanism over any data domain and even under mixed privacy requirements, which have been unexplored so far. In particular, we prove better bounds than those existing in the literature when the effect of preprocessing is taken into account, including for intermediate settings between sequential and parallel composition, and we introduce a new setting that allows for further improvements.

In addition, since the original parallel composition theorem [3] does not generalize to all metrics, we also prove additional hypotheses to obtain the best possible privacy loss when the composed mechanism inputs disjoint sets. We conclude that these conditions are only satisfied for some particular metrics, such as $d_{\mathbb{D}_{\mathcal{X}}}^{\mathcal{U}}$, but not in general. Nevertheless, the common-domain setting allows us to obtain a better bound than $\sum_{i=1}^k d_i$, even when the best bound is not achieved. In particular, we prove a significantly better bound on the privacy loss for bounded DP when the composed mechanisms are applied to disjoint databases (Corollary III.5).

Acknowledgments

Javier Parra-Arnau is a "Ramón y Cajal" fellow (ref. RYC2021-034256-I) funded by the MCIN/AEI/10.13039/501100011033 and the EU "NextGenerationEU"/PRTR. This work was supported by KASTEL Security Research Labs, Karlsruhe, as well as the BMBF project "PROPOLIS" (16KIS1393K). We thank the inhouse textician at KASTEL.

References

[1] P. Guerra-Balboa, À. Miranda-Pascual, J. Parra-Arnau, and T. Strufe, "Composition in differential privacy for general granularity notions," in *Proc. IEEE Comput. Security Found. Symp. (CSF)*, 2024, pp. 48–64. DOI: 10.1109/CSF61375.2024.00004.

[2] C. Dwork and A. Roth, "The algorithmic foundations of differential privacy," *Found. Trends Theor. Comput. Sci.*, 2014. DOI: 10.1561/0400000042.

[3] F. McSherry, "Privacy integrated queries," in *Proc. ACM SIGMOD Int. Conf. Manage. Data (MOD)*, 2009. DOI: 10.1145/1559845.1559850.

[4] D. Kifer and A. Machanavajjhala, "No free lunch in data privacy," in *Proc. ACM SIGMOD Int. Conf. Manage. Data (MOD)*, 2011. DOI: 10.1145/1989323.1989345.

[5] D. Desfontaines and B. Pejó, "SoK: Differential privacies," *Proc. Priv. Enhanc. Technol.*, 2020. DOI: 10.2478/popets-2020-0028.

[6] N. Li, M. Lyu, D. Su, and W. Yang, *Differential Privacy: From Theory to Practice*. Springer, 2017. DOI: 10.1007/978-3-031-02350-7.

[7] K. Chatzikokolakis, M. E. Andrés, N. E. Bordenabe, and C. Palamidessi, "Broadening the scope of differential privacy using metrics," in *Proc. Priv. Enhanc. Technol. (PETS)*, 2013. DOI: 10.1007/978-3-642-39077-7_5.

[8] F. Galli, S. Biswas, K. Jung, T. Cucinotta, and C. Palamidessi, "Group privacy for personalized federated learning," in *Proc. Int. Conf. Inform. Syst. Security Priv. (ICISSP)*, 2023. DOI: 10.5220/0011885000003405.