Composition in Differential Privacy for General Granularity Notions

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

Àlex Miranda-Pascual*
Karlsruhe Institute of Technology, Germany
Universitat Politècnica de Catalunya, Spain
alex.pascual@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*—The composition theorems of differential privacy (DP) allow data curators to combine different algorithms to obtain a new algorithm that continues to satisfy DP. However, new granularity notions (i.e., neighborhood definitions), data domains, and composition settings have appeared in the literature that the classical composition theorems do not cover. For instance, the original parallel composition theorem does not translate well to general granularity notions. This complicates the opportunity of composing DP mechanisms in new settings and obtaining accurate estimates of the incurred privacy loss after composition.**

To overcome these limitations, we study the composability of DP in a general framework and for any kind of data domain or neighborhood definition. We give a general composition theorem in both independent and adaptive versions and we provide analogous composition results for approximate, zero-concentrated, and Gaussian DP. Besides, we study the hypothesis needed to obtain the best composition bounds. Our theorems cover both parallel and sequential composition settings. Importantly, they also cover every setting in between, allowing us to compute the final privacy loss of a composition with greatly improved accuracy.

I. Introduction

Differential privacy (ε-DP) [7] is a well-known privacy notion in the field of data protection. One advantage of DP over other privacy notions, such as, for instance, syntactic notions [20], 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. The DP composition theorems serve as a reliable measure for any privacy loss suffered in the newly composed DP mechanism. For these reasons, the advantages of DP composition are recognized throughout the privacy community. For example, composability is key for the construction of most DP algorithms; further, the privacy protection of adaptive updates (e.g., in a streaming scenario or model learning) could not be computed without composition.

*These authors contributed equally.

Currently, DP composition is represented by two results: *sequential composition* [9] and *parallel composition* [17]. Parallel composition is applied when all combined mechanisms access mutually disjoint databases, the maximum loss before combination determines 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.

DP and the sequential and parallel composition theorems were originally defined for tabular databases in the *unbounded* [12] scenario. Nowadays, however, 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* [9]), such as *bounded DP* [12] or *edge-DP* [11]. Consequently, the mechanisms we compose can be defined for different domains and granularities. There also can be alternatives to accessing either the whole database or disjoint parts of it. Therefore, we need new composition rules for more general settings.

However, the existing composition theorems may not extend directly to these general settings. For instance, Li et al. [16] show that the proof of the parallel composition theorem [17] does not hold if we change the original granularity to bounded DP. Since composition for new domains and new granularity notions may be non-trivial or even impossible, curators need to understand how composition results work for each case and when they yield no significant results. Otherwise, curators risk 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 interpreted and compared with the initial ones, we set up a general mathematical framework based on the notion of d-privacy introduced by Chatzikokolakis et al. [3]. Using this framework we present composition theorems (IV.1

and V.2) for when a mechanism is applied independently of the others (the *independent* scenario) or using the output of a mechanism as input in the following ones (the *adaptive* scenario). Our results allow us to obtain new composition theorems for any domain and granularity notion, both existing and future, and even allow combining different domains and granularity notions. Consequently, we improve the understanding of how different granularity notions affect composition in DP. Furthermore, our results facilitate a more accurate calculation of the privacy loss upon any possible composition of DP mechanisms and showcase the effect that preprocessing has on the computation. For instance, if the mechanisms take as input non-necessarily disjoint subsets of the initial database, it is now possible to obtain better bounds than the sum obtained using sequential composition (see Example IV.5).

Besides, we study the settings that are common in the literature and provide the corresponding privacy estimates obtained by using our composition theorems. Furthermore, we study sufficient conditions to obtain the "$\max \varepsilon_i$" bound when the mechanisms take as input disjoint parts of the initial database. For the cases where this bound cannot be achieved, we provide a new variation on composition (see Section IV-D) that allows us to achieve better results. In particular, we provide a solution to the open problem of Li et al. [16] by giving the lowest possible privacy loss for the composition of bounded DP mechanisms executed on mutually disjoint databases (Corollary IV.13).

To further showcase our results, we extend our composition theorems to other privacy notions based on DP where the granularity can be changed. These other privacy notions are *approximate DP* ((ε, δ)-DP), *zero-concentrated DP* (ρ-zCDP) and *Gaussian DP* (μ-GDP). To the best of our knowledge, we are the first to define the d-private counterparts of (ε, δ)-DP, ρ-zCDP, and μ-GDP in order to gain a more general perspective on these three notions. Besides, we provide the first statement of the zCDP composition over disjoint databases. Moreover, we provide a tighter bound than $\max_{i\in[k]} \mu_i d$ for Gaussian DP over disjoint databases (see Example VI.25).

An overview of the generalized results is given in Figure 1. Our contributions are as follows:

- We prove the independent composition (IC) and the adaptive composition (AC) theorems, two new results that allow for reducing the estimated privacy loss and designing improved DP mechanisms in general contexts. Moreover, our theorems make it possible to mix different granularity mechanisms while controlling the privacy guarantees offered.
- We study particular cases of previous theorems that generalize the sequential and parallel composition to any granularity notion. This allows us to compute the minimum privacy loss for the bounded case when the mechanism processes disjoint parts of the database.
- We define $(d_{\mathbb{D}}, \delta_{\mathbb{D}})$-privacy, $d_{\mathbb{D}}^2$-zCprivacy, and $d_{\mathbb{D}}$-Gprivacy, $d_{\mathbb{D}}$-private versions of (ε, δ)-DP, ρ-zCDP, and μ-GDP. Our definitions allow us to generalize to other domains and to provide general composition bounds. We also adapt our general composition results to $(d_{\mathbb{D}}, \delta_{\mathbb{D}})$-privacy, $d_{\mathbb{D}}^2$-zCprivacy and $d_{\mathbb{D}}$-Gprivacy. Particularly, we show that the parallel composition metric bound can be improved in $d_{\mathbb{D}}$-Gprivacy.

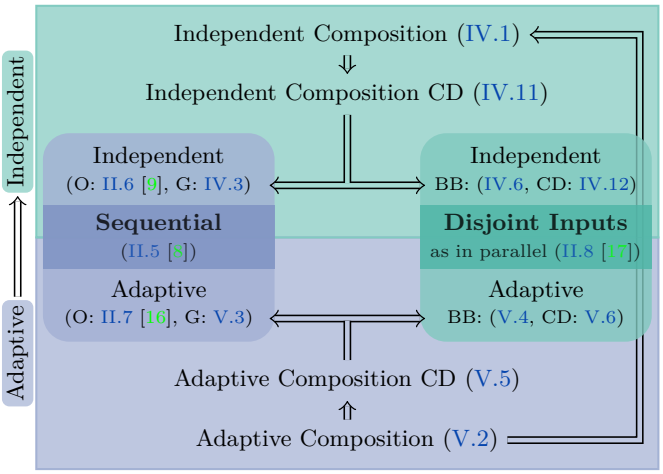

Figure 1: Overview of the theorems proved in this paper, classified according to whether they are adaptive or independent. The theorems represented are the generalizations of sequential composition and the best bound (BB) for disjoint inputs (as in the parallel setting). In the figure, "O" denotes the original theorem, "G" our generalized version, and "CD" common domain. Arrows indicate that a result directly implies the other.

The paper is organized as follows: Preliminaries are explored in Section II, and we formalize the granularities and the generalization to $d_{\mathbb{D}}$-privacy in Section III. We present our independent composition theorem in Section IV, including interesting cases such as a generalization of the independent sequential composition and the setting where the mechanisms take as input disjoint parts of the database. In Section V, we discuss the analogous results for the adaptive scenario. Then we give the composition results for (ε, δ)-DP, ρ-zCDP, and μ-GDP in Section VI. Finally, we discuss post-processing and the reciprocal theorems (Section VII) and conclude with a brief summary of the results (Section VIII). All proofs of our statements can be found in the appendix of the long version of this paper[1].

Related Work: Li et al. [16] analyze the composition theorems in unbounded and bounded DP, and find out that the parallel composition theorem does not necessarily hold for bounded DP mechanisms. However, they do not explore other granularities of the state of the art or attempt to provide a solution for the bounded problem. McSherry [17] gives the first distance-based formulation of DP, later generalized by Chatzikokolakis et al. [3] with the definition of $d_{\mathbb{D}}$-privacy, which we use to set the general framework for composition. However, only sequential composition has been explored for $d_{\mathbb{D}}$-privacy [10]. Therefore, the gener-

[1]Long version: arXiv:2308.14649

Symbol	Meaning
$\mathcal{X}$	Set of possible data records
$\mathbb{D}_{\mathcal{X}}$	The universe of all databases drawn from $\mathcal{X}$
$\mathbb{D}$	Database class
D, D'	A pair of databases
$\lvert D\rvert$	Size of D (number of records)
x	Data record (element of $\mathcal{X}$)
$m_D(x)$	Multiplicity of x in the multiset D
$\mathcal{M}$	A randomized mechanism
$\mathcal{R} \coloneqq \text{Range}(\mathcal{M})$	Range of mechanism $\mathcal{M}$
S	Measurable subset of $\mathcal{R}$
s	Element of $\mathcal{R}$
$\mathcal{G}$	Granularity notion/neighborhood definition
$D \sim_{\mathcal{G}} D'$	D and D' are $\mathcal{G}$-neighboring
$d_{\mathbb{D}}$ (or d)	Metric over $\mathbb{D}$
$d_{\mathbb{D}}^{\mathcal{G}}$	Canonical metric of $\mathcal{G}$ over $\mathbb{D}$
$\mathcal{U}, \mathcal{B}$	Unbounded and bounded granularity (resp.)
$D \triangle D'$	Symmetric difference $((D \cup D') \backslash (D \cap D'))$
$[k]$	Set of indices $\{1, \dots, k\}$
$I_f(D, D')$	For $f = \{f_i\}_{i\in[k]}$, $\lvert\{i \in [k] \mid f_i(D) \neq f_i(D')\}\rvert$

Table I: Summary of the notation used in this paper.

alization of other composition settings, such as parallel, to other granularities (metrics) is still an open question, and to the best of our knowledge, there is no work in the literature, either for DP or for d-privacy, that, in a general manner, computes an accurate privacy loss bound when we have other metrics, domains and composition rules.

II. Preliminaries

In this section, we introduce the main concepts relevant to this work. The main notation used throughout the manuscript is compiled in Table I.

A. Tabular Databases and Differential Privacy

In the original formulation of DP, the database D is assumed to be comprised of a finite number n of rows, where the intuition is that each row contains data related to an individual, drawn from a universe of data records $\mathcal{X}$ [9]. In this case, the data model is a tabular database, and we refer to a single data row as a *record*. We denote the universe of all the possible tabular databases drawn from $\mathcal{X}$ as $\mathbb{D}_{\mathcal{X}}$. In particular, $\mathbb{D}_{\mathcal{X}}$ contains the empty database $\varnothing$ and is closed under subsets (if $D' \subseteq D \in \mathbb{D}_{\mathcal{X}}$, then $D' \in \mathbb{D}_{\mathcal{X}}$) and under basic math operators: $D \cup D'$, $D \cap D', D \backslash D' \in \mathbb{D}_{\mathcal{X}}$ for all $D, D' \in \mathbb{D}_{\mathcal{X}}$. We consider all these operations as defined for multisets [21] for the rest of the paper.

The first definition of ε-DP with precise formulation[2] was introduced by Dwork [7].

Definition II.1 (Differential privacy [7])**.** A randomized mechanism $\mathcal{M}$ with domain $\mathbb{D}_{\mathcal{X}}$ is *ε-differentially private* (ε-DP) if for all $D, D' \in \mathbb{D}_{\mathcal{X}}$ differing on at most one element and all measurable $S \subseteq \text{Range}(\mathcal{M})$,

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

An important part of DP is the concept of *neighborhood*, also referred to as the *granularity notion* of DP [9]. 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., $|D \triangle D'| = |(D \cup D') \backslash (D \cap D')| \leq 1$. In other words, we obtain a neighboring database by removing or adding a single element or row. Assuming each row is linked to a single individual, we get the usual DP interpretation: DP aims to protect the participation of each individual in the original database up to ε.

Two parameters control the privacy of individuals in the DP definition, namely, the *privacy budget* ε and the universe of records $\mathcal{X}$. The former limits the amount of information that an attacker can extract with access to the mechanism's output. The latter encodes what information is considered public. For example, if $\mathcal{X}$ is the set of possible addresses of a city, we can discover (up to ε) that a person lives in a particular city, while if $\mathcal{X}$ is the set of possible addresses of a country, we can discover (up to ε) that an individual lives in the country, but not which exact city.

Furthermore, with the *group privacy* property of DP [9], we also protect the participation of n individuals with the protection degrading linearly with respect to n. More precisely, we have the following result:

Proposition II.2 ([17])**.** *A mechanism $\mathcal{M}$ is ε-DP if and only if for all $D, D' \in \mathbb{D}_{\mathcal{X}}$ and all measurable set $S \subseteq \text{Range}(\mathcal{M})$*

$$\mathrm{P}\{\mathcal{M}(D) \in S\} \leq \mathrm{e}^{\varepsilon |D \triangle D'|}\, \mathrm{P}\{\mathcal{M}(D') \in S\}. \tag{II.2}$$

In this case, $d^{\triangle}(D, D') \coloneqq |D \triangle D'|$ can be thought of as the distance (or metric) between D and D' in $\mathbb{D}_{\mathcal{X}}$. In this regard, McSherry [17] provides the first statement of DP from a metric perspective, which stems from the group privacy property. This laid the foundations of the generalization to d-privacy [3], which we will explore in Section III.

B. Differential Privacy: Unbounded vs. Bounded

Nowadays, many neighborhood definitions for DP exist. A compilation of common granularities is provided in [5]. Among these, unbounded and bounded DP are the most popular ones [12]. The unbounded notion corresponds to the original definition presented by Dwork [7].

Definition II.3 (Unbounded)**.** A pair of databases $D, D' \in \mathbb{D}_{\mathcal{X}}$ are *unbounded neighboring* if D can be obtained from D' by either adding or removing one record (i.e., $|D \triangle D'| = 1$).

Definition II.4 (Bounded)**.** A pair of databases $D, D' \in \mathbb{D}_{\mathcal{X}}$ are *bounded neighboring* if D can be obtained from D' by changing the value of exactly one record (i.e., $|D \triangle D'| = 2$ and $|D| = |D'|$).

These two notions of neighborhood lead to different privacy guarantees. The clearest difference concerns the privacy of the number of records: the unbounded notion protects the number of records in the database, while the bounded notion does not.

[2] For the literature definitions and theorems, we state them as they are defined in the cited reference but using the notation of this manuscript.

C. Introduction to the Composition Theorems

One of the most useful properties of DP mechanisms relates to composition theorems. Sequential and parallel composition are considered key components of DP and are regularly used in the field.

The composition theorems share a common foundation. Simply put, these theorems say that given k ε_i-DP mechanisms $\mathcal{M}_i$, the composed mechanism $\mathcal{M}$ satisfies ε-DP, where ε depends on $\varepsilon_1, \ldots, \varepsilon_k$. In other words, these theorems estimate the privacy loss (i.e., the final privacy budget) of the mechanism $\mathcal{M}$ composed of $\mathcal{M}_i$. However, there are different ways to compose a set of mechanisms, and thus different theorems. We distinguish the following:

Independent vs. adaptive: Composition is *independent* if the outputs of each $\mathcal{M}_i$ are independent of each other. On the other hand, it is *adaptive* if $\mathcal{M}_i$ can use the outputs of any $\mathcal{M}_j$ with $j < i$ as input. More intuitively, $\mathcal{M}$ computes the mechanisms in order (first $\mathcal{M}_1$, then $\mathcal{M}_2$, then $\mathcal{M}_3$, etc.) and can take the output of previous mechanisms as input. Note that adaptive composition is more general than independent composition, i.e., the independent theorems are cases of adaptive results.

Sequential vs. parallel: Orthogonally, if every $\mathcal{M}_i$ takes as input the whole database D in its computation, the composition is *sequential*. 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.

The combination of these variations leads to four clear cases (see Figure 1), which we will refer to as the independent/adaptive sequential/parallel composition settings, due to the lack of consensus[3]. We will also refer to them by the corresponding acronyms: ISC, IPC, ASC, and APC. In the current literature, we frequently find ISC [9], ASC [16], and IPC [17]; while APC remains heavily unused.

D. The Classic Composition Theorems

The sequential and parallel composition theorems were initially stated for the original DP definition [7], unbounded DP, before the introduction of any other granularity. Nevertheless, we specify it in the following theorems.

The first composition result of DP appeared in [8].

Theorem II.5 (Sequential composition [8])**.** *A mechanism that permits T adaptive interactions with an [unbounded] ε-DP mechanism ensures [unbounded] $T\varepsilon$-DP.*

The theorem corresponds to the adaptive definition and includes independent composition as a subcase. Nowadays, these results are sometimes formulated separately with precise hypotheses and allow for different privacy budgets.

Theorem II.6 (Independent sequential composition (ISC) [9])**.** *Let $\mathcal{M}_i \colon \mathbb{D}_\mathcal{X} \to \mathcal{R}_i$ be an [unbounded] ε_i-DP mechanism for each $i \in [k]$. 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 $D \in \mathbb{D}$. Then $\mathcal{M}$ is [unbounded] $(\sum_{i=1}^k \varepsilon_i)$-DP.*

Theorem II.7 (Adaptive sequential composition (ASC) [16])**.** *Let $\mathcal{M}_1, \ldots, \mathcal{M}_k$ be k mechanisms (that take auxiliary inputs) that satisfy [unbounded] ε_1-DP, …, ε_k-DP, respectively, with respect to the input database. Publishing $\boldsymbol{t} = \langle t_1, t_2, \ldots, t_k \rangle$, where $t_1 = \mathcal{M}_1(D)$, $t_2 = \mathcal{M}_2(t_1, D)$, …, $t_k = \mathcal{M}_k(\langle t_1, \ldots, t_{k-1} \rangle, D)$, satisfies [unbounded] $(\sum_{i=1}^k \varepsilon_i)$-DP.*

In search of optimization, the literature has found circumstances for a better bound than the sequential one. Databases can be composed of diverse information and most queries only need to compute values in a proper subset of data. It is these circumstances which, in fact, provide the better bound: The parallel composition theorem.

Theorem II.8 (Parallel composition [17])**.** *Let $\mathcal{M}_i$ each provide [unbounded] ε-DP. Let $\mathcal{X}_i$ be arbitrary disjoint subsets of the universe of records $\mathcal{X}$. The sequence of $\mathcal{M}_i(D_i)$ provides [unbounded] ε-DP, where $D_i \subseteq D$ is the multiset such that element $x \in D$ has multiplicity $m_{D_i}(x) = \mathbf{1}_{\mathcal{X}_i}(x)\, m_D(x)$.*

By abuse of notation, D_i is also often denoted as $D \cap \mathcal{X}_i$. Nowadays, this formulation has also seen modifications. For example, Li et al. [16] use a partitioning function p to define the disjoint subsets in the previous statement, i.e., $p_i(D) = D_i$ for all i and $D \in \mathbb{D}_\mathcal{X}$.

Even though Theorems II.5 to II.7 were initially stated for the unbounded granularity notion, they can easily be translated for other granularities [10]. However, in Theorem II.8, if instead of unbounded, we impose $\mathcal{M}_i$ to be bounded ε-DP, then it is not generally true that the sequence of $\mathcal{M}_i(D_i)$ provides bounded ε-DP. Li et al. [16] show why the proof is not applicable: even if $\mathcal{M}_i$ are bounded ε-DP, $\mathcal{M}'_i$ such that $\mathcal{M}'_i(D) = \mathcal{M}_i(D_i) = \mathcal{M}_i(D \cap \mathcal{X}_i)$ is not necessarily bounded ε-DP. This fact is clear in the following counterexample, which we provide to complete Li et al.'s claim [16]:

Example II.9 (Parallel composition does not hold for bounded DP)**.** Let $\mathbb{D}_\mathcal{X}$ be a database universe and $\mathcal{X}_i$ arbitrary disjoint subsets of $\mathcal{X}$. We show that given $k > 1$ mutually independent bounded ε_i-DP mechanisms $\mathcal{M}_i \colon \mathbb{D}_\mathcal{X} \to \mathcal{R}$, it is not necessarily true that the composed mechanism $\mathcal{M} \colon \mathbb{D}_\mathcal{X} \to \mathcal{R}$ such that $\mathcal{M}(D) = (\mathcal{M}_1(D_1), \ldots, \mathcal{M}_k(D_k))$ is bounded DP, where $D_i \subseteq D$ is the multiset such that element $x \in D$ has multiplicity $m_{D_i}(x) = \mathbf{1}_{\mathcal{X}_i}(x)\, m_D(x)$.

To do so, we prove that we can select $k > 1$ mutually independent bounded ε_i-DP mechanisms $\mathcal{M}_i \colon \mathbb{D}_\mathcal{X} \to \mathcal{R}$ such that mechanism $\mathcal{M} \colon \mathbb{D}_\mathcal{X} \to \mathcal{R}$ with $\mathcal{M}(D) = (\mathcal{M}_1(D_1), \ldots, \mathcal{M}_k(D_k))$ is not bounded ε-DP for any $\varepsilon \geq 0$.

For all $i \in [k]$, we choose $\mathcal{M}_i \colon \mathbb{D}_\mathcal{X} \to \mathcal{R}$ such that they output the number of elements of the input database, i.e., $\mathcal{M}_i(D) = \mathcal{M}^*(D) = |D|$ for all $D \in \mathbb{D}_\mathcal{X}$. It can easily be

[3]For example, ISC and IPC are referred to as independent and sequential composition in [13]; and as sequential and adaptive composition in [5].

checked that this mechanism is bounded 0-DP. Observe that in this case, $\mathcal{M}(D) = (\mathcal{M}^*(D_1), \dots, \mathcal{M}^*(D_k)) = (|D_1|, \dots, |D_k|)$.

Let $D, D' \in \mathbb{D}_{\mathcal{X}}$ be two bounded-neighboring databases such that $D \triangle D' = \{x, x'\}$ with $x \in D_j$ and $x' \in D'_l$, $j \neq l$. It is clear then that $\mathcal{M}^*(D_j) = |D_j| \neq |D'_j| = \mathcal{M}^*(D'_j)$ (analogously for l), so $\mathrm{P}\{\mathcal{M}^*(D_j) = |D_j|\} = 1 \not\leq 0 = \mathrm{P}\{\mathcal{M}^*(D'_j) = |D_j|\}$. Note that this is not a contradiction with $\mathcal{M}^*$ being unbounded DP, since D_j and D'_j are not bounded-neighboring databases.

Consequently, taking $s = (|D_1|, \dots, |D_n|) \in \mathcal{R}$ we obtain $\mathrm{P}\{\mathcal{M}(D) = s\} = 1$, but $\mathrm{P}\{\mathcal{M}^*(D'_j) = |D_j|\} = 0$. Therefore $\mathrm{P}\{\mathcal{M}(D) = s\} = 1 \not\leq 0 = \mathrm{e}^{\varepsilon}\,\mathrm{P}\{\mathcal{M}(D') = s\}$ for all $\varepsilon \geq 0$, so the mechanism $\mathcal{M}$ is not bounded DP.

We want to showcase with this example that the composition results proved for unbounded DP in $\mathbb{D}_{\mathcal{X}}$ cannot be trivially generalized to other data domains or neighborhood definitions. The failure of bounded DP on satisfying $(\max_{i\in[k]} \varepsilon_i)$-DP when composed in parallel opens a new question about how to measure the privacy of composed mechanisms in general.

The main goal of this work is to answer this question by generalizing these composition theorems to more general scenarios, in which the domain of the mechanism is not necessary $\mathbb{D}_{\mathcal{X}}$, and the given granularity notion is not necessarily unbounded. To achieve so, we introduce new more-general composition rules that even allow composing DP mechanisms with different domains and granularity notions. These results are shown in Sections IV and V.

However, to carry out this extension of properties to general settings, we first need to define a formal structural model. Thus, we will begin by generalizing the data domain and the concept of granularity notions in the next section.

III. Generalizing the Granularity Notion of DP

As mentioned earlier, DP was designed to handle aggregated queries on tabular data. However, in many cases, mechanisms impose a maximum or minimum number of elements in the database, are only defined for databases of a fixed size, or are not defined for the empty database, which are incompatible conditions if $\mathbb{D}_{\mathcal{X}}$ is the mechanism domain. Also, the structure of the data is not necessarily tabular, such as graph databases [11]. For instance, in a social network graph, each node is an individual in the database, while the edges represent the social relationship between the nodes. This means that information about individuals is not always encoded in rows or multiset elements.

This motivates the need to generalize DP to different settings. In this section, we provide a mathematical formalization of the granularity notions and the data domain, establishing a framework in which privacy can always be measured and compared between different notions.

Databases are collections of data and can be defined as mathematical objects such as multisets (original case), sets, numbers, functions, streams, or graphs. A collection of databases forms a *database class*[4], which we denote by $\mathbb{D}$. In our setting, we will consider the cases where the domain of $\mathcal{M}$ is a generic class $\mathbb{D}$ instead of $\mathbb{D}_{\mathcal{X}}$ (including the case where $\mathbb{D} \subseteq \mathbb{D}_{\mathcal{X}}$).

Moreover, DP allows many different neighborhood definitions [5], each with its own privacy implications and interpretability. We generalize the definition of granularity notion $\mathcal{G}$ as follows.

Definition III.1 ($\mathcal{G}$-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 $D \sim_{\mathcal{G}} D'$.

We will use calligraphic letters to denote certain granularity notions (e.g., $\mathcal{U}$ for unbounded, $\mathcal{B}$ for bounded). With Definition III.1, we can establish a general framework for DP similar to that in [12]. That is, a mechanism $\mathcal{M}$ with domain $\mathbb{D}$ is *$\mathcal{G}$ ε-DP* ($\varepsilon \geq 0$) if for all $\mathcal{G}$-neighboring $D, D' \in \mathbb{D}$ and all measurable $S \subseteq \mathrm{Range}(\mathcal{M})$,

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

Note that, given a data domain $\mathbb{D}$, we can construct [3] a *canonical metric* $d_{\mathbb{D}}^{\mathcal{G}}$ for each granularity $\mathcal{G}$ over $\mathbb{D}$ by defining the distance between two databases $d_{\mathbb{D}}^{\mathcal{G}}(D, D')$ 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). In particular, note that $d_{\mathbb{D}}^{\mathcal{G}}(D, D') = 0$ if and only if $D = D'$, and $d_{\mathbb{D}}^{\mathcal{G}}(D, D') = 1$ if and only if $D \sim_{\mathcal{G}} D'$ (and $D \neq D'$). See Proposition A.1 (in the long version) for more details and the proof of well-definition.

Then, from the group property of DP (Proposition II.2), $\mathcal{M}$ 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 to measure privacy protection. As mentioned, this idea first appeared in [17] with $d_{\mathbb{D}_{\mathcal{X}}}^{\mathcal{U}}(D, D') = |D \triangle D'|$ (note that $d_{\mathbb{D}_{\mathcal{X}}}^{\mathcal{U}} = d_{\mathbb{D}_{\mathcal{X}}}^{\triangle}$, but not generally over $\mathbb{D}$). Later, a formal generalization called $d_{\mathbb{D}}$-privacy [3] was introduced. We consider the variant [3] modeled by an *extended pseudometric* $d_{\mathbb{D}} \colon \mathbb{D}^2 \to [0, \infty]$, i.e., a metric in which the distance between two different databases can also be 0 and ∞. To simplify the terminology, we will simply refer to $d_{\mathbb{D}}$ as *metrics*. Note that having a metric $d_{\mathbb{D}}$ implies that $(\mathbb{D}, d_{\mathbb{D}})$ is a (pseudo)metric space, which we will call *privacy space*.

Definition III.2 ($d_{\mathbb{D}}$-privacy [3])**.** 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 $S \subseteq \mathrm{Range}(\mathcal{M})$,

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

[4]We define it as a mathematical class instead of a set because a collection of sets does not need to be a well-defined set [15]. We denote the usual inclusion of classes by $\mathbb{D}' \subseteq \mathbb{D}$.

Observe that the metric absorbs the privacy budget (ε), i.e., $d_{\mathbb{D}}$ can be written as $d_{\mathbb{D}} = \varepsilon d'_{\mathbb{D}}$ where $d'_{\mathbb{D}}$ is also a metric. We will also denote it simply as d when possible.

Additionally, we obtain the following result:

Theorem III.3. *Let $\mathcal{G}$ be a granularity notion over the database class $\mathbb{D}$. Then, 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.*

Given any granularity notion we can obtain a metric, but that not all metrics (even up to ε) are the canonical metric for a granularity notion. Therefore, the notion of $d_{\mathbb{D}}$-privacy is more general than $\mathcal{G}$ ε-DP. Besides, note that the restriction of $d_{\mathbb{D}}^{\mathcal{G}}$ to the subclass $\mathbb{D}' \subseteq \mathbb{D}$ is not always $d_{\mathbb{D}'}^{\mathcal{G}}$ (see Remark A.2 in the long version).

To understand the real privacy implications of a $d_{\mathbb{D}}$-privacy, we need to look at the domain and the distance.

The domain, $\mathbb{D}$, encodes what information we consider public knowledge and what we want to protect up to $d_{\mathbb{D}}$. The larger the domain, the greater the privacy, but it also comes with the cost of greater sensitivities and harder-to-achieve privacy protection. The distance, $d_{\mathbb{D}}$, encodes how hard it is to distinguish any pair of databases, and therefore what information we are protecting.

Additionally, it is important to select domains with compatible metrics. For example, information may be disclosed if $d_{\mathbb{D}}(D, D') = \infty$. Therefore, *connected* privacy spaces (i.e., $d_{\mathbb{D}}(D, D') < \infty$ for all $D, D' \in \mathbb{D}$) are preferable because the change across connected components is not guaranteed to be protected by a DP mechanism. For example, when $\mathbb{D}$ is totally disconnected, we can end up with nonsensical privacy guarantees like in the following example.

Example III.4. Consider $\mathbb{D} \coloneqq \{D \in \mathbb{D}_{\mathcal{X}} \mid |D| = N\}$, the class of tabular databases of size N, and choose the unbounded granularity notion. It is clear that unbounded-neighboring databases always differ by one element. Therefore, there are no unbounded-neighboring databases in $\mathbb{D}$ (i.e., the privacy space is totally disconnected).

This privacy space would imply, by *reductio ad absurdum*, that *any* mechanism is unbounded ε-DP for all $\varepsilon \geq 0$ since for all the neighbors (none) the definition holds. In particular, the identity mechanism (such that $\mathcal{M}(D) = D$) defined over $\mathbb{D}$ (which does not provide any protection) is unbounded 0-DP.

Note that choosing $\mathbb{D}_{\mathcal{X}}$ as the domain does not lead to the same problem, but as we mentioned before, relaxing the domain so that it is defined for subsets $\mathbb{D}$ of $\mathbb{D}_{\mathcal{X}}$ and other database types is usually more convenient, coherent, and necessary. Following the same line, the bounded granularity defines a connected privacy space over $\mathbb{D} \coloneqq \{D \in \mathbb{D}_{\mathcal{X}} \mid |D| = N\}$, but defines a disconnected one over $\mathbb{D}_{\mathcal{X}}$.

A. Relationship between Metrics

So far, we have described a mathematical model to understand any metric and granularity notion. This will be necessary for the following sections to define general properties and theorems. However, we need to understand the real privacy implications of metrics given their formal definition.

The notion of $d_{\mathbb{D}}$-privacy allows us to compare the privacy level between metrics over the same domain, which also helps to extend composability notions proven for one to others. Consider two metrics, d_1 and d_2, over $\mathbb{D}$ such that $d_1 \leq d_2$ (pointwise). In this case, we can say that d_1 offers more protection than d_2 because any mechanism $\mathcal{M}\colon \mathbb{D} \to \mathcal{R}$ that satisfies d_1-privacy also satisfies d_2-privacy [3].

In particular, given two canonical metrics $d_{\mathbb{D}}^{\mathcal{G}_1}$ and $d_{\mathbb{D}}^{\mathcal{G}_2}$ such that

$$k = \operatorname{dist}_{\mathbb{D}}(\mathcal{G}_1, \mathcal{G}_2) \coloneqq \max_{\substack{D, D' \in \mathbb{D} \\ D \sim_{\mathcal{G}_2} D'}} d_{\mathbb{D}}^{\mathcal{G}_1}(D, D') < \infty,$$

we obtain $d_{\mathbb{D}}^{\mathcal{G}_1} \leq k d_{\mathbb{D}}^{\mathcal{G}_2}$ (see Proposition A.3 in the long version). Therefore, if $\mathcal{M}\colon \mathbb{D} \to \mathcal{R}$ is $\mathcal{G}_1$ ε-DP, then $\mathcal{M}$ is $\mathcal{G}_2$ $k\varepsilon$-DP. This fact allows us to compare different granularity notions over the same domain, e.g., all information protected by $\mathcal{G}_1$ must also be protected by $\mathcal{G}_2$, while not necessarily the other way around.

From this result, we can deduce the well-known fact that unbounded ε-DP implies bounded 2ε-DP in $\mathbb{D}_{\mathcal{X}}$ [16] since $\operatorname{dist}_{\mathbb{D}_{\mathcal{X}}}(\mathcal{U}, \mathcal{B}) = 2$. However, $\operatorname{dist}_{\mathbb{D}_{\mathcal{X}}}(\mathcal{B}, \mathcal{U}) = \infty$ because $d_{\mathbb{D}_{\mathcal{X}}}^{\mathcal{B}}(D, D') = \infty$ for all $D \sim_{\mathcal{U}} D'$. Note that the privacy-level comparison between two granularity notions directly depends on which class we compare them in. While this result holds in $\mathbb{D}_{\mathcal{X}}$, we saw in Example III.4 that this is not the case for all database classes.

If the diameter of $(\mathbb{D}, d_1)$, $\operatorname{diam}(\mathbb{D}, d_1) \coloneqq \max d_1$, is bounded, we can always compare it to the other metrics over $\mathbb{D}$. For example, the *free-lunch* granularity notion $\mathcal{FL}$ [12] is defined such that all pairs of databases are free-lunch neighboring, i.e., $d_{\mathbb{D}}^{\mathcal{FL}}(D, D') = 1$ for all $D \neq D'$. Therefore, $d_{\mathbb{D}}^{\mathcal{FL}} \leq d_{\mathbb{D}}^{\mathcal{G}}$ verifies for any canonical metric $d_{\mathbb{D}}^{\mathcal{G}}$, and thus free-lunch DP implies all others.

B. Changing the Privacy Space

It is also interesting to understand how queries or other transformations can produce a transition from one privacy space to another and how this change can be reflected in our overall privacy.

Definition III.5 (Sensitivity [3])**.** Let $(\mathbb{D}_1, d_1)$ and $(\mathbb{D}_2, d_2)$ be two privacy spaces and let $f\colon \mathbb{D}_1 \to \mathbb{D}_2$ be a deterministic map. We define the *sensitivity of f* with respect to d_1 and d_2 as the smallest value $\Delta f \in [0, \infty]$ such that $d_2(f(D), f(D')) \leq \Delta f \, d_1(D, D')$ holds for all $D, D' \in \mathbb{D}_1$ with $d_1(D, D') < \infty$.

Proposition III.6 (Preprocessing [3])**.** *Let $(\mathbb{D}_1, d_1)$ and $(\mathbb{D}_2, d_2)$ be two privacy spaces and let f be a deterministic map with sensitivity $\Delta f < \infty$ with respect to d_1 and d_2, and let $\mathcal{M}\colon \mathbb{D}_2 \to \mathcal{R}_2$ be a d_2-private mechanism. Then $\mathcal{M} \circ f$ satisfies $(\Delta f) d_1$-privacy.*

In the case where the metrics are the canonical metrics of granularities $\mathcal{G}_1$ and $\mathcal{G}_2$, we obtain that the sensitivity is $\Delta f \coloneqq \max_{D\sim_{\mathcal{G}_1} D'} d_2(f(D), f(D'))$. If we then choose $f = \mathrm{id}\colon \mathbb{D} \to \mathbb{D}$, we obtain that $\Delta\mathrm{id} = \mathrm{dist}_{\mathbb{D}}(\mathcal{G}_1, \mathcal{G}_2)$.

Remark III.7. The reciprocal of Proposition III.6 is not true. For example, consider $(\mathbb{D}_1, d_1) = (\mathbb{D}_2, d_2) = (\mathbb{R}, d^{\mathcal{FL}})$, the free-lunch metric over $\mathbb{R}$. Take $\mathcal{M}\colon \mathbb{R} \to \mathbb{R}$ such as $\mathcal{M}(x) = x + Z$ where $Z \sim \mathrm{Lap}(\frac{1}{\varepsilon})$. We can easily verify that this mechanism is not free-lunch DP by selecting two numbers $x << y$. In other words, the sensitivity of the identity map over the real numbers is infinite. However, if we take $f\colon \mathbb{R} \to \mathbb{R}$ such that $f(x) = \frac{1}{1+e^x}$, then $\Delta f = \|f(x) - f(y)\|_1 \leq 1$. Therefore, $\mathcal{M} \circ f$ corresponds to a Laplace mechanism, and is free-lunch ε-DP. In conclusion, there exist $\mathcal{M}$ and f such that $\mathcal{M} \circ f$ is $(\varepsilon\Delta f)d_{\mathbb{R}}^{\mathcal{FL}}$-private but $\mathcal{M}$ is not $\varepsilon' d_{\mathbb{R}}^{\mathcal{FL}}$-private for any $\varepsilon' > 0$.

We can also apply multiple preprocessing functions to a mechanism, obtaining the following bound:

Proposition III.8 (Sensitivity of the composition)**.** *Let $(\mathbb{D}_1, d_1)$, $(\mathbb{D}_2, d_2)$ and $(\mathbb{D}_3, d_3)$ be privacy spaces and let $f\colon \mathbb{D}_1 \to \mathbb{D}_2$ and $g\colon \mathbb{D}_2 \to \mathbb{D}_3$ be two deterministic maps. Then $\Delta(g \circ f) \leq \Delta f\, \Delta g$.*

IV. The Independent Composition Theorem

Now that we have a general framework for DP in arbitrary privacy spaces, we can start to explore how we can extend the properties of DP from $(\mathbb{D}_{\mathcal{X}}, d_{\mathbb{D}_{\mathcal{X}}}^{\mathcal{U}})$ to the other privacy spaces. In this section, we focus this analysis on the independent composition. To this end, we present a theorem that models all possible independent compositions of mechanisms over arbitrary privacy spaces. To begin, we first need to understand composability, in its more general form.

A. Composing Mechanisms

Assuming the role of the curator, we have a database $D \in \mathbb{D}$ and we want to publish certain extracted information $s \in \mathcal{R}$. However, we cannot publish s directly because it would compromise privacy. Therefore, we want an attacker with access only to the output $\tilde{s}$ of our mechanism to be unable to distinguish aspects of D from other databases of $\mathbb{D}$. Besides, the information we need to extract can be obtained as a function of some query answers. That is, $s = h(s_1, \ldots, s_k)$ 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). Thus, by trying to get every $\tilde{s}_i$ (private output of s_i) and computing $\tilde{s} = h(\tilde{s}_1, \ldots, \tilde{s}_k)$, we make it possible to discretize our problem. To do this, we use k d_i-private mechanisms $\mathcal{M}_i^*\colon \mathbb{D}_i \to \mathcal{R}_i$ such that $\mathcal{M}_i^*(f_i(D)) = \tilde{s}_i$. Therefore, the question arises whether the composition of the mechanisms $\mathcal{M}$ such that $\mathcal{M}(D) = (\mathcal{M}_1^*(f_1(D)), \ldots, \mathcal{M}_k^*(f_k(D)))$ for all $D \in \mathbb{D}$ is $d_{\mathbb{D}}$-private, and what privacy $d_{\mathbb{D}}$ implies. To answer this question, we state and prove the independent/adaptive composition theorems (IV.1 and V.2). Note that $\mathcal{M}_i \coloneqq \mathcal{M}_i^* \circ f_i$ defines a mechanism over $\mathbb{D}$ for all $i \in [k]$.

In Section IV-D, we will explore the scenario where, instead of imposing $\mathcal{M}_i^*$ to be d_i-private, we directly impose $\mathcal{M}_i$ to be d_i-private. Since each $\mathcal{M}_i$ is defined over the same domain as $\mathcal{M}$, we call this scenario *common domain*. This change is significant because it allows us to prove alternative theorems (IV.11 and V.5) to our composition results (Theorems IV.1 and V.2, respectively), and it ensures that the composed mechanism does not completely lose the privacy guarantee, as it happens in Example II.9. As a result, we can provide tighter bounds on the privacy loss for cases such as bounded DP, which are not covered outside the common domain.

B. Independent Composition

In this section, we introduce our generalized version of independent composition. We will explore its adaptive counterpart in Section V. Note that adaptive composition includes independent composition, but we present the results for the independent case first to simplify the notation.

Formally, independent composition refers to the case where the mechanisms $\mathcal{M}_1, \ldots, \mathcal{M}_k$ are *mutually independent*, i.e., $\mathcal{M}_1(D), \ldots, \mathcal{M}_k(D)$ are mutually independent random elements for all $D \in \mathbb{D}$. In other words, the output of each of these mechanisms does not depend on the others. The *independent-composed mechanism* $\mathcal{M} \coloneqq (\mathcal{M}_1, \ldots, \mathcal{M}_k)_{\mathrm{ind}}$ is then defined as the mechanism with domain $\mathbb{D}$ such that $\mathcal{M}(D) = (\mathcal{M}_1(D), \ldots, \mathcal{M}_k(D))$ for all $D \in \mathbb{D}$.

With this definition, we can state the independent composition (IC) theorem. 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.

Theorem IV.1 (IC 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 mutually independent d_i-private mechanisms. Then mechanism $\mathcal{M} = (\mathcal{M}_1^* \circ f_1, \ldots, \mathcal{M}_k^* \circ f_k)_{\mathrm{ind}}$ is $d_{\mathbb{D}}$-private with*

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

It is important to note that the IC theorem (IV.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 as we see in the following example.

Example IV.2. Let $\mathcal{X} = \mathcal{X}_1 \cup \mathcal{X}_2$ be a set of locations in $\mathbb{R}^2$ of two districts $i \in [2]$, each associated with hospital i in location l_i, and consider $\mathbb{D} = \mathbb{D}_{\mathcal{X}}$, consisting of databases of locations from ambulances in both districts. Assume

that the maximum Euclidean distance between any two points in $\mathcal{X}_1$ and $\mathcal{X}_2$ in the districts is equal and finite, $\operatorname{diam}(\mathcal{X}_1) = \operatorname{diam}(\mathcal{X}_2) = L$. Our goal is to compute the number of locations in each district and determine the closest ambulance to each hospital. To do so, we will compose the following d-private mechanisms: A $d_{\mathbb{D}}^{\mathcal{U}}$-private mechanism $\mathcal{M}_a^* \colon \mathbb{D} \to \mathbb{N}$ that outputs the noisy count of records in $D \in \mathbb{D}$, and a $d_{\mathcal{X}}^{\mathrm{Eu}}$-private mechanism $\mathcal{M}_b^* \colon \mathcal{X} \to \mathcal{X}$, with $d_{\mathcal{X}}^{\mathrm{Eu}}$ the Euclidean distance over $\mathcal{X}$, that given $x \in \mathcal{X}$ outputs a perturbed version of it.

For all $i \in [2]$ and $D \in \mathbb{D}$, let $p_i(D) = D \cap \mathcal{X}_i$ be the subset of locations of D in district i, and let $f_i(D) = \arg\min_{x \in p_i(D)} \{\|x - l_i\|_2\}$ be the closest ambulance to hospital i. Thus, we can obtain the wanted information through the composed mechanism $\mathcal{M}$ such that $\mathcal{M}(D) = (\mathcal{M}_a^*(p_1(D)), \mathcal{M}_a^*(p_2(D)), \mathcal{M}_b^*(f_1(D)), \mathcal{M}_b^*(f_2(D)))$.

Now using the IC theorem (IV.1), we can compute the privacy that $\mathcal{M}$ provides. For all $D, D' \in \mathbb{D}$, we have a protection of $d_{\mathbb{D}}(D, D') \coloneqq \sum_{i=1}^{2} (d_{\mathbb{D}}^{\mathcal{U}}(p_i(D), p_i(D')) + d_{\mathcal{X}}^{\mathrm{Eu}}(f_i(D), f_i(D'))) \leq (d_{\mathbb{D}}^{\mathcal{U}} + 2d_{\mathbb{D}}^{\infty})(D, D') \leq d_{\mathbb{D}}^{\mathcal{U}}(D, D') + 2L$ with $d_{\mathbb{D}}^{\infty}(D, D') = \max_{x \in D, x' \in D'} d_{\mathcal{X}}^{\mathrm{Eu}}(x, x')$ the maximum distance.

Note that in the IC theorem (IV.1), we can 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. However, we can still obtain reasonable $d_{\mathbb{D}}$ in general cases where $d_{\mathbb{D}}$ possesses good privacy properties.

For $\mathbb{D}_i = \mathbb{D}$ and $f_i = \mathrm{id}$, we obtain a result reminiscent of the sequential composition theorem:

Theorem IV.3 (Generalized ISC)**.** *Let $\{(\mathbb{D}, d_i)\}_{i \in [k]}$ be a set of privacy spaces. For all $i \in [k]$, let $\mathcal{M}_i \colon \mathbb{D} \to \mathcal{R}_i$ be mutually independent d_i-private mechanisms. Then $\mathcal{M} = (\mathcal{M}_1, \dots, \mathcal{M}_k)_{\mathrm{ind}}$ is $(\sum_{i=1}^{k} d_i)$-private.*

Note that by choosing $d_i = \varepsilon_i d$, we obtain that $\mathcal{M}$ is εd-private with $\varepsilon = \sum_{i=1}^{k} \varepsilon_i$ (first proven in [10]). Furthermore, by selecting d as $d_{\mathbb{D}}^{\mathcal{G}}$, we obtain the sequential composition theorem for every granularity: If $\mathcal{M}_i \colon \mathbb{D} \to \mathcal{R}_i$ are mutually independent $\mathcal{G}$ ε_i-DP mechanisms, then $\mathcal{M} = (\mathcal{M}_1, \dots, \mathcal{M}_k)_{\mathrm{ind}}$ is $\mathcal{G}$ $(\sum_{i=1}^{k} \varepsilon_i)$-DP. This shows that sequential composition works as expected for every granularity.

On the other hand, the setting 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.8. We can model this setting by taking f_i in the IC theorem (IV.1) so they define a partitioning function. More formally, we define a *k-partitioning function* $p = \{p_1, \dots, p_k\}$ as a function where $p_i \colon \mathbb{D} \to p_i(\mathbb{D}) =: \mathbb{D}_i$ such that $p_i(D) \subseteq D$ with $p_i(D) \cap p_j(D) \neq \varnothing$ for $i \neq j$[5]. Note, therefore, that the domains $\mathbb{D}_i$ of $\mathcal{M}_i$ might be different in this setting by construction. Let us see an example of a partitioning function, based on that of [16].

[5]We do not require that $D = \bigcup_{i=1}^{k} p_i(D)$, i.e., our partition can be *non-exhaustive*.

Example IV.4 (Partitioning function for $\mathbb{D} \subseteq \mathbb{D}_{\mathcal{X}}$)**.** Let $\mathbb{D} \subseteq \mathbb{D}_{\mathcal{X}}$. A partition $\{\mathcal{X}_i\}_{i \in [k]}$ of $\mathcal{X}$, extends naturally as a partition of the elements $D \in \mathbb{D}$, i.e., $p_i(D) \subseteq D$ is the multiset such that element $x \in D$ has multiplicity $m_{p_i(D)}(x) = \mathbf{1}_{\mathcal{X}_i}(x)\, m_D(x)$. In this case, the partitioning function p uses only x to compute the value of $p(x)$, and therefore the result is independent of the other records.

In this setting, the IC theorem (IV.1) yields 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 $I_p(D, D') \coloneqq \#\{i \mid p_i(D) \neq p_i(D')\}$. This fact is coherent with what we know: Assuming a partitioning function of Example IV.4, if we select $\varepsilon_i d_{\mathbb{D}_i}^{\mathcal{U}}$ mechanisms then $d_{\mathbb{D}} \leq (\max_{i \in [k]} \varepsilon_i) d_{\mathbb{D}}^{\mathcal{U}}$, since $\Delta p_i = \varepsilon_i$ and $I_p(D, D') = 1$ for all $D \sim_{\mathcal{U}} D'$. If we select $d_i = \varepsilon_i d_{\mathbb{D}_i}^{\mathcal{B}}$, there exist $D, D' \in \mathbb{D}$, as we saw in Example II.9, such that $d_i(D, D') = d_{\mathbb{D}_i}^{\mathcal{B}}(p_i(D), p_i(D')) = \infty$ for some i and therefore $d_{\mathbb{D}}(D, D') = \infty$. In general, we have no better expression for $d_{\mathbb{D}}$ unless we add extra conditions. In Sections IV-C and IV-D, we will explore conditions to achieve the best bound in this setting.

Furthermore, between accessing the whole database (Theorem IV.3) or a partition of it, the IC theorem (IV.1) allows considering intermediate composition strategies that provide tighter, more-precise bounds, such as shown in the following example:

Example IV.5. We continue with the scenario presented in Example IV.2, but now we have $k > 3$ hospitals and each ambulance has at least three associated hospital locations. The universe of records in this case is $\mathcal{X}' = (\mathcal{X}, [k]^{\leq 3})$ and $\mathbb{D} = \mathbb{D}_{\mathcal{X}'}$, where $[k]^{\leq 3}$ denotes the subsets of at least three elements of $[k]$. We consider the analogous $\mathcal{M}_a^*$ mechanism. We want to know the number of available ambulances for each hospital, so we consider $\mathcal{M}$ such that $\mathcal{M}(D) = (\mathcal{M}_a^*(f_1(D)), \dots, \mathcal{M}_a^*(f_k(D)))$ where $f_i(D)$ is the subdatabase of $D \in \mathbb{D}$ of ambulances assigned to hospital i. Since each ambulance only collaborates with at most three hospitals, $I_f(D, D') \leq 3 d_{\mathbb{D}}^{\mathcal{U}}(D, D')$. Applying then the IC theorem (IV.1), we obtain that $\mathcal{M}$ is $d_{\mathbb{D}}$-private with $d_{\mathbb{D}}(D, D') = \sum_{i=1}^{k} d_{\mathbb{D}}^{\mathcal{U}}(f_i(D), f_i(D')) \leq 3 d_{\mathbb{D}}^{\mathcal{U}}(D, D') < k d_{\mathbb{D}}^{\mathcal{U}}(D, D')$.

In particular, the last example showcases this intermediate setting. Function $f = \{f_i\}_{i \in [k]}$ does not define a partition, so we cannot apply Theorem II.8, but a single change to database D affects at most three databases in $\{f_i(D)\}_{i \in [k]}$, hence the final budget is $d_{\mathbb{D}} \leq 3 d_{\mathbb{D}}^{\mathcal{U}}$ instead of $k d_{\mathbb{D}}^{\mathcal{U}}$ given by the sequential counterpart (Theorem IV.3).

C. A Better Bound for Disjoint Inputs

Following the discussion in Section IV-B, considering as input disjoint subsets of the initial database, we explore the possibility to obtain the best possible bound. For this section, we assume that mechanisms $\mathcal{M}_i$ are d_i-private, with d_i "proportional" to a single metric type (e.g., $d_i = \varepsilon_i d_{\mathbb{D}_i}^{\triangle}$) or over a fixed granularity (i.e., $d_i = \varepsilon_i d_{\mathbb{D}_i}^{\mathcal{G}}$).

Theorem II.8 tells us that if $\mathcal{M}_i$ are $\varepsilon_i d^{\triangle}_{\mathbb{D}_\mathcal{X}}$-private, then the composed mechanism $\mathcal{M} = (\mathcal{M}_1, \dots, \mathcal{M}_k)_{\text{ind}}$ is $(\max_{i\in[k]} \varepsilon_i) d^{\triangle}_{\mathbb{D}_\mathcal{X}}$-private. This privacy bound is the best possible bound we can get in this setting. Note that in the case where mechanisms $\mathcal{M}_i$ satisfy the same privacy guarantee ($\varepsilon d^{\triangle}_{\mathbb{D}_\mathcal{X}}$-privacy) for all $i \in [k]$, then $\mathcal{M}$ also satisfies it. Thus, the composition does not degrade the privacy level at all. However, as we mentioned before, the best bound cannot be obtained for all metrics. Therefore we explore in this section which additional conditions the partition must satisfy (with respect to the metric) to ensure that we obtain the best-case bound, the maximum privacy budget of $\mathcal{M}_i$.

The first case we consider is a metric-type d^* that is well-defined over $\mathbb{D}$ and $\mathbb{D}_i$ for all $i \in [k]$[6]. We can give a sufficient condition for obtaining the best bound: We say that metric d^* *commutes with the partition given by* p if, for all $D, D' \in \mathbb{D}$,

$$\sum_{i=1}^{k} d^*_{\mathbb{D}_i}(p_i(D), p_i(D')) = d^*_{\mathbb{D}}\left(\bigcup_{i=1}^{k} p_i(D), \bigcup_{i=1}^{k} p_i(D')\right) \leq d^*_{\mathbb{D}}(D, D').$$

By the IC theorem (IV.1), if d^* commutes with p and $\mathcal{M}_i$ are $\varepsilon_i d^*_{\mathbb{D}_i}$-private, then $\mathcal{M}$ is $(\max_{i\in[k]} \varepsilon_i) d^*_{\mathbb{D}}$-private. For example, $d^{\triangle}$ commutes with all partitions p of Example IV.4 (see Proposition A.4 in the long version), which relates to the original result of McSherry [17].

Secondly, we can also focus on a fixed granularity notion $\mathcal{G}$, and given $d_i = \varepsilon_i d^{\mathcal{G}}_{\mathbb{D}_i}$ for all $i \in [k]$, we study when we obtain that $\mathcal{M}$ is $\varepsilon d^{\mathcal{G}}_{\mathbb{D}}$-private with $\varepsilon = \max_{i\in[k]} \varepsilon_i$. Recall that different domains define different canonical metrics, so the previous case does not apply, and checking commutativity is not an option. In this case, the corresponding equation translates to $\sum_{i=1}^{k} d^{\mathcal{G}}_{\mathbb{D}_i}(p_i(D), p_i(D')) = d^{\mathcal{G}}_{\mathbb{D}}(D, D')$. This equation can be hard to check in general, but it holds if the partition verifies:

- $d^{\mathcal{G}}_{\mathbb{D}}$*-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., $I_p(D, D') = 1$ for all $D \sim_{\mathcal{G}} D'$; and
- $\mathcal{G}$ is also well-defined over $\mathbb{D}_i$ and the sensitivity of p_i with respect to $d^{\mathcal{G}}_{\mathbb{D}}$ and $d^{\mathcal{G}}_{\mathbb{D}_i}$ is $\Delta p_i \leq 1$ (i.e., $d^{\mathcal{G}}_{\mathbb{D}_i}(p_i(D), p_i(D')) \leq 1$ if $d^{\mathcal{G}}_{\mathbb{D}}(D, D') = 1$).

Under these conditions, we obtain the desired result (where $\mathcal{M}^*_i$ can have different domains):

Theorem IV.6 (IC best bound for disjoint inputs)**.** *Let $\mathbb{D}$ be a database class and $\mathcal{G}$ a granularity over $\mathbb{D}$. Let p be a $d^{\mathcal{G}}_{\mathbb{D}}$-compatible k-partitioning function such that $\Delta p_i \leq 1$. For all $i \in [k]$, let $\mathcal{M}^*_i \colon \mathbb{D}_i \to \mathcal{R}_i$ be mutually independent $\varepsilon_i d^{\mathcal{G}}_{\mathbb{D}_i}$-private mechanisms. Then mechanism $\mathcal{M} = (\mathcal{M}^*_1 \circ p_1, \dots, \mathcal{M}^*_k \circ p_k)_{\text{ind}}$ is $\varepsilon d^{\mathcal{G}}_{\mathbb{D}}$-private with $\varepsilon = \max_{i\in[k]} \varepsilon_i$.*

[6]This means that metrics $d^*_{\mathbb{D}}$ and $d^*_{\mathbb{D}_i}$ are well-defined metrics and that $d^*(D, D')$ is constant for all domains containing $D, D' \in \mathbb{D}$. Examples include $d^{\triangle}$, which is well-defined for all $\mathbb{D} \subseteq \mathbb{D}_\mathcal{X}$.

As discussed, all partitions p of Example IV.4 are $d^{\mathcal{U}}_{\mathbb{D}}$-compatible since the addition/removal of one record can only affect the partition this record belongs to, so $I_p(D, D') = 1$ for all $D \sim_{\mathcal{U}} D'$, and additionally $\Delta p_i \leq 1$. Therefore, Theorem IV.6 can be applied to obtain Theorem II.8.

Even though Theorem IV.6 is stated for any granularity, $d^{\mathcal{G}}_{\mathbb{D}}$-compatibility is a strict condition. For example, no partitioning function of Example IV.4 (with $k > 1$) is $d^{\mathcal{B}}_{\mathbb{D}_\mathcal{X}}$-compatible (see Proposition A.5 in the long version). Nevertheless, we can construct compatible partitioning functions to certain bounded metrics $d^{\mathcal{B}}_{\mathbb{D}}$, as shown in the following example:

Example IV.7 (A $d^{\mathcal{B}}_{\mathbb{D}}$-compatible partition)**.** Consider a database D with ordered elements, i.e., every element $(n, x) \in D$ consists of a record value $x \in \mathcal{X}$ and a unique identifier $n \in [|D|]$. Let $\mathbb{D}^{\text{ord}}_\mathcal{X}$ denote the class of all such databases.

Let p be a k-partitioning function of $\mathbb{N}$, which induces a partition of the elements of $\mathbb{D} \subseteq \mathbb{D}^{\text{ord}}_\mathcal{X}$ that divides the databases only taking the order into account, i.e., such that $p(n, x) = p(n, y)$ for all $x, y \in \mathcal{X}$. Then p is $d^{\mathcal{B}}_{\mathbb{D}}$-compatible and verifies $\Delta p_i \leq 1$ (see proof in Proposition A.6 in the long version). Therefore, we can obtain the best bound for bounded in this case using Theorem IV.6.

D. Common-Domain Setting

The common-domain setting 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 $\mathbb{D}$. This change provides new composition rules that allow us to obtain better privacy bounds. Importantly, when we impose the privacy constraints in $\mathcal{M}^*_i$, in the case where $d_i(D, D')$ are well-defined and finite, we can still end up with $d_{\mathbb{D}}(D, D') = \infty$, as we saw in Example II.9. However, if $\mathcal{M}_i$ are d_i-private, we can bound the privacy loss by at least $d_{\mathbb{D}}(D, D') = \sum_{i=1}^{k} d_i(D, D') < \infty$, avoiding this problem.

In this scenario, while $\mathcal{M}_i$ can protect any database of $D \in \mathbb{D}$, the computation of $\mathcal{M}_i$ depends exclusively on the information of contained $f_i(D)$ and not the total information of D. The exclusive dependence of $\mathcal{M}_i$ on specific information improves the privacy guarantee and gives better privacy-loss bounds. To be able to analyze the composition in this setting, we present a coherent formalization of "depending exclusively on $f_i(D)$" under the notion of *dependency*:

Definition IV.8 (Dependency)**.** Let $\mathcal{M} \colon \mathbb{D} \to \mathcal{R}$ be a randomized mechanism, and let f be a deterministic map with domain $\mathbb{D}$. We say that $\mathcal{M}$ is f*-dependent* if there exists $\mathcal{M}^* \colon f(\mathbb{D}) \to \mathcal{R}$ such that $\mathcal{M} = \mathcal{M}^* \circ f$.

This definition implies that $\mathrm{P}\{\mathcal{M}^*(f(D)) \in S\} = \mathrm{P}\{\mathcal{M}(D) \in S\}$ for all measurable $S \subseteq \mathcal{R}$. Since $\mathcal{M}^*(f(D))$ depends exclusively on $f(D)$, consequently $\mathcal{M}(D)$ depends

exclusively on the information in $f(D)$ for all $D \in \mathbb{D}$ (i.e., only data in $f(D)$ affects the output of $\mathcal{M}(D)$).

Example IV.9 (Dependency)**.** Let us revisit the scenario of Example IV.2. For $i \in [2]$, we define $\mathcal{M}_i \colon \mathbb{D}_{\mathcal{X}} \to \mathbb{N}$ such that it outputs the noisy participants count from district i in D, i.e., $\mathcal{M}_i(D) = \sum_{x \in \mathcal{X}_i} m_D(x) + z$ with $z \sim \mathrm{Lap}(\frac{1}{\varepsilon_i})$ (note it is the Laplace mechanism). Mechanisms $\mathcal{M}_i$ are p_i-dependent, since there exists $\mathcal{M}_i^*(D) = |D| + z$ such that $\mathcal{M}_i = \mathcal{M}_i^* \circ p_i$. This means that, even though $\mathcal{M}$ takes as input the whole database D, it just needs to see the information contained in subset $p_i(D)$ to know how many locations belong to district i.

Under this definition, we arrive at the following result:

Proposition IV.10 (Minimum privacy)**.** *Let $(\mathbb{D}, d_{\mathbb{D}})$ be a privacy space, let f be a deterministic map with domain $\mathbb{D}$, and let $\mathcal{M} \colon \mathbb{D} \to \mathcal{R}$ be a $d_{\mathbb{D}}$-private mechanism. If $\mathcal{M}$ is f-dependent, then $\mathcal{M}$ is $d_{\mathbb{D}}^f$-private* with*

$$d_{\mathbb{D}}^f(D, D') \coloneqq \min_{\substack{\tilde{D}, \tilde{D}' \in \mathbb{D} \\ f(\tilde{D}) = f(D) \\ f(\tilde{D}') = f(D')}} d_{\mathbb{D}}(\tilde{D}, \tilde{D}').$$

Note that $d_{\mathbb{D}}^f$ is not necessarily a metric[7] (thus we call it *d-privacy**). However, it gives an accurate value for the distance between the probability distributions of the output given two input databases. Since $d_{\mathbb{D}}^f \leq d_{\mathbb{D}}$, having the dependency constraint in a mechanism can imply more privacy. This way, the privacy loss is chosen as the minimum with respect to the dependent data $f(D)$, and not D. In particular, if $f(D) = f(D')$ for a pair $D, D' \in \mathbb{D}$, then $d_{\mathbb{D}}^f(D, D') = 0$. Furthermore, it is possible to find metrics d in-between these, i.e., $d_{\mathbb{D}}^f \leq d \leq d_{\mathbb{D}}$.

Applying Proposition IV.10 to the IC theorem (IV.1), we obtain:

Theorem IV.11 (IC theorem for common domain)**.** *For $i \in [k]$, 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 mutually independent mechanisms satisfying d_i-privacy and f_i-dependency. Then mechanism $\mathcal{M} = (\mathcal{M}_1, \ldots, \mathcal{M}_k)_{\mathrm{ind}}$ is $d_{\mathbb{D}}$-private* with $d_{\mathbb{D}} \coloneqq \sum_{i=1}^k d_i^{f_i}$.*

We can also bound the result with

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

which are not metrics, but are better bounds than $\sum_{i=1}^k d_i$ given by the IC theorem (IV.1). Translating this result to the case of 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

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

[7] It does not generally fulfill the triangle inequality.

Theorem IV.11 allows us to obtain the corresponding cases, corollaries, and examples to those we obtained from the IC theorem (IV.1) for this new setting. In some cases, such as taking $f_i = \mathrm{id}$ for all $i \in [k]$, correspond to the same result (Theorem IV.3), since $d_{\mathbb{D}}^{\mathrm{id}} = d_{\mathbb{D}}$. In others, however, the change of setting leads to a different scenario and results, such as when trying to find the best bound for disjoint inputs (i.e., the counterpart of Section IV-C).

The corresponding question of Section IV-C translates as follows: Given k mechanisms $\mathcal{M}_i \colon \mathbb{D} \to \mathcal{R}$ that are d_i-private with $d_i = \varepsilon_i d$ for a metric d over $\mathbb{D}$ and p_i-dependent with p an arbitrary partitioning function, we are interested in studying the conditions such that $\mathcal{M} = (\mathcal{M}_1, \ldots, \mathcal{M}_k)_{\mathrm{ind}}$ is $d_{\mathbb{D}}$-private with $d_{\mathbb{D}} = (\max_{i \in [k]} \varepsilon_i) d$.

The natural approach is to check when metric d verifies

$$\sum_{i=1}^k d^{p_i}(D, D') = d(D, D') \tag{IV.1}$$

for all $D, D' \in \mathbb{D}$, since then $d_{\mathbb{D}} = \max_{i \in [k]} \varepsilon_i d$ follows from Theorem IV.11.

Equation (IV.1) can be hard to check directly, but we can give sufficient conditions for it when $d = d_{\mathbb{D}}^{\mathcal{G}}$, the canonical distance of a granularity notion. Here, it is sufficient to ask that the partition is $d_{\mathbb{D}}^{\mathcal{G}}$-compatible.

Theorem IV.12 (IC best bound for disjoint inputs (common domain))**.** *Let $\mathbb{D}$ be a database class and $\mathcal{G}$ a granularity over $\mathbb{D}$. Let p be a $d_{\mathbb{D}}^{\mathcal{G}}$-compatible k-partitioning function. For all $i \in [k]$, let $\mathcal{M}_i \colon \mathbb{D} \to \mathcal{R}_i$ be mutually independent mechanisms satisfying $\varepsilon_i d_{\mathbb{D}}^{\mathcal{G}}$-privacy and p_i-dependency. Then mechanism $\mathcal{M} = (\mathcal{M}_1, \ldots, \mathcal{M}_k)_{\mathrm{ind}}$ is $\varepsilon d_{\mathbb{D}}^{\mathcal{G}}$-private with $\varepsilon = \max_{i \in [k]} \varepsilon_i$.*

Note that in this case, it is not necessary to impose "$\Delta p_i \leq 1$", which was necessary for our previous theorem (IV.6). Theorem IV.12 is therefore also a consequence of preprocessing (Proposition III.6) applied to Theorem IV.6.

E. A Better Composition for the Bounded Case over Disjoint Databases

The strict conditions necessary to obtain the $\max_{i \in [k]} \varepsilon_i$ bound in Theorems IV.6 and IV.12 cannot be achieved in the bounded case for partitions of Example IV.4, because they are not $d_{\mathbb{D}}^{\mathcal{B}}$-compatible in general. This is also true for other granularities, especially those based on the bounded notion. However, even if Theorems IV.6 and IV.12 do not apply, we can still compute the best-case bound when considering a partition of the database.

In this subsection, we briefly discuss how we can bound the minimum privacy budget consumed when taking a partition of the databases using Theorem IV.11. We thus provide a solution to the problem posed by Li et al. [16], obtaining a tight bound for composition over disjoint databases in bounded DP (when taking a partition of Example IV.4), which was previously missing.

Corollary IV.13. *Let p be a k-partitioning function of Example IV.4. For all $i \in [k]$, let $\mathcal{M}_i \colon \mathbb{D} \to \mathcal{R}_i$ be mutually independent mechanisms satisfying bounded ε_i-DP and p_i-dependent. Then mechanism $\mathcal{M} = (\mathcal{M}_1, \dots, \mathcal{M}_k)_{\text{ind}}$ with domain $\mathbb{D}$ is bounded ε-DP with $\varepsilon = \max_{i,j \in [k];\, i \neq j}(\varepsilon_i + \varepsilon_j)$.*

Note that this result is stated for common domain, and that the non–common-domain counterpart cannot be defined as we prove in Example II.9.

Also, note that returning to the DP formulation increases the tightness of the bound with respect to the direct statement using $d^{\mathcal{B}}_{\mathbb{D}}$-privacy. In this case, the best bound is $\sum_{i=1}^{k} \varepsilon_i d^{\mathcal{B},p_i}_{\mathbb{D}}$. To showcase the improvement we add the following example:

Example IV.14. We continue from Example IV.9 but considering $k > 2$ districts instead of two. We already showed that $\mathcal{M}_i$ are p_i-dependent. Besides, $\mathcal{M}_i$ are $\varepsilon_i d^{\mathcal{B}}_{\mathbb{D}_{\mathcal{X}}}$-private because they are Laplace mechanisms. Applying Corollary IV.13, we have that $\mathcal{M} = (\mathcal{M}_1, \dots, \mathcal{M}_n)_{\text{ind}}$ is $\varepsilon d^{\mathcal{B}}_{\mathbb{D}_{\mathcal{X}}}$-private for $\varepsilon = \max_{i,j \in [k];\, i \neq j}(\varepsilon_i + \varepsilon_j)$. Particularly, given $D \sim_{\mathcal{B}} D'$ with $D \triangle D' = \{x, x'\}$, D and D' are indistinguishable up to ε_i if x and x' are in the same district i, and up to $\varepsilon_i + \varepsilon_j$ if they are in different districts $i \neq j$. Note that this is a much better bound than applying sequential composition directly, which would give us that $D \sim_{\mathcal{B}} D'$ are indistinguishable up to $\sum_{i=1}^{k} \varepsilon_i > \max_{i,j \in [k];\, i \neq j}(\varepsilon_i + \varepsilon_j)$.

We conclude this section with a small result obtained by applying Proposition IV.10 to bounded DP in $\mathbb{D}_{\mathcal{X}}$.

Corollary IV.15. *Let $\mathbb{D}_{\mathcal{X}}$ be a database universe, $\mathcal{Y} \subsetneq \mathcal{X}$ and $f \colon \mathbb{D}_{\mathcal{X}} \to \mathbb{D}_{\mathcal{Y}}$ such that $f(D) = D \cap \mathcal{Y}$. Let $\mathcal{M} \colon \mathbb{D}_{\mathcal{X}} \to \mathcal{R}$ be a $d^{\mathcal{B}}_{\mathbb{D}_{\mathcal{X}}}$-private mechanism that is f-dependent. Then, $\mathcal{M}$ is $d_{\mathbb{D}_{\mathcal{X}}}$-private* with*

$$\begin{aligned} d_{\mathbb{D}_{\mathcal{X}}}(D, D') &\coloneqq \min\{d^{\mathcal{B}}_{\mathbb{D}_{\mathcal{X}}}(D, D'), |f(D) \triangle f(D')|\} \\ &\leq \min\{d^{\mathcal{B}}_{\mathbb{D}_{\mathcal{X}}}(D, D'), d^{\mathcal{U}}_{\mathbb{D}_{\mathcal{X}}}(D, D')\}. \end{aligned}$$

V. The Adaptive Composition Theorem

In the previous section, we elaborated on composability when we apply mechanisms that work independently from each other, obtaining the IC theorem (IV.1). However, the question remains open on how composition works in the adaptive scenario, where each mechanism can also take as input the output of the previous mechanisms. In this section, we discuss adaptive composition, which is a generalization of independent composition, and provide the adaptive counterparts to the theorems of Section IV.

To be precise, we formalize the adaptive-composed mechanism as follows:

Definition V.1 (Adaptive-composed mechanism)**.** For $i \in [k]$, let $\overline{\mathcal{R}}_i \coloneqq \mathcal{R}_1 \times \cdots \times \mathcal{R}_{i-1}$ (where $\overline{\mathcal{R}}_1 = \varnothing$), and let $\mathcal{M}_i \colon \overline{\mathcal{R}}_i \times \mathbb{D} \to \mathcal{R}_i$ be randomized mechanisms. We define the *adaptive-composed mechanism* $\mathcal{M} \coloneqq (\mathcal{M}_1, \dots, \mathcal{M}_k)_{\text{adapt}}$ as the mechanism with domain $\mathbb{D}$ such that $\mathcal{M}(D) = (\mathcal{N}_1(D), \dots, \mathcal{N}_k(D))$ for all $D \in \mathbb{D}$, where $\mathcal{N}_i(D)$ are defined recursively as $\mathcal{N}_i(D) = \mathcal{M}_i(\mathcal{N}_1(D), \dots, \mathcal{N}_{i-1}(D), D)$ for $i \in [k]$ (where $\mathcal{N}_1 = \mathcal{M}_1$).

In other words, given $D \in \mathbb{D}$, $\mathcal{M}$ first draws D_1 following the distribution of $\mathcal{M}_1(D)$; then, $\mathcal{M}$ draws D_i following the distribution of $\mathcal{M}_i(D_1, \dots, D_{i-1}, D)$ for each $i = 2, \dots, k$ in order. At the end, $\mathcal{M}$ outputs $\mathcal{M}(D) = (D_1, \dots, D_k)$.

Note that adaptive-composed mechanisms are more general than independent-composed mechanisms, corresponding to the case where $\mathcal{M}_i$ are mutually independent and, in particular, constant over $\overline{\mathcal{R}}_i$.

We directly define the adaptive composition (AC) theorem. Similar to the independent results, this result does not impose any conditions on the privacy level of the initial mechanisms $\mathcal{M}_i$.

Theorem V.2 (AC theorem)**.** *Let $\mathbb{D}$ be a database class, and, for all $i \in [k]$, let $(\mathbb{D}_i, d_i)$ be a privacy space, $f_i \colon \mathbb{D} \to \mathbb{D}_i$ a deterministic map and $f^*_i = \mathrm{id}_{\overline{\mathcal{R}}_i} \times f_i$ (with $f^*_1 = f_1$).*

For $i \in [k]$, let $\mathcal{M}^*_i \colon \overline{\mathcal{R}}_i \times \mathbb{D}_i \to \mathcal{R}_i$ be a mechanism such that $\mathcal{M}^*_i(\bar{s}_i, \cdot) \colon \mathbb{D}_i \to \mathcal{R}_i$ satisfies d_i-privacy for any $\bar{s}_i \in \overline{\mathcal{R}}_i$.

Then mechanism $\mathcal{M} = (\mathcal{M}^*_1 \circ f^*_1, \dots, \mathcal{M}^*_k \circ f^*_k)_{\text{adapt}}$ is $d_{\mathbb{D}}$-private with

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

Observe that the expression of $d_{\mathbb{D}}$ does not change with respect to the IC theorem (IV.1). Therefore using adaptive composition, which is more general than independent composition, does not affect the privacy bound of the resulting mechanism; or, alternatively, no improvement is gained by considering mechanisms $\mathcal{M}_1, \dots, \mathcal{M}_k$ mutually independent.

Analogously to the independent case, particular composition rules can be derived from Theorem V.2, as well as translated to the common domain. The same consequences are extracted from these adaptive results. We present such results, which also generalize their respective independent cases.

First, if we impose $f_i = \mathrm{id}$ and $\mathbb{D} = \mathbb{D}_i$ for all $i \in [k]$, we obtain a generalization of the sequential setting as expected:

Theorem V.3 (Generalized ASC)**.** *Let $\{(\mathbb{D}, d_i)\}_{i \in [k]}$ be a set of privacy spaces. For $i \in [k]$, let $\mathcal{M}_i \colon \overline{\mathcal{R}}_i \times \mathbb{D} \to \mathcal{R}_i$ be a mechanism such that $\mathcal{M}_k(\bar{s}_i, \cdot) \colon \mathbb{D} \to \mathcal{R}_i$ is d_i-private for all $\bar{s}_i \in \overline{\mathcal{R}}_i$. Then $\mathcal{M} = (\mathcal{M}_1, \dots, \mathcal{M}_k)_{\text{adapt}}$ is $(\sum_{i=1}^{k} d_i)$-private.*

Second, if we study what happens when we apply adaptive composition over disjoint subsets of the input, we obtain the analogous adaptive counterpart of Theorem IV.6:

Theorem V.4 (AC best bound for disjoint inputs)**.** *Let $\mathbb{D}$ be a database class and $\mathcal{G}$ a granularity over $\mathbb{D}$. Let p be a $d^{\mathcal{G}}_{\mathbb{D}}$-compatible k-partitioning function such that $\Delta p_i \leq 1$,*

and $p_i^* = \mathrm{id}_{\overline{\mathcal{R}}_i} \times p_i$ *(with* $p_1^* = p_1$*). For* $i \in [k]$*, let* $\mathcal{M}_i^* \colon \overline{\mathcal{R}}_i \times \mathbb{D}_i \to \mathcal{R}_i$ *be a mechanism such that* $\mathcal{M}_i^*(\bar{s}_i, \cdot) \colon \mathbb{D}_i \to \mathcal{R}_i$ *satisfies* $\varepsilon_i d_{\mathbb{D}_i}^{\mathcal{G}}$*-privacy for any* $\bar{s}_i \in \overline{\mathcal{R}}_i$*. Then mechanism* $\mathcal{M} = (\mathcal{M}_1^* \circ p_1^*, \ldots, \mathcal{M}_k^* \circ p_k^*)_{\text{adapt}}$ *is* $\varepsilon d_{\mathbb{D}}^{\mathcal{G}}$*-private with* $\varepsilon = \max_{i \in [k]} \varepsilon_i$*.*

Note that we cannot get around the problem that $\mathcal{M}_i^*$ being d-privacy does not imply that $\mathcal{M}_i = \mathcal{M}_i^* \circ f_i$ is d-private in the adaptive setting. Therefore, we show the common-domain results that show what happens if we impose $\mathcal{M}_i$ to be d_i-private directly (i.e., the counterparts to Theorems IV.11 and IV.12):

Theorem V.5 (AC theorem for common domain)**.** *For* $i \in [k]$*, let* $(\mathbb{D}, d_i)$ *be a privacy space, and let* f_i *be a deterministic map over* $\mathbb{D}$*. For* $i \in [k]$*, let* $\mathcal{M}_i \colon \overline{\mathcal{R}}_i \times \mathbb{D} \to \mathcal{R}_i$ *be a mechanism such that* $\mathcal{M}_k(\bar{s}_i, \cdot) \colon \mathbb{D} \to \mathcal{R}_i$ *satisfies* d_i*-privacy and* f_i*-dependency for any* $\bar{s}_i \in \overline{\mathcal{R}}_i$*. Then mechanism* $\mathcal{M} = (\mathcal{M}_1, \ldots, \mathcal{M}_k)_{\text{adapt}}$ *is* $d_{\mathbb{D}}$*-private* with* $d_{\mathbb{D}} \coloneqq \sum_{i=1}^{k} d_i^{f_i}$*.*

Theorem V.6 (AC best bound for disjoint inputs (common domain))**.** *Let* $\mathbb{D}$ *be a database class and* $\mathcal{G}$ *a granularity over* $\mathbb{D}$*. Let* p *be a* $d_{\mathbb{D}}^{\mathcal{G}}$*-compatible* k*-partitioning function. For* $i \in [k]$*, let* $\mathcal{M}_i \colon \overline{\mathcal{R}}_i \times \mathbb{D} \to \mathcal{R}_i$ *be a mechanism such that* $\mathcal{M}_k(\bar{s}_i, \cdot) \colon \mathbb{D} \to \mathcal{R}_i$ *satisfies* $\varepsilon_i d_{\mathbb{D}}^{\mathcal{G}}$*-privacy and* p_i*-dependency for any* $\bar{s}_i \in \overline{\mathcal{R}}_i$*. Then mechanism* $\mathcal{M} = (\mathcal{M}_1, \ldots, \mathcal{M}_k)_{\text{adapt}}$ *is* $\varepsilon d_{\mathbb{D}}^{\mathcal{G}}$*-private with* $\varepsilon = \max_{i \in [k]} \varepsilon_i$*.*

Observe that, since all results are a consequence of the AC theorem (V.2), which has the same bound as its IC counterpart, none of the results degrade their bound with respect to their IC versions.

VI. Extending to Other DP-Based Notions

Given that composability is not an exclusive property of ε-DP, but also of other DP-based notions, it is interesting to understand how composition extends to other DP-based notions. In this section, we present $d_{\mathbb{D}}$-privacy formulations of approximate DP [8], zero-concentrated DP [1], and Gaussian DP [6], and study the corresponding adaptive composition theorems. Note that since each notion has its own group property, each extension behaves differently than that of $d_{\mathbb{D}}$-privacy, although similar patterns are present.

A. Extending to Approximate DP

Approximate DP [8], also known as (ε, δ)-DP, is an important and popular extension of DP. In this section, we introduce an adapted version of $d_{\mathbb{D}}$-privacy for the approximate scenario, $(d_{\mathbb{D}}, \delta_{\mathbb{D}})$-privacy, which generalizes (ε, δ)-DP in the same way that $d_{\mathbb{D}}$-privacy generalizes ε-DP. Afterward, we present the composition results for this notion.

From the original definition of (ε, δ)-DP [8, 9], defined for unbounded neighboring databases, we present the definition of approximate DP for any granularity:

Definition VI.1 ($\mathcal{G}$ (ε, δ)-DP)**.** Let $\varepsilon, \delta \geq 0$. A randomized mechanism $\mathcal{M}$ with domain $\mathbb{D}_{\mathcal{X}}$ is $\mathcal{G}$ (ε, δ)*-DP* if for all measurable $S \subseteq \mathrm{Range}(\mathcal{M})$ and for all $\mathcal{G}$-neighboring $D, D' \in \mathbb{D}_{\mathcal{X}}$,

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

Different privacy interpretations of δ can be found in [2, 4, 9, 18]. Note that having $\delta \geq 1$ is meaningless and provides no privacy since any mechanism, including one that releases the raw data, is $\mathcal{G}$ (ε, δ)-DP for $\delta \geq 1$.

Our definition of $(d_{\mathbb{D}}, \delta_{\mathbb{D}})$-privacy is formulated so that Theorem VI.3 verifies, which is analogous to Theorem III.3 for the pure-DP case. The construction of $d_{\mathbb{D}}$-privacy from ε-DP uses the fact that the privacy budget ε scales linearly with respect to distance $d_{\mathbb{D}}$.

Definition VI.2 ($(d_{\mathbb{D}}, \delta_{\mathbb{D}})$-privacy)**.** Let $d_{\mathbb{D}}$ be a metric over $\mathbb{D}$ and $\delta_{\mathbb{D}} \colon \mathbb{D}^2 \to [0, \infty]$. Then, a randomized mechanism $\mathcal{M}$ with domain $\mathbb{D}$ is $(d_{\mathbb{D}}, \delta_{\mathbb{D}})$*-private* if for all $D, D' \in \mathbb{D}$ and all measurable $S \subseteq \mathrm{Range}(\mathcal{M})$,

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

Analogously to (ε, δ)-DP, if $\delta_{\mathbb{D}}(D, D') \geq 1$, the indistinguishability (up to ε) between $D, D' \in \mathbb{D}$ is no longer guaranteed. Moreover, we recover $d_{\mathbb{D}}$-privacy when $\delta_{\mathbb{D}} = 0$.

Note that $\delta_{\mathbb{D}}$ does not need to be a metric. Furthermore, in (ε, δ)-DP, δ does not scale linearly under group privacy, but rather ends up as $\delta \frac{\mathrm{e}^{\varepsilon n} - 1}{\mathrm{e}^{\varepsilon} - 1}$ (which can be larger than 1). Parameter $\delta_{\mathbb{D}}$ scales in the same way, which is shown in our next result, where we denote as $[d]_{\varepsilon} \colon \mathbb{D}^2 \to [0, \infty]$ the function such that $[d]_{\varepsilon}(D, D') = \frac{1}{\mathrm{e}^{\varepsilon} - 1}(\mathrm{e}^{\varepsilon d(D, D')} - 1)$.

Theorem VI.3. *Let* $\mathcal{G}$ *be a granularity notion over the database class* $\mathbb{D}$*. Then, a mechanism* $\mathcal{M}$ *with domain* $\mathbb{D}$ *is* $(\varepsilon d_{\mathbb{D}}^{\mathcal{G}}, \delta[d_{\mathbb{D}}^{\mathcal{G}}]_{\varepsilon})$*-private if and only if it is* $\mathcal{G}$ (ε, δ)*-DP.*

However, please note that $\delta[d_{\mathbb{D}}^{\mathcal{G}}]_{\varepsilon}$ can scale to numbers greater than 1. This can lead to weak privacy models since such values result in no privacy, as we said before. For instance with $\delta = 10^{-5}$ and $\varepsilon = 1$ we have $\delta[d_{\mathbb{D}}^{\mathcal{B}}]_{\varepsilon}(D, D') > 1$ for all $D, D' \in \mathbb{D}$ such that $d_{\mathbb{D}}^{\mathcal{B}}(D, D') \geq 13$. Therefore a $(d_{\mathbb{D}}^{\mathcal{B}}, 10^{-5}[d_{\mathbb{D}}^{\mathcal{B}}]_1)$-private mechanism can allow an attacker to likely distinguish outputs of two databases in which we have changed more than thirteen records.

We now present the AC result for $(d_{\mathbb{D}}, \delta_{\mathbb{D}})$-privacy.

Theorem VI.4 (Approximate AC theorem)**.** *Let* $\mathbb{D}$ *be a database class, and, for all* $i \in [k]$*, let* $(\mathbb{D}, d_i)$ *be a privacy space and* $\delta_i \colon \mathbb{D}^2 \to [0, \infty]$*. Let* $f_i \colon \mathbb{D} \to \mathbb{D}_i$ *be a deterministic map and* $f_i^* = \mathrm{id}_{\overline{\mathcal{R}}_i} \times f_i$ *(with* $f_1^* = f_1$*).*

For $i \in [k]$*, let* $\mathcal{M}_i^* \colon \overline{\mathcal{R}}_i \times \mathbb{D}_i \to \mathcal{R}_i$ *be a mechanism such that* $\mathcal{M}_i^*(\bar{s}_i, \cdot) \colon \mathbb{D}_i \to \mathcal{R}_i$ *satisfies* (d_i, δ_i)*-privacy for any* $\bar{s}_i \in \overline{\mathcal{R}}_i$*.*

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

$$d_{\mathbb{D}}(D, D') \coloneqq \sum_{i=1}^{k} d_i(f_i(D), f_i(D')) \quad \textit{and}$$

$$\delta_{\mathbb{D}}(D, D') \coloneqq \sum_{i=1}^{k} \delta_i(f_i(D), f_i(D')) \quad \textit{for all } D, D' \in \mathbb{D}.$$

Note that from this result here, we are able to derive all the results so far in this paper. In addition, Theorem VI.4 can be used to define the approximate variations of all our main composition results, where the same consequences can be extracted as in Section IV:

Theorem VI.5 (Generalized approximate ASC)**.** *Let $\mathbb{D}$ be a database class, and, for all $i \in [k]$, let $(\mathbb{D}, d_i)$ be a privacy space and $\delta_i \colon \mathbb{D}^2 \to [0, \infty]$. For $i \in [k]$, let $\mathcal{M}_i \colon \overline{\mathcal{R}}_i \times \mathbb{D} \to \mathcal{R}_i$ be a mechanism such that $\mathcal{M}_k(\bar{s}_i, \cdot) \colon \mathbb{D} \to \mathcal{R}_i$ is (d_i, δ_i)-private for any $\bar{s}_i \in \overline{\mathcal{R}}_i$. Then $\mathcal{M} = (\mathcal{M}_1, \dots, \mathcal{M}_k)_{\text{adapt}}$ is $(\sum_{i=1}^k d_i, \sum_{i=1}^k \delta_i)$-private.*

It is important to remark that $\sum_{i=1}^k d_i < \infty$ if and only if $d_i < \infty$, but we can still end up with no privacy guarantee if $\sum_{i=1}^k \delta_i \geq 1$, which can happen even if all $\delta_i < 1$. This fact motivates further the search for tighter bounds and the introduction of the approximate counterpart of Theorem IV.6:

Theorem VI.6 (Approximate best bound for disjoint inputs)**.** *Let $\mathbb{D}$ be a database class and $\mathcal{G}$ a granularity over $\mathbb{D}$. Let p be a $d_{\mathbb{D}}^{\mathcal{G}}$-compatible k-partitioning function such that $\Delta p_i \leq 1$, and $p_i^* = \mathrm{id}_{\overline{\mathcal{R}}_i} \times p_i$ (with $p_1^* = p_1$). For $i \in [k]$, let $\mathcal{M}_i^* \colon \overline{\mathcal{R}}_i \times \mathbb{D}_i \to \mathcal{R}_i$ be a mechanism such that $\mathcal{M}_i^*(\bar{s}_i, \cdot) \colon \mathbb{D}_i \to \mathcal{R}_i$ satisfies $(\varepsilon_i d_{\mathbb{D}_i}^{\mathcal{G}}, \delta_i [d_{\mathbb{D}_i}^{\mathcal{G}}]_{\varepsilon_i})$-privacy for any $\bar{s}_i \in \overline{\mathcal{R}}_i$. Then mechanism $\mathcal{M} = (\mathcal{M}_1^* \circ p_1^*, \dots, \mathcal{M}_k^* \circ p_k^*)_{\text{adapt}}$ is $(\varepsilon d_{\mathbb{D}}^{\mathcal{G}}, \delta [d_{\mathbb{D}}^{\mathcal{G}}]_{\varepsilon})$-private with $\varepsilon = \max_{i \in [k]} \varepsilon_i$ and $\delta = \max_{i \in [k]} \delta_i$.*

Furthermore, the common-domain setting imposing $\mathcal{M}_i^* \circ f_i$ to be (d_i, δ_i)-private also leads to interesting composition results for this DP variation:

Theorem VI.7 (Approximate AC theorem for common domain)**.** *For $i \in [k]$, let $(\mathbb{D}, d_i)$ be a privacy space, $\delta_i \colon \mathbb{D}^2 \to [0, \infty]$, and let f_i be a deterministic map over $\mathbb{D}$. For $i \in [k]$, let $\mathcal{M}_i \colon \overline{\mathcal{R}}_i \times \mathbb{D} \to \mathcal{R}_i$ be a mechanism such that $\mathcal{M}_k(\bar{s}_i, \cdot) \colon \mathbb{D} \to \mathcal{R}_i$ satisfies (d_i, δ_i)-privacy and f_i-dependency for any $\bar{s}_i \in \overline{\mathcal{R}}_i$. Then mechanism $\mathcal{M} = (\mathcal{M}_1, \dots, \mathcal{M}_k)_{\text{ind}}$ is $(\sum_{i=1}^k d_i^{f_i}, \sum_{i=1}^k \delta_i^{f_i})$-private* with*

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

Example VI.8. We continue from Example IV.5, where we want to know the number of available ambulances for each hospital. However, we instead consider $\mathcal{M} = (\mathcal{M}_1, \dots, \mathcal{M}_k)_{\text{ind}}$ such that $\mathcal{M}_i = \mathcal{M}_a^* \circ f_i$ are bounded $(1, 10^{-5})$-DP (i.e., $(d_{\mathbb{D}}^{\mathcal{B}}, \delta_0 [d_{\mathbb{D}}^{\mathcal{B}}]_1)$-private with $\delta_0 = 10^{-5}$). By construction, for all $i \in [k]$, mechanism $\mathcal{M}_i$ outputs the perturbed number of ambulances linked to i and is f_i-dependent (where $f_i(D)$ is the subdatabase of $D \in \mathbb{D}$ of ambulances assigned to hospital i).

Applying Theorem VI.7, we obtain that mechanism $\mathcal{M}$ is $(\sum_{i=1}^k (d_{\mathbb{D}}^{\mathcal{B}})^{f_i}, \sum_{i=1}^k (\delta_0 [d_{\mathbb{D}}^{\mathcal{B}}]_1)^{f_i})$-private*. Note that under the bounded metric, we have that $I_f(D, D') \leq 6$ for all $D \sim_{\mathcal{B}} D'$. Therefore, we can bound the privacy parameters as follows: $\sum_{i=1}^k (d_{\mathbb{D}}^{\mathcal{B}})^{f_i}(D, D') \leq \sum_{i \in I_f(D, D')} d_{\mathbb{D}}^{\mathcal{B}}(D, D') \leq 6 d_{\mathbb{D}}^{\mathcal{B}}(D, D')$ and, analogously, $\sum_{i=1}^k (\delta_0 [d_{\mathbb{D}}^{\mathcal{B}}]_1)^{f_i}(D, D') \leq \delta_0 \sum_{i \in I_f(D, D')} [d_{\mathbb{D}}^{\mathcal{B}}]_1(D, D') \leq 6 \delta_0 [d_{\mathbb{D}}^{\mathcal{B}}]_1(D, D')$.

In conclusion, $\mathcal{M}$ is $(6 d_{\mathbb{D}}^{\mathcal{B}}, 6 \delta_0 [d_{\mathbb{D}}^{\mathcal{B}}]_1)$-private (i.e., bounded $(6, 6 \cdot 10^{-5})$-DP).

The approximate variant of Theorem V.6 can also be enunciated (see Theorem A.13 in the long version).

B. Extending to Zero-Concentrated DP

Another common adaptation of DP is *zero-concentrated DP* (zCDP) [1]. This privacy metric is based on a bound on the Rényi divergence:

Definition VI.9 (Rényi divergence [1, 22])**.** Given two probability distributions P and Q defined over $\mathcal{R}$, the *Rényi divergence of order* $\alpha \in (1, \infty)$ is defined as

$$D_\alpha(P \,\|\, Q) \coloneqq \frac{1}{\alpha - 1} \ln \int_{\mathcal{R}} p^\alpha q^{1-\alpha} \, \mathrm{d}\mu$$

where p and q are the densities of P and Q with respect to measure μ[8], respectively. For order $\alpha = \infty$, it is defined as

$$D_\infty(P \,\|\, Q) \coloneqq \lim_{\alpha \to \infty} D_\alpha(P \,\|\, Q) = \ln \sup_{S \text{ meas.}} \frac{P(S)}{Q(S)}$$

The previous integral notation will be useful to represent both continuous and discrete cases, i.e., if P and Q are continuous, the integral equals $\int_{\mathcal{R}} p(s)^\alpha q(s)^{1-\alpha} \, \mathrm{d}s$ with p and q the corresponding density functions, and if P and Q are discrete, it equals $\sum_{s \in \mathcal{R}} p(s)^\alpha q(s)^{1-\alpha}$ with p and q the corresponding probability mass functions.

Note that the Rényi divergence is not a metric for $\alpha \in (1, \infty)$, since it does not satisfy the symmetry property and the triangle inequality. It does, however, satisfy the weaker triangle inequality [1]: For all probability distributions P, Q and R, and all $\alpha, k \geq 1$, we have

$$D_\alpha(P \,\|\, R) \leq \frac{k\alpha}{k\alpha - 1} D_{\frac{k\alpha - 1}{k - 1}}(P \,\|\, Q) + D_{k\alpha}(Q \,\|\, R).$$

In the subsequent results, we denote $D_\alpha(\mathcal{M}(D) \,\|\, \mathcal{M}(D'))$ as the Rényi divergence of the distributions of $\mathcal{M}(D)$ and $\mathcal{M}(D')$. Observe that the case $\alpha = \infty$ can be used to define $d_{\mathbb{D}}$-privacy (and DP), i.e., $\mathcal{M}$ with domain $\mathbb{D}$ is $d_{\mathbb{D}}$-private if and only if for all $D, D' \in \mathbb{D}$

$$D_\infty(\mathcal{M}(D) \,\|\, \mathcal{M}(D')) \leq d(D, D').$$

We can state now the definition of zero-concentrated DP [1] directly extended for any possible granularity $\mathcal{G}$.

Definition VI.10 (Zero-concentrated DP)**.** Let $\rho \geq 0$. A randomized mechanism $\mathcal{M}$ with domain $\mathbb{D}_{\mathcal{X}}$ is $\mathcal{G}$ *ρ-zero-concentrated DP* ($\mathcal{G}$ ρ-zCDP) if, for all $\mathcal{G}$-neighboring $D, D' \in \mathbb{D}$ and all $\alpha \in (1, \infty)$:

$$D_\alpha(\mathcal{M}(D) \,\|\, \mathcal{M}(D')) \leq \rho\alpha.$$

[8] Measure μ always exists in this case and its choice does not affect the results [22].

The extension to metric zCDP is not trivial, since the bound of the Rényi divergence does not scale linearly for group privacy, but instead quadratically (i.e., $D_\alpha(\mathcal{M}(D)\,\|\,\mathcal{M}(D')) \leq (d_{\mathbb{D}}^{\mathcal{G}}(D,D'))^2\rho\alpha$). In this case, bounding the divergence by a metric would be too restrictive with regard to the original notion. In particular, known zCDP mechanisms, such as the Gaussian mechanism, would not satisfy a linear privacy degradation. Therefore, knowing that the Rényi divergence scales quadratically, we define the following notion:

Definition VI.11 ($d_{\mathbb{D}}^2$-zCprivacy)**.** Let $(\mathbb{D}, d_{\mathbb{D}})$ be a privacy space. Then, a randomized mechanism $\mathcal{M}$ with domain $\mathbb{D}$ is $d_{\mathbb{D}}^2$-*zCprivacy* if for all $D, D' \in \mathbb{D}$ and all $\alpha \in (1,\infty)$,

$$D_\alpha(\mathcal{M}(D)\,\|\,\mathcal{M}(D')) \leq d_{\mathbb{D}}^2(D,D')\alpha \qquad \text{(VI.1)}$$

where $d_{\mathbb{D}}^2(D,D') \coloneqq (d_{\mathbb{D}}(D,D'))^2$.

With this definition, we obtain once again the analogous to Theorem III.3 for zCDP:

Theorem VI.12. *Let $\mathcal{G}$ be a granularity notion over the database class $\mathbb{D}$. Then, a mechanism $\mathcal{M}$ with domain $\mathbb{D}$ is $\rho(d_{\mathbb{D}}^{\mathcal{G}})^2$-private if and only if it is $\mathcal{G}$ ρ-zCDP.*

We now present the AC theorem for $d_{\mathbb{D}}^2$-zCprivacy:

Theorem VI.13 (Zero-concentrated AC 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 $f_i\colon \mathbb{D} \to \mathbb{D}_i$ a deterministic map and $f_i^* = \mathrm{id}_{\overline{\mathcal{R}}_i} \times f_i$ (with $f_1^* = f_1$).*

For $i \in [k]$, let $\mathcal{M}_i^*\colon \overline{\mathcal{R}}_i \times \mathbb{D}_i \to \mathcal{R}_i$ be a mechanism such that $\mathcal{M}_i^*(\bar{s}_i,\cdot)\colon \mathbb{D}_i \to \mathcal{R}_i$ satisfies d_i^2-zCprivacy for any $\bar{s}_i \in \overline{\mathcal{R}}_i$.

Then mechanism $\mathcal{M} = (\mathcal{M}_1^* \circ f_1^*, \ldots, \mathcal{M}_k^* \circ f_k^*)_{\text{adapt}}$ is $d_{\mathbb{D}}^2$-zCprivate with

$$d_{\mathbb{D}}^2(D,D') \coloneqq \sum_{i=1}^k d_i^2(f_i(D), f_i(D')) \quad \text{for all } D, D' \in \mathbb{D}.$$

As in the previous cases, Theorem VI.13 can be used to formulate the corresponding corollaries.

Theorem VI.14 (Zero-concentrated AC theorem for common domain)**.** *For $i \in [k]$, let $(\mathbb{D}, d_i)$ be a privacy space, and let f_i be a deterministic map over $\mathbb{D}$. For $i \in [k]$, let $\mathcal{M}_i\colon \overline{\mathcal{R}}_i \times \mathbb{D} \to \mathcal{R}_i$ be a mechanism such that $\mathcal{M}_k(\bar{s}_i,\cdot)\colon \mathbb{D} \to \mathcal{R}_i$ satisfies d_i^2-zCprivacy and f_i-dependency for any $\bar{s}_i \in \overline{\mathcal{R}}_i$. Then mechanism $\mathcal{M} = (\mathcal{M}_1, \ldots, \mathcal{M}_k)_{\text{adapt}}$ is $d_{\mathbb{D}}^2$-zCprivate* with $d_{\mathbb{D}}^2 \coloneqq \sum_{i=1}^k (d_i^{f_i})^2$.*

Theorem VI.15 (Generalized zero-concentrated ASC)**.** *Let $\mathbb{D}$ be a database class, and, for all $i \in [k]$, let $(\mathbb{D}, d_i)$ be a privacy space. For $i \in [k]$, let $\mathcal{M}_i\colon \overline{\mathcal{R}}_i \times \mathbb{D} \to \mathcal{R}_i$ be a mechanism such that $\mathcal{M}_k(\bar{s}_i,\cdot)\colon \mathbb{D} \to \mathcal{R}_i$ is d_i^2-zCprivate for any $\bar{s}_i \in \overline{\mathcal{R}}_i$. Then $\mathcal{M} = (\mathcal{M}_1, \ldots, \mathcal{M}_k)_{\text{adapt}}$ is $(\sum_{i=1}^k d_i^2)$-zCprivate.*

When $d_i = \rho_i(d_{\mathbb{D}}^{\mathcal{U}})^2$ we recover the original composition bound $\sum_{i=1}^k \rho_i$ established for unbounded zCDP in [1], which generalizes to all granularities. However, to the best of our knowledge, no analysis on the privacy loss has previously been performed for zCDP when mechanism $\mathcal{M}_i$ input disjoint subsets. Therefore, we give the two first results about how zCDP degrades when composed, similar to parallel composition:

Theorem VI.16 (Zero-concentrated best bound for disjoint inputs)**.** *Let $\mathbb{D}$ be a database class and $\mathcal{G}$ a granularity over $\mathbb{D}$. Let p be a $d_{\mathbb{D}}^{\mathcal{G}}$-compatible k-partitioning function such that $\Delta p_i \leq 1$, and $p_i^* = \mathrm{id}_{\overline{\mathcal{R}}_i} \times p_i$ (with $p_1^* = p_1$). For $i \in [k]$, let $\mathcal{M}_i^*\colon \overline{\mathcal{R}}_i \times \mathbb{D}_i \to \mathcal{R}_i$ be a mechanism such that $\mathcal{M}_i^*(\bar{s}_i,\cdot)\colon \mathbb{D}_i \to \mathcal{R}_i$ satisfies $\rho_i(d_{\mathbb{D}_i}^{\mathcal{G}})^2$-zCprivacy for any $\bar{s}_i \in \overline{\mathcal{R}}_i$. Then mechanism $\mathcal{M} = (\mathcal{M}_1^* \circ p_1^*, \ldots, \mathcal{M}_k^* \circ p_k^*)_{\text{adapt}}$ is $\rho(d_{\mathbb{D}}^{\mathcal{G}})^2$-zCprivate with $\rho = \max_{i\in[k]} \rho_i$.*

For the common-domain setting, we find the analogous theorem (see Theorem A.15 in the long version).

C. Extending to Gaussian DP

Finally, we extend our results to Gaussian DP (GDP) [6]. GDP uses the hypothesis testing interpretation of DP to bound the privacy loss. This way, we understand that an attacker is trying to solve a hypothesis testing problem for two neighboring databases D and D' as [6]

$$\begin{cases} H_0\colon \text{The input database is } D, \\ H_1\colon \text{The input database is } D'. \end{cases}$$

Specifically, given an output s, an attacker will use a rejection rule ϕ to decide whether D or D' was the initial database. The difficulty in distinguishing between the two hypotheses is then described by the optimal trade-off between the *type I error* (i.e., rejecting H_0 when it is true) and the *type II error* (i.e., failing to reject H_0 when it is false). If P and Q are the distribution functions of $\mathcal{M}(D)$ and $\mathcal{M}(D')$ respectively, then the type I and type II errors are defined respectively as $\alpha_\phi \coloneqq \mathbb{E}_P[\phi]$ and $\beta_\phi \coloneqq 1 - \mathbb{E}_Q[\phi]$, given a rejection rule $0 \leq \phi \leq 1$. This motivates the definition of trade-off function [6].

Definition VI.17 (Trade-off function [6])**.** Let P and Q be two probability distributions on the same measurable space. A *trade-off function* is defined as $T(P,Q)\colon [0,1] \to [0,1]$ such that

$$T(P,Q)(\alpha) = \inf_\phi \{\beta_\phi \mid \alpha_\phi \leq \alpha\},$$

where the infimum is taken over all (measurable) rejection rules ϕ.

A trade-off function $T(P,Q)(\alpha)$ represents the minimum achievable type II error β for a given level of type I error α. Note that the minimum β_ϕ can be achieved by the likelihood-ratio test, since it is the test with the highest *power* (i.e., lowest type II error for a prespecified type I error α) according to the Neyman–Pearson lemma [14].

The larger the trade-off function, the harder it is to distinguish between the two hypotheses. This idea of "hard to distinguish" leads us to the definition of Gaussian DP (GDP) [6], which we directly define for any neighborhood notion:

Definition VI.18 (Gaussian DP)**.** Let $\mu \geq 0$. A mechanism $\mathcal{M}$ with domain $\mathbb{D}$ is said to be $\mathcal{G}$ *μ-GDP* if, for all $\mathcal{G}$-neighboring $D, D' \in \mathbb{D}$,

$$T(\mathcal{M}(D), \mathcal{M}(D'))(\alpha) \geq T(\mathcal{N}(0,1), \mathcal{N}(\mu,1))(\alpha)$$

for all $\alpha \in [0,1]$. We denote $G_\mu \coloneqq T(\mathcal{N}(0,1), \mathcal{N}(\mu,1))$.

First, note that $T(\mathcal{M}(D), \mathcal{M}(D'))$ is the trade-off function of the distribution of $\mathcal{M}(D)$ and $\mathcal{M}(D')$ (by abuse of notation). GDP establishes that distinguishing between $\mathcal{M}(D)$ and $\mathcal{M}(D')$ is at least as hard as distinguishing between the normal distributions $\mathcal{N}(0,1)$ and $\mathcal{N}(\mu,1)$. By the Neyman–Pearson lemma, we can explicitly express G_μ as $G_\mu(\alpha) = \Phi(\Phi^{-1}(1-\alpha) - \mu)$ for all $\alpha \in [0,1]$, where Φ is the distribution function of $\mathcal{N}(0,1)$. Note that this trade-off function decreases with respect to μ, i.e., $G_\mu \leq G_{\mu'}$ if $\mu \geq \mu'$.

GDP satisfies a group privacy property that establishes that privacy degrades linearly with respect to the number of changes between the two databases [6]. Consequently, we use this property to define the $d_{\mathbb{D}}$-privacy adaptation of GDP:

Definition VI.19 ($d_{\mathbb{D}}$-Gaussian privacy)**.** Let $d_{\mathbb{D}} \colon \mathbb{D}^2 \to [0,\infty]$ be a metric. A mechanism $\mathcal{M}$ with domain $\mathbb{D}$ is said to be *$d_{\mathbb{D}}$-Gprivate* if, for all $D, D' \in \mathbb{D}$,

$$T(M(D), M(D')) \geq G_{d_{\mathbb{D}}(D,D')},$$

where $G_\infty(\alpha) \coloneqq \lim_{\mu\to\infty} G_\mu(\alpha) = 0$.

Our definition of $d_{\mathbb{D}}$-Gprivacy generalizes the original notion of Gaussian DP:

Theorem VI.20. *Let $\mathcal{G}$ be a granularity notion over the database class $\mathbb{D}$. Then, a mechanism $\mathcal{M}$ with domain $\mathbb{D}$ is $\mu d_{\mathbb{D}}^{\mathcal{G}}$-Gprivate if and only if it is $\mathcal{G}$ μ-GDP.*

We can now present the AC theorem for $d_{\mathbb{D}}$-Gprivacy.

Theorem VI.21 (Gaussian AC 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 $f_i \colon \mathbb{D} \to \mathbb{D}_i$ a deterministic map and $f_i^* = \mathrm{id}_{\overline{\mathcal{R}}_i} \times f_i$ (with $f_1^* = f_1$).*

For $i \in [k]$, let $\mathcal{M}_i^* \colon \overline{\mathcal{R}}_i \times \mathbb{D}_i \to \mathcal{R}_i$ be a mechanism such that $\mathcal{M}_i^*(\bar{s}_i, \cdot) \colon \mathbb{D}_i \to \mathcal{R}_i$ satisfies d_i-Gprivacy for any $\bar{s}_i \in \overline{\mathcal{R}}_i$. Then mechanism $\mathcal{M} = (\mathcal{M}_1^* \circ f_1^*, \dots, \mathcal{M}_k^* \circ f_k^*)_{\mathrm{adapt}}$ is $d_{\mathbb{D}}$-Gprivate with

$$d_{\mathbb{D}}(D, D') \coloneqq \sqrt{\sum_{i=1}^{k} d_i(f_i(D), f_i(D'))^2} \quad \text{for all } D, D' \in \mathbb{D}.$$

Note that unlike the AC theorem (V.2), $d_{\mathbb{D}}$ is not the sum of the distances (i.e., the ℓ_1-norm), but actually the sum of the squares of the distances (i.e., the ℓ_2-norm). Recall that $\|(d_1, \dots, d_k)\|_2 \leq \|(d_1, \dots, d_k)\|_1$. In this case, we can notice improvements in GDP to the composition results. We also see the same improvements in the common-domain counterpart.

Theorem VI.22 (Gaussian AC theorem for common domain)**.** *For $i \in [k]$, let $(\mathbb{D}, d_i)$ be a privacy space, and let f_i be a deterministic map over $\mathbb{D}$. For $i \in [k]$, let $\mathcal{M}_i \colon \overline{\mathcal{R}}_i \times \mathbb{D} \to \mathcal{R}_i$ be a mechanism such that $\mathcal{M}_k(\bar{s}_i, \cdot) \colon \mathbb{D} \to \mathcal{R}_i$ satisfies d_i-Gprivacy and f_i-dependency for any $\bar{s}_i \in \overline{\mathcal{R}}_i$. Then mechanism $\mathcal{M} = (\mathcal{M}_1, \dots, \mathcal{M}_k)_{\mathrm{adapt}}$ is $d_{\mathbb{D}}$-Gprivate* with $d_{\mathbb{D}} \coloneqq \sqrt{\sum_{i=1}^{k} (d_i^{f_i})^2}$.*

As in the previous subsections, we recover the generalized ASC results when $f_i = \mathrm{id}$:

Theorem VI.23 (Generalized Gaussian ASC)**.** *Let $\mathbb{D}$ be a database class, and d a metric defined in $\mathbb{D}$. For $i \in [k]$, let $\mathcal{M}_i^* \colon \overline{\mathcal{R}}_i \times \mathbb{D}_i \to \mathcal{R}_i$ be a mechanism such that $\mathcal{M}_i^*(\bar{s}_i, \cdot) \colon \mathbb{D} \to \mathcal{R}_i$ satisfies d_i-DP for any $\bar{s}_i \in \overline{\mathcal{R}}_i$. Then mechanism $\mathcal{M} = (\mathcal{M}_1^*, \dots, \mathcal{M}_k^*)_{\mathrm{adapt}}$ is $d_{\mathbb{D}}$-Gprivate with with $d_{\mathbb{D}} = \sqrt{d_1^2 + \cdots + d_k^2}$.*

Choosing $d_i = \mu_i d_{\mathbb{D}}^{\mathcal{G}}$, we obtain from this theorem the already-known [6] sequential bound $\|(\mu_1, \dots, \mu_k)\|_2$.

For d-Gprivacy, as for the other notions, it is interesting to find cases where we can obtain better bounds than the sequential one using our result. We explore these cases in the following corollaries. For example, we can also obtain the best bound for when f defines a partitioning function:

Theorem VI.24. *Let $\mathbb{D}$ be a database class, and let p be k-partitioning function of $\mathbb{D}$ in $\mathbb{D}_i$ and $p_i^* = \mathrm{id}_{\overline{\mathcal{R}}_i} \times p_i$ (with $p_1^* = p_1$). Let d^* be well-defined over $\mathbb{D}$ and $\mathbb{D}_i$. For $i \in [k]$, let $\mathcal{M}_i^* \colon \overline{\mathcal{R}}_i \times \mathbb{D}_i \to \mathcal{R}_i$ be a mechanism such that $\mathcal{M}_i^*(\bar{s}_i, \cdot) \colon \mathbb{D}_i \to \mathcal{R}_i$ satisfies $\mu_i d_{\mathbb{D}_i}^*$-Gprivacy. If d^* commutes with p then mechanism $\mathcal{M} = (\mathcal{M}_1^* \circ p_1^*, \dots, \mathcal{M}_k^* \circ p_k^*)_{\mathrm{adapt}}$ is $\tilde{d}_{\mathbb{D}}$-Gprivate with*

$$\tilde{d}_{\mathbb{D}} \coloneqq \sqrt{\sum_{i=1}^{k} (\mu_i d_{\mathbb{D}_i}^*(p_i(D), p_i(D')))^2} \leq \max_{i\in[k]} \mu_i d_{\mathbb{D}}^*(D, D'). \tag{VI.2}$$

Note that the inequality is in fact an equality when $d_{\mathbb{D}_i}^*(p_i(D), p_i(D')) = 0$ for all but one $i \in [k]$. Therefore in some cases, the Gaussian AC theorem (VI.21) can give us a tighter bound than $\max_{i\in[k]} \mu_i d_{\mathbb{D}}^*$. We see this in the following example:

Example VI.25. Let $\mathbb{D} \subseteq \mathbb{D}_{\mathcal{X}}$, let $\mathbb{D}_i = \mathbb{D}_{\mathcal{X}_i}$ where $\{\mathcal{X}_i\}_{i\in[k]}$ defines a partition, and consider $d^{\triangle}$, which commutes with the previous partition (see Proposition A.4 in the long version). If $\mathcal{M}_i \colon \mathbb{D}_i \to \mathcal{R}_i$ are $d_{\mathbb{D}_i}^{\triangle}$-Gprivate, then mechanism $\mathcal{M} = (\mathcal{M}_1, \dots, \mathcal{M}_k)_{\mathrm{adapt}}$ is $\tilde{d}_{\mathbb{D}}$-Gprivate with $\tilde{d}_{\mathbb{D}} \leq d_{\mathbb{D}}^{\triangle}$. For instance, if $D = D' \backslash \{x_i, x_j\}$ with $x_i \in \mathcal{X}_i$ and $x_j \in \mathcal{X}_j$ $(i \neq j)$, we have that $d_{\mathbb{D}}^{\triangle}(D, D') = 2$, while

$$\tilde{d}_{\mathbb{D}}(D, D') = \sqrt{d_{\mathbb{D}_i}^{\triangle}(p_i(D), p_i(D'))^2 + d_{\mathbb{D}_j}^{\triangle}(p_j(D), p_j(D'))^2}$$

$$= \sqrt{|\{x_i\}|^2 + |\{x_j\}|^2} = \sqrt{1+1} = \sqrt{2} < 2.$$

The Gaussian version of Theorem V.4 also holds. However, in this case, a compatible partition implies $d^{\mathcal{G}}_{\mathbb{D}_i}(p_i(D), p_i(D')) = 0$ for all but one $i \in [k]$, so the inequality in Equation (VI.2) becomes an equality and the AC theorem does not provide a tighter bound.

Theorem VI.26 (Gaussian best bound for disjoint inputs)**.** *Let $\mathbb{D}$ be a database class and $\mathcal{G}$ a granularity over $\mathbb{D}$. Let p be a $d^{\mathcal{G}}_{\mathbb{D}}$-compatible k-partitioning function such that $\Delta p_i \leq 1$, and $p^*_i = \mathrm{id}_{\overline{\mathcal{R}}_i} \times p_i$ (with $p^*_1 = p_1$). For $i \in [k]$, let $\mathcal{M}^*_i \colon \overline{\mathcal{R}}_i \times \mathbb{D}_i \to \mathcal{R}_i$ be a mechanism such that $\mathcal{M}^*_i(\bar{s}_i, \cdot) \colon \mathbb{D}_i \to \mathcal{R}_i$ satisfies $\mu_i d^{\mathcal{G}}_{\mathbb{D}_i}$-Gprivacy for any $\bar{s}_i \in \overline{\mathcal{R}}_i$. Then mechanism $\mathcal{M} = (\mathcal{M}^*_1 \circ p^*_1, \ldots, \mathcal{M}^*_k \circ p^*_k)_{\text{adapt}}$ is $\mu d^{\mathcal{G}}_{\mathbb{D}}$-Gprivate with $\mu = \max_{i \in [k]} \mu_i$.*

The common-domain setting of this theorem for GDP is analogous (see Theorem A.17 in the long version).

VII. Post-Processing and Reciprocal Results

Finally, we study post-processing in the privacy notions we have introduced that leads to reciprocal results. All the $d_{\mathbb{D}}$-privacy adaptations of DP notions we introduced, as well as $d_{\mathbb{D}}$-privacy, are robust to post-processing:

Theorem VII.1 (Post-processing)**.** *The privacy notions of $d_{\mathbb{D}}$-privacy, $(d_{\mathbb{D}}, \delta_{\mathbb{D}})$-privacy, $d^2_{\mathbb{D}}$-zCprivacy and $d_{\mathbb{D}}$-Gprivacy are robust to post-processing.*

Moreover, we obtain reciprocal results for the composition theorems for common domain for any privacy notion $\mathfrak{P}$ that is robust to post-processing. More precisely, Theorem IV.11 has a reciprocal result.

Theorem VII.2 (Reciprocal to the IC theorem (common domain))**.** *Let $\mathfrak{P}$ be a privacy notion that is robust to post-processing. For all $i \in [k]$, let $\mathcal{M}_i \colon \mathbb{D} \to \mathcal{R}_i$ be mutually independent randomized mechanisms. Let $\mathcal{M} = (\mathcal{M}_1, \ldots, \mathcal{M}_k)_{\text{ind}}$ be a mechanism that satisfies $\mathfrak{P}$. Then $\mathcal{M}_i$ must satisfy $\mathfrak{P}$ for all $i \in [k]$.*

Even though it is not useful in constructing new mechanisms, this result makes it clear that we cannot obtain a $\mathfrak{P}$ mechanism by independently composing mechanisms that do not satisfy $\mathfrak{P}$, and can serve as a first check to ensure whether a mechanism satisfies $\mathfrak{P}$ or not. For instance, Example II.9 fails because $\mathcal{M}_i = \mathcal{M}^*_i \circ f_i$ do not satisfy $\mathfrak{P}$. Also, for the adaptive case, we have the following result:

Theorem VII.3 ("Reciprocal" to the AC theorem (common domain))**.** *Let $\mathfrak{P}$ be a privacy notion that is robust to post-processing. Let $\mathcal{M}_i \colon \overline{\mathcal{R}}_i \times \mathbb{D} \to \mathcal{R}_i$ for $i \in [k]$ be randomized mechanisms. Let $\mathcal{M} = (\mathcal{M}_1, \ldots, \mathcal{M}_k)_{\text{adapt}}$ be a mechanism satisfying $\mathfrak{P}$. Recall that by definition $\mathcal{M}(D) = (\mathcal{N}_1(D), \ldots, \mathcal{N}_k(D))$ for all $D \in \mathbb{D}$, where $\mathcal{N}_i(D)$ are defined recursively as $\mathcal{N}_i(D) = \mathcal{M}_i(\mathcal{N}_{i-1}(D), \ldots, \mathcal{N}_1(D), D)$ for $i \in [k]$. Then $\mathcal{N}_i$ must satisfy $\mathfrak{P}$ for all $i \in [k]$.*

Note that this result tells us that all $\mathcal{N}_i$ satisfy $\mathfrak{P}$, but this is not the exact reciprocal of Theorem V.5. Given the same hypotheses, it is not necessarily true that $\mathcal{M}_i(\bar{s}_i, \cdot)$ satisfy $\mathfrak{P}$ for all $\bar{s}_i \in \overline{\mathcal{R}}_i$.

Furthermore, no result for $\mathcal{M}^*_i$ can be generally stated. For example, in Remark III.7, we provide a case where $\mathcal{M}^*_i \circ f_i$ is free-lunch DP while $\mathcal{M}^*_i$ is not.

VIII. Conclusions

In this paper, we study the composability properties of DP in the new settings of the literature, including new granularities and data domains. We show that composability can be defined independently of the neighborhood definition. Our results can be used to directly obtain specific composition rules when new granularity notions (or metrics) are proposed, without having to prove these same rules for each case.

Moreover, our IC and AC theorems (IV.1 and V.2) are defined for $d_{\mathbb{D}}$-privacy. The notion of $d_{\mathbb{D}}$-privacy not only generalizes the original DP setting, but also provides more precise information about the protection given. Therefore, we facilitate the computation of the final privacy guarantee of any composed mechanism over any desired data domain and even under mixed privacy requirements, which was not previously defined. In particular, we prove the existence of a significantly better bound to the privacy loss for bounded DP when the composed mechanisms are applied to disjoint databases (Corollary IV.13).

Besides, we study particularly interesting composition settings in the literature such as the case in which each composed mechanism inputs the whole database or just disjoint subsets, and we compare them with the original sequential and parallel composition results. Since the original parallel composition theorem [17] does not generalize to all metrics, we also investigate the additional hypotheses necessary to obtain the best possible privacy loss when we work over a partitioned database. We provide the hypotheses under which we obtain the best bound and conclude that these conditions are easily satisfied for some metrics, such as $d^{\triangle}$; while others metrics only work for specific partitions, such as the bounded metric.

Furthermore, we extend our results to other DP-based privacy notions: namely, approximate DP, zero-concentrated DP, and Gaussian DP. To this end, we present $d_{\mathbb{D}}$-privacy variants that simultaneously include both the original definition and their group privacy property. Also, we provide the corresponding composition theorems for each of these notions.

Finally, we discuss reciprocal versions of the composition, which can be used to check when a mechanism fails to guarantee DP.

Future work: In this paper, we limit ourselves to some DP-based notions that can be directly expressed with a metric. Extending our composition theorems to other DP-based semantic privacy notions, such as Rényi DP [19] or f-DP [6], could be interesting future work. Moreover, it will be interesting to explore the advanced composition versions

of the presented theorems for such semantic notions that allow advanced composition.

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 also supported by the COMPROMISE (PID2020-113795RB-C31) and MOBILYTICS (TED2021-129782B-I00) projects, funded by the same two institutions above. The authors at KIT are supported by the KASTEL Security Research Labs (Topic 46.23 of Helmholtz Association) and EXC 2050/1 'CeTI' (ID 390696704), as well as the BMBF project "PROPOLIS" (16KIS1393K).

The authors also thank the inhouse textician at KASTEL Security Research Labs.

References

[1] M. Bun and T. Steinke. "Concentrated Differential Privacy: Simplifications, Extensions, and Lower Bounds". In: *Proc. Theory Cryptogr. Conf. (TCC)*. 2016, pp. 635–658. ISBN: 978-3-662-53641-4. DOI: 10.1007/978-3-662-53641-4_24.

[2] C. L. Canonne, G. Kamath, and T. Steinke. "The Discrete Gaussian for Differential Privacy". In: *Proc. Int. Conf. Neural Inform. Process. Syst. (NeurIPS)*. Vol. 33. 2020, pp. 15676–15688. URL: https://doi.org/10.29012/jpc.784.

[3] 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, pp. 82–102. ISBN: 978-3-642-39077-7. DOI: 10.1007/978-3-642-39077-7_5.

[4] D. Desfontaines. *The Privacy Loss Random Variable*. Ted is writing things (personal blog). 2020. URL: https://desfontain.es/privacy/privacy-loss-random-variable.html (visited on 22/8/2023).

[5] D. Desfontaines and B. Pejó. "SoK: Differential Privacies". In: *Proc. Priv. Enhanc. Technol.* 2020.2 (2020), pp. 288–313. ISSN: 2299-0984. DOI: 10.2478/popets-2020-0028.

[6] J. Dong, A. Roth, and W. J. Su. "Gaussian Differential Privacy". In: *J. Royal Stat. Soc. Ser. B* 84.1 (Feb. 2022), pp. 3–37. ISSN: 1369-7412. DOI: 10.1111/rssb.12454.

[7] C. Dwork. "Differential Privacy". In: *Proc. Int. Colloq. Automata, Lang., Program. (ICALP)*. 2006, pp. 1–12. ISBN: 978-3-540-35908-1. DOI: 10.1007/11787006_1.

[8] C. Dwork, K. Kenthapadi, F. McSherry, I. Mironov, and M. Naor. "Our Data, Ourselves: Privacy via Distributed Noise Generation". In: *Proc. Adv. Cryptology – Annual Int. Conf. Theory Appl. Cryptogr. Techniques (EUROCRYPT)*. 2006, pp. 486–503. ISBN: 978-3-540-34547-3. DOI: 10.1007/11761679_29.

[9] C. Dwork and A. Roth. "The Algorithmic Foundations of Differential Privacy". In: *Found. Trends Theor. Comput. Sci.* 9.3–4 (Aug. 2014), pp. 211–407. ISSN: 1551-305X. DOI: 10.1561/0400000042.

[10] 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)*. Apr. 2023, pp. 252–263. ISBN: 978-989-758-624-8. DOI: 10.5220/0011885000003405.

[11] M. Hay, C. Li, G. Miklau, and D. Jensen. "Accurate Estimation of the Degree Distribution of Private Networks". In: *Proc. IEEE Int. Conf. Data Min. (ICDM)*. IEEE, Dec. 2009, pp. 169–178. DOI: 10.1109/ICDM.2009.11.

[12] D. Kifer and A. Machanavajjhala. "No Free Lunch in Data Privacy". In: *Proc. ACM SIGMOD Int. Conf. Manage. Data (MOD)*. SIGMOD '11. June 2011, pp. 193–204. ISBN: 978-1-4503-0661-4. DOI: 10.1145/1989323.1989345.

[13] D. Kifer, S. Messing, A. Roth, A. Thakurta, and D. Zhang. *Guidelines for Implementing and Auditing Differentially Private Systems*. May 2020. DOI: 10.48550/arXiv.2002.04049.

[14] E. L. Lehmann and J. P. Romano. *Testing Statistical Hypotheses*. 3rd ed. Springer Texts in Statistics. New York: Springer, 2005. ISBN: 978-0-387-98864-1.

[15] A. Lévy. *Basic Set Theory*. Dover Publications, 2002. ISBN: 978-0-486-42079-0.

[16] N. Li, M. Lyu, D. Su, and W. Yang. *Differential Privacy: From Theory to Practice*. Vol. 8. Synthesis Lectures on Information Security, Privacy, and Trust. San Rafael, California: Morgan & Claypool, Oct. 2016. ISBN: 1-62705-297-6.

[17] F. McSherry. "Privacy Integrated Queries". In: *Proc. ACM SIGMOD Int. Conf. Manage. Data (MOD)*. June 2009. DOI: 10.1145/1559845.1559850.

[18] S. Meiser. *Approximate and Probabilistic Differential Privacy Definitions*. 2018. eprint: 2018/277. URL: https://eprint.iacr.org/2018/277.

[19] I. Mironov. "Rényi Differential Privacy". In: *Proc. IEEE Comput. Security Found. Symp. (CSF)*. Aug. 2017, pp. 263–275. DOI: 10.1109/CSF.2017.11.

[20] J. Soria-Comas and J. Domingo-Ferrer. "Big Data Privacy: Challenges to Privacy Principles and Models". In: *Data Sci. Eng.* 1.1 (Mar. 2016), pp. 21–28. ISSN: 2364-1541. DOI: 10.1007/s41019-015-0001-x.

[21] A. Syropoulos. "Mathematics of Multisets". In: *Proc. Multiset Process.* 2001, pp. 347–358. ISBN: 978-3-540-45523-3. DOI: 10.1007/3-540-45523-X_17.

[22] T. van Erven and P. Harremoës. "Rényi Divergence and Kullback-Leibler Divergence". In: *IEEE Trans. Inform. Theory* 60.7 (July 2014), pp. 3797–3820. ISSN: 1557-9654. DOI: 10.1109/TIT.2014.2320500.

See the proofs and further remarks in the appendix of the long version of this paper (arXiv:2308.14649).