Information-theoretic Bounds for Differentially Private Mechanisms

Gilles Barthe
IMDEA Software
gilles.barthe@imdea.org

Boris Köpf
IMDEA Software
boris.koepf@imdea.org

Abstract—**There are two active and independent lines of research that aim at quantifying the amount of information that is disclosed by computing on confidential data. Each line of research has developed its own notion of confidentiality: on the one hand, differential privacy is the emerging consensus guarantee used for privacy-preserving data analysis. On the other hand, information-theoretic notions of leakage are used for characterizing the confidentiality properties of programs in language-based settings.**

The purpose of this article is to establish formal connections between both notions of confidentiality, and to compare them in terms of the security guarantees they deliver. We obtain the following results. First, we establish upper bounds for the leakage of every ϵ-differentially private mechanism in terms of ϵ and the size of the mechanism's input domain. We achieve this by identifying and leveraging a connection to coding theory.

Second, we construct a class of ϵ-differentially private channels whose leakage grows with the size of their input domains. Using these channels, we show that there cannot be domain-size-independent bounds for the leakage of all ϵ-differentially private mechanisms. Moreover, we perform an empirical evaluation that shows that the leakage of these channels almost matches our theoretical upper bounds, demonstrating the accuracy of these bounds.

Finally, we show that the question of providing optimal upper bounds for the leakage of ϵ-differentially private mechanisms in terms of rational functions of ϵ is in fact decidable. Our work provides the first analysis of the leakage of differentially private channels as end-to-end mechanisms.

I. Introduction

Confidentiality is a property that captures that no secret information is exposed to unauthorized parties; it is one of the most fundamental security properties and an essential requirement for most security-critical applications.

Unfortunately, perfect confidentiality is often difficult or even impossible to achieve in practice. In some cases, perfect confidentiality is in conflict with the functional requirements of a system. For example, the result of a statistical query on a medical database necessarily reveals some information about the individual entries in the database. In other cases, perfect confidentiality is in conflict with non-functional requirements such as bounds on the resource-usage. For example, variations in the execution time of a program may reveal partial information about the program's input; however a (perfectly secure) implementation with constant execution time may have inacceptable performance.

Because such conflicting requirements are ubiquitous, there is a need for tools that enable formal reasoning about imperfect confidentiality. Quantitative approaches to confidentiality can provide such tools: First, quantitative notions of confidentiality can express a continuum of degrees of security, making them an ideal basis for reasoning about the trade-off between security and conflicting requirements such as utility [10] or performance [19]. Second, despite their flexibility, a number of quantitative notions of confidentiality are backed up by rigorous operational security guarantees such as lower bounds on the effort required for brute-forcing a secret.

While convergence has been achieved for definitions of perfect confidentiality (they are subsumed under the cover term *noninterference* and differ mainly in the underlying system and adversary models) this is not the case for their quantitative counterparts: There is large a number of proposals for quantitative confidentiality properties, and their relationships (e.g. in terms of the assumptions made and the guarantees provided) are often not well-understood.

In particular, there are two active and independent lines of research dealing with quantitative notions of confidentiality. The first line is motivated by the privacy-preserving publishing of data, with *differential privacy* [10] as the emerging consensus definition. The second line is motivated by tracking the information-flow in arbitrary programs, where most approaches quantify *leakage* as reduction in entropy about the program's input, see e.g. [32].

There have been efforts to understand the connections between the different notions of confidentiality proposed within each line of research (see [12] and [15], [32], respectively). The first studies of the relationship between differential privacy and quantitative notions of information-flow are emerging [1], [7], however, they do not directly compare leakage and differential privacy in terms of the security guarantees they deliver (see also the section on related work). Such a comparison could be highly useful, as it could enable one to transfer existing analysis techniques and enforcement mechanisms from one line of research to the other.

It is not difficult to see that there can be no upper bound for differential privacy in terms of leakage.[1] However, it has been an open question whether it is possible to give upper bounds for the leakage in terms of differential privacy.

In this paper, we will address this open question. To begin with, we identify information-theoretic channels as a common model for casting differential privacy and leakage, where assume (w.l.o.g, as we will argue) that the input domain is fixed to $\{0, 1\}^n$. Based on this model, we formally contrast the compositionality properties of differential privacy and leakage under sequential and parallel composition.

We observe a difference in the behavior of leakage and differential privacy under parallel composition, and we exploit this difference to construct, for every n, a channel that is ϵ-differentially private and that leaks an amount of information that grows linearly with n. This result implies there can be no general (i.e. independent of the domain size) upper bound for the leakage of all ϵ-differentially private channels.

The situation changes, however, if we consider channels on input domains of bounded size. For such channels, we exhibit the following connections between leakage and differential privacy.

For the case $n = 1$, we give a complete characterization of leakage in terms of differential privacy. More precisely, we prove an upper bound for the leakage of every ϵ-differentially private channel. Moreover, we show that this bound is tight in the sense that, for every ϵ, there is an ϵ-differentially channel whose leakage matches the bound.

For the case $n > 1$, we prove upper bounds for the leakage of every ϵ-differentially private channel in terms of n and ϵ. Technically, we achieve this by covering the channel's input domain of by spheres of a fixed radius (with respect to the Hamming metric). The definition of ϵ-differential privacy ensures that the elements within each sphere produce similar output, where similarity is quantified in terms of ϵ and the sphere radius. Based on this similarity and a recent characterization of the maximal leakage of channels [3], [21], we establish upper bounds for the information leaked about the elements of each sphere. By summing over all spheres, we obtain bounds for the information leaked about the entire input domain.

Our bounds are parametric in the number and the radius of the spheres used for covering the domain. We show how coding theory can be used for obtaining good instantiations of these parameters. In particular, we give examples where we derive bounds based on different classes of covering codes. We also exhibit limits for the bounds that can be obtained using our proof technique in terms of the sphere-packing bound. We perform an empirical evaluation that shows that the bounds we derived are close to this theoretical limit; moreover, we give an example channel whose leakage is only slightly below this limit, demonstrating the accuracy of our analysis.

Finally, although an explicit formula that precisely characterizes leakage in terms of privacy for finite input domains is still elusive, we show that such a characterization is in fact decidable. More precisely, we show that, for all n and all rational functions r (i.e. all quotients of polynomials with integer coefficients), one can decide whether the leakage of all ϵ-differentially channels is upper-bounded by $\log_2 r(\epsilon)$. This result gives a perspective for future work on the automatic derivation of such a characterization.

In summary, our main contribution is to prove formal connections between leakage and differential privacy, the two most influential quantitative notions of confidentiality to date. In particular, (i) we prove upper bounds for the leakage in terms of differential privacy for channels with bounded input domain, and (ii) we show that there can be no such bounds that hold for unbounded input domains. Finally, (iii) we show that the question of a precise characterization of leakage in terms of differential privacy is in fact decidable.

The remainder of this paper is structured as follows. In Section II we introduce differential privacy and leakage. We cast both properties in a common model and compare their compositionality in Section III. We prove bounds for the leakage in terms of differential privacy in Sections IV and V, and we show their decidability in Section VI. We present related work in Section VII before we conclude in Section VIII.

II. Preliminaries

In this section we review the definitions of min-entropy leakage and differential privacy, the two confidentiality properties of interest for this paper. For completeness, we also briefly review the most important analysis tools and enforcement mechanisms for each property.

A. Min-entropy Leakage

In quantitative information-flow, one characterizes the security of a program in terms of the difficulty of guessing the input to the program when only the output is known.[2] The difficulty of guessing can be captured in terms of information-theoretic entropy, where different notions of entropy correspond to different kinds of guessing [4]. In this paper, we focus on min-entropy as a measure, because it is associated with strong security guarantees, see [32]. However, instead of characterizing the security of a program in terms of the remaining entropy, we characterize the amount of leaked information in terms of the reduction in entropy about the program's input when the output is observed. Both viewpoints are informally related by the equation

[1] Intuitively, the leakage of a single sensitive bit (e.g. from a medical record) can entirely violate an individual's privacy; on a technical level, no deterministic program satisfies differential privacy even if it leaks only a small amount of information, because differentially private programs are necessarily probabilistic.

[2] See Section VII for alternative characterizations.

initial entropy = leaked information + remaining entropy.

Formally, we model the input to a probabilistic program as a random variable X and the output as a random variable Y. The dependency between X and Y is formalized as a conditional probability distribution $P_{Y|X}$ and is determined by the program's semantics. Such a conditional probability distribution $P_{Y|X}$ forms an information-theoretic *channel* from X to Y. As is standard in the literature on quantitative information-flow, we will use the notion of a channel as the basis for our analysis.

We consider an adversary that receives the outcomes Y of a channel $P_{Y|X}$ and wants to determine the corresponding value of X, where we assume that X is distributed according to P_X. The initial uncertainty about the chosen element of X is given by the *min-entropy* [29]

$$H_\infty(X) = -\log_2 \max_x P_X(x)$$

of X, which captures the probability of correctly guessing the outcome of X in one shot.

The *conditional min-entropy* $H_\infty(X|Y)$ of X given Y is defined by

$$H_\infty(X|Y) = -\log_2 \sum_y P_Y(y) \max_x P_{X|Y}(x, y)$$

and captures the probability of guessing the value of X in one shot when the outcome of Y is known.

The *(min-entropy) leakage* L of a channel $P_{Y|X}$ with respect to the input distribution P_X characterizes the reduction in uncertainty about X when Y is observed,

$$L = H_\infty(X) - H_\infty(X|Y) \ ,$$

and is the logarithm of the factor by which the probability of guessing the value of X is reduced by observing Y. Note that L is not a property of the channel $P_{Y|X}$ alone as it also depends on P_X. We eliminate this dependency by considering the maximal leakage over all input distributions.

Definition 1 (Maximal Leakage). *The* maximal leakage ML *of a channel $P_{Y|X}$ is the maximal reduction in uncertainty about X when Y is observed*

$$ML(P_{Y|X}) = \max_{P_X}(H_\infty(X) - H_\infty(X|Y)) \ , \qquad (1)$$

where the maximum is taken over all possible input distributions.

The following result appears in [3], [21] and shows how the maximal leakage can be computed from the channel $P_{Y|X}$. For completeness, we include its proof in the appendix.

Theorem 1. *The maximal leakage of a channel $P_{Y|X}$ can be computed by*

$$ML(P_{Y|X}) = \log_2 \sum_x \max_y P_{Y|X}(y, x) \ ,$$

where the maximum is assumed (e.g.) for uniformly distributed input.

If the channel is deterministic, i.e. if for every x there is a y such that $P_{Y|X}(y, x) = 1$, we obtain $ML(P_{Y|X}) = \log_2 |Range(Y)|$. A direct consequence of this observation is that, for deterministic programs, any algorithm for computing the size of $Range(Y)$ (which corresponds to the size of the set of reachable final states of the program) can be used for computing ML, see also [20].

B. Differential Privacy

In research privacy-preserving data publishing, differential privacy by Dwork et al. [10] is the most popular definition of privacy to date; it quantifies the influence of individual records of an input dataset on the output of a publishing algorithm.

Informally, a probabilistic algorithm satisfies ϵ-differential privacy if its output is robust (i.e. bounded in terms of ϵ) with respect small changes in the input. This robustness ensures privacy because if two datasets differ only in one individual record, the algorithm's output on both datasets will be almost indistinguishable. An adversary will not be able to deduce from the output whether an individual record is present in the dataset or not.

Formally, one consider algorithms M that take as input subsets of a set D of elements without further substructure. The distance between two input sets $D_1, D_2 \subseteq D$ is the size of their symmetric difference $D_1 \oplus D_2$ defined as

$$D_1 \oplus D_2 = (D_1 \setminus D_2) \cup (D_2 \setminus D_2) \ .$$

The following definition is due to Dwork et al. [10].

Definition 2 (Differential Privacy). *A randomized algorithm M satisfies* ϵ-differential privacy *if, for all input sets $D_1, D_2 \subseteq D$ and for all $S \subseteq Range(A)$*

$$P[M(D_1) \in S] \le e^{\epsilon|D_1 \oplus D_2|} P[M(D_2) \in S] \ . \qquad (2)$$

It follows from the triangle inequality that, for determining ϵ, it is sufficient to consider only input sets D_1, D_2 with $|D_1 \oplus D_2| = 1$.

Note that Definition 2 requires that the secret data is released by a randomized algorithm: a deterministic algorithm with non-constant output will not satisfy ϵ-differential privacy for any finite ϵ. Dwork et al. [11] have proposed a method for making deterministic programs differentially private by adding a certain amount of noise to the algorithm's output. The amount of noise that is required for achieving ϵ-differential privacy depends on ϵ and the degree of sensitivity of the program, which corresponds to a form of Lipschitz-continuity.

Definition 3 (Sensitivity). *For a function F that maps subsets of D to real numbers, one defines the* sensitivity

$S(F)$ of F by

$$\max_{D_1,D_2\subseteq D} \frac{|f(D_1)-f(D_2)|}{|D_1 \oplus D_2|} \ .$$

The so-called *Laplacian mechanism* turns a deterministic algorithms F into an ϵ-differentially private probabilistic algorithm. This is achieved by adding symmetric, exponentially distributed noise to the output of of F. The center of the noise is the output of F and the standard deviation of the noise is calibrated to $S(F)$ and ϵ. One obtains the following theorem, which is fundamental for most practical applications of differential privacy.

Theorem 2 (Laplacian Mechanism [11])**.** *For a function F mapping subsets of D to real numbers and a random variable N distributed according to $N \sim Lap(S(f)/\epsilon)$, the probabilistic algorithm*

$$M = F + N$$

defined by $M(D) = F(D) + N$ satisfies ϵ-differential privacy.

While the Laplacian mechanism is the most influential enforcement mechanism to date, there are several variants and alternative mechanisms emerging, see e.g. [22], [28], [30], [34]. In this paper, we analyze the leakage of differentially private algorithms without making any assumptions about the mechanism used to achieve differential privacy. In particular, this implies that our results apply to all existing and future differentially private mechanisms.

III. Contrasting Leakage and Differential Privacy

In this section, we cast leakage and differential privacy in a common model. Based on this model, we will formally compare the basic assumptions made, and guarantees delivered by, both notions. Moreover, the model enables us to compare both notions in terms of their the behavior under (sequential and parallel) composition. This comparison serves two purposes. First, it systematizes and completes existing knowledge. Second, the exposed differences form the basis for our results in Section V.

A. A Common Model for Leakage and Differential Privacy

Differential privacy and leakage are properties of objects of different types. Differential privacy (Definition 2) is a property of algorithms that take as input sets of (a possibly infinite number of) basis elements, whereas leakage (Definition 1) is a property of programs that take as input the elements of a fixed finite set.

We bridge this gap by assuming that the base set D of the data-publishing algorithm has finite size n, i.e. $D = \{d_1, \ldots, d_n\}$. This assumption seems reasonable, because it corresponds to an upper bound on the number of participants in a dataset. With this assumption, we can describe every subset $D' \subseteq D$ by a vector $x' \in \{0,1\}^n$, where the kth component $\pi_k(x)$ is set to 1 if and only if $d_k \in D'$. Notice that for two sets $D', D'' \subseteq D$ with vector representations $x', x'' \in \{0,1\}^n$, the set distance $|D' \oplus D''|$ corresponds to the number of positions in which x' and x'' differ, i.e. to their Hamming distance.

Furthermore, we will assume that the range of the algorithm is also finite. Although in theory, differentially private algorithms typically map to a continuum of real numbers, the active range of most practical implementations will be discrete, bounded, and finite.

For a given algorithm M according to definition 2, we hence set $Range(X) = \{0,1\}^n$ and define an channel $P_{Y|X}$ by $P_{Y|X}(y,x) = P[M(D') = y]$, where x is the characteristic vector of $D' \subseteq D$. In the remainder of this paper we will hence always assume that $Range(X) = \{0,1\}^n$, that $Range(Y)$ is finite, and that the program is given in terms of a channel $P_{Y|X}$.

B. Security Guarantees

The structure of Definitions 1 and 2 does not bear much resemblance at first sight. We will recast both definitions in a form that better exhibits their differences and similarities.

We begin by introducing some additional notation. For $x \in \{0,1\}^n$, $i \in \{1,\ldots,n\}$, and $\dot{x} \in \{0,1\}$, let $x[i/\dot{x}]$ denote the bit vector that is equal to x except that the ith component $\pi_i(x)$ is replaced by $\dot{x}$, i.e. $\pi_i(x[i/\dot{x}]) = \dot{x}$ and $\pi_j(x[i/\dot{x}]) = \pi_j(x)$ for $i \neq j$. Let $\dot{X}$ denote a random variable that models the choice of one bit, i.e. $Range(\dot{X}) = \{0,1\}$. Then $x[i/\dot{X}]$ denotes the random variable with range $\{x[i/0], x[i/1]\}$ that is distributed according to $\dot{X}$.

The following lemma states that a differentially private mechanism protects every individual bit of every possible input vector, in the sense that even if the input is completely known except for a single bit, the result of the mechanism does not significantly influence the probability of that bit being set to 1. The result and the proof are inspired by the one sketched in the appendix of [11].

Lemma 1. *A channel $P_{Y|X}$ guarantees ϵ-differential privacy if and only if, for all $x \in \{0,1\}^n$, all $i \in \{1,\ldots,n\}$, all distributions $P_{\dot{X}}$ of $\dot{X}$, and all $y \in Range(Y)$,*

$$P[\dot{X} = 1] \leq e^{\epsilon}\, P[\dot{X} = 1 \mid Y = y \wedge X = x[i/\dot{X}]] \ . \qquad (3)$$

A proof of Lemma 1 can be found in the appendix.

The following lemma is obtained by a direct reformulation of Definition 1; it exhibits the structural similarity behind Definition 1 and the statement in Lemma 1.

Lemma 2. *The leakage of channel $P_{Y|X}$ is upper-bounded by $ML(P_{Y|X}) \leq l$ if and only if, for all distributions P_X we have that*

$$\max_x P[X = x] \leq 2^l\, E_y[\max_x P[X = x|Y = y]] \ , \qquad (4)$$

where E_y denotes the expected value over y.

The comparison of the statements in Lemmas 1 and 2 exhibits the differences in the guarantees provided by leakage and differential privacy.

Security guarantees based on leakage (Lemma 2) protect entire bit-vectors from being guessed, under the assumption that they are hard to guess a priori, i.e. before observing the system's output. Leakage-based guarantees are hence the adequate tool for protecting high-entropy secrets, such as cryptographic keys, passwords, or biometric data. However, leakage-based guarantees do not make any assertions about the difficulty of guessing individual bits, which is fundamental for the protection of privacy.

In contrast, differential privacy protects each individual bit in a bit-vector, even if all other bits are known (Lemma 1). This guarantee is adequate for protecting secrets in contexts where one cannot make any reasonable assumptions about an adversary's background knowledge, such as the value of other bits, or the distribution from which the secret bit is drawn.

C. Compositionality

Well-behavedness under compositionality is an essential prerequisite for any meaningful notion of security. Without good composition properties, two secure systems may become insecure when combined.[3] We next compare the compositionality properties of leakage and differential privacy.

1) Sequential Composition: We define sequential composition as the subsequent application of two queries to the same dataset, where the second query may also take the output of the first query into account.[4] Formally, we model this by requiring that the domain of the second channel is the cartesian product of the range of the first channel and the dataset. It was already known that the differential privacy of two sequentially composed channels is upper-bounded by the sum of the differential privacy of the individual channels [25]. We next show that the leakage shows the same additive behavior under this notion of sequential composition.

We begin by defining the sequential composition of channels.

Definition 4. *The* sequential composition C_1+C_2 *of channels* $C_1 = P_{Y_1|X}$ *and* $C_2 = P_{Y_2|Y_1\times X}$ *is defined as* $C_1+C_2 = P_{Y_1\times Y_2|X}$*, where*

$$P_{Y_1\times Y_2|X}((y_1,y_2),x) = P_{Y_1|X}(y_1,x)P_{Y_2|Y_1\times X}(y_2,(y_1,x)) \ .$$

We obtain the following properties of leakage and privacy of the sequential composition of two channels.

Lemma 3. *Let* C_i *be channels that are* ϵ_i*-differentially private* ($i = 1,2$)*. Then*

1) $C_1 + C_2$ *is* $\epsilon_1 + \epsilon_2$*-differentially private, and*
2) $ML(C_1 + C_2) \leq ML(C_1) + ML(C_2)$.

Proof: For the original proof of Assertion 1), refer to [25]. For a proof on basis of our model, see the appendix. For Assertion 2), consider

$$\begin{aligned}
&ML(C_1 + C_2)\\
&\overset{(*)}{=} \log \sum_{(y_1,y_2)} \max_x P_{Y_1\times Y_2|X}((y_1,y_2),x)\\
&\overset{(**)}{=} \log \sum_{(y_1,y_2)} \max_x P_{Y_1|X}(y_1,x)P_{Y_2|Y_1\times X}(y_2,(y_1,x))\\
&\leq \log \sum_{y_1}\sum_{y_2} \max_x P_{Y_1|X}(y_1,x) \max_{x,y_1} P_{Y_2|Y_1\times X}(y_2,(y_1,x))\\
&= \log \left(\sum_{y_1} \max_x P_{Y_1|X}(y_1,x)\right)\left(\sum_{y_2} \max_{y_1,x} P_{Y_2|Y_1\times X}(y_2,(y_1,x))\right)\\
&=ML(C_1) + ML(C_2),
\end{aligned}$$

where x_1, x_2, y_1, and y_2 range over X_1, X_2, Y_1, and Y_2 respectively. Note that (*) follows from Theorem 1 and (**) follows from Definition 4. ■

2) Parallel Composition: We define the parallel composition of channels as their application to disjoint subsets of the same dataset. It was already known that the maximum of the differential privacy bounds of the individual channels is also a bound for the differential privacy of their parallel composition [25]. As we will show next, the situation is different for leakage, which adds up under parallel composition.

Definition 5. *The* parallel composition $C_1 \times C_2$ *of channels* $C_1 = P_{Y_1|X_1}$ *and* $C_2 = P_{Y_2|X_2}$ *is defined as* $C_1 \times C_2 = P_{Y_1\times Y_2|X_1\times X_2}$*, where*

$$P_{Y_1\times Y_2|X_1\times X_2}((y_1,y_2),(x_1,x_2)) = P_{Y_1|X_1}(y_1,x_1)P_{Y_2|X_2}(y_2,x_2) \ .$$

For casting differentially private mechanisms as information-theoretic channels, we represent data sets as bit-vectors, as discussed in Section III-A. Notice that the cartesian product of such bit-vector-representations corresponds to the disjoint union of the corresponding sets. That is, Definition 5 indeed captures the application of channels to disjoint subsets of a dataset.

We can prove the following properties about leakage and privacy of the parallel composition of two channels.

Lemma 4. *Let* C_i *be channels that are* ϵ_i*-differentially private* ($i = 1,2$)*. Then*

1) $C_1 \times C_2$ *is* $\max\{\epsilon_1,\epsilon_2\}$*-differentially private, and*
2) $ML(C_1 \times C_2) = ML(C_1) + ML(C_2)$.

Proof: For the original proof of Assertion 1), see [25]. For a proof in our channel-based model, see the appendix.

[3] In fact, well-behavedness under composition is one of the key advantages of differential privacy over alternative notions of privacy, see e.g. [13].

[4] For an alternative notion of sequential composition, see e.g. [21].

For Assertion 2) consider

$$
\begin{aligned}
ML(C_1 \times C_2) &\overset{(*)}{=} \log \sum_{(y_1,y_2)} \max_{(x_1,x_2)} P_{Y_1\times Y_2|X_1\times X_2}((y_1,y_2),(x_1,x_2)) \\
&= \log \sum_{y_1}\sum_{y_2} \max_{x_1}\max_{x_2} P_{Y_1|X_1}(y_1,x_1) P_{Y_2|X_2}(y_2,x_2) \\
&= \log \sum_{y_1} \max_{x_1} P_{Y_1|X_1}(y_1,x_1) \sum_{y_2} \max_{x_2} P_{Y_2|X_2}(y_2,x_2) \\
&= ML(C_1) + ML(C_2) \ ,
\end{aligned}
$$

where x_1, x_2, y_1 and y_2 range over X_1, X_2, Y_1, and Y_2 respectively. Note that (*) follows from Theorem 1. ■

Lemma 4 exhibits a difference in the behavior under parallel composition of differential privacy and leakage. In Section V, we will exploit this difference for proving the impossibility of general bounds for the leakage in terms of differential privacy.

IV. Characterizing the Leakage of Differentially Private Mechanisms – The Case $n = 1$

In this section, we provide a characterization of the leakage of differentially private channels that take only a single bit of input. In particular, we prove upper bounds in terms of ϵ for the leakage of any ϵ-differentially channel. Moreover, we show that there is a channel chose leakage matches this bound. In Section V we consider the general case.

A. Channels with 1-bit range

We first consider the case of channels whose range is also a single bit, i.e. $Range(X) = Range(Y) = \{0,1\}$. For simplicity of presentation, we will assume that all channels $P_{Y|X}$ are given in one of the two canonical forms $C_1(\epsilon,p)$ and $C_2(\epsilon,p)$ defined by the matrix in Figure 1. As the following Lemma shows, this assumption is not a restriction.

Lemma 5. *Let $P_{Y|X}$ be a channel. Then there are $p \in [0,1]$, $\epsilon > 0$, and $i \in \{1,2\}$ such that*

$$P_{Y|X} = C_i(\epsilon,p) \ .$$

Proof: Choose $i = 1,2$ depending on which of the probabilities $P_{Y|X}(0,0)$ and $P_{Y|X}(0,1)$ is larger. There is always an $\epsilon > 0$ such that the ratio between the entries in the first column is given by e^ϵ. The statement follows from the observation that the entries of each row must sum to 1. ■

For what follows it is irrelevant whether we are dealing with the canonical representation C_1 or C_2 of a channel. We will hence drop the index and refer to the canonical representation as $C(\epsilon,p)$.

We obtain the following privacy guarantees for the canonical representation $C(\epsilon,p)$.

$C_1(\epsilon,p) = P_{Y\|X}$	$Y=0$	$Y=1$
$X=0$	p	$1-p$
$X=1$	pe^ϵ	$1-pe^\epsilon$

$C_2(\epsilon,p) = P_{Y\|X}$	$Y=0$	$Y=1$
$X=0$	pe^ϵ	$1-pe^\epsilon$
$X=1$	p	$1-p$

where $0 \le p \le 1$ and $pe^\epsilon \le 1$.

Figure 1. The canonical representations $C_1(\epsilon,p)$ and $C_2(\epsilon,p)$ of channels $P_{Y|X}$.

Lemma 6. *The channel $C(\epsilon,p)$ is ϵ-differentially private if and only if*

$$p \le \frac{1}{e^\epsilon + 1} \ .$$

Proof: It suffices to show that $P_{Y|X}(y,y) \le e^\epsilon P_{Y|X}(y,x')$ for all $x,x',y \in \{0,1\}$. A direct calculation shows that this condition is satisfied if and only if $p \le \frac{e^\epsilon-1}{e^{2\epsilon}-1} = \frac{1}{e^\epsilon+1}$. ■

The following theorem characterizes the leakage of ϵ-differentially private channels with 1-bit domain and range: It gives an upper bound for the leakage of every channel, and it shows that this bound can be matched.

Theorem 3. *Let $Range(X) = Range(Y) = \{0,1\}$.*

1) *If a channel $P_{Y|X}$ is ϵ-differentially private, then*

$$ML(P_{Y|X}) \le \log_2 \frac{2e^\epsilon}{e^\epsilon + 1} \ .$$

2) *The channel $P_{Y|X} = C(\epsilon, \frac{1}{e^\epsilon+1})$ is ϵ-differentially private and*

$$ML(P_{Y|X}) = \log_2 \frac{2e^\epsilon}{e^\epsilon + 1} \ .$$

Proof: For the proof of Assertion 1), consider an ϵ-differentially private channel $P_{Y|X}$. From Lemma 6 it follows that there is a $p \le \frac{1}{e^\epsilon+1}$ such that $P_{Y|X} = C(\epsilon,p)$. We apply Theorem 1 to compute the maximal leakage of $C(\epsilon,p)$ as the sum of the column maximums of the matrix in Figure 1. We obtain

$$ML(C(\epsilon,p)) = \log_2(1 + p\ (e^\epsilon - 1)) \ .$$

Since $p \le \frac{1}{e^\epsilon+1}$, we conclude

$$ML(C(\epsilon,p)) \le \log_2 \frac{2e^\epsilon}{e^\epsilon + 1} \ .$$

For the proof of Assertion 2), we compute the maximal leakage of $C(\epsilon, \frac{1}{e^\epsilon+1})$ along the same lines as in the proof of Assertion 1. The ϵ-differential privacy of the channel follows from Lemma 6. ■

B. Channels of arbitrary range

We now show that the characterization of the leakage of channels binary range extends to channels with arbitrary (but finite) range.

Formally, we let $X = \{0, 1\}$, $Y = \{y_1 \dots y_k\}$ and let $P_{Y|X}$ be a a channel. Then $P_{Y|X}$ can be represented by a matrix

$$\begin{array}{c|ccc} P_{Y|X} & Y = y_1 & \dots & Y = y_k \\ \hline X = 0 & p_1 & \dots & p_k \\ X = 1 & q_1 & \dots & q_k \end{array}$$

where $p_1 + \ldots + p_k = q_1 + \ldots + q_k = 1$.

The following corollary states that the upper bounds for the leakage given in Theorem 3 also hold for channels of arbitrary range.

Corollary 1. *If $P_{Y|X}$ is ϵ-differentially private, then the maximal leakage ML of $P_{Y|X}$ verifies*

$$ML(P_{Y|X}) \leq \log_2 \frac{2e^\epsilon}{e^\epsilon + 1}$$

Proof: We construct an ϵ-differentially private channel $P_{\bar{Y}|X}$ with $Range(\bar{Y}) = \{0, 1\}$ such that the leakage of $P_{\bar{Y}|X}$ matches that of $P_{Y|X}$. Technically, let $I = \{i \mid p_i \leq q_i\}$, $\bar{p} = \sum_{i \notin I} p_i$ and $\bar{q} = \sum_{i \in I} q_i$. Then we define $P_{\bar{Y}|X}$ as:

$$\begin{array}{c|cc} P_{\bar{Y}|X} & Y = 0 & Y = 1 \\ \hline X = 0 & \bar{p} & 1 - \bar{p} \\ X = 1 & 1 - \bar{q} & \bar{q} \end{array}$$

Note that $1 - \bar{p} = \sum_{i \in I} p_i$, $1 - \bar{q} = \sum_{i \notin I} q_i$, hence we have $\bar{p} \geq 1 - \bar{q}$ and $\bar{q} \geq 1 - \bar{p}$. Applying Theorem 1 to $P_{Y|X}$, we obtain

$$\begin{aligned} ML(P_{Y|X}) &= \log_2 \sum_y \max_x P_{Y|X}(y, x) \\ &= \log_2 (\sum_{i \notin I} p_i + \sum_{i \in I} q_i) \\ &= \log_2 (\bar{p} + \bar{q}) \\ &= ML(P_{\bar{Y}|X}) \end{aligned}$$

Moreover, as $P_{Y|X}$ is ϵ-differentially private, we know that, for every $i \in I$, $q_i \leq e^\epsilon \ p_i$, and for every $i \notin I$, $p_i \leq e^\epsilon q_i$. We obtain $\sum_{i \in I} q_i \leq e^\epsilon \ \sum_{i \in I} p_i$ and $\sum_{i \notin I} p_i \leq e^\epsilon \ \sum_{i \notin I} q_i$. Hence $P_{\bar{Y}|X}$ is also differentially private, which concludes this proof. ■

Tightness of the bound, as expressed in Theorem 3.2, immediately extends to channels with arbitrary output size, since one can view any channel with output of size 2 as a channel with output of size k.

V. Bounds on the Leakage of Differentially Private Mechanisms on Arbitrary Input Domains

In Section IV, we have characterized the maximal leakage of differentially private channels with input domains of size 2. In this section, we consider the case of differentially private channels with input domains of arbitrary size. It turns out that a complete characterization of the leakage of such channels is a challenging problem. We take the following two steps towards such a characterization: First, we establish upper bounds for the leakage of any ϵ-differentially private channel with an n-bit input domain. We obtain this result by exhibiting a connection to coding theory, through which we derive both concrete bounds for the leakage, and limits for the bounds that can be achieved using this connection.

Second, we construct an ϵ-differentially channel that takes inputs of n bits and leaks at least $n \log_2 \frac{2e^\epsilon}{e^\epsilon+1}$ of those bits. By increasing n, one can hence leak an arbitrary amount of information, while retaining the privacy guarantee. This result implies that there can be no general (i.e. independent of the size of the input domain) upper bound for the leakage of ϵ-differentially private channels.

A. Bounds for Input Domains of Fixed Size

To obtain our bounds, we first cover the channel's input domain by spheres of a fixed radius with respect to the Hamming metric Δ. Using Theorem 1 we obtain upper bounds for the information leaked about the elements of each sphere. By summing over all spheres, this yields bounds for the information leaked about the entire input domain. These bounds are parametric in the number of spheres and their radius. We observe that the used sphere covering corresponds to a covering code, and we show how a given code can be used to instantiate the parameters and obtain a concrete bound for the leaked information; moreover, we show how the Hamming bound leads to limits for the bounds that can be obtained in this way.

Formally, a subset $\{x_1, \ldots, x_m\} \subseteq \{0, 1\}^n$ is a *d-covering code of length n and size m* if, for every $x \in \{0, 1\}^n$ there is an $i \in \{1, \ldots, m\}$ such that $\Delta(x, x_i) \leq d$. This definition corresponds to the requirements that the spheres $U_d(x_i)$ with radius d and center x_i, defined by

$$U_d(x_i) = \{x \in \{0, 1\}^n \mid \Delta(x, x_i) \leq d\} \ ,$$

cover the whole space. For an overview of research on covering codes, see [8].

For a given d-covering code, we obtain the following bounds for the leakage of differentially private mechanisms.

Theorem 4. *Let $P_{Y|X}$ be a channel with $Range(X) = \{0, 1\}^n$ and let $\{x_1, \ldots, x_m\}$ be a d-covering code of length n. If $P_{Y|X}$ satisfies ϵ-differential privacy, then its maximal leakage is upper-bounded by*

$$ML(P_{Y|X}) \leq \epsilon d \log_2 e + \log_2 m \ .$$

Proof of Theorem 4.:

$$
\begin{aligned}
ML(P_{Y|X}) &\overset{(*)}{=} \log_2 \sum_y \max_x P_{Y|X}(y,x) \\
&\leq \log_2 \sum_y \sum_{i=1}^{m} \max_{x \in U_d(x_i)} P_{Y|X}(y,x) \\
&\overset{(**)}{\leq} \log_2 \sum_y \sum_{i=1}^{m} P_{Y|X}(y,x_i) e^{\epsilon d} \\
&= \log_2 e^{\epsilon d} \sum_{i=1}^{m} \sum_y P_{Y|X}(y,x_i) \\
&= \epsilon d \log_2 e + \log_2 m \ ,
\end{aligned}
$$

where $(*)$ follows from Theorem 1 and $(**)$ follows from the definition of differential privacy (Definition 2). ■

Note that the bounds obtained using Theorem 4 are of interest only for instantiations with $\epsilon d \log_2 e + \log_2 m \leq n$ since, trivially, $ML(P_{Y|X}) \leq n$ for a channel with inputs of n bits. Moreover, note that a concrete covering code is required for obtaining the desired bounds for $ML(P_{Y|X})$ in terms of n and ϵ. We next give two example instantiations with simple covering codes.

For the first instantiation, we consider a trivial n-covering code of size 1. With this code, we obtain the following corollary of Theorem 4.

Corollary 2. *Let $P_{Y|X}$ be a channel with $Range(X) = \{0,1\}^n$. If $P_{Y|X}$ satisfies ϵ-differential privacy, then*

$$ML(P_{Y|X}) \leq n \ \epsilon \ \log_2 e \ .$$

For the second instantiation, we consider an n-ary repetition code of size 2. More precisely, we consider the code $\{0 \cdots 0, 1 \cdots 1\} \subseteq \{0,1\}^n$. Observe that for this code $d = \left\lfloor \frac{n}{2} \right\rfloor$, because each $x \in \{0,1\}^n$ has a Hamming distance of at most $\left\lfloor \frac{n}{2} \right\rfloor$ to either one of the codewords. Using this code, we obtain the following corollary of Theorem 4.

Corollary 3. *Let $P_{Y|X}$ be a channel with $Range(X) = \{0,1\}^n$. If $P_{Y|X}$ satisfies ϵ-differential privacy, then*

$$ML(P_{Y|X}) \leq \left\lfloor \frac{n}{2} \right\rfloor \ \epsilon \ \log_2 e + 1 \ .$$

For a fixed n, Figures 2 and 3 depict the bounds obtained by trivial covering codes (Corollary 2) and repetition codes (Corollary 3) as functions of ϵ. The corresponding curves cross, which illustrates that each of the codes gives tighter (i.e. lower) upper bounds for the leakage for different ranges of ϵ.

While there is a large number of codes with which Theorem 4 can be instantiated and with which the bounds from Corollaries 2 and 3 could potentially be improved [8], the so-called Hamming-bound puts a theoretical limit on what can be achieved using our proof technique.

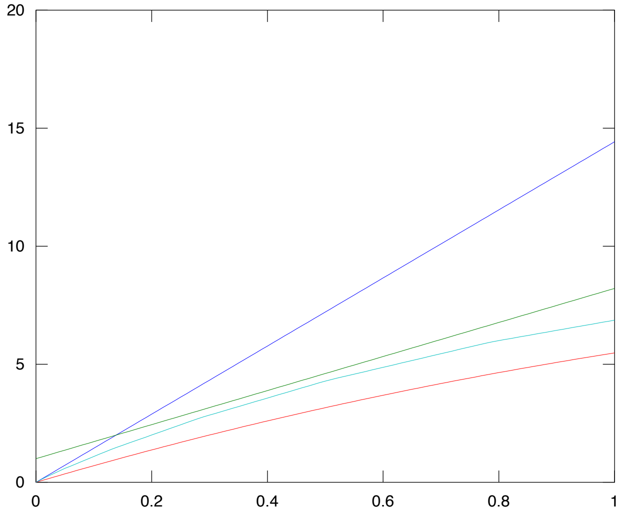

Figure 2. For a fixed $n = 10$, this figure depicts different upper and lower bounds for the leakage as functions of $\epsilon \in [0,1]$ (horizontal axis). The upper (blue) curve depicts the upper bounds obtained using trivial covering codes in Corollary 2. The second (green) curve depicts the upper bounds obtained using repetition codes in Corollary 3. The intersection of the curves shows that the each code leads to better bounds for different ranges of ϵ. The third (cyan) curve depicts the limit of what we can achieve using our proof technique, as stated in Corollary 4. Finally, the lower (red) curve depicts the leakage of the channel constructed in Theorem 5. The small gap between the green and the cyan curves illustrates the good quality of the bounds obtained using repetition codes; the small gap between the red and the green curves illustrates that, for small n, the leakage of the channel constructed in Theorem 5 is not too far from optimal.

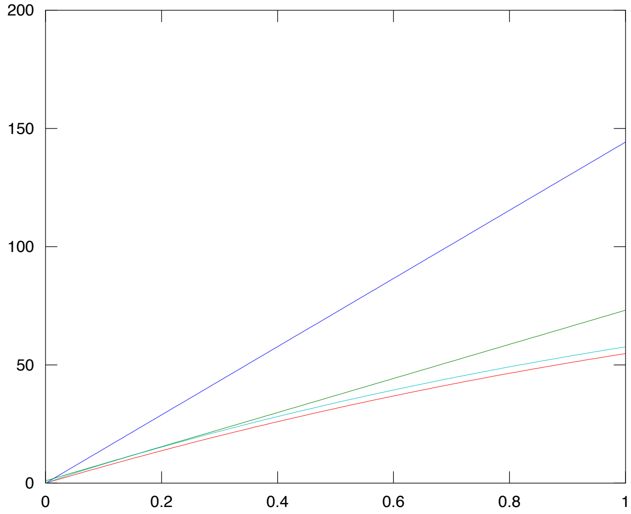

Figure 3. The curves in this figure depict the bounds as described in Figure 2, but for the case $n = 100$.

Corollary 4. *For given n and $\epsilon > 0$, any bound $B(\epsilon, n)$ for the maximal leakage obtained by Theorem 4 will satisfy*

$$B(\epsilon, n) \geq \min_{d=0}^{n} \left(\epsilon d \log_2 e + n - \log_2 \sum_{i=0}^{d} \binom{n}{i} \right) .$$

Proof: Formally the Hamming bound

$$m \sum_{i=0}^{d} \binom{n}{i} \geq 2^n$$

is obtained by observing that a necessary requirement for any d-covering code of length n and size m is that the number of elements in the corresponding spheres sums to 2^n. Inserting this bound into Theorem 4 yields the assertion. ■

Figures 2 and 3 depict the limit given by Corollary 4 as a function of ϵ. They also illustrate that, for $n = 10$ and $n = 100$, the bounds obtained by repetition codes are close to this limit.

B. Impossibility of General Bounds for the Leakage of Differentially Private Mechanisms

Theorem 4 and Corollary 2 give bounds for the leakage in terms of n and ϵ. In this section, we show that there can be no such bounds that are independent of n.

To this end we construct, for every $\epsilon > 0$ and every $n \in \mathbb{N}$, a channel that is ϵ-differentially private and that leaks $n \log_2 \frac{2e^\epsilon}{e^\epsilon+1}$ bits. Technically, we achieve this by n-fold parallel composition of the 1-bit channel from Theorem 3.2. The composition results developed in Section III-C then show that differential privacy remains invariant under composition, but that the leakage increases by a factor of n. We obtain the following theorem.

Theorem 5. *There is an ϵ-differentially private channel $P_{Y|X}$ with $Range(X) = \{0, 1\}^n$ whose maximal leakage satisfies*

$$ML(P_{Y|X}) = n \log_2 \frac{2e^\epsilon}{e^\epsilon + 1} .$$

Proof: Theorem 3.2 shows that there is a 1-bit channel C that leaks $\log_2 \frac{2e^\epsilon}{e^\epsilon+1}$ bits. Consider the n-fold parallel composition $C^n = C \times \cdots \times C$ of C. Lemma 4.1 shows that C^n is still ϵ-differentially private. Lemma 4.2 shows that C^n leaks $n \log_2 \frac{2e^\epsilon}{e^\epsilon+1}$ bits, which concludes this proof. ■

In Figures 2 and 3, we depict the leakage of the channel constructed in Theorem 5 as a function of ϵ. It should be noted that for $\epsilon \to \infty$, $n \log_2 \frac{2e^\epsilon}{e^\epsilon+1}$ converges from below to $n \log_2 2 = n$, which is the maximal leakage of a channel with input set $\{0, 1\}^n$.

An important consequence of Theorem 5 is that there cannot be domain-size-independent bounds for the leakage in terms of differential privacy.

Corollary 5. *For every $\epsilon > 0$ and $l > 0$, there exists an ϵ-differentially private channel C such that $ML(C) > l$.*

It is sufficient to consider the channel C^n with a sufficiently large value for n, since the channel is ϵ-differentially private for all n and $ML(C^n) \to \infty$ as $n \to \infty$.

VI. Decidability of Bounds for the Leakage of Differentially Private Channels

In contrast to the case $n = 1$ (see Theorem 3), our upper bounds for the leakage of differentially private mechanisms are not tight for the case $n > 1$. In particular, there is a gap between the upper bounds derived in Theorem 4 and the leakage of the channel constructed in Theorem 5, see Figures 2 and 3. Although a closed expression characterizing the leakage for $n > 1$ is still elusive, we next show that it is decidable whether a given rational function constitutes such a characterization. Our proof proceeds by reducing the problem to a system of polynomial inequalities over the reals.

A. Background on decidability of real closed fields

For completeness, we provide a brief overview of the decidability results that are needed for our reduction; we refer the interested reader to [16] for further details.

A real closed field is an ordered field $\mathbb{F}$ such that every positive element of $\mathbb{F}$ is a square, and every polynomial P with coefficients in $\mathbb{F}$ that is of odd degree has at least one root. Formally, the theory of ordered fields is obtained from the theory of closed fields by adding a binary relation $\leq$ that satisfies the axioms of partial orders: reflexivity, transitivity, and anti-symmetry, plus the axiom of totality

$$\forall a\ b \in \mathbb{F}.\ a \leq b \vee b \leq a$$

and compatibility axioms for addition and multiplication:

$$\forall a\ b\ c \in \mathbb{F}.\ a \leq b \Rightarrow a + c \leq b + c$$
$$\forall a\ b \in \mathbb{F}.\ (0 \leq a \Rightarrow 0 \leq b) \Rightarrow 0 \leq a\ b$$

Then, the theory of real closed fields is obtained from the theory of ordered fields by adding an axiom for the existence of square roots

$$\forall x \in \mathbb{F}.\ x > 0 \Longrightarrow \exists y \in \mathbb{F}.\ x = y^2$$

and an axiom scheme for the existence of a root for polynomials Q of odd order

$$\exists x \in \mathbb{F}.\ Q(x) = 0$$

Any sentence of the theory of real closed fields is valid if and only if it is valid for the real numbers. Moreover, Tarski [33] proved that the theory admits elimination of quantifiers, i.e. for every sentence ϕ of the theory, there exists an equivalent sentence ψ that is quantifier-free, and then concluded that the theory of real closed fields is decidable. However, the decision procedure that can be extracted from Tarski's result is highly inefficient. More efficient algorithms have been devised subsequently, such as the Cylindrical Algebraic Decomposition method [9], or Hörmander's method [17].

It should be noted that many useful constructions can be expressed in the theory of real closed fields. For instance, one can use 1 and addition to encode any natural number, and hence any polynomial expression with integer coefficients. Moreover, one can encode the maximum of two expressions e and e' by introducing a fresh variable x and requiring that $e \leq x \wedge e' \leq x \wedge (x = e \vee x = e')$. One could use such an encoding to express that the leakage of a channel is bounded by the logarithm of an expression that depends on ϵ. For the sake of readability, we prefer to give a more direct formalization.

B. Reduction

We prove the decidability of the existence of (optimal) rational upper bounds by characterizing them in the theory of real closed fields. In the sequel, we assume channels $P_{Y|X}$ with $Range(X) = \{0,1\}^n$ and $Range(Y) = \{y_1, \ldots, y_m\}$ arbitrary but finite. The following theorem formalizes our results.

Theorem 6 (Decidability of Rational Bounds)**.** *Let r and s be polynomials with coefficients in $\mathbb{Z}$, such that r and s are strictly positive over* $[1,\infty)$*. Then the following questions are decidable.*

1) *For all $\epsilon > 0$ and for all ϵ-differentially private channels $P_{Y|X}$ it holds that*

$$ML(P_{Y|X}) \leq \log_2 \frac{r(e^\epsilon)}{s(e^\epsilon)}$$

2) *For all $\epsilon > 0$ there exists an ϵ-differentially private channel $P_{Y|X}$ such that*

$$ML(P_{Y|X}) = \log_2 \frac{r(e^\epsilon)}{s(e^\epsilon)} \ .$$

Proof: It is sufficient to find formulae of the theory of reals expressing that $\log_2 r/s$ is an upper bound (resp. a tight upper bound) of the leakage of every ϵ-differentially private channel. As a first step, it is convenient to rephrase the problems by considering $\lambda = e^\epsilon$ instead of ϵ. Formally, Assertion 1) of Theorem 6 is equivalent to the following assertion: for all $\lambda > 1$ and for all $\log_2 \epsilon$-differentially private channels $P_{Y|X}$ it holds that

$$ML(P_{Y|X}) \leq \frac{r(\lambda)}{s(\lambda)} \ .$$

As a further step, for all $x \in \{0,1\}^n$ and $y \in \{y_1, \ldots, y_m\}$ we introduce variables $p_{x,y}$ for representing the entries of a matrix representation of channels $P_{Y|X}$. The requirement that a matrix represents a valid channel can be expressed by the following formula

$$\mathsf{VC} \equiv \left(\bigwedge_{x,y} 0 \leq p_{x,y} \right) \wedge \left(\bigwedge_{x} \sum_{y} p_{x,y} = 1 \right) \ ,$$

where variables x, y range over $Range(X)$ and $Range(Y)$, respectively. The requirement that a channel is $\log_2 \lambda$-differentially private can be expressed by the following formula

$$\mathsf{DP} \equiv \bigwedge_{\{x,x',y|\Delta(x,x')=1\}} p_{x',y} \leq \lambda p_{x,y}$$

where Δ denotes the Hamming distance.

Finally, we can express the requirement that the sum of the maximal elements of each column of a channel is upper-bounded by $r(\lambda)/s(\lambda)$, or equivalently that the maximal leakage of this channel is upper-bounded by $\log_2 (r(e^\epsilon)/s(e^\epsilon))$, for $\epsilon = \log_2 \lambda$, using the following formula

$$\mathsf{UB} \equiv \bigwedge_{0 \leq x_1, \ldots, x_m \leq 2^n - 1} \left(\sum_{1 \leq i \leq m} p_{x_i, y_i} \right) s(\lambda) \leq r(\lambda) \ .$$

Notice that UB bounds the sum over all possible combinations of column entries. In particular, it bounds the sum over the largest entries in each column, whose logarithm corresponds to the maximal leakage (see Theorem 1).

It follows immediately that the following formula is equivalent to Assertion 1) of Theorem 6

$$\forall \lambda \ \forall (p_{x,y}) : \ \lambda \geq 1 \Rightarrow \mathsf{VC} \Rightarrow \mathsf{DP} \Rightarrow \mathsf{UB}$$

In other words, the validity of a rational upper bound to the exponential of the leakage can be expressed in terms of polynomial inequalities, which concludes the proof for Assertion 1).

By a similar reasoning, one can establish that the following formula is equivalent to Assertion 2) of Theorem 6

$$\forall \lambda : \ \lambda \geq 1 \Rightarrow \exists (p_{x,y}) : \ \mathsf{VC} \wedge \mathsf{DP} \wedge \mathsf{EB}$$

where the formula EB indicates that the rational function is reached for some values.

$$\mathsf{EB} \equiv \bigvee_{0 \leq x_1, \ldots, x_m \leq 2^n - 1} \left(\sum_{1 \leq i \leq m} p_{x_i, y_i} \right) s(\lambda) = r(\lambda) \ .$$

■

A particular consequence of Theorem 6 is that, for every fixed n, it is decidable whether for every $\epsilon \geq 0$, $n \ \log \frac{2e^\epsilon}{e^\epsilon + 1}$ is an upper bound of the leakage for ϵ-differentially private channels. An important target for future work is to explore existing methods for resolving systems of polynomial inequalities over reals, in the aim of settling whether the above inequality holds for small n.

VII. Related work

We studied the relationship between two quantitative notions of confidentiality that have proposed in the context of two independent lines of research. The first line focusses on privacy-preserving data publishing; the second line focusses on information-flow analysis.

Privacy-preserving data publishing is a long-standing and well-studied problem, see e.g. the survey by Fung et al. [12].

Differential privacy by Dwork et al. [10] is the emerging consensus definition of privacy in the field. This is due to the fact that differential privacy enjoys desirable properties such as independence of adversary knowledge, well-behavedness under composition of queries, and that it comes with practical enforcement mechanisms such as the Laplacian mechanism [11], [22], [28]. Despite the strong security guarantees it provides, differential privacy has proven to be useful in practical applications [25], [31].

Information-theoretic notions of confidentiality have emerged from research on information-flow security and were initially targeted towards the analysis of covert channels [14], [23], [26]. More recently, they have also been applied for the analysis of side-channels in cryptographic algorithms [18], [19], [32], and for the quantitative analysis of anonymity protocols [5]. Moreover, a number of automatic analysis techniques have recently been proposed [2], [6], [24], [27].

We are aware of two independent and concurrent studies that aim at understanding the connections between differential privacy and information-theoretic notions of confidentiality. Alvim et al. [1] perform an information-theoretic analysis of the Laplacian mechanism (see Section II-B). In particular, they give upper bounds for the mutual information between the outputs of the original query and the outputs of the query that is perturbed using the Laplacian mechanism. Our analysis is more general in that it does not focus on a particular mechanism for achieving differential privacy; rather, it yields an information-theoretic characterization of the end-to-end confidentiality guarantees provided by *any* differentially private query.

Clarkson and Schneider [7] consider differential privacy in their study of quantitative integrity. In particular, they characterize differential privacy in terms of the mutual information between the output of a query and each single bit of the input. This characterization bears resemblance to our analysis of the case $n = 1$ in Section IV; the difference lies in the security guarantees provided and the notions of entropy used: Clarkson and Schneider give bounds on the transmission of information using Shannon entropy, whereas our approach gives bounds on the probability of guessing the input using min-entropy. Our approach goes beyond the single-bit-case in that we give upper and lower bounds, together with decidability results, for the leakage of differentially private queries on arbitrary input domains. Clarkson and Schneider also capture *utility* in their model; this is an interesting starting point for a future investigation of the trade-off between leakage and utility.

VIII. Conclusion

We performed an information-theoretic analysis of differential privacy. In particular, we established the first upper bounds for the leakage of differentially private mechanisms and provided empirical evidence of their accuracy. On a practical level, our contributions pave the way for applying tools from differential privacy for bounding the leakage of programs. On a conceptual level, we unveiled a connection to coding theory, and we demonstrated how this connection can be leveraged for deriving precise bounds from covering codes.

Our prime target for future work is to close the gap between the upper bounds for leakage which we derived using coding theory, and the lower bounds which we obtained using the n-fold compositions of an optimal 1-bit channel. Thanks to our decidability results, it may be possible to use proof tools to validate rational bounds for small n.

References

[1] M. S. Alvim, K. Chatzikokolakis, P. Degano, and C. Palamidessi. Differential privacy versus quantitative information flow. *CoRR*, abs/1012.4250, 2010.

[2] M. Backes, B. Köpf, and A. Rybalchenko. Automatic Discovery and Quantification of Information Leaks. In *Proc. IEEE Symp. on Security and Privacy (S&P '09)*, pages 141–153. IEEE, 2009.

[3] C. Braun, K. Chatzikokolakis, and C. Palamidessi. Quantitative notions of leakage for one-try attacks. *Electr. Notes Theor. Comput. Sci.*, 249:75–91, 2009.

[4] C. Cachin. *Entropy Measures and Unconditional Security in Cryptography*. PhD thesis, ETH Zürich, 1997.

[5] K. Chatzikokolakis, C. Palamidessi, and P. Panangaden. Anonymity protocols as noisy channels. *Inf. Comput.*, 206(2-4):378–401, 2008.

[6] D. Clark, S. Hunt, and P. Malacaria. A static analysis for quantifying information flow in a simple imperative language. *Journal of Computer Security*, 15(3):321–371, 2007.

[7] M. R. Clarkson and F. B. Schneider. Quantification of integrity. Cornell Computing and Information Science Technical Reports, 2011. http://hdl.handle.net/1813/22012.

[8] G. Cohen, I. Honkala, S. Litsyn, and A. Lobstein. *Covering Codes*. Elsevier Science, 1997.

[9] G. E. Collins. Hauptvortrag: Quantifier elimination for real closed fields by cylindrical algebraic decomposition. In H. Barkhage, editor, *Automata Theory and Formal Languages*, volume 33 of *Lecture Notes in Computer Science*, pages 134–183. Springer, 1975.

[10] C. Dwork. Differential Privacy. In *Proc. 33rd Intl. Colloquium on Automata, Languages and Programming (ICALP '06)*, volume 4052 of *LNCS*, pages 1–12. Springer, 2006.

[11] C. Dwork, F. McSherry, K. Nissim, and A. Smith. Calibrating Noise to Sensitivity in Private Data Analysis. In *Proc. 3rd Theory of Cryptography Conference (TCC '06)*, volume 3876 of *LNCS*, pages 265–284. Springer, 2006.

[12] B. C. M. Fung, K. Wang, R. Chen, and P. S. Yu. Privacy-preserving data publishing: A survey of recent developments. *ACM Comput. Surv.*, 42(4), 2010.

[13] S. R. Ganta, S. P. Kasiviswanathan, and A. Smith. Composition attacks and auxiliary information in data privacy. In *Proc. 14th ACM Conference on Knowledge Discovery and Data Mining (KDD '08)*, pages 265–273. ACM, 2008.

[14] J. W. Gray. Toward a Mathematical Foundation for Information Flow Security. *Journal of Computer Security*, 1(3-4):255–294, 1992.

[15] S. Hamadou, V. Sassone, and C. Palamidessi. Reconciling belief and vulnerability in information flow. In *Proc. 31st IEEE Symposium on Security and Privacy (S&P '10)*, pages 79–92. IEEE Computer Society, 2010.

[16] J. Harrison. *Handbook of Practical Logic and Automated Reasoning*. Cambridge University Press, 2009.

[17] L. Hörmander. *The Analysis of Linear Partial Differential Operators II: Differential Operators with Constant Coefficients.* Springer, 1983.

[18] B. Köpf and D. Basin. An Information-Theoretic Model for Adaptive Side-Channel Attacks. In *Proc. ACM Conf. on Computer and Communications Security (CCS '07)*, pages 286–296. ACM, 2007.

[19] B. Köpf and M. Dürmuth. A Provably Secure and Efficient Countermeasure against Timing Attacks. In *Proc. IEEE Computer Security Foundations Symposium (CSF '09)*, pages 324–335. IEEE, 2009.

[20] B. Köpf and A. Rybalchenko. Approximation and Randomization for Quantitative Information-Flow Analysis. In *Proc. 23rd IEEE Computer Security Foundations Symposium (CSF '10)*, pages 3–14. IEEE, 2010.

[21] B. Köpf and G. Smith. Vulnerability Bounds and Leakage Resilience of Blinded Cryptography under Timing Attacks. In *Proc 23rd. IEEE Computer Security Foundations Symposium (CSF '10)*, pages 44–56. IEEE, 2010.

[22] C. Li, M. Hay, V. Rastogi, G. Miklau, and A. McGregor. Optimizing linear counting queries under differential privacy. In *Proc. 29th ACM Symposium on Principles of Database Systems (PODS '10)*, pages 123–134. ACM, 2010.

[23] G. Lowe. Quantifying Information Flow. In *Proc. IEEE Computer Security Foundations Workshop (CSFW '02)*, pages 18–31. IEEE, 2002.

[24] S. McCamant and M. D. Ernst. Quantitative information flow as network flow capacity. In *Proc. ACM Conf. on Programming Language Design and Implementation (PLDI '08)*, pages 193–205. ACM, 2008.

[25] F. McSherry. Privacy integrated queries: an extensible platform for privacy-preserving data analysis. In *Proc. International Conference on Management of Data (SIGMOD '09)*, pages 19–30. ACM, 2009.

[26] J. K. Millen. Covert Channel Capacity. In *Proc. IEEE Symp. on Security and Privacy (S&P '87)*, pages 60–66. IEEE, 1987.

[27] J. Newsome, S. McCamant, and D. Song. Measuring channel capacity to distinguish undue influence. In *Proc. ACM Workshop on Programming Languages and Analysis for Security (PLAS '09)*, pages 73–85. ACM, 2009.

[28] K. Nissim, S. Raskhodnikova, and A. Smith. Smooth sensitivity and sampling in private data analysis. In *Proc. 39th Annual ACM Symposium on Theory of Computing (STOC '07)*, pages 75–84. ACM, 2007.

[29] A. Rényi. On measures of entropy and information. In *Proceedings of the 4th Berkeley Symposium on Mathematics, Statistics and Probability 1960*, pages 547–561, 1961.

[30] A. Roth and T. Roughgarden. Interactive privacy via the median mechanism. In *Proc. 42nd ACM Symposium on Theory of Computing (STOC '10)*, pages 765–774. ACM, 2010.

[31] I. Roy, S. T. V. Setty, A. Kilzer, V. Shmatikov, and E. Witchel. Airavat: Security and Privacy for MapReduce. In *Proc. 7th USENIX Symposium on Networked Systems Design and Implementation (NSDI '10)*, pages 297–312. USENIX Association, 2010.

[32] G. Smith. On the foundations of quantitative information flow. In *Proc. Intl. Conf. of Foundations of Software Science and Computation Structures (FoSSaCS '09)*, LNCS 5504, pages 288–302. Springer, 2009.

[33] A. Tarski. *A Decision Method for Elementary Algebra and Geometry*. Univ. of California Press, 2nd edition, 1951.

[34] X. Xiao, G. Wang, and J. Gehrke. Differential privacy via wavelet transforms. In *Proc. 26th International Conference on Data Engineering (ICDE '10)*, pages 225–236. IEEE, 2010.

Appendix

Theorem 1 ([21]). *The maximal leakage of a channel $P_{Y|X}$ can be computed by*

$$ML(P_{Y|X}) = \log_2 \sum_x \max_y P_{Y|X}(y, x) \ ,$$

where the maximum is assumed (e.g.) for uniformly distributed input.

Proof: Assume a fixed distribution P_X. Then

$$\begin{aligned} L &= \log_2 \frac{\sum_y P_Y(y) \max_x P_{X|Y}(x, y)}{\max_x P_X(x)} \\ &\overset{(*)}{=} \log_2 \frac{\sum_y \max_x (P_X(x) P_{Y|X}(y, x))}{\max_x P_X(x)} \\ &\overset{(**)}{\leq} \log_2 \frac{\sum_y \max_x P_{Y|X}(y, x)(\max_x P_X(x))}{\max_x P_X(x)} \\ &= \log_2 \sum_y \max_x P_{Y|X}(y, x)) \ , \end{aligned}$$

where (*) is Bayes' rule. Note that (**) is an equality if P_X is uniformly distributed, from which the assertion follows. ■

Lemma 1. *A channel $P_{Y|X}$ guarantees ϵ-differential privacy if and only if, for all $x \in \{0, 1\}^n$, all $i \in \{1, \ldots, n\}$, all distributions $P_{\dot{X}}$ of $\dot{X}$, and all $y \in Range(Y)$,*

$$P[\dot{X} = 1] \leq e^{\epsilon} P[\dot{X} = 1 \mid Y = y \wedge X = x[i/\dot{X}]] \ . \quad (5)$$

Proof: Assume that $P_{Y|X}$ is ϵ-differentially private but does not satisfy (5). Then there are $x \in \{0, 1\}^n$, $i \in \{0, \ldots, n\}$, $P_{\dot{X}}$ and $y \in Range(Y)$ such that

$$\frac{P[\dot{X} = 1]}{P[\dot{X} = 1 \mid Y = y \wedge X = x[i/\dot{X}]]} > e^{\epsilon}$$

Using Bayes' rule we obtain

$$\frac{P[Y = y \wedge X = x[i/\dot{X}]]}{P[Y = y \wedge X = x[i/\dot{X}] \mid \dot{X} = 1]} > e^{\epsilon} \ . \quad (6)$$

We observe that the denominator in (6) is equivalent to

$$P[Y = y \wedge X = x[i/1]]$$

and conclude that

$$\frac{P[Y = y \wedge X = x[i/0]]}{P[Y = y \wedge X = x[i/1]]} > e^{\epsilon} \ . \quad (7)$$

Here (7) follows by observing that the numerator in (6) is equivalent to

$$P[\dot{X} = 0] P[Y = y \wedge X = x[i/0]] + P[\dot{X} = 1] \, P[Y = y \wedge X = x[i/1]] \ .$$

Note that (7) contradicts the assumption that $P_{Y|X}$ is ϵ-differentially private because $x[i/0]$ and $x[i/1]$ differ only in one bit.

For the other direction, assume that $P_{Y|X}$ satisfies (5) but is not ϵ-differentially private. I.e. there are $x \in \{0, 1\}^n$, $i \in \{1, \ldots, n\}$, and $y \in Range(Y)$ such that

$$\frac{P[Y = y \mid X = x[i/0]]}{P[Y = y \mid X = x[i/1]]} > e^{\epsilon} \quad (8)$$

We obtain a contradiction by showing that

$$\frac{P[\dot{X} = 1]}{P[\dot{X} = 1 \mid Y = y \wedge X = x[i/\dot{X}]]} > e^{\epsilon} \quad (9)$$

for some distribution $P_{\dot{X}}$. To this end, observe that the left-hand side of (9) is equal to

$$\frac{P[Y = y \wedge X = x[i/\dot{X}]]}{P[Y = y \wedge X = x[i/1]]} \quad (10)$$

which is obtained by applying Bayes rule to the denominator of (9). We define $P_{\dot{X}}$ such that $P_{\dot{X}}(0) = \alpha$. Then

$$\begin{aligned} P[Y = y \wedge X = x[i/\dot{X}]] = (1 - \alpha) \, & P[Y = y \wedge X = x[i/1]] \\ & + \alpha \, P[Y = y \wedge X = x[i/0]] \end{aligned}$$

By inserting this expansion into (10) and applying (8) we obtain

$$\frac{P[Y = y \mid X = x[i/0]]}{P[Y = y \mid X = x[i/1]]} > (1 - \alpha) + \alpha e^{\epsilon}$$

which becomes larger than e^{ϵ} if α converges to 1 and concludes this proof. ■

Lemma 3. *Let C_i be channels that are ϵ_i-differentially private ($i = 1, 2$). Then*

1) *$C_1 + C_2$ is $\epsilon_1 + \epsilon_2$-differentially private, and*
2) *$ML(C_1 + C_2) \leq ML(C_1) + ML(C_2)$.*

Proof: For 1), choose an arbitrary $(y_1, y_2) \in Y_1 \times Y_2$ and consider $x, x' \in X$ that differ exactly in one position. Then we have

$$\begin{aligned} & \left| \log \frac{P_{Y_1 \times Y_2 | X = x}(y_1, y_2)}{P_{Y_1 \times Y_2 | X = x'}(y_1, y_2)} \right| \\ &= \left| \log \frac{P_{Y_1 | X = x}(y_1) P_{Y_2 | Y_1, X = y_1, x}(y_2)}{P_{Y_1 | X = x'}(y_1) P_{Y_2 | Y_1, X = y_1, x'}(y_2)} \right| \\ &\leq \left| \log \frac{P_{Y_1 | X = x}(y_1)}{P_{Y_1 | X = x'}(y_1)} \right| + \left| \log \frac{P_{Y_2 | Y_1, X = y_1, x}(y_2)}{P_{Y_2 | Y_1, X = y_1, x'}(y_2)} \right| \\ &\leq \epsilon_1 + \epsilon_2 \ , \end{aligned}$$

from which the assertion follows directly. A proof of 2) is given in the body of the paper. ■

Lemma 4. *Let C_i be channels that are ϵ_i-differentially private ($i = 1, 2$). Then*

1) *$C_1 \times C_2$ is $\max\{\epsilon_1, \epsilon_2\}$-differentially private, and*
2) *$ML(C_1 \times C_2) = ML(C_1) + ML(C_2)$.*

Proof: For the proof of 1), consider arbitrary $(y_1, y_2) \in Y$ and $x = (x_1, x_2), x' = (x'_1, x'_2) \in X_1 \times X_2 = \{0, 1\}^n$ that differ in one single bit. Consider first the case that the differing bit is in the first component, i.e. $x_1 \neq x_1'$ and $x_2 = x_2'$. Then

$$\begin{aligned}\left|\log \frac{P_{Y|X=x}(y)}{P_{Y|X=x'}(y)}\right| &= \left|\log \frac{P_{Y_1|X_1=x_1}(y_1)P_{Y_2|X_2=x_2}(y_2)}{P_{Y_1|X_1=x_1'}(y_1)P_{Y_2|X_2=x_2'}(y_2)}\right| \\ &\overset{(*)}{=} \left|\log \frac{P_{Y_1|X_1=x_1}(y_1)}{P_{Y_1|X=x_1'}(y_1)}\right| \\ &\overset{(**)}{\leq} \epsilon_1 \ ,\end{aligned}$$

where $(*)$ follows because x and x' coincide on their second component and $(**)$ follows because C_1 is ϵ_1-differentially private. For the case that x and x' differ only in their second component, we obtain a symmetric bound in terms of ϵ_2. Combining both bounds yields the assertion. A proof of 2) can be found in the body of the paper. ■