Privacy with Imperfect Randomness

Yevgeniy Dodis[1] and Yanqing Yao[2] **

[1] Department of Computer Science, New York University, New York, USA
dodis@cs.nyu.edu
[2] School of Computer Science and Engineering, Beihang University, Beijing, China
yaoyanqing1984@gmail.com

Abstract. We revisit the impossibility of a variety of cryptographic tasks including privacy and differential privacy with imperfect randomness. For traditional notions of privacy, such as security of encryption, commitment or secret sharing schemes, dramatic impossibility results are known [MP90, DOPS04] for several concrete sources $\mathcal{R}$, including a (seemingly) very "nice and friendly" Santha-Vazirani (SV) source. Somewhat surprisingly, Dodis et al. [DLMV12] showed that non-trivial *differential* privacy is possible with the SV sources. This suggested a qualitative gap between traditional and differential privacy, and left open the question of whether differential privacy is possible with more realistic (i.e., less structured) sources than the SV sources.

Motivated by this question, we introduce a new, modular framework for showing strong impossibility results for (both traditional and differential) privacy under a *general* imperfect source $\mathcal{R}$. As direct corollaries of our framework, we get the following new results:

(1) Existing, but *quantitatively improved*, impossibility results for traditional privacy, but under a wider variety of sources $\mathcal{R}$.
(2) First impossibility results for *differential* privacy for a variety of realistic sources $\mathcal{R}$ (including most "block sources", but not the SV source).
(3) Any imperfect source allowing (either traditional or differential) privacy under $\mathcal{R}$ admits a certain type of deterministic bit extraction from $\mathcal{R}$.

1 Introduction

Traditional cryptographic tasks take for granted the availability of perfect random sources, i.e., sources that output unbiased and independent random bits. However, in many situations it seems unrealistic to expect a source to be perfectly random, and one must deal with various imperfect sources of randomness. Some well known examples of such imperfect random sources are physical sources [BST03, BH05], biometric data [BDK+05, DORS08], secrets with partial leakage, and group elements from Diffie-Hellman key exchange [GKR04, Kra10].

Imperfect Sources. To abstract this concept, several formal models of imperfect sources have been described (e.g., [vN51,CFG+85,B86,SV86,CG88,LLS89,Z-

** Most of this work was done while the author visited New York University.

uc96,ACRT99,D01]). Roughly, they can be divided into extractable and non-extractable. Extractable sources (e.g., [vN51, CFG+85, Blu86, LLS89]) allow for deterministic extraction of nearly perfect randomness. And, while the question of optimizing the extraction rate and efficiency has been very interesting, from the qualitative perspective such sources are good for any application where perfect randomness is sufficient. Unfortunately, it was quickly realized many imperfect sources are non-extractable [SV86, CG88, Dod01]. The simplest example is the Santha-Vazirani (SV) source [SV86], which produces an infinite sequence of bits $r_1, r_2, \ldots$, with the property that $\Pr[r_i = 0 \mid r_1 \ldots r_{i-1}] \in [\frac{1}{2}(1-\gamma), \frac{1}{2}(1+\gamma)]$, for any setting of the prior bits $r_1, \ldots, r_{i-1}$. Namely, each bit has almost one bit of fresh entropy, but can have a small bias $\gamma < 1$. Santha and Vazirani [SV86] showed that there exists no deterministic extractor $\mathsf{Enc} : \{0,1\}^n \to \{0,1\}$ capable of extracting even a *single* bit of bias *strictly* less than γ from the γ-SV source, irrespective of how many SV bits $r_1, \ldots, r_n$ it is willing to wait for.

Despite this pessimistic result, ruling out the "black-box compiler" from imperfect (e.g., SV) to perfect randomness for *all* applications, one may still hope that specific "non-extractable" sources, such as SV-sources, might be sufficient for *concrete* applications, such as simulating probabilistic algorithms or cryptography. Indeed, a series of results [VV85, SV86, CG88, Zuc96, ACRT99] showed that very "weak" sources (including SV-sources and even much more realistic "weak" and "block" sources) are sufficient for simulating probabilistic polynomial-time algorithms; namely, for problems which do not inherently need randomness, but which could potentially be sped up using randomization. Moreover, even in the area of cryptography — where randomness is *essential* (e.g., for key generation) — it turns out that many "non-extractable" sources (again, including SV sources and more) are sufficient for *authentication* applications, such as the designs of MACs [MW97,DKRS06] and even signature schemes [DOPS04, ACM+14] (under appropriate hardness assumptions). Intuitively, the reason for the latter "success story" is that authentication applications only require that it is hard for the attacker to completely guess (i.e., "forge") some long string, so having min-entropy in our source should be sufficient to achieve this goal.

Negative Results for Privacy with Imperfect Randomness. In contrast, the situation appears to be much less bright when dealing with *privacy* applications, such as encryption, commitment, zero-knowledge, and a few others. First, McInnes and Pinkas [MP90] showed that unconditionally secure symmetric encryption cannot be based on SV sources, even if one is restricted to encrypting a single bit. This result was subsequently strengthened by Dodis et al. [DOPS04], who showed that SV sources are not sufficient for building even computationally secure encryption (again, even of a single bit), and, in fact, essentially any other cryptographic task involving "privacy" (e.g., commitment, zero-knowledge, secret sharing and others). This was again strengthened by Austrin et al. [ACM+14], who showed that the negative results still hold even if the SV source is efficiently samplable. Finally, Bosley and Dodis [BD07] showed an even more negative result: if a source of randomness $\mathcal{R}$ is "good enough" to generate a secret key capable of encrypting k bits, then one can deterministically

extract nearly k almost uniform bits from $\mathcal{R}$, suggesting that traditional privacy *requires* an "extractable" source of randomness.[1]

What about Differential Privacy? While the above series of negative results seem to strongly point in the direction that privacy inherently requires extractable randomness, a recent work of Dodis et al. [DLMV12] put a slight dent into this consensus, by showing that SV sources are provably sufficient for achieving a more recent notion of privacy, called *differential privacy* [DMNS06]. Intuitively, a differentially private mechanism $M(D, \mathbf{r})$ uses its randomness $\mathbf{r}$ to add some "noise" to the true answer $q(D)$, where D is some sensitive database of users, and q is some useful aggregate information (query) about the users of D. This noise is added in a way as to satisfy the following two conflicting properties (see Definitions 6 and 7 for formalism):

(a) *ε-differential privacy* (ε-DP): up to "advantage" ε, the returned value $z = M(D, \mathbf{r})$ does not tell any information about the value $D(i)$ of any individual user i, which was not already known to the attacker before z was returned;
(b) *ρ-utility*: on average (over $\mathbf{r}$), $|z - q(D)|$ is upper bounded by ρ, meaning that perturbed answer is not too far from the true answer.

Since we will be mainly talking about negative results, for the rest of this work we will restrict our attention to the simplest concrete example of differential privacy, where a "record" $D(i)$ is a single bit, and q is the Hamming weight $wt(D)$ of the corresponding bit-vector D (i.e., $wt(D) = \sum D(i)$). In this case, a very simple ε-DP mechanism [DMNS06] $M(D, \mathbf{r})$ would simply return $wt(D)+e(\mathbf{r})$ (possibly truncated to always be between 0 and $|D|$), where $e(\mathbf{r})$ is an appropriate noise[2] with $\rho = \mathbb{E}[|q(\mathbf{r})|] \approx 1/\varepsilon$. Intuitively, this setting ensures that when $D(i)$ changes from 0 to 1, the answer distribution $M(D, \mathbf{r})$ does not "change" by more than ε.

Coming back to Dodis et al. [DLMV12], the authors show that although no "additive noise" mechanism of the form $M(D, \mathbf{r}) = wt(D) + e(\mathbf{r})$ can simultaneously withstand all γ-SV-distributions $\mathbf{r} \leftarrow R$, a better designed mechanism (that they also constructed) is capable of working with all such distributions, provided that the utility ρ is now relaxed to be polynomial in $1/\varepsilon$, whose degree and coefficients depend on γ, but *not* on the size of the database D. Moreover, *the value ε can be made an arbitrarily small constant* (e.g., $\varepsilon \ll \gamma$). This should be contrasted with the impossibility results for the traditional privacy [MP90, DOPS04] with SV sources, where it was shown that $\varepsilon = \Omega(\gamma)$, meaning that even a fixed *constant* (let alone "negligible") security is impossible. Hence, the result of [DLMV12] suggested a *qualitative gap between traditional and differential privacy*, but left open the question of whether differential privacy is possible with more realistic (i.e., less structured) sources than the SV sources. Indeed, the SV sources seem to be primarily interesting from the perspective of negative results,

[1] On the positive side, [DS02] and [BD07] showed that extractable sources are not strictly necessary for encrypting a "very small" number of bits. Still, for natural "non-extractable" sources, such as SV sources, it is known that encrypting even a single bit is impossible [SV86, DOPS04, ACM+14].

[2] So called Laplacian distribution, but the details do not matter here.

since real-world distributions are unlikely to produce a sequence of bits, each of which has almost a full unit of fresh entropy.

Our Results In Brief. In part motivated by solving this question, we abstract and generalize prior techniques for showing impossibility results for achieving privacy with various imperfect sources of randomness. Unlike prior work (with the exception of [BD07]), which focused on specific imperfect sources $\mathcal{R}$ (e.g., SV sources), we obtain most of our results for *general* sources $\mathcal{R}$, but then use various natural sources (namely, SV sources [SV86], weak/block sources [CG88], and Bias-Control Limited sources [Dod01]) as specific examples to illustrate our technique. In particular, we introduce the concepts of *expressiveness* and *separability* of a given imperfect source $\mathcal{R}$ as a measure of its "imperfectness", and show the following results:

- Low levels of expressiveness generically imply strong impossibility results for *differential* as well as traditional privacy.
- We reduce expressiveness to separability and prove the equivalence between "weak bit extraction" and NON-separability.
- Though the separability of some concrete (e.g., SV) sources $\mathcal{R}$ was implicitly known, we show new separability results for several important sources, including general "block sources".

We stress that the first two results are completely generic, and reduce the question of feasibility of privacy under $\mathcal{R}$ to a much easier and self-contained question of separability of $\mathcal{R}$. And establishing the latter is the only "source-specific" technical work which remains. In particular, after explicitly stating known separability results for weak and SV sources, and establishing our new separability results for block and Bias-Control Limited (BCL) sources, we obtain the following direct corollaries:

- Existing, but *quantitatively improved*, impossibility results for traditional privacy, but under a wider variety of sources $\mathcal{R}$ (i.e., weak, block, SV, BCL).
- First impossibility results for *differential* privacy. Although, unsurprisingly, these results (barely) miss the highly structured SV sources, they come back *extremely quickly* once the source becomes slightly more realistic (e.g., a very "constrained" weak/block/BCL source).
- Any imperfect source allowing (either traditional or differential) privacy admits a certain type of deterministic bit extraction. (This result is incomparable to the result of [BD07].)

We briefly expand on these results below, but conclude that, despite the result of [DLMV12], our results seem to unify and strengthen the belief that, for the most part, privacy with imperfect randomness is impossible, unless the source is (almost) deterministically extractable. More importantly, they provide an intuitive, modular and unified picture elucidating the (im)possibility of privacy with *general* imperfect sources.

1.1 Our Results in More Detail

At a high level, our results follow the blueprint of [DOPS04] (who concentrated exclusively on the SV sources), but in significantly more modular and quantitatively optimized way (making our proofs somewhat more illuminating, in our opinion). In essence, they establish an impossibility of a given privacy task P under a source $\mathcal{R}$ using three steps:

STEP 1: IMPOSSIBILITY OF TASK P UNDER $\mathcal{R} \longrightarrow$ EXPRESSIVENESS OF $\mathcal{R}$.
Intuitively, *expressiveness* of $\mathcal{R}$ means that $\mathcal{R}$ is rich enough to "distinguish" any functions f and g which are not point-wise equal almost everywhere (see Definition 1): there exists $R \in \mathcal{R}$ s.t. $\mathsf{SD}(f(R), g(R))$ is "noticeable", where SD is the statistical distance between distributions.[3] With this clean abstraction, we almost trivially show (see Theorem 1) that most traditional privacy tasks P (extraction, encryption, secret sharing, commitment) imply the existence of sufficiently-distinct functions f and g that violate the expressiveness of $\mathcal{R}$. For example, such $f(\mathbf{r})$ and $g(\mathbf{r})$ are simply the encryptions of two different plaintexts under key $\mathbf{r}$ when P is encryption, and similar arguments hold for commitment, extraction and secret sharing schemes.

More interestingly, we show expressiveness is again sufficient to rule out even *differential* privacy (Theorem 2). The proof follows the same high-level intuition as for the traditional privacy, but is somewhat more involved. This is because DP only gives us security for "close" databases, while the utility guarantees are only meaningful for "far" databases. In particular, for this reason it will turn out that the expressiveness requirement on $\mathcal{R}$ for ruling out differential privacy will be slightly higher than that for traditional privacy (Theorem 2 vs. Theorem 1).[4] Still, aside from this quantitative difference, there is no qualitative difference between our arguments for traditional and differential privacy.

Overall, the deceptive simplicity of our "privacy-to-expressiveness" arguments is actually a *feature* of our framework, as these arguments are the *only place when the specific details of P matter*, as the rest of the framework — described below — will only concentrate on the expressiveness of $\mathcal{R}$!

STEP 2: EXPRESSIVENESS OF $\mathcal{R} \longrightarrow$ SEPARABILITY OF $\mathcal{R}$.
Intuitively, *separability* of $\mathcal{R}$ means that $\mathcal{R}$ is rich enough to "separate" any sufficiently large disjoint sets G and B (see Definition 8; wlog, assume that $|G| \geq |B|$): there exists $R \in \mathcal{R}$ s.t. $(\Pr[R \in G] - \Pr[R \in B])$ is "noticeable".[5] A moment reflection shows that separability is closely related to expressiveness, but restricted to *boolean* functions f and g of disjoint support (i.e., the characteristic

[3] Like in [DOPS04] and unlike [MP90], our distinguishers between $f(R)$ and $g(R)$ will be very efficient, but we will not require this in order not to clutter the notation.

[4] Jumping ahead, this will be the reason although our new impossibility results for DP will (barely) miss the SV sources, they will come back very quickly once the source becomes more realistic.

[5] For example, if $\mathcal{R}$ only consists of the uniform distribution U_n, the latter is impossible when $|G| = |B|$. In contrast, we will see that natural "non-extractable" sources (i.e., weak, block, SV, and BCL sources) are separable.

functions of G and B), which makes it noticeably easier to work with (as we will see).

Nevertheless, we show that *separability generically implies expressiveness*, with nearly identical parameters (see Theorem 3). This is where we differ and quantitatively improve the argument implicit in [DOPS04]: while [DOPS04] used a bit-by-bit hybrid argument to show expressiveness (for the SV source), our proof of Theorem 3 used a more clever "universal hashing trick", [6] allowing us to obtain results which are independent of the ranges of f and g (which, in turn, will later correspond to bit sizes of ciphertexts, commitments, secret shares, etc.)

Of independent interest, we also show that NON-separability of $\mathcal{R}$ is equivalent to some type of "weak bit extraction" from $\mathcal{R}$ (see Theorem 4): (a) when produced, the extracted bit is guaranteed to be almost unbiased, (b) although the extractor is allowed to fail, it will typically succeed at least on the uniform distribution.[7]

Coupled with Step 1, we get the following two implications. First, we reduce the impossibility of many privacy tasks P under $\mathcal{R}$ to a much easier question of separability of $\mathcal{R}$ (which is independent of P). Second, we generically show that the feasibility of P under $\mathcal{R}$ implies deterministic weak bit extraction from $\mathcal{R}$, incomparably complementing the prior result of [BD07]. Namely, [BD07] showed that several traditional privacy primitives, including (only multi-bit) encryption and commitment (but not secret sharing) imply the existence of multi-bit deterministic extraction schemes capable of extracting almost the same number of bits as the plaintext. On the positive, our result applies to a much wider set of primitives P (e.g., secret-sharing, as well as even *single-bit* encryption and commitment). On the negative, we can only argue a rather weak kind of single-bit extraction, where the extractor is allowed to fail, while [BD07] showed traditional, and possibly multi-bit, extraction.

Step 3: separability of various sources $\mathcal{R}$.
Unlike the prior results in [MP90, DOPS04, ACM+14], all the above results are true for any imperfect source $\mathcal{R}$. To get concrete impossibility results for natural sources, though, we finally must establish good separability bounds for specific $\mathcal{R}$. Such bounds were already implicitly known [DOPS04] (or trivial to see) for the SV and general weak sources, but we show how they can also be demonstrated for other natural sources: block sources [CG88] and Bias-Control Limited sources [Dod01]. In particular, our separability bounds for block sources turned out to be quite non-trivial, and form one of the more technical contributions of this work. See the proof of Lemma 2(b).

Aside from being natural and interesting in their own right, the new separability results for block/BCL sources are especially interesting from the perspective of differential privacy (see below). Indeed, both of them can be viewed as realistic relaxations of highly-structured (and unrealistic!) SV sources, but yet not

[6] Similar trick with randomness extractors was used, in a slightly different context, by [ACM+14].

[7] Unfortunately, we demonstrate that the limitation of part (b) holding only for the uniform distribution is somewhat inherent *in this great level of generality.*

as general/unstructured as weak sources. And since we already know that DP *is* possible with SV sources [DLMV12], it is interesting to know how soon it will take for the impossibility results to come back, once the source slowly becomes more realistic/unstructured, but before going "all the way" to being weak.

Putting them all together: New and Old Impossibility Results. Applying Steps 1-3 to specific sources of interest (i.e., weak, block, SV, and BCL sources), we immediately derive a variety of impossibility results for traditional privacy (see Table 1). Although these results were derived mainly as a "warm-up" to our (completely new) impossibility results for differentially privacy, they offer quantitative improvements to the results of [DOPS04] (due to stronger expressiveness-to-separability reduction). For example, they rule out even constant (as opposed to negligible) security for encryption/commitment/secret sharing, irrespective of the sizes of ciphertexts/commitments/shares. Relatedly, we unsurprisingly get stronger impossibility results for block/BCL sources than the more structured SV sources.

More interestingly, we obtain first impossibility results for differential privacy with imperfect randomness. In light of the positive result of [DLMV12], our separability result for SV sources is (barely) not strong enough to rule out differential privacy under SV sources. As we explained, this failure happened not because our framework was too weak to apply to SV sources or differential privacy, but rather due to a "local-vs-global gap" between the privacy and utility requirements for differential privacy.

However, once we consider general weak sources, or even much more structured BCL/block sources, the impossibility results come back *extremely quickly*! For example, when studying ε-DP with utility ρ, n-bit weak sources of min-entropy k are ruled out the moment $k = n - \log(\varepsilon\rho) - O(1)$ (Theorem 6(a)),[8] while BCL sources are ruled out the moment the number of "SV bits" b the attacker can fix completely (instead of only bias by γ) is just $b = \Omega(\log(\varepsilon\rho)/\gamma)$ (Theorem 6(c)). As $\varepsilon\rho$ is typically desired to be a constant, $\log(\varepsilon\rho)$ is an even smaller constant, which means we even rule out *constant* entropy deficiency $(n-k)$ (or $m-k$ for block source) or number of "interventions" b, respectively. We also compare impossibility results for traditional and differential privacy in Table 2, and observe that the latter are only marginally weaker than the former. This leads us to the conclusion that differential privacy is still rather demanding to achieve with realistic imperfect sources of randomness.

Due to space limitations, most proofs are deferred to the full version [DY14].

2 Preliminaries

Let U_S be the uniform distribution over a set S. For simplicity, $U_n \stackrel{def}{=} U_{\{0,1\}^n}$. For a distribution or random variable R, let $\mathbf{r} \leftarrow R$ denote the operation of

[8] More generally, even n-bit block sources with block length m and fresh min-entropy k per block are ruled out when $k = m - \log(\varepsilon\rho) - O(1)$, irrespective of the number of blocks n/m. See Theorem 6(b).

sampling a random $\mathbf{r}$ according to R , and $\mathbf{H}_\infty(R) \stackrel{def}{=} \min_{\mathbf{r}\in\mathrm{supp}(R)} \log \frac{1}{\Pr[R=\mathbf{r}]}$ denote the min-entropy of R. We call a family of distributions over $\{0,1\}^n$ a source, denoted as $\mathcal{R}_n$. All logarithms are to the base 2.

For two random variables R and R' over $\{0,1\}^n$, the statistical distance between R and R' is defined as $\mathsf{SD}(R,R') \stackrel{def}{=} \frac{1}{2} \sum_{\mathbf{r}\in\{0,1\}^n} |\Pr[R=\mathbf{r}] - \Pr[R'=\mathbf{r}]|$. One can observe that $\mathsf{SD}(R,R') = \max_{\mathsf{Eve}} |\Pr[\mathsf{Eve}(R)=1] - \Pr[\mathsf{Eve}(R')=1]|$, where Eve is a distinguisher. We say that the relative distance between R and R' is ε, denoted as $\mathsf{RD}(R,R')=\varepsilon$, if ε is the smallest number such that $e^{-\varepsilon}\cdot\Pr[R'=\mathbf{r}] \le \Pr[R=\mathbf{r}] \le e^{\varepsilon}\cdot\Pr[R'=\mathbf{r}]$ for all $\mathbf{r}\in\{0,1\}^n$. It's easy to see that $\mathsf{RD}(R,R')\le\varepsilon$ implies $\mathsf{SD}(R,R')\le e^{\varepsilon}-1$.

3 Expressiveness and its Implications to Privacy

In this section, we introduce the concept of expressiveness of a source. Then we study its implications to both traditional and differential privacy.

Informally, an expressive source $\mathcal{R}_n$ can separate two distributions $f(R)$ and $g(R)$, unless the functions f and g are point-wise equal almost everywhere.

Definition 1. *We say that a source $\mathcal{R}_n$ is $(t,\delta)-$expressive if for any functions $f,g:\{0,1\}^n\to\mathcal{C}$, where $\mathcal{C}$ is any universe, such that $\Pr_{\mathbf{r}\leftarrow U_n}[f(\mathbf{r})\neq g(\mathbf{r})] \ge \frac{1}{2^t}$ for some $t\ge 0$, there exists a distribution $R\in\mathcal{R}_n$ such that $\mathsf{SD}(f(R),g(R))\ge\delta$.*

3.1 Implications to Traditional Privacy

We recall (or define) some cryptographic primitives related to traditional privacy: bit extractor, bit encryption scheme, weak bit commitment, and bit T-secret sharing as follows.

Definition 2. *We say that $\mathsf{Ext}:\{0,1\}^n\to\{0,1\}$ is $(\mathcal{R}_n,\delta)$-secure bit extractor if for every distribution $R\in\mathcal{R}_n$, $|\Pr_{\mathbf{r}\leftarrow R}[\mathsf{Ext}(\mathbf{r})=1] - \Pr_{\mathbf{r}\leftarrow R}[\mathsf{Ext}(\mathbf{r})=0]| < \delta$ (equivalently, $\mathsf{SD}(\mathsf{Ext}(R),U_1)<\delta/2$).*

In the following, we consider the simplest encryption scheme, where the plaintext is composed of a single bit x.

Definition 3. *A $(\mathcal{R}_n,\delta)-$secure bit encryption scheme is a tuple of functions $\mathsf{Enc}:\{0,1\}^n\times\{0,1\}\to\{0,1\}^\lambda$ and $\mathsf{Dec}:\{0,1\}^n\times\{0,1\}^\lambda\to\{0,1\}$, where, for convenience, $\mathsf{Enc}(\mathbf{r},x)$ (resp. $\mathsf{Dec}(\mathbf{r},\mathbf{c})$) is denoted as $\mathsf{Enc}_{\mathbf{r}}(x)$ (resp. $\mathsf{Dec}_{\mathbf{r}}(\mathbf{c})$), satisfying the following two properties:*

(a) Correctness: for all $\mathbf{r}\in\{0,1\}^n$ and $x\in\{0,1\}$, $\mathsf{Dec}_{\mathbf{r}}(\mathsf{Enc}_{\mathbf{r}}(x))=x$;
(b) Statistical Hiding: $\mathsf{SD}(\mathsf{Enc}_R(0),\mathsf{Enc}_R(1))<\delta$, for every distribution $R\in\mathcal{R}_n$.

Commitment schemes allow the sender Alice to commit a chosen value (or statement) while keeping it secret from the receiver Bob, with the ability to reveal the committed value in a later stage. Binding and hiding properties are essential to any commitment scheme. Informally, "binding" means that it's "hard" for Alice to alter her commitment after she has made it; "hiding" means that it's "hard" for Bob to find out the committed value without Alice revealing it.

Each of them can be computational or information theoretical. However, we can't achieve information theoretically binding and information theoretically hiding properties at the same time. Instead of defining computational notions, we relax binding to some very weak property, so that hiding and this new (very weak) binding properties both can be information theoretical. Since we aim to show an impossibility result, such relaxation is justified.

Definition 4. *A $(\mathcal{R}_n, \delta)-$secure weak bit commitment is a function $\mathsf{Com} : \{0,1\}^n \times \{0,1\} \to \{0,1\}^\lambda$ satisfying that: for any distribution $R \in \mathcal{R}_n$,*

(a) Weak Binding: $\Pr_{\mathbf{r} \leftarrow U_n}[\mathsf{Com}(0;\mathbf{r}) \neq \mathsf{Com}(1;\mathbf{r})] \geq \frac{1}{2}$;

(b) Statistical Hiding: $\mathsf{SD}(\mathsf{Com}(0;R), \mathsf{Com}(1;R)) < \delta$.

Note that in the traditional notion of commitment, the binding property holds if it is "hard" to find $\mathbf{r}_1$ and $\mathbf{r}_2$ such that $\mathsf{Com}(0;\mathbf{r}_1) = \mathsf{Com}(1;\mathbf{r}_2)$. Here we give a much weaker binding notion. We only require that the attacker can not win with probability $\geq \frac{1}{2}$ by choosing $\mathbf{r}_1 = \mathbf{r}_2$ uniformly at random. For example, $\mathsf{Com}(x;r) = x \oplus r$, where $x, r \in \{0,1\}$ can be easily verified to be a weak bit commitment for any $\delta > 0$ (despite not being a standard commitment).

In the notion of T-party Secret Sharing, two thresholds T_1 and T_2, where $1 \leq T_1 < T_2 \leq T$, are involved such that (a) any T_1 parties have "no information" about the secret, (b) any T_2 parties enable to recover the secret. Because our purpose is to show an impossibility result, we restrict to $T_1 = 1$ and $T_2 = T$, and only consider one bit secret x.

Definition 5. *A $(\mathcal{R}_n, \delta)-$secure bit $T-$Secret Sharing scheme is a tuple $(\mathsf{Share}_1, \mathsf{Share}_2, \ldots, \mathsf{Share}_T, \mathsf{Rec})$ satisfying the following two properties:*

(a) Correctness: $\mathsf{Rec}(\mathsf{Share}_1(x,\mathbf{r}), \ldots, \mathsf{Share}_T(x,\mathbf{r})) = x$ for all $\mathbf{r} \in \{0,1\}^n$ and each $x \in \{0,1\}$;

(b) Statistical Hiding: $\mathsf{SD}(\mathsf{Share}_j(0;R), \mathsf{Share}_j(1;R)) < \delta$, for every index $j \in [T]$ and any distribution $R \in \mathcal{R}_n$.

Now we abstract and generalize the results of [MP90, DOPS04] to show that expressiveness implies the impossibility of security involving traditional privacy. See [DY14] for the proof.

Theorem 1.

(a) When $\mathcal{R}_n$ is $(0,\delta)-$expressive, no $(\mathcal{R}_n, \delta)$-secure bit extractor exists.

(b) When $\mathcal{R}_n$ is $(0,\delta)-$expressive, no $(\mathcal{R}_n, \delta)$-secure bit encryption scheme exists.

(c) *When $\mathcal{R}_n$ is $(1,\delta)-$expressive, no $(\mathcal{R}_n,\delta)$-secure weak bit commitment exists.*
(d) *When $\mathcal{R}_n$ is $(\log T,\delta)-$expressive, no $(\mathcal{R}_n,\delta)$-secure bit T-secret sharing exists.*

3.2 Implications to Differential Privacy

Dodis et al. [DLMV12] have shown how to do differential privacy with respect to the γ-SV source for all "queries of low sensitivity". Since we aim to show impossibility results, henceforth we only consider the simplest case: let $\mathcal{D} = \{0,1\}^N$ be the space of all databases and for $D \in \mathcal{D}$, the query function q is the Hamming weight function $wt(D) = |\{i \mid D(i) = 1\}|$, where $D(i)$ means the i-th bit ("record") of D. If the source $\mathcal{R}_n$ has only one distribution U_n, $\mathcal{R}_n$ is denoted by U_n for simplicity. For any $D, D' \in \mathcal{D}$, the discrete distance function between them is defined by $\Delta(D,D') \stackrel{def}{=} wt(D \oplus D')$, where $\oplus$ is the bitwise exclusive OR operator. We say that D and D' are neighboring if $\Delta(D,D') = 1$. A mechanism M is an algorithm that takes as input a database $D \in \mathcal{D}$ and a distribution $R \in \mathcal{R}_n$, and outputs a random value z. Informally, we wish $z = M(D,R)$ to approximate the true value $wt(D)$ without revealing too much information about any individual $D(i)$. More formally, a mechanism is differentially private for the Hamming weight queries if replacing an entry in the database with one containing fake information only changes the output distribution of the mechanism by a small amount. In other words, evaluating the mechanism on two neighboring databases, does not change the outcome distribution by much. On the other hand, we define its utility to be the expected difference between the true answer $wt(D)$ and the output of the mechanism. More formally,

Definition 6. *Let $\varepsilon \geq 0$ and $\mathcal{R}_n$ be a source. A mechanism M (for the Hamming weight queries) is $(\mathcal{R}_n,\varepsilon)$-differentially private if for all neighboring databases $D_1, D_2 \in \mathcal{D}$, and all distributions $R \in \mathcal{R}_n$, we have $RD(M(D_1,R), M(D_2,R)) \leq \varepsilon$. Equivalently, for any possible output z:*

$$\frac{\Pr_{\mathbf{r}\leftarrow R}[M(D_1,\mathbf{r}) = z]}{\Pr_{\mathbf{r}\leftarrow R}[M(D_2,\mathbf{r}) = z]} \leq e^{\varepsilon}.$$

Note that for $\varepsilon < 1$, we can rather accurately approximate e^{ε} by $1+\varepsilon$.

Definition 7. *Let $0 < \rho \leq N/4$ and $\mathcal{R}_n$ be a source. A mechanism M has $(\mathcal{R}_n,\rho)$-utility for the Hamming weight queries, if for all databases $D \in \mathcal{D}$ and all distributions $R \in \mathcal{R}_n$, we have $\mathbb{E}_{\mathbf{r}\leftarrow R}[|M(D,\mathbf{r}) - wt(D)|] \leq \rho$.*

We show that, much like with traditional privacy, expressiveness implies impossibility of differential privacy with imperfect randomness, albeit with slightly more demanding parameters.

Theorem 2. *Assume $1/(8\rho) \leq \varepsilon \leq 1/4$ and the source $\mathcal{R}_n$ is $(\log(\frac{\rho\varepsilon}{\delta}) + 4, \delta)-$expressive, for some $2\varepsilon \leq \delta \leq 1$. Then no $(\mathcal{R}_n,\varepsilon)-$differentially private and*

(U_n, ρ)-accurate mechanism for the Hamming weight queries exists. In particular, plugging $\delta = 2\varepsilon$ and $\delta = \frac{1}{2}$, respectively, this holds if either
(a) $\mathcal{R}_n$ is $(3 + \log(\rho), 2\varepsilon)-$expressive; or (b) $\mathcal{R}_n$ is $(5 + \log(\rho\varepsilon), \frac{1}{2})-$expressive.

The high-level idea is as follows. For two databases D and D', define two functions $f(\mathbf{r}) \stackrel{def}{=} M(D, \mathbf{r})$ and $g(\mathbf{r}) \stackrel{def}{=} M(D', \mathbf{r})$. Intuitively, for all $R \in \mathcal{R}_n$, since $\mathsf{RD}(f(R), g(R)) \leq \varepsilon \cdot \Delta(D, D')$ implies $\mathsf{SD}(f(R), g(R)) \leq e^{\varepsilon \cdot \Delta(D,D')} - 1$, we could use expressiveness to argue that $f(\mathbf{r}) = g(\mathbf{r})$ almost everywhere, which must eventually contradict utility (even for uniform distribution). However, we can't use this technique directly, because if $\varepsilon \cdot \Delta(D, D')$ is large enough, then $e^{\varepsilon \cdot \Delta(D,D')} - 1 > 1$, which is greater than the general upper bound 1 of the statistical distance. Instead, we simply use this trick on close-enough databases D and D', and then use a few "jumps" from D_0 to D_1, etc., until eventually we must violate the ρ-utility.

Proof. Assume for contradiction that there exists such a mechanism M. Let $\mathcal{D}' \stackrel{def}{=} \{D \mid wt(D) \leq 4\rho\}$. Denote

$$\mathsf{Trunc}(x) \stackrel{def}{=} \begin{cases} 0, & if\ x < 0; \\ x, & if\ x \in \{0, 1, \ldots, 4\rho\}; \\ 4\rho, & otherwise. \end{cases}$$

For any $D \in \mathcal{D}'$, define the truncated mechanism $M' \stackrel{def}{=} \mathsf{Trunc}(M)$ by $M'(D, \mathbf{r}) \stackrel{def}{=} \mathsf{Trunc}(M(D, \mathbf{r}))$. Since for every $D \in \mathcal{D}'$, we have $wt(D) \in \{0, 1, \ldots, 4\rho\}$, M' still has $(U_n, \rho)-$utility on $\mathcal{D}'$. Additionally, from Definition 6, it's straightforward that M' is $(\mathcal{R}_n, \varepsilon)$-differentially private on $\mathcal{D}'$. In the following, we only consider the truncated mechanism M' on $\mathcal{D}'$.

Let $t = \log(\frac{\rho\varepsilon}{\delta}) + 4$ and $s = \frac{\delta}{2\varepsilon}$. Notice, $1 \leq s \leq 1/(2\varepsilon) \leq 4\rho$, $e^{\varepsilon s} - 1 < \delta$, and $2^t = 8\rho/s$.

We start with the following claim:

Claim. Consider any databases $D, D' \in \mathcal{D}'$, s.t. $\Delta(D, D') \leq s$, and denote $f(\mathbf{r}) \stackrel{def}{=} M'(D, \mathbf{r})$ and $g(\mathbf{r}) \stackrel{def}{=} M'(D', \mathbf{r})$. Then $\Pr_{\mathbf{r} \leftarrow U_n}[f(\mathbf{r}) \neq g(\mathbf{r})] < \frac{1}{2^t}$.

Proof. Since M' is $(\mathcal{R}_n, \varepsilon)$-differentially private, then for all $R \in \mathcal{R}_n$, we have $\mathsf{RD}(f(R), g(R)) \leq \varepsilon \cdot \Delta(D, D') \leq \varepsilon \cdot s$. Hence, $\mathsf{SD}(f(R), g(R)) \leq e^{\varepsilon \cdot s} - 1 < \delta$, by our choice of s. Since this holds for all $R \in \mathcal{R}_n$ and $\mathcal{R}_n$ is $(t, \delta)-$expressive, we conclude that it must be the case that $\Pr_{\mathbf{r} \leftarrow U_n}[f(\mathbf{r}) \neq g(\mathbf{r})] < \frac{1}{2^t}$. □

Coming back to the main proof, consider a sequence of databases $D_0, D_1, \cdots, D_{4\rho/s}$ such that $wt(D_i) = i \cdot s$ and $\Delta(D_i, D_{i+1}) = s$. Denote $f_i(R) \stackrel{def}{=} M'(D_i, R)$ for all $i \in \{0, 1, \ldots, 4\rho/s\}$. From the above Claim, we get that $\Pr_{\mathbf{r} \leftarrow U_n}[f_i(\mathbf{r}) \neq f_{i+1}(\mathbf{r})] < \frac{1}{2^t}$. By the union bound and our choice of s and t,

$$\Pr_{\mathbf{r} \leftarrow U_n}[f_0(\mathbf{r}) \neq f_{4\rho/s}(\mathbf{r})] < \frac{4\rho}{2^t \cdot s} \leq \frac{1}{2} \tag{1}$$

Let $\alpha \stackrel{def}{=} \mathbb{E}_{\mathbf{r}\leftarrow U_n}[\ f_{4\rho/s}(\mathbf{r}) - f_0(\mathbf{r})\ ]$. From (U_n, ρ)-utility, we get that

$$\alpha \geq (wt(D_{4\rho/s}) - \rho) - (wt(D_0) + \rho) = (4\rho - \rho) - (0 + \rho) = 2\rho.$$

On the other hand, from Inequation (1),

$$\alpha \leq \Pr_{\mathbf{r}\leftarrow U_n}[f_0(\mathbf{r}) \neq f_{4\rho/s}(\mathbf{r})] \cdot \max_{\mathbf{r}} |f_{4\rho/s}(\mathbf{r}) - f_0(\mathbf{r})| < \frac{1}{2} \cdot 4\rho = 2\rho,$$

which is a contradiction.

□

4 Separability and its Implications

Expressiveness is a powerful tool, but it's hard for us to use it directly. In this section, we introduce the concept of separability and show that it implies expressiveness, and also has its own applications to (weak) coin flipping. Several typical examples can been seen in Section 5.

Intuitively, separable sources $\mathcal{R}_n$ allow one to choose a distribution $R \in \mathcal{R}_n$ capable of "separating" any sufficiently large, disjoint sets G and B: increasing a relative weight of one set w.r.t. R without doing the same for the counterpart of the other one.

Definition 8. *We say that a source $\mathcal{R}_n$ is $(t, \delta)-$separable if for all $G, B \subseteq \{0,1\}^n$, where $G \cap B = \emptyset$ and $|G \cup B| \geq 2^{n-t}$, there exists a distribution $R \in \mathcal{R}_n$ such that $|\Pr_{\mathbf{r}\leftarrow R}[\mathbf{r} \in G] - \Pr_{\mathbf{r}\leftarrow R}[\mathbf{r} \in B]| \geq \delta$.*

4.1 Separability Implies Expressiveness

We investigate the relationship between separability and expressiveness. We show that separable sources must be expressive. The high-level idea of the proof comes from the work of [DOPS04] (who only applied it to SV sources), but we quantitatively improve the technique of [DOPS04], by making the gap between expressiveness and separability independent of the range $\mathcal{C}$ of the functions f and g. See [DY14] for the proof.

Theorem 3. *If a source $\mathcal{R}_n$ is $(t+1, \delta)-$separable, then it's (t, δ)-expressive.*

Remark 1. Note that if the universe $\mathcal{C}$ is a subset of $\{0,1\}^{poly(n)}$, then the universal hash function family in the proof of Theorem 3 can be made efficient (in n). Hence, the distinguisher Eve can be made efficient as well. Therefore, there exists an efficient distinguisher Eve such that $|\Pr_{\mathbf{r}\leftarrow R}[\mathsf{Eve}(f(\mathbf{r})) = 1] - \Pr_{\mathbf{r}\leftarrow R}[\mathsf{Eve}(g(\mathbf{r})) = 1]| \geq \delta$. Namely, $f(R)$ is "*δ-computationally distinguishable*" from $g(R)$.

Combining Theorem 3 with Theorems 1 and 2, we get

Corollary 1.

(a) *If* $\mathcal{R}_n$ *is* $(1,\delta)-$*separable, then no* $(\mathcal{R}_n,\delta)$*-secure bit extractor exists.*
(b) *If* $\mathcal{R}_n$ *is* $(1,\delta)-$*separable, then no* $(\mathcal{R}_n,\delta)$*-secure bit encryption exists.*
(c) *If* $\mathcal{R}_n$ *is* $(2,\delta)-$*separable, then no* $(\mathcal{R}_n,\delta)$*-secure weak bit commitment exists.*
(d) *If* $\mathcal{R}_n$ *is* $(\log T+1,\delta)-$*separable, then no* $(\mathcal{R}_n,\delta)$*-secure bit* T*-secret sharing exists.*
(e) *Assume* $1/(8\rho)\le\varepsilon\le 1/4$ *and* $\mathcal{R}_n$ *is* $(\log(\frac{\rho\varepsilon}{\delta})+5,\delta)-$*separable, for some* $2\varepsilon\le\delta\le 1$*. Then no* $(\mathcal{R}_n,\varepsilon)-$*differentially private and* (U_n,ρ)*-accurate mechanism for the Hamming weight queries exists. In particular, plugging* $\delta=2\varepsilon$ *and* $\delta=\frac{1}{2}$*, respectively, this holds if either (e.1)* $\mathcal{R}_n$ *is* $(4+\log(\rho),2\varepsilon)-$*separable; or (e.2)* $\mathcal{R}_n$ *is* $(6+\log(\rho\varepsilon),\frac{1}{2})-$*separable.*

The above results are illustrated by several typical sources in Section 5.

4.2 Separability and Weak Bit Extraction

In this section, we define weak bit extraction and show that weak bit extraction is equivalent to NON-separability. Then we propose its implications to privacy.

Recall, Bosley and Dodis [BD07] initiated the study of the general question: *does privacy inherently require "extractable" source of randomness*? A bit more formally, if a primitive P admits $(\mathcal{R}_n,\delta)$-secure implementation, does it mean one can construct a (deterministic, single- or multi-) bit extractor from $\mathcal{R}_n$?

They also obtained very strong affirmative answers to this question for several traditional privacy primitives, including (only multi-bit) encryption and commitment (but not secret sharing, for example). Here we make the observation that our impossibility results give an incomparable (to [BD07]) set of affirmative answers to this question. On the positive, our results apply to a much wider set of primitives P (e.g., secret-sharing, as well as even single-bit encryption and commitment). On the negative, we can only argue a rather weak kind of single-bit extraction (as opposed to [BD07], who showed traditional, and possibly multi-bit extraction). Our weak notion of extraction is defined below.

Definition 9. *We say that* $\mathsf{Ext}:\{0,1\}^n\to\{0,1,\perp\}$ *is* $(\mathcal{R}_n,\delta,\tau)$*-secure weak bit extractor if*

(a) *for every distribution* $R\in\mathcal{R}_n$, $|\Pr_{\mathbf{r}\leftarrow R}[\mathsf{Ext}(\mathbf{r})=1]-\Pr_{\mathbf{r}\leftarrow R}[\mathsf{Ext}(\mathbf{r})=0]|<\delta$;
(b) $\Pr_{\mathbf{r}\leftarrow U_n}[\mathsf{Ext}(\mathbf{r})\neq\perp]\ge\tau$.

We briefly discuss this notion, before showing our results. First, we notice that setting $\tau=1$ recovers the notion of traditional bit-extractor given in Definition 2. And, even for general $\tau<1$, the odds of outputting 0 or 1 are roughly the same, for any distribution R in the source. However, now the extractor is also allowed to output a failure symbol $\perp$, which means that each of the above two probabilities can occur with probabilities noticeably smaller than $1/2$. Hence, to make it interesting, we also add the requirement that Ext does not output $\perp$ all the time. This is governed by the second parameter τ requiring that $\Pr_{\mathbf{r}\leftarrow R}[\mathsf{Ext}(\mathbf{r})\neq\perp]\ge\tau$. Ideally, we would like this to be true for any distribution R in the source.

Unfortunately, such a desirable guarantee will not be achievable in our setting (see Remark 2). Thus, to salvage a meaningful and realizable notion, we will only require that this non-triviality guarantee at least holds for $R \equiv U_n$. Namely, while we do not rule out the possibility that some particular distributions R might force Ext to fail the extraction with high probability, we still ensure that: (a) when the extraction succeeds, the extracted bit is unbiased for *any* R in the source; (b) the extraction succeeds with noticeable probability at least when R is ("close to") the uniform distribution U_n.

We now observe (and prove in [DY14]) that the notion of weak bit-extraction is simply a different way to express (the negation of) our notion of separability!

Lemma 1. *$\mathcal{R}_n$ has a $(\mathcal{R}_n, \delta, 2^{-t})$-secure weak bit extractor if and only if $\mathcal{R}_n$ is <u>not</u> (t, δ)-separable.*

Combining Lemma 1 with the counter-positive of Corollary 1, we get

Theorem 4.

(a) If $(\mathcal{R}_n, \delta)$-secure bit encryption scheme exists, then $(\mathcal{R}_n, \delta, \frac{1}{2})$-secure weak bit-extraction exists.

(b) If $(\mathcal{R}_n, \delta)$-secure weak bit commitment exists, then $(\mathcal{R}_n, \delta, \frac{1}{4})$-secure weak bit extraction exists.

(c) If $(\mathcal{R}_n, \delta)$-secure bit T-secret-sharing exists, then $(\mathcal{R}_n, \delta, \frac{1}{2T})$-secure weak bit extraction exists.

(d) If $(\mathcal{R}_n, \varepsilon)-$differentially private and (U_n, ρ)-accurate mechanism for the Hamming weight queries exists, then $(\mathcal{R}_n, 2\varepsilon, \frac{1}{16\rho})$-secure weak bit extraction exists.

It is also instructive to see the explicit form of our weak bit extractor. For example, in the case of bit encryption (part (a), other examples similar), we get

$$\mathsf{Ext}(\mathbf{r}) \stackrel{def}{=} \begin{cases} 1, & \text{if } h^*(\mathsf{Enc}_{\mathbf{r}}(1)) = 1 \text{ and } h^*(\mathsf{Enc}_{\mathbf{r}}(0)) = 0, \\ 0, & \text{if } h^*(\mathsf{Enc}_{\mathbf{r}}(1)) = 0 \text{ and } h^*(\mathsf{Enc}_{\mathbf{r}}(0)) = 1, \\ \perp, & \text{otherwise (i.e., if } h^*(\mathsf{Enc}_{\mathbf{r}}(1)) = h^*(\mathsf{Enc}_{\mathbf{r}}(0))), \end{cases}$$

where h^* is the boolean universal hash function from the proof of Theorem 3, chosen as to ensure $\Pr_{\mathbf{r} \leftarrow U_n}[\mathsf{Ext}(\mathbf{r}) \neq \perp] = \Pr_{\mathbf{r} \leftarrow U_n}[h^*(\mathsf{Enc}_{\mathbf{r}}(0)) \neq h^*(\mathsf{Enc}_{\mathbf{r}}(1))] \geq \frac{1}{2}$. When the bit encryption (resp. commitment, secret sharing, DP mechanism) is computationally efficient (in n), our bit extractor is efficient too. This means that even computationally secure analogs of encryption (commitment, secret sharing, DP mechanism) imply efficient, statistically secure weak bit extraction.

Remark 2. As we mentioned, the major weakness of our weak bit extraction definition comes from the fact that the non-triviality condition $\Pr_{\mathbf{r} \leftarrow R}[\mathsf{Ext}(\mathbf{r}) \neq \perp] \geq \tau$ is only required for $R \equiv U_n$. Unfortunately, we observe that the analog of Theorem 4.(a)-(c) is no longer true if we require the extraction non-triviality to hold for all $R \in \mathcal{R}_n$. Indeed, this stronger notion of $(\mathcal{R}_n, \delta, \tau)$-secure weak

bit extraction clearly implies traditional $(\mathcal{R}_n, 1+\delta-\tau)$-secure bit extraction (by mapping $\perp$ to 1). On the other hand, Dodis and Spencer [DS02] gave an example of a source $\mathcal{R}_n$ for which, for any $\varepsilon > 0$, there exists $(\mathcal{R}_n, \varepsilon)$-secure bit encryption (and hence, weak commitment and 2-secret sharing) scheme, but no $(\mathcal{R}_n, 1-2^{1-n/2})$-secure bit-extraction. Thus, the only analogs of Theorem 4.(a)-(c) we could hope to prove using the strengthened notion of weak bit extraction would have to satisfy $\tau \leq \delta + 2^{1-n/2}$, which is not a very interesting weak bit extraction scheme (e.g., if δ is "negligible", then the extraction succeeds with "negligible" probability as well). [9]

5 Privacy with Several Typical Imperfect Sources

Now we define several imperfect sources $\mathcal{R}_n$: the $(k,n)-$source [CG88], n-bit (k,m)-block source [CG88], n-bit γ-Santha-Vazirani (SV) source [SV86], and (γ, b, n)-Bias-Control Limited (BCL) source [Dod01] below. Then we prove all these sources are separable. Based on this result, we show they are all expressive. Afterwards, we study the impossibility of traditional and differential privacy with weak, block and BCL sources, and explain why the SV source does not work. Finally, we compare the impossibility of traditional and differential privacy.

Definition 10. *The (k,n)-source (or n-bit weak source with min-entropy at least k) is defined by $\mathcal{W}eak(k,n) \stackrel{def}{=} \{R \mid \mathbf{H}_\infty(R) \geq k$, where R is over $\{0,1\}^n\}$.*

Block sources are generalizations of weak sources, allowing n/m blocks $R_1, \ldots, R_{n/m}$ each having k fresh bits of entropy.[10]

Definition 11. *Let m divide n, and $R_1, \ldots, R_{n/m}$ be a sequence of Boolean random variables over $\{0,1\}^m$. A probability distribution $R = (R_1, \ldots, R_{n/m})$ over $\{0,1\}^n$ is an n-bit (k,m)-block distribution, denoted by $\mathcal{B}lock(k,m,n)$, if for all $i \in [n/m]$ and for every $s_1, \ldots, s_{i-1} \in \{0,1\}^m$, we have*

$$\mathbf{H}_\infty(R_i \mid R_1 \ldots R_{i-1} = s_1 \ldots s_{i-1}) \geq k.$$

We define the n-bit (k,m)-block source $\mathcal{B}lock(k,m,n)$ to be the set of all n-bit (k,m)-block distributions.

Hence, weak sources correspond to $m = n$ (i.e., one block). From the other extreme, SV sources as shown in Definition 12 correspond to 1-bit blocks (i.e., $m = 1$). In this case, it is customary to express the imperfectness of the source as the function of its "bias" γ instead of min-entropy k. Of course, for 1-bit random variables bias and min-entropy are related by $2^{-k} = (1+\gamma)/2$.

[9] For differential privacy (part (d)), we do not have an analog of the counter-example in [DS02], and anyway the value $\tau = O(1/\rho) \ll \delta = O(\varepsilon)$ (so no contradiction). Of course, this does not imply that a stronger bit extraction result should be true; only that it is not definitely false.

[10] For consistency with prior work, we only assume that R_i has k fresh bits conditioned on the prior blocks, but our impossibility results easily extend to the case when we condition on both the past and the future blocks.

Definition 12. *Let $r_1, \ldots, r_n$ be a sequence of Boolean random variables and $0 \leq \gamma < 1$. A probability distribution $R = (r_1, \ldots, r_n)$ over $\{0,1\}^n$ is an n-bit γ-Santha-Vazirani distribution, denoted by $SV(\gamma, n)$, if for all $i \in \{1, \ldots, n\}$ and every string $s \in \{0,1\}^{i-1}$, $\frac{1-\gamma}{2} \leq \Pr[r_i = 1 \mid r_1 \ldots r_{i-1} = s] \leq \frac{1+\gamma}{2}$ holds. We define the n-bit γ-SV source $\mathcal{SV}(\gamma, n)$ to be the set of all n-bit γ-SV distributions.*

Finally, we define BCL sources [Dod01].

Definition 13. *Assume that $0 \leq \gamma < 1$. The (γ, b, n)-Bias-Control Limited (BCL) source $\mathcal{BCL}(\gamma, b, n)$ generates n bits $r_1, \ldots, r_n$, where for all $i \in \{1, \ldots, n\}$, the value of r_i can depend on $r_1, \ldots, r_{i-1}$ in one of the following two ways:*
(a) r_i is determined by $r_1, \ldots, r_{i-1}$, but this can happen for at most b bits. This rule of determining a bit is called an intervention.
(b) $\frac{1-\gamma}{2} \leq \Pr[r_i = 1 \mid r_1 r_2 \ldots r_{i-1}] \leq \frac{1+\gamma}{2}$.

Every distribution over $\{0,1\}^n$ generated from $\mathcal{BCL}(\gamma, b, n)$ is called a (γ, b, n)-BCL distribution $BCL(\gamma, b, n)$.

In particular, if $b = 0$, $\mathcal{BCL}(\gamma, b, n)$ degenerates into $\mathcal{SV}(\gamma, n)$ [SV86]; if $\gamma = 0$, it yields the sequential-bit-fixing source of Lichtenstein, Linial, and Saks [LLS89].

5.1 Separability Results

In the following, we propose that the above sources are separable. It should be noted that: (a) The results for the weak and SV sources are implicitly known; (b) The BCL source was not considered before, but it is not hard to prove its separability given careful application of prior work; (c) The separability of the block source is new. It was not considered before because the SV source is a block source with each block of length 1, and [MP90, DOPS04] showed traditional privacy impossible even with the SV source (hence with the block source). But in light of [DLMV12], where differential privacy is possible with the SV source, we find it important to precisely figure out the separability of the block source. A naive approach would be to employ the so called γ-biased half-space source (see [DY14]), introduced by [RVW04] and [DOPS04], which is both γ-SV and $(m - \log \frac{1+\gamma}{1-\gamma}, m)$-block sources. We can easily conclude that (1) $\mathcal{SV}(\gamma, n)$ is $(t, \frac{\gamma}{2^{t+1}})$−separable, and (2) $\mathcal{B}lock(k, m, n)$ is $(t, \frac{2^{m-k}-1}{2^{t+1} \cdot (2^{m-k}+1)})$−separable. However, these results are somewhat sub-optimal. Instead, we introduce a new separability bound for block sources in Lemma 2 (b), and use it to get an improved result about the SV sources as well (see [DY14] for the proof).

Lemma 2.
(a) Assume that $k \leq n-1$. Then $\mathcal{W}eak(k, n)$ is $(t, 1)$−separable when $k \leq n - t - 1$, and $(t, 2^{n-t-k-1})$−separable when $n - t - 1 < k \leq n - 1$. In particular, it's $(t, \frac{1}{2})$−separable when $k \leq n - t$.
(b) $\mathcal{B}lock(k, m, n)$ is $\left(t, \frac{1}{1+2^{t+1} \cdot \left(\frac{2^k-1}{2^m-2^k}\right)}\right)$ −separable. In particular, it is $(t, 1/(1+2^{2+t+k-m}))$−separable when $k \leq m - 1$ (and, hence, $(t, \frac{1}{2})$−separable when $k \leq m - t - 2$).

(c) $\mathcal{SV}(\gamma, n)$ *is* $(t, \frac{\gamma}{2^t})-$*separable.*

(d) $\mathcal{BCL}(\gamma, b, n)$ *is* $(t, 1 - \frac{2^{t+2}}{(1+\gamma)^b})-$*separable. In particular, it is* $(t, \frac{1}{2})-$*separable for* $b \geq \frac{t+3}{\log(1+\gamma)} = \Theta(\frac{t+1}{\gamma})$.

5.2 Implications to Traditional and Differential Privacy

Impossibility of Traditional Privacy. From Lemma 2 and Corollary 1 (a)-(d), we conclude:

Theorem 5. *For the following values of* δ*, shown in Table 1, no* $(\mathcal{R}_n, \delta)-$*secure cryptographic primitive* P *exists, where* $\mathcal{R}_n \in \{\mathcal{W}eak(k, n), \mathcal{B}lock(m-1, m, n), \mathcal{SV}(\gamma, n), \mathcal{BCL}(\gamma, b, n)\}$ *and* $P \in \{$*bit extractor, bit encryption scheme, weak bit commitment, bit* T*-secret sharing*$\}$.

$\mathcal{R}_n$ \ P	bit extractor	bit encryption scheme	weak bit commitment	bit T-secret sharing
$\mathcal{W}eak(k, n)$	1, if $k \leq n-2$	1, if $k \leq n-2$	1, if $k \leq n-3$	1, if $k \leq n - \log T - 2$
$\mathcal{W}eak(n-1, n)$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{2T}$
$\mathcal{B}lock(m-1, m, n)$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{9}$	$\frac{1}{4T+1}$
$\mathcal{SV}(\gamma, n)$	$\frac{\gamma}{2}$	$\frac{\gamma}{2}$	$\frac{\gamma}{4}$	$\frac{\gamma}{2T}$
$\mathcal{BCL}(\gamma, b, n)$	$\frac{1}{2}$, if $b \geq \frac{4}{\log(1+\gamma)}$	$\frac{1}{2}$, if $b \geq \frac{4}{\log(1+\gamma)}$	$\frac{1}{2}$, if $b \geq \frac{5}{\log(1+\gamma)}$	$\frac{1}{2}$, if $b \geq \frac{\log T+4}{\log(1+\gamma)}$

Table 1: Values of δ for which no $(\mathcal{R}_n, \delta)-$secure cryptographic primitive P exists.

We notice that, while the impossibility results for the block and BCL sources are new, the prior work of [MP90, DOPS04] already obtained similar results for the weak and SV sources. However, our results still offer some improvements over the works of [MP90, DOPS04]. First, unlike the work of [MP90], our distinguisher is efficient (see Remark 1), ruling out even computationally secure encryption, commitment, and secret sharing schemes. Second, unlike the work of [DOPS04], our lower bound on δ does not depend on the sizes of ciphertext/commitment/shares. In particular, while [DOPS04] used a bit-by-bit hybrid argument to show their impossibility results, our proof of Theorem 3 used a more clever "universal hashing trick". More importantly, instead of focusing the entire proof on some specific weak/block/SV sources [MP90, DOPS04], our impossibility results for such sources were obtained in a more modular manner, making these proofs somewhat more illuminating.

Impossibility of Differential Privacy with the Weak, Block and BCL sources. Now we apply the impossibility results of differential privacy to the sources $\mathcal{W}eak(k, n)$, $\mathcal{B}lock(k, m, n)$, and $\mathcal{BCL}(\gamma, b, n)$. In particular, by combining Corollary 1 (e.2) with Lemma 2 (a), (b), and (d), respectively, we get

Theorem 6. *For the following sources* $\mathcal{R}_n$*, no* $(\mathcal{R}_n, \varepsilon)-$*differentially private and* (U_n, ρ)*-accurate mechanisms for the Hamming weight queries exist:*

(a) $\mathcal{W}eak(k,n)$ *where* $k \leq n - \log(\varepsilon\rho) - 6$*;*
(b) $\mathcal{B}lock(k,m,n)$ *where* $k \leq m - \log(\varepsilon\rho) - 8$*;*
(c) $\mathcal{BCL}(\gamma,b,n)$ *where* $b \geq \frac{\log(\varepsilon\rho)+9}{\log(1+\gamma)} = \Omega(\frac{\log(\varepsilon\rho)+1}{\gamma})$.

We discuss the (non-)implications to the SV source below, but notice the strength of these negative results the moment the source becomes a little bit more "adversarial" as compared to the SV source. In particular, useful mechanisms in differential privacy (called "non-trivial" by [DLMV12]) aim to achieve utility ρ (with respect to the uniform distribution) which only depends on the differential privacy ε, and not on the size N of the database D. This means that the value $\log(\varepsilon\rho)$ is typically upper bounded by some constant $c = O(1)$. For such "non-trivial" mechanisms, our negative results say that differential privacy is impossible with (1) weak sources even when the min-entropy $k = n - O(1)$; (2) block sources even when the min-entropy $k = m - O(1)$; (3) BCL sources even when the number of interventions $b = \Omega(1)$. So what prevented us from strong impossibility for the SV sources, as is expected given the feasibility results of [DLMV12]? The short answer is that the separability of the SV sources given by Lemma 2 (c) is just not good enough to yield very strong results. We explain it in more detail in [DY14].

5.3 Comparing Impossibility Results for Traditional and Differential Privacy

In this section, we compare the impossibility of traditional privacy and differential privacy (see Table 2). For traditional privacy, we consider bit extractor, bit encryption scheme, weak bit commitment, and bit T-secret sharing (i.e., set $T = 2$ for concreteness). We observe that the impossibility results for differential privacy are only marginally weaker than those for traditional privacy.

Source	Traditional Privacy δ	Differential Privacy ε & Utility ρ
$\mathcal{B}lock(k,m,n)$	Impossible if $\delta \leq \frac{1}{9}$, even if $k = m - 1$	Impossible if $k \leq m - \log(\varepsilon\rho) - O(1)$
$\mathcal{SV}(\gamma,n)$	Impossible if $\delta = O(\gamma)$	Impossible if $\rho = O(\frac{1}{\varepsilon})$, even for U_n (Possible if $\rho = poly_{1/(1-\gamma)}(\frac{1}{\varepsilon}) \gg \frac{1}{\varepsilon}$)
$\mathcal{BCL}(\gamma,b,n)$	Impossible if $\delta = O(\gamma)$, even if $b = 0$; Impossible if $\delta \leq \frac{1}{2}$ and $b = \Omega(\frac{1}{\gamma})$	Impossible if $b = \Omega(\frac{\log(\varepsilon\rho)+1}{\gamma})$

Table 2: Comparison about the Impossibility of Traditional Privacy and Differential Privacy.

In particular, while a very "structured" (and, hence, rather unrealistic) SV source is sufficient to guarantee loose, but non-trivial differential privacy, without guaranteeing (strong-enough) traditional privacy, once the source becomes more realistic (e.g., number of interventions b becomes super-constant, or one removes the conditional entropy guarantee within different blocks), both notions of privacy become impossible *extremely quickly*. In other words, despite the surprising feasibility result of [DLMV12] regarding differential privacy with SV sources, the prevalent opinion that "privacy is impossible with realistic weak randomness" appears to be rather accurate.

Acknowledgments. The authors would like to thank Benjamin Fuller, Sasha Golovnev, Hamidreza Jahanjou, Zhoujun Li, Umut Orhan, and Abhishek Samanta. In particular, the authors thank Prof. Zhoujun Li very much for his great help about this paper. The authors also thank the anonymous reviewers for their helpful comments. Yevgeniy Dodis was partially supported by gifts from VMware Labs and Google, and NSF grants 1319051, 1314568, 1065288, and 1017471. Yanqing Yao was supported by NSFC grants 61170189 and 61370126, the Fund for the Doctoral Program of Higher Education of China 20111102130003, the Scholarship Award for Excellent Doctoral Student granted by Ministry of Education 400618, and CSC grant 201206020063.

References

[ACM+14] Austrin, P., Chung, K.M., Mahmoody, M., Pass, R., Seth, K.: On the Impossibility of Cryptography with Tamperable Randomness. CRYPTO 2014, LNCS 8616, 462-479 (2014)

[ACRT99] Andreev, A.E., Clementi, A.E.F., Rolim, J.D.P., Trevisan, L.: Weak random sources, hitting sets, and BPP simulations. SIAM J. Comput. 28(6), 2103-2116 (1999)

[Blu86] Blum, M.: Independent unbiased coin-flips from a correlated biased source-a finite state Markov chain. Combinatorica 6(2), 97-108 (1986)

[BD07] Bosley, C., Dodis, Y.: Does privacy require true randomness? TCC 2007, LNCS 4392, 1-20 (2007)

[BDK+05] Boyen, X., Dodis, Y., Katz, J., Ostrovsky, R., Smith, A.: Secure remote authentication using biometric data. EUROCRYPT 2005, LNCS 3494, 147-163 (2005)

[BH05] Barak, B., Halevi, S.: A model and architecture for pseudo-random generation with applications to /dev/random. Proceedings of the 12th ACM conference on Computer and communications security, 203-212 (2005)

[BST03] Barak, B., Shaltiel, R., Tromer, E.: True random number generators secure in a changing environment. In Proceedings of the 5th Cryptographic Hardware and Embedded Systems, LNCS 2779, 166-180 (2003)

[CFG+85] Chor, B., Friedman, J., Goldreich, O., Håstad, J., Rudich, S., Smolensky, R.: The Bit Extraction Problem or t-resilient Functions. In the 26th Annual Symposium on Foundations of Computer Science, 396-407 (1985)

[CG88] Chor, B., Goldreich, O.: Unbiased bits from sources of weak randomness and probabilistic communication complexity. SIAM J. Comput. 17(2), 230-261 (1988)

[CW79] Carter, J.L., Wegman, M.N.: Universal Classes of Hash Functions. J. Comput. Syst. Sci. 18(2), 143-154 (1979)

[Dod01] Dodis, Y.: New Imperfect Random Source with Applications to Coin-Flipping. ICALP 2001, LNCS 2076, 297-309 (2001)

[DKRS06] Dodis, Y., Katz, J., Reyzin, L., Smith, A.: Robust fuzzy extractors and authenticated key agreement from close secrets. CRYPTO 2006, LNCS 4117, 232-250 (2006)

[DLMV12] Dodis, Y., López-Alt, A., Mironov, I., Vadhan, S.P.: Differential Privacy with Imperfect Randomness. CRYPTO 2012, LNCS 7417, 497-516 (2012)

[DMNS06] Dwork, C., McSherry, F., Nissim, K., Smith, A.: Calibrating noise to sensitivity in private data analysis. TCC 2006, LNCS 3876, 265-284 (2006)

[DOPS04] Dodis, Y., Ong, S.J., Prabhakaran, M., Sahai, A.: On the (im)possibility of cryptography with imperfect randomness. 45th Symposium on Foundations of Computer Science, 196-205 (2004)

[DORS08] Dodis, Y., Ostrovsky, R., Reyzin, L., Smith, A.: Fuzzy extractors: How to generate strong keys from biometrics and other noisy data. SIAM J. Comput. 38(1), 97-139 (2008)

[DS02] Dodis, Y., Spencer, J.: On the (non)Universality of the One-Time Pad. 43rd Symposium on Foundations of Computer Science, 376-385 (2002)

[DY14] Dodis, Y., Yao, Y.Q.: Privacy with Imperfect Randomness. IACR Cryptology ePrint Archive 2014: 623 (2014)

[GKR04] Gennaro, R., Krawczyk, H., Rabin, T.: Secure hashed diffie-hellman over non-ddh groups. In Christian Cachin and Jan Camenisch, editors, EUROCRYPT 2004, LNCS 3027, 361-381 (2004)

[Kra10] Krawczyk, H.: Cryptographic Extraction and Key Derivation: The HKDF Scheme. CRYPTO 2010, LNCS 6223, 631-648 (2010)

[LLS89] Lichtenstein, D., Linial, N., Saks, M.E.: Some extremal problems arising form discrete control processes. Combinatorica 9(3), 269-287 (1989)

[MP90] McInnes, J.L., Pinkas, B.: On the impossibility of private key cryptography with weakly random keys. CRYPTO 1990, LNCS 537, 421-435 (1991)

[MW97] Maurer, U.M., Wolf, S.: Privacy amplification secure against active adversaries. CRYPTO 1997, LNCS 1294, 307-321 (1997)

[RVW04] Reingold, O., Vadhan, S., Widgerson, A.: No Deterministic Extraction from Santha-Vazirani Sources: a Simple Proof. http://windowsontheory.org/2012/02/21/nodeterministic-extraction-from-santha-vazirani-sources-a-simple-proof/ (2004)

[SV86] Santha, M., Vazirani, U.V.: Generating quasi-random sequences from semi-random sources. J. Comput. Syst. Sci. 33(1), 75-87 (1986)

[vN51] Neumann, J.v.: Various techniques used in connection with random digits. In National Bureau of Standards, Applied Math. Series 12, 36-38 (1951)

[VV85] Vazirani, U.V., Vazirani, V.V.: Random polynomial time is equal to slightly random polynomial time. 26th Annual Symposium on Foundations of Computer Science, 417-428 (1985)

[Zuc96] Zuckerman, D.: Simulating BPP using a general weak random source. Algorithmica 16(4/5), 367-391 (1996)