Privacy and Imperfect Randomness

Yevgeniy Dodis[a] Yanqing Yao [b,a]

[a]Department of Computer Science, New York University, New York 10012, USA
[b]State Key Laboratory of Software Development Environment, Beihang University, Beijing 100191, China
dodis@cs.nyu.edu,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]. In fact, they are true even if the imperfect source is modeled as a seemingly very "nice and friendly" Santha-Vazirani (SV) source. The SV source outputs a sequence of bits $r_1, r_2, \ldots$, where each r_i has almost 1 full bit of fresh entropy conditioned on the previous bits $r_1, \ldots, r_{i-1}$. Moreover, Bosley and Dodis [BD07] gave strong evidence that many traditional privacy tasks (e.g., encryption) inherently require an "extractable" source of randomness.

The common interpretation of these negative results is that traditional privacy is impossible even with "very structured" imperfect sources. Somewhat surprisingly, Dodis et al. [DLMV12] put a slight dent in this belief, by showing that non-trivial *differential* privacy is possible with SV sources. This suggested a qualitative gap between traditional and differential privacy, and left open the question if differential privacy is possible with more realistic (i.e., less structured) sources than the SV sources. Motivated by solving this question, we abstract and generalize prior techniques for showing impossibility results for achieving privacy with various imperfect sources of randomness. In particular, we introduce the concepts of separability and expressivity of a given imperfect source as a measure of its "imperfectness", and show the following results:

- Separability implies expressivity;
- Low levels of expressivity (and, thus, separability) generically imply strong impossibility results for both traditional and *differential* privacy;
- Existing (and quantitatively improved!) impossibility results for traditional privacy with respect to known imperfect sources easily follow as corollaries of our unified framework; New results follow equally easily.
- Although, unsurprisingly, our new impossibility results for differential privacy (barely) miss the highly structured SV sources, they come back *extremely quickly* once the source becomes slightly more realistic. E.g., if a small number of bits r_i can be fully determined by the previous bits;
- 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].)

Overall, our results unify and strengthen the belief that, for the most part, privacy with imperfect randomness is impossible, unless the source is (almost) deterministically extractable.

Keywords: imperfect randomness, entropy sources, Santha-Vazirani sources, Bias-Control Limited sources, randomness extraction, privacy, differential privacy

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 realistic imperfect sources have been described (e.g., [vN51,CFG+85,Blu86,SV86,CG88,LLS89,Zuc96,ACRT99,Dod01]). 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 that many realistic sources are non-extractable [SV86,CG88,Dod01]. The simplest example of such a source is the Santha-Vazirani (SV) source [SV86], which produces an infinite sequence of (possibly correlated) 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 $r_1 \ldots r_{i-1}$. However, despite the fact that each bit has almost one bit of fresh entropy, 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 perfect to imperfect (e.g., SV) 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" 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 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, if 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

[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].

$M(D;\mathbf{r})$ 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 11 and 12 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; and
(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 the value $D(i)$ 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 to work with all such distributions, provided that the utility ρ is now relaxed to be polynomial of $1/\varepsilon$, whose degree and coefficients depend on γ, but *not* on the size of the database D. Coupled with the impossibility of traditional privacy with SV-sources, this result suggested a qualitative gap between traditional and differential privacy, but left open the question if differential privacy is possible with more realistic (i.e., less structured) sources than the SV sources. Indeed, SV sources seem to be primarily interested from the perspective of negative results, 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 realistic imperfect sources $\mathcal{R}$ (e.g., SV sources), we obtain most of our results for general sources $\mathcal{R}$, but then use various realistic sources (namely, SV sources [SV86], weak sources [CG88] and bias-control limited sources [Dod01]) as specific examples illustrating our technique. In particular, we introduce the concepts of separability and expressivity of a given imperfect source $\mathcal{R}$ as a measure of its "imperfectness", and show the following results:

(1) Separability implies expressivity;
(2) Low levels of expressivity (and, thus, separability) generically imply strong impossibility results for both traditional and *differential* privacy;
(3) Existing (and quantitatively improved!) impossibility results for traditional privacy with respect to known imperfect sources easily follow as corollaries of our unified framework; New results follow equally easily.
(4) Although, unsurprisingly, our new impossibility results for differential privacy (barely) miss the highly structured SV sources, they come back *extremely quickly* once the source becomes slightly more realistic. E.g., if a small number of bits r_i can be fully determined by the previous bits;
(5) Any imperfect source allowing (either traditional or differential) privacy admits a certain type of deterministic bit extraction.

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.

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. First, we introduce notions of separability and expressivity (Result (1)). Intuitively, separability of $\mathcal{R}$ means that $\mathcal{R}$ is rich enough to

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

"separate" any sufficiently large disjoint sets G and B (where $|G| \geq |B|$; see Definition 4): there exists $R \in \mathcal{R}$ s.t. $(\Pr[R \in G] - \Pr[R \in B])$ is "noticeable". For example, if $\mathcal{R}$ only consists of the uniform distribution U, the latter is impossible when $|G| = |B|$. In contrast, all natural "non-extractable" sources are separable. This is known (or trivial to see) for the SV and general weak sources, but we show how it can be easily demonstrated for other sources as well.

In particular, we concentrate on the bias-control-limited (BCL) source of Dodis [Dod01]. BCL source generates n bits $r_1, r_2, \ldots, r_n$, where for $i = 1, 2, \ldots, n$, the value of r_i can depend on $r_1, r_2, \ldots, r_{i-1}$ in one of the following two ways: (A) r_i is determined by $r_1, r_2, \ldots, r_{i-1}$, but this happens for at most b bits, or (B) $\frac{1-\gamma}{2} \leq \Pr[r_i = 1 \mid r_1 r_2 \ldots r_{i-1}] \leq \frac{1+\gamma}{2}$. In particular, when $b = 0$, it degenerates into the γ-SV source [SV86]; when $\gamma = 0$, it yields the b-sequential-bit-fixing source of Lichtenstein, Linial and Saks [LLS89]. The BCL source models the setting that each of the bits produced by a realistic streaming source is unlikely to be perfectly random: slight errors of the source are inevitable almost always, and, rarely, some of the bits could have non-trivial dependencies on the previous bits, to the point of being completely determined by them. Hence, BCL source appears much more realistic than the SV source, especially if the number of interventions b is somewhat moderate. From our perspective, the BCL source will be especially interesting when we deal with differential privacy. Indeed, since it naturally (and realistically!) relaxes the SV source, for which non-trivial differential privacy is possible, it will be interesting to see the minimal value of b when the impossibility results come back.

Returning to our results, after showing simple separability claims for weak, SV and BCL sources (see Lemma 1), we define the notion of expressivity. Intuitively, expressivity 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 6): 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] We then show that separability generically implies expressivity, with nearly identical parameters (see Theorem 1). This is where we differ and quantitatively improve the argument from [DOPS04]: while [DOPS04] used a bit-by-bit hybrid argument to show expressivity (for SV source), our proof of Theorem 1 used a more clever "universal hashing trick".[4] This allowed 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.) As a consequence, we get simple and elegant expressivity statements for a variety of natural sources (Corollary 1).

We then use very similar technique to [DOPS04] to show that most traditional privacy tasks are impossible with any "mildly expressive" source $\mathcal{R}$ (Theorem 2 and first part of Result (2)). Applying this to specific separable/expressive source (weak, SV, BCL), we immediately derive a variety of impossibility results for traditional privacy (Table 1 and Result (3)). 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 expressivity bounds), and also allow immediate applications to other imperfect sources. E.g., we get the following new result for BCL sources: even constant security $1/2$ for traditional privacy is impossible to achieve when the number of interventions $b = \Omega(1/\gamma)$. More importantly, instead of focusing the entire proof on some specific SV/weak sources [MP90,DOPS04], our privacy impossibility results for such sources were obtained in a more modular manner, making these proofs somewhat more illuminating.

More interestingly, despite the positive result of [DLMV12] regarding the SV sources, we show that expressivity is again sufficient to rule out even *differential* privacy (Theorem 4 and second part of Result (2)). The slight catch is that the expressivity requirement on $\mathcal{R}$ for ruling out differential privacy will be slightly higher than for traditional privacy (Theorem 4 vs. Theorem 2). As a result, the impossibility results will (barely) miss the Santha-Vazirani sources. However, once we consider general weak sources, or even much more structured BCL sources with $b > 0$, 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 5), while BCL sources are ruled out the moment $b = \Omega(\log(\varepsilon\rho)/\gamma)$ (Theorem 6). 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 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

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

marginally weaker than the former. This gives us our Result (4), and the conclusion that differential privacy is still rather demanding to achieve with realistic imperfect sources of randomness.

Finally, we show that any imperfect source allowing (either traditional or differential) privacy admits a certain type of deterministic bit extraction (Result (5), formalized in Theorem 7): (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. This result is similar in spirit, but incomparable to the result of Bosley and Dodis [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.

2 Preliminaries

For a positive integer n, let $[n] \stackrel{def}{=} \{1, 2, \ldots, n\}$. For a set S, we write U_S to denote the uniform distribution over S. For simplicity, denote $U_n \stackrel{def}{=} U_{\{0,1\}^n}$. For a distribution or a random variable R, let $\mathbf{r} \leftarrow R$ denote the operation of sampling a random $\mathbf{r}$ according to R. All logarithms are to the base 2. The min-entropy of a random variable R is defined as $\mathbf{H}_\infty(R) \stackrel{def}{=} \min_{\mathbf{r} \in \mathrm{supp}(R)} \log \frac{1}{\Pr[R=\mathbf{r}]}$.

For two random variables R and R' over $\{0,1\}^n$, the statistical distance between R and R' is denoted as $\mathsf{SD}(R, R') \stackrel{def}{=} \frac{1}{2} \sum_{\mathbf{r} \in \{0,1\}^n} |\Pr[R = \mathbf{r}] - \Pr[R' = \mathbf{r}]| = \max_{\mathsf{Eve}} |\Pr[\mathsf{Eve}(R) = 1] - \Pr[\mathsf{Eve}(R') = 1]|$, where each 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 $\Pr[R = \mathbf{r}] \in [e^{-\varepsilon} \cdot \Pr[R' = \mathbf{r}], e^{\varepsilon} \cdot \Pr[R' = \mathbf{r}]]$ for all $\mathbf{r} \in \{0,1\}^n$. It's easy to see that $\mathsf{RD}(R, R') \leq \varepsilon$ implies $\mathsf{SD}(R, R') \leq e^{\varepsilon} - 1$.

We call a family of distributions over $\{0,1\}^n$ a source, denoted as $\mathcal{R}_n$. Now we define several imperfect sources $\mathcal{R}_n$: the $(n,k)-$source [CG88], γ-Santha-Vazirani source [SV86], and (γ, b, n)-Bias-Control Limited source [Dod01] as follows.

Definition 1. *The $(n,k)-$source is defined by $\mathcal{W}eak(k,n) \stackrel{def}{=} \{X \in \{0,1\}^n \mid \mathbf{H}_\infty(X) \geq k\}$.*

Definition 2. *Let $r_1, r_2, \ldots, r_n$ be a sequence of Boolean random variables and $0 \leq \gamma < 1$. A probability distribution $R = (r_1, r_2, \ldots, r_n)$ over $\{0,1\}^n$ is an n-bit γ-Santha-Vazirani (SV) distribution, denoted by $SV(\gamma, n)$, if for all $i \in [n]$ and for every string s of length $i-1$, we have*

$$\frac{1-\gamma}{2} \leq \Pr[r_i = 1 \mid r_1 r_2 \ldots r_{i-1} = s] \leq \frac{1+\gamma}{2}.$$

We define the n-bit γ-SV source $\mathcal{SV}(\gamma, n)$ to be the set of all n-bit γ-SV distributions.

Definition 3. *Assume that $0 \leq \gamma < 1$. The (γ, b, n)-Bias-Control Limited (BCL) source $\mathcal{BCL}(\gamma, b, n)$ generates n bits $r_1, r_2, \ldots, r_n$, where for all $i \in [n]$, the value of r_i can depend on $r_1, r_2, \ldots, r_{i-1}$ in one of the following two ways:*

(A) r_i is determined by $r_1, r_2, \ldots, r_{i-1}$, 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, when $b = 0$, $\mathcal{BCL}(\gamma, b, n)$ degenerates into $\mathcal{SV}(\gamma, n)$ [SV86]; when $\gamma = 0$, it yields the sequential-bit-fixing source of Lichtenstein, Linial and Saks [LLS89].

3 Separability and Expressivity of Imperfect Sources of Randomness

In this section, we introduce the concept of separability of a source. Then we prove that several weak sources (i.e., $\mathcal{W}eak(k,n)$, $\mathcal{SV}(\gamma,n)$, and $\mathcal{BCL}(\gamma,b,n)$) are separable. Afterwards, we introduce another concept called expressivity of a source. Then we investigate the relationship between separability and expressivity. Based on this result, we show that the $\mathcal{W}eak(\gamma,n)$, $\mathcal{SV}(\gamma,n)$, and $\mathcal{BCL}(\gamma,b,n)$ sources are all expressive.

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, where $|G| \geq |B|$: increasing a relative weight of a G w.r.t. R without doing the same for the counterpart B.

Definition 4. *We say that a source $\mathcal{R}_n$ is $(t,\delta)-$separable if for all $G, B \subseteq \{0,1\}^n$, where $G \cap B = \emptyset$, $|G \cup B| \geq 2^{n-t}$ and $|G| \geq |B|$, there exists a distribution $R \in \mathcal{R}_n$ such that*

$$|\Pr_{\mathbf{r}\leftarrow R}[\mathbf{r} \in G] - \Pr_{\mathbf{r}\leftarrow R}[\mathbf{r} \in B]| \geq \delta.$$

In the following, we enumerate several natural sources which are separable.

Lemma 1.
(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) $\mathcal{SV}(\gamma,n)$ is $(t, \frac{\gamma}{2^{t+1}})-$separable.
(c) $\mathcal{BCL}(\gamma,b,n)$ is $(t, 1-\frac{2^{t+2}}{(1+\gamma)^b})-$separable. In particular, it's $(t,\frac{1}{2})-$separable for $b \geq \frac{t+3}{\log(1+\gamma)} = \Theta(\frac{t+1}{\gamma})$.

Proof. Assume that $G, B \subseteq \{0,1\}^n$ where $G \cap B = \emptyset$, $|G \cup B| \geq 2^{n-t}$ and $|G| \geq |B|$. Then $|G| \geq 2^{n-t-1}$.

(a) **Case 1**: Assume that $k \leq n-t-1$. Pick any $S \subset \{0,1\}^n$ of size $|S| = 2^k$ such that $S \subseteq G$. Then

$$\Pr_{\mathbf{r}\leftarrow U_S}[\mathbf{r} \in G] - \Pr_{\mathbf{r}\leftarrow U_S}[\mathbf{r} \in B] = 1 - 0 = 1.$$

Case 2: Assume that $n-t-1 < k \leq n-1$.
Case 2.1: Suppose that $|G| \leq 2^k$. Then $|B| + 2^k \leq |G| + 2^k \leq 2^k + 2^k \leq 2^n$. Choose a set $S \subset \{0,1\}^n$ of size $|S| = 2^k$ such that $G \subseteq S$ and $B \cap S = \emptyset$. Then

$$\Pr_{\mathbf{r}\leftarrow U_S}[\mathbf{r} \in G] - \Pr_{\mathbf{r}\leftarrow U_S}[\mathbf{r} \in B] = \frac{1}{2^k} \cdot |G| - 0 \geq 2^{n-t-k-1}.$$

Case 2.2: Now assume that $|G| > 2^k$. Then pick any $S \subset \{0,1\}^n$ of size $|S| = 2^k$ such that $S \subset G$. Then

$$\Pr_{\mathbf{r}\leftarrow U_S}[\mathbf{r} \in G] - \Pr_{\mathbf{r}\leftarrow U_S}[\mathbf{r} \in B] = 1 - 0 = 1.$$

Assume that $k \leq n-t$. If $k \leq n-t-1$, it can be reduced to Case 1. Otherwise, it can be reduced to Case 2.

(b) In proving Lemma 1(b), we use a notion called the $\gamma-$biased halfspace source [DOPS04], which was implicitly defined by [RVW04].

Definition 5. *Given $S \subset \{0,1\}^n$ of size $|S| = 2^{n-1}$, and $0 \leq \gamma < 1$. The distribution $H_S(\gamma,n)$ over $\{0,1\}^n$ is defined as*

$$R \equiv H_S(\gamma,n) \stackrel{def}{=} \begin{cases} \Pr[R=\mathbf{r}] = (1+\gamma)\cdot 2^{-n}, & if\ \mathbf{r} \in S; \\ \Pr[R=\mathbf{r}] = (1-\gamma)\cdot 2^{-n}, & otherwise. \end{cases}$$

The γ-biased halfspace source $\mathcal{H}(\gamma,n)$ is defined as

$$\mathcal{H}(\gamma,n) \stackrel{def}{=} \{H_S(\gamma,n) \mid S \subseteq \{0,1\}^n \text{ and } |S| = 2^{n-1}\}.$$

Claim. ([DOPS04,RVW04]) *For any $n \in \mathbb{Z}^+$ and $0 \leq \gamma < 1$, $\mathcal{H}(\gamma,n) \subset \mathcal{SV}(\gamma,n)$.*

Therefore, we only need to choose a subset S such that $\Pr_{\mathbf{r}\leftarrow H_S(\gamma,n)}[\mathbf{r}\in G] - \Pr_{\mathbf{r}\leftarrow H_S(\gamma,n)}[\mathbf{r}\in B] \geq \frac{\gamma}{2^t}$.

Case 1: Suppose that $|G| \leq 2^{n-1}$. Then $|B| + 2^{n-1} \leq |G| + 2^{n-1} \leq 2^n$. Choose a set $S \subset \{0,1\}^n$ of size $|S| = 2^{n-1}$ such that $G \subseteq S$ and $B \cap S = \emptyset$. Then

$$\Pr_{\mathbf{r}\leftarrow H_S(\gamma,n)}[\mathbf{r}\in G] - \Pr_{\mathbf{r}\leftarrow H_S(\gamma,n)}[\mathbf{r}\in B] = \frac{1+\gamma}{2^n}\cdot|G| - \frac{1-\gamma}{2^n}\cdot|B| = \frac{|G|-|B|}{2^n} + \gamma\cdot\frac{|G|+|B|}{2^n} \geq \gamma\cdot\frac{2^{n-t}}{2^n} = \frac{\gamma}{2^t}.$$

Case 2: Now assume that $|G| > 2^{n-1}$. Pick any $S \subset \{0,1\}^n$ of size $|S| = 2^{n-1}$ such that $S \subset G$. Then $|S| = |\{0,1\}^n \setminus S| \geq |G \setminus S|$.

$$\begin{aligned}
\Pr_{\mathbf{r}\leftarrow H_S(\gamma,n)}[\mathbf{r}\in G] - \Pr_{\mathbf{r}\leftarrow H_S(\gamma,n)}[\mathbf{r}\in B] &= \frac{1+\gamma}{2^n}\cdot|S| + \frac{1-\gamma}{2^n}\cdot|G\setminus S| - \frac{1-\gamma}{2^n}\cdot|B| \\
&= \frac{1-\gamma}{2^n}\cdot\left(\frac{1+\gamma}{1-\gamma}\cdot|S| + |G\setminus S| - |B|\right) \\
&= \frac{1-\gamma}{2^n}\cdot\left(|S| + |G\setminus S| - |B| + \frac{2\gamma}{1-\gamma}\cdot|S|\right) \\
&\geq \frac{1-\gamma}{2^n}\cdot\left[(|S| + |G\setminus S|) - |B| + \frac{\gamma}{1-\gamma}\cdot(|S| + |G\setminus S|)\right] \\
&= \frac{1-\gamma}{2^n}\cdot\left[(|G| - |B|) + \frac{\gamma}{1-\gamma}\cdot|G|\right] \\
&\geq \frac{1-\gamma}{2^n}\cdot\frac{\gamma}{1-\gamma}\cdot 2^{n-t-1} \\
&= \frac{\gamma}{2^{t+1}}.
\end{aligned}$$

(c) We start by recalling the following auxiliary result from [Dod01].

Given a Boolean function $f_e : \{0,1\}^n \to \{0,1\}$, it is associated with an event E such that "E happens $\iff f_e(\mathbf{x}) = 1$", where $\mathbf{x} \in \{0,1\}^n$. The *natural probability* p of E is defined as the probability that E happens for an ideal source (in our case, emitting n perfect unbiased bits). More formally,

$$p = \Pr_{\mathbf{r}\leftarrow U_n}[f_e(\mathbf{r}) = 1] = \Pr_{\mathbf{r}\leftarrow U_n}[E \textit{ happens}].$$

We then say that E (or f_e) is p-sparse. We define the set of all p-sparse events (or Boolean functions) as $\mathcal{E}$.

We view the source $\mathcal{BCL}(\gamma, b, n)$ as an adversary $\mathcal{A}$ who can influence the ideal behavior of the source by applying rules (A) and (B) of Definition 3. Our goal is to see whether our adversary $\mathcal{A}$ has enough power to significantly influence the occurrence of the event E. For a given number of interventions b, to obtain the largest probability of "success" that $\mathcal{A}$ can achieve (i.e., the largest probability that any p-sparse event E happens for $\mathcal{BCL}(\gamma, b, n)$), we first study the complement notion of "smallest probability of failure" and get the following claim.

Claim. ([Dod01]) *Let* $F(p,n,b) \stackrel{def}{=} \max_{e\in\mathcal{E}} \min_{R\in\mathcal{BCL}(\gamma,b,n)} \Pr_{\mathbf{r}\leftarrow R}[f_e(\mathbf{r}) = 0]$. *Then* $F(p,n,b) \leq \frac{1}{p\cdot(1+\gamma)^b} = 2^{\log\frac{1}{p} - \Theta(\gamma b)}$.

In other words, if b is "high enough" (i.e., $b \gg \frac{1}{\gamma}\log\frac{1}{p}$), then the imperfect source attacker $\mathcal{A}$ can force any p-sparse event to happen with probability very close to 1.

Now let's come back to our Lemma. Define the function $f_e : \{0,1\}^n \to \{0,1\}$ as follows.

$$f_e(\mathbf{r}) = \begin{cases} 1, & if\ \mathbf{r}\in G; \\ 0, & otherwise. \end{cases}$$

Then from the above claim, we have $\min_{R\in\mathcal{BCL}(\gamma,b,n)} \Pr_{\mathbf{r}\leftarrow R}[f_e(\mathbf{r}) = 0] \leq \frac{1}{\frac{|G|}{2^n}\cdot(1+\gamma)^b}$.

Thus, there exists a $(\gamma, b, n)-$BCL distribution R_0 such that

$$\Pr_{\mathbf{r}\leftarrow R_0}[f_e(\mathbf{r})=0] = \min_{R\in\mathcal{BCL}(\gamma,b,n)} \Pr_{\mathbf{r}\leftarrow R}[f_e(\mathbf{r})=0] \leq \frac{1}{\frac{|G|}{2^n}\cdot(1+\gamma)^b}.$$

Hence,

$$\Pr_{\mathbf{r}\leftarrow R_0}[\mathbf{r}\in G] = \Pr_{\mathbf{r}\leftarrow R_0}[f_e(\mathbf{r})=1] \geq 1 - \frac{1}{\frac{|G|}{2^n}\cdot(1+\gamma)^b}.$$

$$\Pr_{\mathbf{r}\leftarrow R_0}[\mathbf{r}\in B] \leq \Pr_{\mathbf{r}\leftarrow R_0}[f_e(\mathbf{r})=0] \leq \frac{1}{\frac{|G|}{2^n}\cdot(1+\gamma)^b}.$$

Correspondingly, $\Pr_{\mathbf{r}\leftarrow R_0}[\mathbf{r}\in G] - \Pr_{\mathbf{r}\leftarrow R_0}[\mathbf{r}\in B] \geq 1 - \frac{2}{\frac{|G|}{2^n}\cdot(1+\gamma)^b} \geq 1 - \frac{2^{t+2}}{(1+\gamma)^b}$. Therefore, $\mathcal{BCL}(\gamma,b,n)$ is $(t, 1-\frac{2^{t+2}}{(1+\gamma)^b})-$separable.

Let $\frac{2^{t+2}}{(1+\gamma)^b} \leq \frac{1}{2}$, that is, $b \geq \frac{t+3}{\log(1+\gamma)}$. Therefore, $\mathcal{BCL}(\gamma,b,n)$ is $(t,\frac{1}{2})-$separable if $b \geq \frac{t+3}{\log(1+\gamma)}$. □

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 6. *We call 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 $\Pr_{\mathbf{r}\leftarrow U_n}[f(\mathbf{r})\neq g(\mathbf{r})] \geq \frac{1}{2^t}$ for some $t\geq 0$, there exists a distribution $R\in\mathcal{R}_n$ such that $\mathsf{SD}(f(R),g(R))\geq\delta$.*

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 we quantitatively improve the technique of [DOPS04], by making the gap between expressivity and separability independent of the range $\mathcal{C}$ of the functions f and g.

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

Proof. Suppose that $f,g:\{0,1\}^n\to\mathcal{C}$ are two arbitrary functions such that $\Pr_{\mathbf{r}\leftarrow U_n}[f(\mathbf{r})\neq g(\mathbf{r})]\geq\frac{1}{2^t}$. Let $S=\{\mathbf{r}\in\{0,1\}^n \mid f(\mathbf{r})\neq g(\mathbf{r})\}$. By assumption, $|S|\geq 2^{n-t}$.

To build intuition, let's start with the special case where $\mathcal{C}=\{0,1\}$, in which case we will even show that (t,δ)-separability is enough (i.e., no need to increase t by 1). For $\alpha,\beta\in\{0,1\}$, denote $S_{\alpha\beta}=\{\mathbf{r}\in\{0,1\}^n \mid f(\mathbf{r})=\alpha \text{ and } g(\mathbf{r})=\beta\}$.

The distinguisher Eve is defined as $\mathsf{Eve}(x)=1 \Leftrightarrow x=0$. Without loss of generality, assume that $|S_{01}|\geq|S_{10}|$. Denote $G\stackrel{def}{=}S_{01}$ and $B\stackrel{def}{=}S_{10}$. Since $\mathcal{R}_n$ is $(t,\delta)-$separable and $|G\cup B|\geq 2^{n-t}$, there exists a distribution $R\in\mathcal{R}_n$ such that $|\Pr_{\mathbf{r}\leftarrow R}[\mathbf{r}\in G]-\Pr_{\mathbf{r}\leftarrow R}[\mathbf{r}\in B]|\geq\delta$. That is, $|\Pr_{\mathbf{r}\leftarrow R}[\mathbf{r}\in S_{01}]-\Pr_{\mathbf{r}\leftarrow R}[\mathbf{r}\in S_{10}]|\geq\delta$. Therefore,

$$\begin{aligned}
\mathsf{SD}(f(R),g(R)) &\geq |\Pr_{\mathbf{r}\leftarrow R}[\mathsf{Eve}(f(\mathbf{r}))=1]-\Pr_{\mathbf{r}\leftarrow R}[\mathsf{Eve}(g(\mathbf{r}))=1]| \\
&= |\Pr_{\mathbf{r}\leftarrow R}[f(\mathbf{r})=0]-\Pr_{\mathbf{r}\leftarrow R}[g(\mathbf{r})=0]| \\
&= |\{\Pr_{\mathbf{r}\leftarrow R}[\mathbf{r}\in S_{00}]+\Pr_{\mathbf{r}\leftarrow R}[\mathbf{r}\in S_{01}]\}-\{\Pr_{\mathbf{r}\leftarrow R}[\mathbf{r}\in S_{00}]+\Pr_{\mathbf{r}\leftarrow R}[\mathbf{r}\in S_{10}]\}| \\
&= |\Pr_{\mathbf{r}\leftarrow R}[\mathbf{r}\in S_{01}]-\Pr_{\mathbf{r}\leftarrow R}[\mathbf{r}\in S_{10}]| \\
&\geq \delta
\end{aligned}$$

In the following, we analyze the general case. We'll need to use the notion of universal hash function family [CW79] with a single bit output. Recall that $\mathcal{H}=\{h \mid h:\mathcal{C}\to\{0,1\}\}$ is a family of universal hash

functions if for all $z \neq z'$, $\Pr_{h \xleftarrow{\$} \mathcal{H}}[h(z) \neq h(z')] = \frac{1}{2}$. Such families are known to exist for any universe $\mathcal{C}$ and can be made efficient in n if $\mathcal{C} \subseteq \{0,1\}^{poly(n)}$.

For $\alpha, \beta \in \{0,1\}$ and $h \in \mathcal{H}$, denote $S_{\alpha\beta}(h) = \{\mathbf{r} \in S \mid h(f(\mathbf{r})) = \alpha \text{ and } h(g(\mathbf{r})) = \beta\}$. Then

$$\begin{aligned}
\mathbb{E}_{h \leftarrow U_{\mathcal{H}}}[|S_{01}(h)| + |S_{10}(h)|] &= \mathbb{E}_{h \leftarrow U_{\mathcal{H}}}[\sum_{\mathbf{r} \in S} \chi_{S_{01}(h) \cup S_{10}(h)}(\mathbf{r})] \\
&= \sum_{\mathbf{r} \in S} \Pr_{h \leftarrow U_{\mathcal{H}}}[\mathbf{r} \in S_{01}(h) \cup S_{10}(h)] \\
&= \sum_{\mathbf{r} \in S} \Pr_{h \leftarrow U_{\mathcal{H}}}[h(f(\mathbf{r})) \neq h(g(\mathbf{r}))] \\
&= \frac{|S|}{2},
\end{aligned}$$

where $\chi_{S_{01}(h) \cup S_{10}(h)}$ denotes the characteristic function of the set $S_{01}(h) \cup S_{10}(h)$.

Hence, there exists a fixed hash function $h^* \in \mathcal{H}$ such that $|S_{01}(h^*) \cup S_{10}(h^*)| \geq \frac{|S|}{2} \geq 2^{n-t-1}$.

Eve is defined as $\mathsf{Eve}(C) = 1 \Leftrightarrow h^*(C) = 0$, for all $C \in \mathcal{C}$. Without loss of generality, assume that $|S_{01}(h^*)| \geq |S_{10}(h^*)|$. Denote $G \stackrel{def}{=} S_{01}(h^*)$ and $B \stackrel{def}{=} S_{10}(h^*)$. Since $\mathcal{R}_n$ is $(t+1, \delta)-$separable, there exists a distribution $R' \in \mathcal{R}_n$ such that $|\Pr_{\mathbf{r} \leftarrow R'}[\mathbf{r} \in G] - \Pr_{\mathbf{r} \leftarrow R'}[\mathbf{r} \in B]| \geq \delta$. That is, $|\Pr_{\mathbf{r} \leftarrow R'}[\mathbf{r} \in S_{01}(h^*)] - \Pr_{\mathbf{r} \leftarrow R'}[\mathbf{r} \in S_{10}(h^*)]| \geq \delta$. Hence,

$$\begin{aligned}
\mathsf{SD}(f(R'), g(R')) &\geq |\Pr_{\mathbf{r} \leftarrow R'}[\mathsf{Eve}(f(\mathbf{r})) = 1] - \Pr_{\mathbf{r} \leftarrow R'}[\mathsf{Eve}(g(\mathbf{r})) = 1]| \\
&= |\Pr_{\mathbf{r} \leftarrow R'}[h^*(f(\mathbf{r})) = 0] - \Pr_{\mathbf{r} \leftarrow R'}[h^*(g(\mathbf{r})) = 0]| \\
&= |\{\Pr_{\mathbf{r} \leftarrow R'}[\mathbf{r} \in S_{00}(h^*)] + \Pr_{\mathbf{r} \leftarrow R'}[\mathbf{r} \in S_{01}(h^*)]\} - \{\Pr_{\mathbf{r} \leftarrow R'}[\mathbf{r} \in S_{00}(h^*)] + \Pr_{\mathbf{r} \leftarrow R'}[\mathbf{r} \in S_{10}(h^*)]\}| \\
&= |\Pr_{\mathbf{r} \leftarrow R'}[\mathbf{r} \in S_{01}(h^*)] - \Pr_{\mathbf{r} \leftarrow R'}[\mathbf{r} \in S_{10}(h^*)]| \\
&\geq \delta
\end{aligned}$$

Therefore, the source $\mathcal{R}_n$ is $(t, \delta)-$expressive.

□

Combining Theorem 1 and Lemma 1, we immediately get:

Corollary 1.
(a) $\mathcal{W}eak(k,n)$ is $(t,1)-$expressive when $k \leq n-t-2$, and $(t, 2^{n-t-k-2})-$expressive when $n-t-2 < k \leq n-1$. In particular, it's $(t, \frac{1}{2})-$expressive when $k \leq n-t-1$.
(b) $\mathcal{SV}(\gamma, n)$ is $(t, \frac{\gamma}{2^{t+2}})-$expressive.
(c) $\mathcal{BCL}(\gamma, b, n)$ is $(t, 1 - \frac{2^{t+3}}{(1+\gamma)^b})-$expressive. In particular, it's $(t, \frac{1}{2})-$expressive for $b \geq \frac{t+4}{\log(1+\gamma)} = \Theta(\frac{t+1}{\gamma})$.

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 1 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 $|\Pr[\mathsf{Eve}(f(R)) = 1] - \Pr[\mathsf{Eve}(g(R)) = 1]| \geq \delta$. Namely, $f(R)$ is "*δ-computationally distinguishable*" from $g(R)$.

4 On the Impossibility of 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 7. *We say that $\mathsf{Ext} : \{0,1\}^n \to \{0,1\}$ is $(\mathcal{R}_n, \delta)$-secure bit extractor if for every distribution $R \in \mathcal{R}_n$, $|\Pr[\mathsf{Ext}(R) = 1] - \Pr[\mathsf{Ext}(R) = 0]| < \delta$ (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 8. *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: it is "hard" for Alice to alter her commitment after she has made it;
- Hiding: it is "hard" for Bob to find out the commitment 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 9. *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 the following two properties:*

(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: $\mathsf{SD}(\mathsf{Com}(0;R), \mathsf{Com}(1;R)) < \delta$, for every distribution $R \in \mathcal{R}_n$.

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 weak 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, $\mathsf{Com}(b;r) = b \oplus r$, where $b, r \in \{0,1\}$ 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 10. *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: for all $\mathbf{r} \in \{0,1\}^n$ and $x \in \{0,1\}$, $\mathsf{Rec}(\mathsf{Share}_1(x,\mathbf{r}), \mathsf{Share}_2(x,\mathbf{r}), \ldots, \mathsf{Share}_T(x,\mathbf{r})) = x$;

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

Now we abstract and generalize the results of [MP90,DOPS04] to show that expressivity implies impossibility of security involving traditional privacy.

Theorem 2.
(a) If a source $\mathcal{R}_n$ is $(0,\delta)-$expressive, then no $(\mathcal{R}_n, \delta)$-secure bit extractor exists.
(b) If a source $\mathcal{R}_n$ is $(0,\delta)-$expressive, then no $(\mathcal{R}_n, \delta)$-secure bit encryption scheme exists.
(c) If a source $\mathcal{R}_n$ is $(1,\delta)-$expressive, then no $(\mathcal{R}_n, \delta)$-secure weak bit commitment exists.
(d) If a source $\mathcal{R}_n$ is $(\log T,\delta)-$expressive, then no $(\mathcal{R}_n, \delta)$-secure bit T-secret sharing exists.

Proof.

(a) Assume that there exists a $(\mathcal{R}_n, \delta)$-secure bit extractor Ext. Define $f(\mathbf{r}) \stackrel{def}{=} \mathsf{Ext}(\mathbf{r})$ and $g(\mathbf{r}) \stackrel{def}{=} 1 - \mathsf{Ext}(\mathbf{r})$. Since for all $\mathbf{r} \in \{0,1\}^n$, it holds that $\mathsf{Ext}(\mathbf{r}) \neq 1 - \mathsf{Ext}(\mathbf{r})$, we get $\Pr_{\mathbf{r} \leftarrow U_n}[f(\mathbf{r}) \neq g(\mathbf{r})] = 1 = \frac{1}{2^0}$. Definition 6 implies that there exists a distribution $R \in \mathcal{R}_n$ such that $\mathsf{SD}(f(R), g(R)) \geq \delta$. Therefore,

$$|\Pr[\mathsf{Ext}(R) = 1] - \Pr[\mathsf{Ext}(R) = 0]| = \mathsf{SD}(f(R), g(R)) \geq \delta,$$

which is a contradiction.

(b) Assume that there exists a $(\mathcal{R}_n, \delta)$-secure bit encryption scheme. Define $f(\mathbf{r}) \stackrel{def}{=} \mathsf{Enc}_{\mathbf{r}}(0)$ and $g(\mathbf{r}) \stackrel{def}{=} \mathsf{Enc}_{\mathbf{r}}(1)$. Since for all secret keys $\mathbf{r} \in \{0,1\}^n$, it holds that $\mathsf{Enc}_{\mathbf{r}}(0) \neq \mathsf{Enc}_{\mathbf{r}}(1)$, we have $\Pr_{\mathbf{r} \leftarrow U_n}[f(\mathbf{r}) \neq g(\mathbf{r})] = 1 = \frac{1}{2^0}$. Definition 6 implies that there exists a distribution $R \in \mathcal{R}_n$ such that $\mathsf{SD}(f(R), g(R)) \geq \delta$, which is in contradiction to $\mathsf{SD}(f(R), g(R)) < \delta$.

(c) Assume that there exists a $(\mathcal{R}_n, \delta)$-secure weak bit commitment. Define $f(\mathbf{r}) \stackrel{def}{=} \mathsf{Com}(0; \mathbf{r})$ and $g(\mathbf{r}) \stackrel{def}{=} \mathsf{Com}(1; \mathbf{r})$. Since $\Pr_{\mathbf{r} \leftarrow U_n}[\mathsf{Com}(0; \mathbf{r}) \neq \mathsf{Com}(1; \mathbf{r})] \geq \frac{1}{2}$, there exists a distribution $R \in \mathcal{R}_n$ such that $\mathsf{SD}(f(R), g(R)) \geq \delta$, which is in contradiction to $\mathsf{SD}(f(R), g(R)) < \delta$.

(d) Assume that there exists a $(\mathcal{R}_n, \delta)$-secure bit T-secret sharing. Let $t = \log T$. Then for all $\mathbf{r} \in \{0,1\}^n$,

$$\begin{aligned}
&(\mathsf{Share}_1(0; \mathbf{r}), \mathsf{Share}_2(0; \mathbf{r}), \ldots, \mathsf{Share}_T(0; \mathbf{r})) \neq (\mathsf{Share}_1(1; \mathbf{r}), \mathsf{Share}_2(1; \mathbf{r}), \ldots, \mathsf{Share}_T(1; \mathbf{r})) \\
&\Rightarrow \textit{there exists } j = j(\mathbf{r}) \textit{ such that } \mathsf{Share}_j(0; \mathbf{r}) \neq \mathsf{Share}_j(1; \mathbf{r}). \\
&\Rightarrow \textit{there exists } j^* \in [T] \textit{ such that } |\{\mathbf{r} \mid j(\mathbf{r}) = j^*\}| \geq \frac{2^n}{T} = 2^{n-t}.
\end{aligned}$$

Define $f(\mathbf{r}) \stackrel{def}{=} \mathsf{Share}_{j^*}(0; \mathbf{r})$ and $g(\mathbf{r}) \stackrel{def}{=} \mathsf{Share}_{j^*}(1; \mathbf{r})$. Then $\Pr_{\mathbf{r} \leftarrow U_n}[f(\mathbf{r}) \neq g(\mathbf{r})] \geq \frac{1}{2^t}$. Therefore, there exists a distribution $R \in \mathcal{R}_n$ such that $\mathsf{SD}(f(R), g(R)) \geq \delta$, which is in contradiction to $\mathsf{SD}(f(R), g(R)) < \delta$. □

From Theorem 2 and Corollary 1, we conclude:

Theorem 3. *For the following values of δ, shown in Table 1, no $(\mathcal{R}_n, \delta)-$secure cryptographic primitive P exists, where $\mathcal{R}_n \in \{\mathcal{W}eak(k,n), \mathcal{SV}(\gamma,n), \mathcal{BCL}(\gamma,b,n)\}$ 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 $k \leq n - \log T - 2$
$\mathcal{W}eak(n-1,n)$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{2T}$
$\mathcal{SV}(\gamma,n)$	$\frac{\gamma}{4}$	$\frac{\gamma}{4}$	$\frac{\gamma}{8}$	$\frac{\gamma}{4T}$
$\mathcal{BCL}(\gamma,b,n)$	$\frac{1}{2}$, 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 BCL source 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 1 used a more clever "universal hashing trick". More importantly, instead of focusing the entire proof on some specific SV/weak sources [MP90,DOPS04], our impossibility results for such sources were obtained in a more modular manner, making these proofs somewhat more illuminating.

5 On the Impossibility of 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, denote $\mathcal{R}_n$ as 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 two databases 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 random variable $R \in \mathcal{R}_n$, and outputs a random value z. Informally, we wish $z = M(D, R)$ to approximate the true Hamming weight $wt(D)$ 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 11. *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 $\mathsf{RD}(M(D_1, R), M(D_2, R)) \leq \varepsilon$. Equivalently, for any possible output z:*

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

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

Definition 12. *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, expressivity implies impossibility of differential privacy with imperfect randomness, albeit with slightly more demanding parameters. As a high-level idea, 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 expressivity 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. Details follow.

Theorem 4. *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.*

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 11, 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)-security, 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 Equation (1),

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

which is a contradiction.

□

Implications for Weak and BCL sources. Now we apply the impossibility results of differential privacy to the sources $\mathcal{W}eak(k,n)$ and $\mathcal{BCL}(\gamma,b,n)$. In particular, by combining Theorem 4.(b) with Corollary 1.(a) and Corollary 1.(c), respectively, we get

Theorem 5. *If* $k \leq n - \log(\varepsilon\rho) - 6$, *then no* $(\mathcal{W}eak(k,n),\varepsilon)-$*differentially private and* (U_n,ρ)*-accurate mechanism for the Hammimg weight queries exists.*

Theorem 6. *If* $b \geq \frac{\log(\varepsilon\rho)+9}{\log(1+\gamma)} = \Omega(\frac{\log(\varepsilon\rho)+1}{\gamma})$, *then no* $(\mathcal{BCL}(\gamma,b,n),\varepsilon)-$*differentially private and* (U_n,ρ)*-accurate mechanism for the Hammimg weight queries exists.*

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) 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 expressivity of the SV sources given by Corollary 1.(b) is just not good enough to yield very strong results, as we explain now.

(Non-)Implications for the SV source. We observe that Theorem 4 can't be applied to the SV source, as $\mathcal{SV}(\gamma, n)$ is only (t, δ)-expressive for $\delta = \frac{\gamma}{2^{t+2}}$, which means that $2^t\delta = O(\gamma)$. In contrast, to apply Theorem 4 we need $2^t\delta \geq \Omega(\rho\varepsilon)$. Thus, to have any hope, we need, $\rho = O(\gamma/\varepsilon)$, but this violates our pre-condition (used in the proof) that $\rho \geq 1/(8\varepsilon)$. In fact, a simple reworking of the proof of Theorem 4 (omitted) can be used to show that if there exists a $(\mathcal{SV}(\gamma, n), \varepsilon)-$differentially private and (U_n, ρ)-accurate mechanism for the Hammimg weight queries, then $\rho > \frac{\gamma}{64\cdot\varepsilon} = \Omega(\frac{\gamma}{\varepsilon})$.

Unfortunately, this implication that we get is quite weak, because we can get a *stronger* result, even if $\mathcal{R}_n$ consists only of the uniform distribution U_n. We present this well known folklore result for completeness.

Lemma 2. *Assume that the mechanism M is $(U_n, \varepsilon)-$differentially private and (U_n, ρ)-accurate for the Hammimg weight queries. Then $\rho \geq \frac{1}{e+1} \cdot \frac{1}{\varepsilon} = \Omega(\frac{1}{\varepsilon})$.*

Proof. For any $D, D' \in \mathcal{D}$, let $\beta \stackrel{def}{=} \frac{\mathbb{E}_{\mathbf{r}\leftarrow R}[M(D',R)]}{\mathbb{E}_{\mathbf{r}\leftarrow R}[M(D,R)]}$. From Definition 12, we have $\beta \geq \frac{wt(D')-\rho}{wt(D)+\rho}$. By Definition 11, we obtain

$$\beta = \frac{\sum_z z \Pr[M(D',R)=z]}{\sum_z z \Pr[M(D,R)=z]} \leq \frac{\sum_z z e^{\varepsilon\cdot\Delta(D,D')} \Pr[M(D,R)=z]}{\sum_z z \Pr[M(D,R)=z]} = e^{\varepsilon\cdot\Delta(D,D')}.$$

Therefore, $\frac{wt(D')-\rho}{wt(D)+\rho} \leq e^{\varepsilon\cdot\Delta(D,D')}$.

Take a specific D such that $wt(D) = 0$. Let $Ball(D, \alpha) = \{D' : \Delta(D, D') \leq \alpha\}$. Then

$$\forall D' \in Ball\left(D, \frac{1}{\varepsilon}\right) \Rightarrow \frac{wt(D')-\rho}{\rho} \leq e^{\varepsilon\cdot\Delta(D,D')} \leq e \Rightarrow \rho \geq \frac{1}{e+1} wt(D').$$

Taking D' such that $wt(D') = \frac{1}{\varepsilon}$, we get $\rho \geq \frac{1}{e+1} \cdot \frac{1}{\varepsilon}$.

□

Thus, our technique cannot yield any results for the γ-SV source, which we even didn't already know for the uniform distribution. Of course, this is not surprising, because Dodis et al. [DLMV12] have shown that we can get $(\mathcal{SV}(\gamma, n), \varepsilon)-$differentially private and $(\mathcal{SV}(\gamma, n), \rho)$-accurate mechanism for all counting queries (including the Hamming weight queries), where $\rho = poly_{1/(1-\gamma)}(\frac{1}{\varepsilon}) \gg \frac{1}{\varepsilon}$ and $poly_{1/(1-\gamma)}(x)$ denotes a polynomial whose degree and coefficients are fixed (and rather large) functions of $1/(1-\gamma)$.

6 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 2-secret sharing (e.g., set $T = 2$ for concreteness).

In particular, while a very "structured" (and, hence, rather unrealistic) SV source was 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.

Source	Traditional Privacy δ	Differential Privacy ε & Utility ρ
$\mathcal{W}eak(k,n)$	Impossible if $\delta \leq \frac{1}{4}$, even if $k = n-1$	Impossible if $k \leq n - \log(\varepsilon\rho) - O(1)$
$\mathcal{SV}(\gamma,n)$	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.

7 Privacy Implies Weak Bit Extraction

Recall, Bosley and Dodis [BD07] initiated the study of the following general question: *does privacy inherently require "extractable" source of randomness*? A bit more formally, if a given primitive P admits $(\mathcal{R}_n, \delta)$-secure implementation, does it mean that 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 13. *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 $R \in \mathcal{R}_n$, $|\Pr[\mathsf{Ext}(R) = 1] - \Pr[\mathsf{Ext}(R) = 0]| < \delta$;
(b) $\Pr[\mathsf{Ext}(U_n) \neq \perp] \geq \tau$.

We briefly discuss this notion, before showing our results. Like traditional bit-extractor in Definition 7, the odds of outputting 0 or 1 are roughly the same for any distribution R in the source. However, the 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 $\Pr[\mathsf{Ext}(R) \neq \perp] \geq \tau$. Ideally, we would like this to be true for any distribution R in the source. Unfortunately, we will see shortly that 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 that the notion of weak bit-extraction is simply a different way to express (the negation of) our notion of separability!

Lemma 3. *$\mathcal{R}_n$ 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.*

Proof. We only prove that non-separability implies weak bit extraction, as the converse is clear because all our steps will be "if and only if".

Since $\mathcal{R}_n$ is <u>not</u> (t,δ)-separable, then there are two sets G and B such that $G \cap B = \emptyset$, $|G \cup B| \geq 2^{n-t}$ and for all $R \in \mathcal{R}_n$, we have $|\Pr[R \in G] - \Pr[R \in B]| \leq \delta$. Define

$$\mathsf{Ext}(\mathbf{r}) \stackrel{def}{=} \begin{cases} 1, & \text{if } \mathbf{r} \in G; \\ 0, & \text{if } \mathbf{r} \in B; \\ \perp, & \text{otherwise.} \end{cases}$$

This is well defined since $G \cap B = \emptyset$, and satisfies properties (a) and (b) of weak bit extractor, since $\delta \geq |\Pr[R \in G] - \Pr[R \in B]| = |\Pr[\mathsf{Ext}(R) = 1] - \Pr[\mathsf{Ext}(R) = 1]|$, while $\Pr \mathsf{Ext}[(U_n) \neq \perp] = |G \cup B|/2^n \geq 2^{n-t}/2^n = 2^{-t}$.

□

We can now combine Lemma 3 with counter-positives of Theorem 1, Theorem 2 (for traditional privacy) and Theorem 4.(a) (for differential privacy), to get the following result:

Theorem 7.
(a) If $(\mathcal{R}_n, \delta)$-secure bit encryption 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 1, chosen as to ensure that

$$\Pr[\mathsf{Ext}(U_n) \neq \perp] = \Pr_{\mathbf{r} \leftarrow U_n}[h^*(\mathsf{Enc}_{\mathbf{r}}(0)) \neq h^*(\mathsf{Enc}_{\mathbf{r}}(1))] \geq \frac{1}{2}$$

In particular, when the bit encryption (resp. commitment, secret sharing, DP mechanism) is computationally efficient (in n), our bit extractor is efficient as well. 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 $\Pr[\mathsf{Ext}(R) \neq \perp] \geq \tau$ is only required for $R \equiv U_n$. Unfortunately, we observe that the analog of Theorem 7.(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 7.(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). [5]

Acknowledgments. The authors would like to thank Benjamin Fuller, Sasha Golovnev, Hamidreza Jahanjou, Umut Orhan, and Abhishek Samanta.

Yevgeniy Dodis was partially supported by gifts from VMware Labs and Google, and NSF grants 1319051, 1314568, 1065288, and 1017471. Yanqing Yao is supported by the Scholarship Award for Excellent Doctoral Student granted by Ministry of Education 400618, and CSC grant 201206020063.

[5] 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.

References

[ACM+14] Per Austrin, Kai-Min Chung, Mohammad Mahmoody, Rafael Pass, and Karn Seth. On the Impossibility of Cryptography with Tamperable Randomness. *CRYPTO*, volume 8616 of *LNCS*, pages 462-479. Springer, 2014.

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

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

[BD07] Carl Bosley and Yevgeniy Dodis. Does privacy require true randomness? In Salil P. Vadhan, editor, *TCC*, volume 4392 of *LNCS*, pages 1-20. Springer, 2007.

[BDK+05] Xavier Boyen, Yevgeniy Dodis, Jonathan Katz, Rafail Ostrovsky, and Adam Smith. Secure remote authentication using biometric data. In Ronald Cramer, editor, *Advances in Cryptology EUROCRYPT 2005*, volume 3494 of *LNCS*, pages 147-163. Springer-Verlag, 2005.

[BH05] Boaz Barak and Shai Halevi. A model and architecture for pseudo-random generation with applications to /dev/random. In *Proceedings of the 12th ACM Conference on Computer and Communication Security*, pages 203-212, 2005.

[BST03] Boaz Barak, Ronen Shaltiel, and Eran Tromer. True random number generators secure in a changing environment. In *Proceedings of the 5th Cryptographic Hardware and Embedded Systems*, pages 166-180, 2003.

[CFG+85] B. Chor, J. Friedman, O. Goldreich, J. Håstad, S. Rudich, R. Smolensky. The Bit Extraction Problem or t-resilient Functions. In *Proc. of 26th FOCS*, pages 396-407, 1985.

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

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

[Dod01] Yevgeniy Dodis. New Imperfect Random Source with Applications to Coin-Flipping. *ICALP 2001*, pages 297-309.

[DKRS06] Yevgeniy Dodis, Jonathan Katz, Leonid Reyzin, and Adam Smith. Robust fuzzy extractors and authenticated key agreement from close secrets. In Cynthia Dwork, editor, *CRYPTO*, volume 4117 of *LNCS*, pages 232-250. Springer, 2006.

[DLMV12] Yevgeniy Dodis, Adriana López-Alt, Ilya Mironov, and Salil P. Vadhan. Differential Privacy with Imperfect Randomness. *CRYPTO 2012*, pages 497-516.

[DMNS06] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In Shai Halevi and Tal Rabin, editors, *TCC*, volume 3876 of *LNCS*, pages 265-284. Springer, 2006.

[DOPS04] Yevgeniy Dodis, Shien Jin Ong, Manoj Prabhakaran, and Amit Sahai. On the (im)possibility of cryptography with imperfect randomness. *FOCS 2004*, pages 196-205.

[DORS08] Yevgeniy Dodis, Rafail Ostrovsky, Leonid Reyzin, and Adam Smith. Fuzzy extractors: How to generate strong keys from biometrics and other noisy data. *SIAM Journal on Computing*, 38(1): 97-139, 2008.

[DS02] Yevgeniy Dodis and Joel Spencer. On the (non)Universality of the One-Time Pad. *Foundations of Computer Science (FOCS)*, 376-385, 2002.

[GKR04] Rosario Gennaro, Hugo Krawczyk, and Tal Rabin. Secure hashed diffie-hellman over non-ddh groups. In Christian Cachin and Jan Camenisch, editors, *Advances in Cryptology EUROCRYPT 2004*, volume 3027 of *LNCS*, pages 361-381. Springer-Verlag, 2004.

[Kra10] Hugo Krawczyk. Cryptographic Extraction and Key Derivation: The HKDF Scheme. In Tal Rabin, editor, *Advances in Cryptology-CRYPTO 2010*, volume 6223 of *LNCS*, pages 631-648. Springer-Verlag, 2010.

[LLS89] David Lichtenstein, Nathan Linial, and Michael E. Saks. Some extremal problems arising form discrete control processes. *Combinatorica*, 9(3): 269-287, 1989.

[MP90] James L. McInnes and Benny Pinkas. On the impossibility of private key cryptography with weakly random keys. In Alfred Menezes and Scott A. Vanstone, editors, *CRYPTO*, volume 537 of *LNCS*, pages 421-435. Springer, 1990.

[MW97] Ueli M. Maurer and Stefan Wolf. Privacy amplification secure against active adversaries. In Burton S. Kaliski, Jr., editor, *CRYPTO*, volume 1294 of *LNCS*, pages 307-321. Springer, 1997.

[RVW04] Omer Reingold, Salil Vadhan, and Avi Widgerson. 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] Miklos Santha and Umesh V. Vazirani. Generating quasi-random sequences from semirandom sources. *J. Comput. Syst. Sci.*, 33(1): 75-87, 1986.

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

[VV85] Umesh V. Vazirani and Vijay V. Vazirani. Random polynomial time is equal to slightly random polynomial time. *FOCS*, pages 417-428, 1985.

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