Differential Privacy with Bias-Control Limited Sources

Yanqing Yao[a,b], Zhoujun Li[a,c]

[a]State Key Laboratory of Software Development Environment, Beihang University, Beijing 100191, China
[b]Department of Computer Science, New York University, New York 10012, USA
[c]Beijing Key Laboratory of Network Technology, Beihang University, Beijing, China
yaoyanqing1984@gmail.com,lizj@buaa.edu.cn

Abstract. Differential privacy aims at allowing the owner of a sensitive database D to securely release some "aggregate statistics" $f(D)$ while protecting the privacy of individual users whose data is in D. Some randomness $\mathbf{r}$ is used to "add enough noise" to the true answer $f(D)$. Traditionally, most papers assume that perfect randomness is available by default. However, in reality, 'imperfect' random sources can only be obtained. Dodis et al. [CRYPTO'12] proposed that differential privacy can be achieved with Santha-Vazirani (SV) source via adding a stronger property called SV-consistent sampling and left open the question if differential privacy is possible with more realistic (i.e., less structured) sources than the SV sources. Bias-Control Limited (BCL) source, introduced by Dodis [ICALP'01], is a generalization of the SV-source and bit-fixing source, which is more realistic. Unfortunately, if we nationally expand SV-consistent sampling to the BCL source, the expansion is hopeless to achieve differential privacy. One main reason is that SV-consistent sampling requires consecutive strings, while some strings can't be generated from "non-trivial" BCL source. Motivated by these questions, we obtain the following results.

- A new concept, called compact BCL-consistent sampling, is introduced to study differentially private mechanisms. It should be noted that when $b = 0$, the degenerated BCL-consistent sampling is not the same as the SV-consistent sampling proposed by [DLMV12].
- If the BCL source satisfies compact BCL-consistent sampling, then the corresponding mechanism is differentially private. Even if the BCL source is degenerated into the SV-source, compared with [DLMV12], our theorem is more intuitive and our proof is much simpler.
- A new truncation technique is proposed in designing a finite-precision mechanism to satisfy this property.
- Rigorous proofs about differential privacy and accuracy of this kind of mechanism are shown.
- While the result of [DY14] implies a *possible* result about differentially private mechanisms: If there exist $(\mathcal{BCL}(\delta, b), \xi)-$differentially private and $(\mathcal{U}, \rho)$-accurate mechanisms for the Hammimg weight queries, then $\rho > \frac{2^{b\cdot\log(1+\delta)-9}}{\xi}$. We build *explicit* such mechanisms and the parameters match the above condition.

Keywords: Differential Privacy; Bias-Control Limited Source; Consistent Sampling; Laplace Distribution; Finite-Precision Mechanism

1 Introduction

Traditionally, cryptographic issues are studied under the assumption of the availability of perfect randomness, i.e., sources that output unbiased and independent random bits. However, in many settings this assumption seems unrealistic, 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]. 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 speaking, 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. Moreover, 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 is the Santha-Vazirani (SV) source [SV86]. 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 $\mathsf{Ext} : \{0,1\}^n \to \{0,1\}$ capable of extracting even a *single* bit of bias *strictly* less than δ from the δ-SV source, irrespective of how many SV bits $x_1, x_2, \ldots, x_n$ 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") a certain long string, so having (min-)entropy in our source should be sufficient to achieve this goal.

Unfortunately, the situation appears to be much less bright when dealing with *privacy* applications, such as encryption, commitment, zero-knowledge, and some 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. 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.

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]. The motivating scenario of differential privacy is a statistical database. The purpose of a privacy-preserving statistical database is to enable the user to learn released statistical facts without compromising the privacy of the individual users whose data is in the database. Differential privacy ensures the removal or addition of a single database item does not (substantially) affect the outcome of any analysis [Dwo08]. More formally, 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. On one hand, to preserve individual users' privacy, we want M to satisfy ξ-differential privacy, that is, for any neighboring databases D_1 and D_2 (i.e., D_1 and D_2 differ on a single record), and for any possible output z, $e^{-\xi} \leq \Pr_{\mathbf{r}}[M(D_1, f; \mathbf{r}) = z] / \Pr_{\mathbf{r}}[M(D_2, f; \mathbf{r}) = z] \leq e^{\xi}$

for small $\xi > 0$. On the other hand, to keep the utility (or accuracy) of M, we hope the expected value of $|f(D) - M(D, f; \mathbf{r})|$ over random $\mathbf{r}$ to be as small as possible. Usually, we should make a tradeoff between differential privacy and utility.

Additive-noise mechanisms [DLMV12,DMNS06,GRS09,HT10] with perfect randomness were introduced in the original work of [BDMN05,DN03,DMNS06,DN04]. Dwork et al. [DMNS06] designed an additive-noise mechanism $M(D, f; r) = f(D) + X(\mathbf{r})$, where $X(\mathbf{r})$ is the Laplace distribution, which is private and accurate. However, we can not generate a "good enough" sample of the Laplace distribution with the Santha-Vazirani (SV) sources. In fact, any accurate and private additive-noise mechanism for a source R implies the existence of a randomness extractor for R, essentially collapsing the notion of differential privacy to that of traditional privacy, and showing the impossibility of accurate and private additive-noise mechanisms for SV sources [DLMV12]. From another perspective, an additive-noise mechanism must satisfy $T_1 \cap T_2 = \emptyset$, based on which an SV adversary can always succeed in amplifying the ratio $\Pr[\mathbf{r} \in T_1]/\Pr[\mathbf{r} \in T_2]$ (see [DLMV12]), or $|\Pr[\mathbf{r} \in T_1] - \Pr[\mathbf{r} \in T_2]|$ (see [DY14]), where T_i is the set of coins $\mathbf{r}$ with $M(D, f; \mathbf{r}) = z$.

Dodis et al. [DLMV12] observed a necessary condition, called consistent sampling (i.e., informally, $|T_1 \cap T_2| \approx |T_1| \approx |T_2|$), to build SV-robust mechanisms. This property is similar to the concept of "consistent sampling" [Man94,Hol07] with applications in differential privacy [MMP+10], web search [BGMZ97] and parallel repetition theorems [Hol07], among others. They also introduced another condition to match the bit-by-bit property of SV sources. The combination of the above two conditions is called SV-consistent sampling (see Definition 9). They builded a concrete accurate and private Laplace mechanism by using some rounding and arithmetic coding techniques. Such a mechanism 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 δ, 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 the following open problem.

Open Question. *Is differential privacy possible with more realistic (i.e., less structured) sources than the SV sources?*

Dodis et al. [Dod01] introduced more realistic source, called Bias-Control Limited (BCL) source, which realistically generalizes the SV source of Santha and Vazirani [SV86] and the bit-fixing source of Lichtenstein, Linial, and Saks [LLS89]. The BCL source generates a sequence of bits $x_1, x_2, \ldots, x_n$, where for $i = 1, 2, \ldots, n$, the value of x_i can depend on $x_1, x_2, \ldots, x_{i-1}$ in one of the following two ways: (A) x_i is determined by $x_1, \ldots, x_{i-1}$, but this happens for at most b bits, or (B) $\frac{1-\delta}{2} \leq \Pr[x_i = 1 \mid x_1, x_2, \ldots, x_{i-1}] \leq \frac{1+\delta}{2}$, where $0 \leq \delta < 1$. (See Definition 2 for details.) In particular, when $b = 0$, it degenerates into the SV source; when $\delta = 0$, it yields the bit-fixing source; when $b = 0$ and $\delta = 0$, it corresponds to the perfect randomness. If $b \neq 0$ and $\delta \neq 0$, we say the BCL source is non-trivial. The BCL source models the problem that each of the bits produced by a streaming source is unlikely to be perfectly random: slight errors (due to noise, measurement errors, and imperfections) of the source are inevitable, and the situation that some of the bits could have non-trivial dependencies on the previous bits (due to internal correlations, poor measurement or improper setup), to the point of being completely determined by them. Hence, compared with the SV source, the BCL source appears much more realistic, 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 whether existing results can be expanded by using BCL source. However, SV-consistent sampling is useless to BCL source as some strings can't be generated from BCL source, which is crucial to the SV-consistent sampling property. Recently, Dodis and Yao [DY14] have shown an impossibility result for BCL source: when $b \geq \Omega(\log(\xi\rho)/\delta)$, it's impossible to achieve $(\mathcal{BCL}(\delta, b), \xi)$-differentially private and $(\mathcal{U}, \rho)$-accurate mechanism for Hamming weight queries. In other words, if there exists a $(\mathcal{BCL}(\delta, b), \xi)$-differentially private and $(\mathcal{U}, \rho)$-accurate mechanism for Hamming weight queries, then $b \leq O(\log(\xi\rho)/\delta)$. This result gives us a bit hope to design differentially private and accurate mechanisms for some b.

Our Results and Techniques.

We try to naturally expand SV-consistent sampling to BCL-consistent sampling, but can't get optimistic results. Due to space limitations, we omit the concrete pessimistic results. It's not surprising, as the "interval"

property (see Definition 9) is used to achieve SV-differential privacy, while the mechanism based on $\mathcal{BCL}(\delta, b)$ with $b \neq 0$ can't be an interval one.

Essentially, in order to achieve differential privacy, we need to restrict the upper bound of $\Pr_{\mathbf{r}\leftarrow BCL(\delta,b,n)}[\mathbf{r} \in T_1\backslash T_2]/\Pr_{\mathbf{r}\leftarrow BCL(\delta,b,n)}[\mathbf{r} \in T_2]$. Similar to [DLMV12], consistent sampling is a necessary condition for building BCL-robust, differentially private mechanisms. From the generation procedure of $BCL(\delta, b, n)$, we can upper bound the numerator and lower bound the denominator by introducing the common prefix $\mathbf{u}$ of T_1 and T_2. Instead of limiting $|\text{SUFFIX}(\mathbf{u}, n)|/|T_1 \cup T_2| = 2^{n-|\mathbf{u}|}/|T_1 \cup T_2|$ like in [DLMV12], we limit $n - |\mathbf{u}|$. The concept of compact BCL-consistent sampling (Definition 10) emerges from this motivation. Even if $BCL(\delta, b, n)$ degenerates into $\mathcal{SV}(\delta, n)$, compared with [DLMV12], our idea is more intuitive and the proof is much simpler (Theorem 1 and Corollary 1).

However, we are confronted with some difficulties to construct explicit differentially private mechanisms. According to the method of yielding finite precision mechanisms in [DLMV12], we can't upper bound $n-|\mathbf{u}|$ as a constant! To solve this problem, we find a new truncation trick (see step 2 of Section 4) and prove that this kind of mechanism can achieve the upper bound of $n - |\mathbf{u}|$ to be a constant (see Theorem 3). Our contributions are as follows.

- We introduce a new concept, called compact BCL-consistent sampling (see Definition 10) to study differentially private mechanisms. It should be noted that the degenerated BCL-consistent sampling is not the same as the SV-consistent sampling proposed by [DLMV12].
- We prove that if the BCL source satisfies this property, then the corresponding mechanism is differentially private (see Theorem 1). Even if the BCL source is degenerated into the SV-source, compared with [DLMV12], our theorem is more intuitive and our proof is much simpler(see Corollary 1 and Theorem 4.4 of [DLMV12]).
- We use a new truncation technique for the design of a finite-precision mechanisms to satisfy this property (see Step 2 of Section 4.1).
- We also give rigorous proofs about differential privacy and accuracy of this kind of mechanism(Theorems 4 and 5).
- While the result of [DY14] implies that if there exists a $(\mathcal{BCL}(\delta, b), \xi)-$differentially private and $(\mathcal{U}, \rho)$-accurate mechanism for the Hammimg weight queries, then the parameters should satisfy $\rho > \frac{2^{b\cdot\log(1+\delta)-9}}{\xi}$, we build *explicit* such mechanisms and the parameters match the above condition (Theorem 6). Thus we have made some progress.

Organization. The remainder of the paper is organized as follows. In Section 2, we review some notations and concepts. In Section 3, we introduce the concept of compact BCL-consistent sampling, and prove that it's sufficient to achieve differential privacy. In Section 4, we show the concrete construction of finite-precision mechanisms, and give rigorous proofs about differential privacy and accuracy of this kind of mechanism.

2 Preliminaries

In this section, we present some notations and definitions that will be used later.

Let $\{0,1\}^* \stackrel{def}{=} \bigcup_{m\in\mathbb{Z}^+} \{0,1\}^m$. We consider a distribution over $\{0,1\}^*$ as continuously outputting (possibly correlated) bits. We call a family $\mathcal{R}$ of distributions over $\{0,1\}^*$ a source. Denote $\mathcal{U}$ as the uniform source, which is the set containing only the distribution U on $\{0,1\}^*$ that samples each bit independently and uniformly at random. For a set S, we write U_S to denote the uniform distribution over S. 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. For a positive integer n, let $[n] \stackrel{def}{=} \{1, 2, \ldots, n\}$. Denote $\lfloor\cdot\rceil$ as the nearest integer function.

Definition 1. *([SV86]) Let $x_1, x_2, \ldots$ be a sequence of Boolean random variables and $0 \leq \delta < 1$. A probability distribution $X = x_1x_2\ldots$ over $\{0,1\}^*$ is a δ-Santha-Vazirani (SV) distribution, denoted by $SV(\delta)$, if for all $i \in \mathbb{Z}^+$ and for every string s of length $i-1$, we have*

$$\frac{1-\delta}{2} \leq \Pr[x_i = 1 \mid x_1x_2\ldots x_{i-1} = s] \leq \frac{1+\delta}{2}.$$

We define the δ-Santha-Vazirani source $\mathcal{SV}(\delta)$ to be the set of all δ-SV distributions. For a distribution $SV(\delta) \in \mathcal{SV}(\delta)$, we define $SV(\delta, n)$ as the distribution $SV(\delta)$ restricted to the first n coins $x_1x_2 \dots x_n$. We let $\mathcal{SV}(\delta, n)$ be the set of all distributions $SV(\delta, n)$.

Definition 2. *([Dod01]) Let $x_1, x_2, \dots$ be a sequence of Boolean random variables and $0 \leq \delta < 1$. A probability distribution $X = x_1x_2 \dots$ over $\{0,1\}^*$ is a (δ, b)-Bias-Control Limited (BCL) distribution, denoted by $BCL(\delta, b)$, if for all $i \in \mathbb{Z}^+$ and for every string s of length $i-1$, the value of x_i can depend on $x_1, x_2, \dots, x_{i-1}$ in one of the following two ways:*

(A) x_i is determined by $x_1, \dots, x_{i-1}$, but this happens for at most b bits. This process of determining a bit is called intervention.

(B) $\frac{1-\delta}{2} \leq \Pr[x_i = 1 \mid x_1x_2 \dots x_{i-1} = s] \leq \frac{1+\delta}{2}$.

We define the (δ, b)-Bias-Control Limited source $\mathcal{BCL}(\delta, b)$ to be the set of all (δ, b)-BCL distributions. For a distribution $BCL(\delta, b) \in \mathcal{BCL}(\delta, b)$, we define $BCL(\delta, b, n)$ as the distribution $BCL(\delta, b)$ restricted to the first n coins $x_1x_2 \dots x_n$. We let $\mathcal{BCL}(\delta, b, n)$ be the set of all distributions $BCL(\delta, b, n)$.

This source models the facts that physical sources can never produce completely perfect bits and some of the bits generated by a physical source could be determined from the previous bits.

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

Consider a statistical database as an array of rows from some countable set. Two databases are neighboring if they differ in exactly one row. Let $\mathcal{D}$ be the space of all databases. For simplicity, we only consider the query function $f : \mathcal{D} \to \mathbb{Z}$. Recall some concepts mentioned in [DLMV12] as follows.

Definition 3. *Let $\xi \geq 0$, $\mathcal{R}$ be a source, and $\mathcal{F} = \{f : \mathcal{D} \to \mathbb{Z}\}$ be a family of functions. A mechanism M is $(\mathcal{R}, \xi)$-differentially private for $\mathcal{F}$ if for all neighboring databases $D_1, D_2 \in \mathcal{D}$, all $f \in \mathcal{F}$, all possible outputs $z \in \mathbb{Z}$, and all distributions $R \in \mathcal{R}$:*

$$\frac{\Pr_{\mathbf{r} \leftarrow R}[M(D_1, f; \mathbf{r}) = z]}{\Pr_{\mathbf{r} \leftarrow R}[M(D_2, f; \mathbf{r}) = z]} \leq 1 + \xi.$$

In what follows we employ the upper bound of the ratio of probabilities introduced in [DLMV12] other than the traditional upper bound "e^{ξ}" to make later calculations a little simpler. It's reasonable since when $\xi \in [0, 1]$, which is the main useful range, we have $e^{\xi} \approx 1 + \xi$, and when $\xi \geq 0$, we always have $1 + \xi \leq e^{\xi}$.

Remark 1. As has been observed by Dodis et al. [DLMV12], in this paper we assume that the randomness $\mathbf{r}$ as input of the mechanism M is in $\{0,1\}^*$, i.e., M has at its disposal a possibly infinite number of random bits, but for two neighboring databases $D_1, D_2 \in \mathcal{D}$, a query $f \in \mathcal{F}$, and a fixed outcome z, M needs only a finite number of coins $n \stackrel{def}{=} \tilde{\tau}(D_1, D_2, f, z)$, where $\tilde{\tau}$ is a function, to determine whether $M(D_1, f) = z$ and $M(D_2, f) = z$. Furthermore, we assume that if $M(D_1, f; \mathbf{r}) = z$ and $M(D_2, f; \tilde{\mathbf{r}}) = z$ where $\mathbf{r}, \tilde{\mathbf{r}} \in \{0,1\}^n$, then for any $\mathbf{r}'$ with $\mathbf{r}$ as its prefix and $\tilde{\mathbf{r}}'$ with $\tilde{\mathbf{r}}$ as its prefix, we still have $M(D_1, f; \mathbf{r}') = z$ and $M(D_2, f; \tilde{\mathbf{r}}') = z$, that is, providing M with extra coins does not change the output.

Definition 4. *Let $\rho > 0$, $\mathcal{R}$ be a source, and $\mathcal{F} = \{f : \mathcal{D} \to \mathbb{Z}\}$ be a family of functions. A mechanism M has $(\mathcal{R}, \rho)$-utility if for all databases $D \in \mathcal{D}$, all queries $f \in \mathcal{F}$, and all distributions $R \in \mathcal{R}$:*

$$\mathbb{E}_{\mathbf{r} \leftarrow R}[|M(D, f; \mathbf{r}) - f(D)|] \leq \rho.$$

One core problem in the area of differential privacy is to design accurate and private mechanisms.

Definition 5. *We say a function family $\mathcal{F}$ admits accurate and private mechanisms w.r.t. $\mathcal{R}$ if there exists a function $g(\cdot)$ such that for all $\xi > 0$ there exists a mechanism $M_{(\xi)}$ that is $(\mathcal{R}, \xi)$-differentially private and has $(\mathcal{R}, g(\xi))$-utility. $\mathcal{M} = \{M_{(\xi)}\}$ is called a class of accurate and private mechanisms for $\mathcal{F}$ w.r.t. $\mathcal{R}$.*

Though there are already some infinite additive mechanisms based on gaussian, binomial, and Laplace distributions, we must specify how to approximate them under finite precision in practice. When perfect randomness $\mathcal{U}$ is available, we can simply approximate a continuous sample within some "good enough" finite precision, which is often omitted in most differential privacy papers [DLMV12]. Dodis et al. [DLMV12] builded finite-precision mechanisms under imperfect randomness $\mathcal{SV}(\delta)$.

Definition 6. *For $f : \mathcal{D} \to \mathbb{Z}$, the sensitivity of f is defined as*

$$\Delta f \stackrel{def}{=} \max_{D_1, D_2} \|f(D_1) - f(D_2)\|$$

for all neighboring databases $D_1, D_2 \in \mathcal{D}$. For $d \in \mathbb{Z}^+$, denote $\mathcal{F}_d = \{f : \mathcal{D} \to \mathbb{Z} \mid \Delta f \leq d\}$ as the class of functions with sensitivity at most d.

For clarity, in this paper we only consider the case $d = 1$. It's straightforward to extend all our results to any sensitivity bound d.

Definition 7. *The Laplace (or double exponential) distribution with mean μ and standard deviation $\frac{\sqrt{2}}{\varepsilon}$, denoted as $\mathsf{Lap}_{\mu,\frac{1}{\varepsilon}}$, has probability density function*

$$PDF^{\mathsf{Lap}}_{\mu,\frac{1}{\varepsilon}}(x) = \frac{\varepsilon}{2} \cdot e^{-\varepsilon|x-\mu|}.$$

The cumulative distribution function is given by

$$CDF^{\mathsf{Lap}}_{\mu,\frac{1}{\varepsilon}}(x) = \begin{cases} \frac{1}{2} \cdot e^{\varepsilon \cdot (x-\mu)}, & if\ x < \mu; \\ 1 - \frac{1}{2} \cdot e^{-\varepsilon \cdot (x-\mu)}, & if\ x \geq \mu. \end{cases}$$

(Equivalently, $CDF^{\mathsf{Lap}}_{\mu,\frac{1}{\varepsilon}}(x) = \frac{1}{2} + \frac{1}{2} \cdot sgn(x-\mu) \cdot (1 - e^{-\varepsilon \cdot |x-\mu|})$.)
If a random variable X has this distribution, denote $X \sim \mathsf{Lap}_{\mu,\frac{1}{\varepsilon}}$.

In what follows, we consider the case that $\frac{1}{\varepsilon} \in \mathbb{Z}$.

3 Compact BCL-Consistent Sampling and SV-Consistent Sampling

Dodis et al. [DLMV12] introduced the concept of SV-consistent sampling. However, the proof of "SV-consistent sampling implies differential privacy" (see Theorem 4.4 in [DLMV12] for details) is complex. Moreover, its natural expansion to BCL source is difficult and unknown to achieve differential privacy, as the proof of Theorem 4.4 in [DLMV12] depends on the fact that the values in T_2 (resp. T_1) constitutes an interval (see Definition 9), while it may not be the case for BCL sources.

In this section, we introduce the concept of compact (ζ, c)-BCL-consistent sampling. When $b = 0$, we get the concept of compact (ζ', c)-SV-consistent sampling. Then we observe that these concepts are sufficient to design finite-precision differentially private and accurate mechanisms based on BCL and SV sources.

Consider a mechanism M with randomness space $\{0,1\}^*$. Denote $\widetilde{T}(D_i, f, z) \stackrel{def}{=} \{\mathbf{r} \in \{0,1\}^* \mid z = M(D_i, f; \mathbf{r})\}$, where $i \in \{1,2\}$, as the set of all coins such that M outputs z when running on two neighboring databases D_1 and D_2, query f, and randomness $\mathbf{r}$. It should be noted that in our model only $n \stackrel{def}{=} \tilde{\tau}(D_1, D_2, f, z)$ coins need to be sampled to determine if $M(D_1, f) = z$ and $M(D_2, f) = z$. Therefore, let $T(D_i, f, z) \stackrel{def}{=} \{\mathbf{r} \in \{0,1\}^n \mid z = M(D_i, f; \mathbf{r})\}$ for $i \in \{1,2\}$, we can assume that $\widetilde{T}(D_i, f, z) = T(D_i, f, z)$ for $i \in \{1,2\}$ without loss of generality.

For $m \in \mathbb{Z}^+$ and a bit sequence $\mathbf{x} = x_1, \ldots, x_m \in \{0,1\}^m$, denote $\mathrm{SUFFIX}(\mathbf{x}) \stackrel{def}{=} \{\mathbf{y} = y_1, y_2, \ldots \in \{0,1\}^* \mid x_i = y_i\ for\ all\ i \in [m]\}$ as the set of all bit strings that have $\mathbf{x}$ as a prefix. For $n \in \mathbb{Z}^+$ such that $m \leq n$, we define $\mathrm{SUFFIX}(\mathbf{x}, n) \stackrel{def}{=} \mathrm{SUFFIX}(\mathbf{x}) \cap \{0,1\}^n$.

Now consider two neighboring databases D_1 and D_2, $f \in \mathcal{F}$, and a possible outcome z. Denote $n \stackrel{def}{=} \tilde{\tau}(D_1, D_2, f, z)$. Let $T_1 \stackrel{def}{=} T(D_1, f, z)$, $T_2 \stackrel{def}{=} T(D_2, f, z)$, and

$$\mathbf{u} \stackrel{def}{=} \arg\max\{|\mathbf{u}'| \mid \mathbf{u}' \in \{0,1\}^{\leq n}\ and\ T_1 \cup T_2 \subseteq \mathrm{SUFFIX}(\mathbf{u}', n)\}.$$

Then the ratio is

$$\frac{\Pr\limits_{\mathbf{r}\leftarrow BCL(\delta,b,n)}[\mathbf{r}\in T_1\backslash T_2]}{\Pr\limits_{\mathbf{r}\leftarrow BCL(\delta,b,n)}[\mathbf{r}\in T_2]}=\frac{\Pr\limits_{\mathbf{r}\leftarrow BCL(\delta,b,n)}[\mathbf{r}\in T_1\backslash T_2\mid \mathbf{r}\in \mathrm{SUFFIX}(\mathbf{u})]}{\Pr\limits_{\mathbf{r}\leftarrow BCL(\delta,b,n)}[\mathbf{r}\in T_2\mid \mathbf{r}\in \mathrm{SUFFIX}(\mathbf{u})]}.$$

Since SV source and BCL source both generate strings bit by bit, the calculation of the ratio can be simplified.

Recall that the concepts of consistent sampling and SV-consistent sampling [DLMV12] are as follows.

Definition 8. *Let $\zeta>0$. A mechanism M has ζ-consistent sampling if for all potential outputs $z\in\mathbb{Z}$, all queries $f\in\mathcal{F}$, all neighboring databases $D_1,D_2\in\mathcal{D}$:*

$$\frac{|T_1\setminus T_2|}{|T_2|}\le\zeta,$$

where $T_1\stackrel{def}{=}T(D_1,f,z)$, $T_2\stackrel{def}{=}T(D_2,f,z)\neq\emptyset$.

Definition 9. *Let $\tilde{c}>1$ be a constant and $\zeta>0$. We say that a mechanism M is an interval mechanism (see [DLMV12]) if for all queries $f\in\mathcal{F}$, all databases $D\in\mathcal{D}$, and all possible outcomes $z\in\mathbb{Z}$, $T\stackrel{def}{=}T(D,f,z)$, the set $\{\sum\limits_{i=1}^{n}r_i\cdot 2^{n-i}\mid r_1\dots r_n\in T\}$ contains consecutive integers. An interval mechanism has $(\zeta,\tilde{c})$-SV-consistent sampling if it has ζ-consistent sampling and for all queries $f\in\mathcal{F}$, all neighboring databases $D_1,D_2\in\mathcal{D}$, and all possible outcomes $z\in\mathbb{Z}$, which define T_1,T_2 and $\mathbf{u}$ as above, we have $\frac{|SUFFIX(\mathbf{u},n)|}{|T_1\cup T_2|}\le\tilde{c}$.*

It should be noted that when $b\neq 0$, $\mathcal{BCL}(\delta,b,n)$ can't generate all n-bit strings. The corresponding mechanism can't be an interval mechanism. Though Dodis et al. [DLMV12] proposed that if M has $(\zeta,\tilde{c})$-SV-consistent sampling, then M is $(\mathcal{SV}(\delta),\xi)$-differentially private. In that proof, the "interval" property is a basic condition, we can't follow that thought. We resort to a new property instead.

Definition 10. *Let c be a constant and $\zeta>0$. A mechanism is a compact (ζ,c)-BCL-consistent sampling mechanism if it has ζ-consistent sampling and for all queries $f\in\mathcal{F}$, all neighboring databases D_1, $D_2\in\mathcal{D}$, and all possible outcomes $z\in\mathbb{Z}$, which define T_1,T_2 and $\mathbf{u}$ as above, we have $n-|\mathbf{u}|\le c$.*

Definition 11. *Let c be a constant and $\zeta'>0$. A mechanism is a compact (ζ',c)-SV-consistent sampling mechanism if it has ζ'-consistent sampling and for all queries $f\in\mathcal{F}$, all neighboring databases D_1, $D_2\in\mathcal{D}$, and all possible outcomes $z\in\mathbb{Z}$, which define T_1,T_2 and $\mathbf{u}$ as above, we have $n-|\mathbf{u}|\le c$.*

Theorem 1. *If Mechanism M is a compact (ζ,c)-BCL-consistent sampling mechanism for (δ,b)-BCL-sources, then M is $(\mathcal{BCL}(\delta,b),\xi)$-differentially private, where $\xi\le(\frac{1+\delta}{1-\delta})^c\cdot[\frac{1}{2}(1+\delta)]^{-b}\cdot\zeta$. In particular, for $\delta\in[0,1)$, and $c=O(1)$, we have $\lim\limits_{\zeta\to 0}(\frac{1+\delta}{1-\delta})^c\cdot[\frac{1}{2}(1+\delta)]^{-b}\cdot\zeta=0$.*

Proof. Assume that $\frac{|T_1\backslash T_2|}{|T_2|}\le\zeta$.

For any $\mathbf{r}\in\{0,1\}^n$ and $\mathbf{r}'\in\{0,1\}^n$, denote $\mathbf{r}=r_1r_2\dots r_n$ where $r_i\in\{0,1\}$ for $i\in[n]$ and $\mathbf{r}'=r'_1r'_2\dots r'_n$ where $r'_i\in\{0,1\}$ for $i\in[n]$.

Assume that $n-|\mathbf{u}|\le c$, then

$$\begin{aligned}
&\frac{\Pr\limits_{\mathbf{r}\leftarrow BCL(\delta,b,n)}[\mathbf{r}\in T_1\backslash T_2]}{\Pr\limits_{\mathbf{r}\leftarrow BCL(\delta,b,n)}[\mathbf{r}\in T_2]}=\frac{\Pr\limits_{\mathbf{r}\leftarrow BCL(\delta,b,n)}[\mathbf{r}\in T_1\backslash T_2\mid \mathbf{r}\in \mathrm{SUFFIX}(\mathbf{u})]}{\Pr\limits_{\mathbf{r}\leftarrow BCL(\delta,b,n)}[\mathbf{r}\in T_2\mid \mathbf{r}\in \mathrm{SUFFIX}(\mathbf{u})]}\\
&=\frac{\sum\limits_{\mathbf{r}'\in T_1\backslash T_2}\Pr\limits_{\mathbf{r}\leftarrow BCL(\delta,b,n)}[\mathbf{r}=\mathbf{r}'\mid \mathbf{r}'\in\mathrm{SUFFIX}(\mathbf{u})]}{\sum\limits_{\mathbf{r}'\in T_2}\Pr\limits_{\mathbf{r}\leftarrow BCL(\delta,b,n)}[\mathbf{r}=\mathbf{r}'\mid \mathbf{r}'\in\mathrm{SUFFIX}(\mathbf{u})]}\\
&=\frac{\sum\limits_{\mathbf{r}'\in T_1\backslash T_2}\Pr\limits_{\mathbf{r}\leftarrow BCL(\delta,b,n)}[r_{|\mathbf{u}|+1}=r'_{|\mathbf{u}|+1}\mid r_1\dots r_{|\mathbf{u}|}=\mathbf{u}]\dots\Pr\limits_{\mathbf{r}\leftarrow BCL(\delta,b,n)}[r_n=r'_n\mid r_1\dots r_{n-1}=\mathbf{u}r'_{|\mathbf{u}|+1}\dots r'_{n-1}]}{\sum\limits_{\mathbf{r}'\in T_2}\Pr\limits_{\mathbf{r}\leftarrow BCL(\delta,b,n)}[r_{|\mathbf{u}|+1}=r'_{|\mathbf{u}|+1}\mid r_1\dots r_{|\mathbf{u}|}=\mathbf{u}]\dots\Pr\limits_{\mathbf{r}\leftarrow BCL(\delta,b,n)}[r_n=r'_n\mid r_1\dots r_{n-1}=\mathbf{u}r'_{|\mathbf{u}|+1}\dots r'_{n-1}]}\\
&\le\frac{[\frac{1}{2}(1+\delta)]^{n-|\mathbf{u}|-b}}{[\frac{1}{2}(1-\delta)]^{n-|\mathbf{u}|}}\cdot\frac{|T_1\backslash T_2|}{|T_2|}\le(\frac{1+\delta}{1-\delta})^{n-|\mathbf{u}|}\cdot[\frac{1}{2}(1+\delta)]^{-b}\cdot\zeta\le(\frac{1+\delta}{1-\delta})^{c}\cdot[\frac{1}{2}(1+\delta)]^{-b}\cdot\zeta
\end{aligned}$$

□

Let $b = 0$. We get that

Corollary 1. *If Mechanism M is a compact (ζ', c)-SV-consistent sampling mechanism, then M is $(\mathcal{SV}(\delta), \xi)$-differentially private, where $\xi \leq (\frac{1+\delta}{1-\delta})^c \cdot \zeta'$. In particular, for $\delta \in [0,1)$, and $c = O(1)$, we have $\lim_{\zeta' \to 0} (\frac{1+\delta}{1-\delta})^c \cdot \zeta' = 0$.*

4 Accurate and Private BCLCS Mechanisms

In this section, we construct finite-precision mechanisms that achieve compact $(\zeta, O(1))$-BCL-consistent sampling with sensitivity 1. We also propose that the precision of the specific mechanism based on Laplace distribution introduced by Dodis et al. [DLMV12] can be modified via this technique such that it becomes a compact SV-consistent sampling mechanism. Then by Theorems 1 and 2, the mechanism here and the modified mechanism of [DLMV12] are $(\mathcal{BCL}(\delta, b), \xi)$-differentially private and $(\mathcal{SV}(\delta), \xi')$-differentially private, where ξ' is a specific ξ by letting $b = 0$. We also show that these mechanisms have good bound on utility when the random sampling is generated from the BCL source.

4.1 Explicit Construction

We construct an infinite-precision mechanism, called $M_\varepsilon^{\mathsf{CBCLCS}}$, then modify it to a finite precision one, denoted as $\overline{M}_\varepsilon^{\mathsf{CBCLCS}}$. Recall that some truncation method was proposed in [DLMV12] in order to get a finite mechanism, which leads to the nonintuitive notion of SV-consistent sampling. However, it can't be transplanted to BCL sources. In this section, we develop another truncation technique. The finite-precision mechanism is designed as follows.

Explicit Construction of the Mechanism:

Step 1 *On input any neighboring databases $D_1, D_2 \in \mathcal{D}$, $f \in \mathcal{F}$, the infinite-precision mechanism $M_\varepsilon^{\mathsf{CBCLCS}}$ computes $f(D_1)$ and $f(D_2)$. Without loss of generality, assume that $f(D_1) = y$ and $f(D_2) = y-1$. $M_\varepsilon^{\mathsf{CBCLCS}}(D_1, f)$ (resp. $M_\varepsilon^{\mathsf{CBCLCS}}(D_2, f)$) outputs $z_1 \leftarrow \frac{1}{\varepsilon} \cdot \lfloor \varepsilon \cdot (y + \mathsf{Lap}_{0,\frac{1}{\varepsilon}}) \rceil$ (resp. $z_2 \leftarrow \frac{1}{\varepsilon} \cdot \lfloor \varepsilon \cdot (y - 1 + \mathsf{Lap}_{0,\frac{1}{\varepsilon}}) \rceil$). Denote Z_y (resp. Z_{y-1}) as the output distribution of $M_\varepsilon^{\mathsf{CBCLCS}}(D_1, f)$ (resp. $M_\varepsilon^{\mathsf{CBCLCS}}(D_2, f)$) using arithmetic coding (see [DLMV12]).*

Step 2 *Suppose that y is fixed. Let $s_y(k) \stackrel{def}{=} \mathsf{CDF}_{y,\frac{1}{\varepsilon}}^{\mathsf{Lap}}(\frac{k+\frac{1}{2}}{\varepsilon})$ and $s_{y-1}(k) \stackrel{def}{=} \mathsf{CDF}_{y-1,\frac{1}{\varepsilon}}^{\mathsf{Lap}}(\frac{k+\frac{1}{2}}{\varepsilon})$ for all $k \in \mathbb{Z}$. Denote $I_y(k) = [s_y(k-1), s_y(k))$ and $I_{y-1}(k) = [s_{y-1}(k-1), s_{y-1}(k))$. Let $\bar{s}_{y-1}(k-1)$ (resp. $\bar{s}_{y-1}(k)$) be $s_{y-1}(k-1)$ (resp. $s_{y-1}(k)$), rounded to the first $n \stackrel{def}{=} \tau(\min(f(D_1), f(D_2)), k/\varepsilon) = \tau(y-1, k/\varepsilon)$ bits after the binary point. We round $s_y(k-1)$ (resp. $s_y(k)$) to the first $n \stackrel{def}{=} \tau(y-1, k/\varepsilon)$ bits after the binary point. Assume the binary decimal representation of the rounded $s_y(k-1)$ (resp. $s_y(k)$) is $0.r_1r_2\ldots r_n$ (resp. $0.q_1q_2\ldots q_n$), then let $\bar{s}_y(k-1) = 0.r_1r_2\ldots r_n + 0.r'_1r'_2\ldots r'_n$ (resp. $\bar{s}_y(k) = 0.q_1q_2\ldots q_n + 0.q'_1q'_2\ldots q'_n$), where $r'_i = 0$ for $i \in [n-1]$, and $r'_n = 1$ (resp. $q'_i = 0$ for $i \in [n-1]$ and $q'_n = 1$). Denote $\bar{I}_{y-1}(k) = [\bar{s}_{y-1}(k-1), \bar{s}_{y-1}(k))$ and $\bar{I}_y(k) = [\bar{s}_y(k-1), \bar{s}_y(k))$.*

Step 3 *Denote $\overline{Z}_y$ (resp. $\overline{Z}_{y-1}$) as the output distribution of $\overline{M}_\varepsilon^{\mathsf{CBCLCS}}(D_1, f)$ (resp. $\overline{M}_\varepsilon^{\mathsf{CBCLCS}}(D_2, f)$), which approximates Z_y (resp. Z_{y-1}). For any sequence $\mathbf{r} = r_1, r_2, \ldots, r_n \in \{0,1\}^n$, the real representation of $\mathbf{r}$ is $REAL(\mathbf{r}) \stackrel{def}{=} 0.r_1r_2\ldots r_n \in [0,1]$. We obtain distribution $\overline{Z}_y$ (resp. $\overline{Z}_{y-1}$) by sampling a sequence of bits $\mathbf{r} \in \{0,1\}^n$ (resp. $\mathbf{r}' \in \{0,1\}^n$) from a (δ, b, n)-BCL distribution and outputting $\frac{k_1}{\varepsilon}$ (resp. $\frac{k_2}{\varepsilon}$) where $k_1 \in \mathbb{Z}$ (resp. $k_2 \in \mathbb{Z}$) is the unique integer such that $REAL(\mathbf{r}) \in \bar{I}_y(k_1)$ (resp. $REAL(\mathbf{r}') \in \bar{I}_{y-1}(k_2)$).*

□

From the above construction, for all $k \in \mathbb{Z}$, we have

$$\frac{\Pr[M_\varepsilon^{\mathsf{CBCLCS}}(D_1, f) = \frac{k}{\varepsilon}]}{\Pr[M_\varepsilon^{\mathsf{CBCLCS}}(D_2, f) = \frac{k}{\varepsilon}]} = \frac{\Pr[\frac{k-\frac{1}{2}}{\varepsilon} \leq y + \mathsf{Lap}_{0,\frac{1}{\varepsilon}} < \frac{k+\frac{1}{2}}{\varepsilon}]}{\Pr[\frac{k-\frac{1}{2}}{\varepsilon} \leq y - 1 + \mathsf{Lap}_{0,\frac{1}{\varepsilon}} < \frac{k+\frac{1}{2}}{\varepsilon}]} = \frac{\Pr[\frac{k-\frac{1}{2}}{\varepsilon} \leq \mathsf{Lap}_{y,\frac{1}{\varepsilon}} < \frac{k+\frac{1}{2}}{\varepsilon}]}{\Pr[\frac{k-\frac{1}{2}}{\varepsilon} \leq \mathsf{Lap}_{y-1,\frac{1}{\varepsilon}} < \frac{k+\frac{1}{2}}{\varepsilon}]}.$$

$$\frac{\Pr[\overline{M}_{\varepsilon}^{\mathsf{CBCLCS}}(D_1,f)=\frac{k}{\varepsilon}]}{\Pr[\overline{M}_{\varepsilon}^{\mathsf{CBCLCS}}(D_2,f)=\frac{k}{\varepsilon}]}=\frac{\Pr[\overline{Z}_y=\frac{k}{\varepsilon}]}{\Pr[\overline{Z}_{y-1}=\frac{k}{\varepsilon}]}=\frac{|\bar{I}_y(k)|}{|\bar{I}_{y-1}(k)|}.$$

Remark 2. It's easy to prove that $I_{y-1}(k)\cap I_y(k)\neq\emptyset$; in fact, we can view $I_{y-1}(k)$ as having "shifted" $I_y(k)$ slightly to the right. The set of points $\{s_y(k)\}_{k\in\mathbb{Z}}$ partitions the interval [0, 1] into infinitely many intervals $\{I_y(k)\overset{def}{=}[s_y(k-1),s_y(k))\}_{k\in\mathbb{Z}}$. Similarly, the set of points $\{s_{y-1}(k)\}_{k\in\mathbb{Z}}$ partitions the interval [0, 1] into infinitely many intervals $\{I_{y-1}(k)\overset{def}{=}[s_{y-1}(k-1),s_{y-1}(k))\}_{k\in\mathbb{Z}}$.

Remark 3. When $b=0$, the above construction degenerates into the mechanism based on SV source. Correspondingly, $\overline{M}_{\varepsilon}^{\mathsf{CBCLCS}}$ will be replaced with $\overline{M}_{\varepsilon}^{\mathsf{CSVCS}}$.

4.2 Proofs of Differential Privacy and Accuracy

In this section, we first show that our construction satisfies compact $(\zeta,O(1))$-BCL-consistent sampling, and then prove its accuracy. Finally, we give a relation between the result of [DY14] and ours.

The lemma below is one core step to achieve consistent sampling. Though it has essentially been proved by Dodis et al. [DLMV12], there still exist some typos and the upper bound is not tight prior our work. Hence, we modify the Lemma A.1 of [DLMV12] and prove it for completeness. The comparison of them can be seen in the Appendix.

Lemma 1. *Denote* $I'_y(k)\overset{def}{=}I_y(k)\setminus I_{y-1}(k)=[s_y(k-1),s_{y-1}(k-1))$. *For all* $y,k\in\mathbb{Z}$ *and* $\varepsilon\in(0,1)$, *we have*

$$\frac{|I'_y(k)|}{|I_{y-1}(k)|}<e\cdot\varepsilon.$$

Proof. Note that if $x<y$, $\mathsf{CDF}^{\mathsf{Lap}}_{y,\frac{1}{\varepsilon}}(x)<\frac{1}{2}$; otherwise, $\mathsf{CDF}^{\mathsf{Lap}}_{y,\frac{1}{\varepsilon}}(x)\geq\frac{1}{2}$.

$$\frac{|I'_y(k)|}{|I_{y-1}(k)|}=\frac{s_{y-1}(k-1)-s_y(k-1)}{s_{y-1}(k)-s_{y-1}(k-1)}=\frac{\mathsf{CDF}^{\mathsf{Lap}}_{y-1,\frac{1}{\varepsilon}}(\frac{k-\frac{1}{2}}{\varepsilon})-\mathsf{CDF}^{\mathsf{Lap}}_{y,\frac{1}{\varepsilon}}(\frac{k-\frac{1}{2}}{\varepsilon})}{\mathsf{CDF}^{\mathsf{Lap}}_{y-1,\frac{1}{\varepsilon}}(\frac{k+\frac{1}{2}}{\varepsilon})-\mathsf{CDF}^{\mathsf{Lap}}_{y-1,\frac{1}{\varepsilon}}(\frac{k-\frac{1}{2}}{\varepsilon})}.$$

We consider four cases:
Case 1: If $\frac{1}{2}\leq s_y(k-1)<s_{y-1}(k-1)<s_{y-1}(k)$, then $\frac{|I'_y(k)|}{|I_{y-1}(k)|}=\frac{e^{\varepsilon+1}-e}{e-1}$.
Case 2: If $s_y(k-1)<\frac{1}{2}\leq s_{y-1}(k-1)<s_{y-1}(k)$, then

$$\frac{|I'_y(k)|}{|I_{y-1}(k)|}=\frac{1-\frac{1}{2}\cdot e^{-\varepsilon[\frac{k-\frac{1}{2}}{\varepsilon}-(y-1)]}-\frac{1}{2}\cdot e^{\varepsilon(\frac{k-\frac{1}{2}}{\varepsilon}-y)}}{1-\frac{1}{2}\cdot e^{-\varepsilon[\frac{k+\frac{1}{2}}{\varepsilon}-(y-1)]}-\{1-\frac{1}{2}\cdot e^{-\varepsilon[\frac{k-\frac{1}{2}}{\varepsilon}-(y-1)]}\}}.$$

For simplicity, denote $v\overset{def}{=}\frac{k-\frac{1}{2}}{\varepsilon}-y$. By the assumption, we have that $-1\leq v<0$. Correspondingly,

$$\frac{|I'_y(k)|}{|I_{y-1}(k)|}=\frac{1-\frac{1}{2}e^{-\varepsilon(v+1)}-\frac{1}{2}e^{\varepsilon v}}{-\frac{1}{2}e^{-\varepsilon(v+1+\frac{1}{\varepsilon})}+\frac{1}{2}e^{-\varepsilon(v+1)}}=\frac{-(e^{\varepsilon v}-1)^2-e^{-\varepsilon}+1}{-e^{-1-\varepsilon}+e^{-\varepsilon}}\leq\frac{-e^{-\varepsilon}+1}{-e^{-1-\varepsilon}+e^{-\varepsilon}}=\frac{e^{\varepsilon+1}-e}{e-1}.$$

Case 3: If $s_y(k-1)<s_{y-1}(k-1)<\frac{1}{2}\leq s_{y-1}(k)$, then

$$\frac{|I'_y(k)|}{|I_{y-1}(k)|}=\frac{\frac{1}{2}\cdot e^{\varepsilon[\frac{k-\frac{1}{2}}{\varepsilon}-(y-1)]}-\frac{1}{2}\cdot e^{\varepsilon(\frac{k-\frac{1}{2}}{\varepsilon}-y)}}{1-\frac{1}{2}\cdot e^{-\varepsilon[\frac{k+\frac{1}{2}}{\varepsilon}-(y-1)]}-\frac{1}{2}\cdot e^{\varepsilon[\frac{k-\frac{1}{2}}{\varepsilon}-(y-1)]}}.$$

For simplicity, denote $v\overset{def}{=}\frac{k-\frac{1}{2}}{\varepsilon}-y$. By the assumption, we have that $-1-\frac{1}{\varepsilon}\leq v<-1$. Correspondingly,

$$\begin{aligned}\frac{|I'_y(k)|}{|I_{y-1}(k)|} &= \frac{\frac{1}{2}\cdot e^{\varepsilon(v+1)} - \frac{1}{2}\cdot e^{\varepsilon v}}{1-\frac{1}{2}\cdot e^{-\varepsilon(v+\frac{1}{\varepsilon}+1)} - \frac{1}{2}\cdot e^{\varepsilon(v+1)}} \\ &= \frac{e^{\varepsilon}-1}{2\cdot e^{-\varepsilon v} - e^{-2\varepsilon v-\varepsilon-1} - e^{\varepsilon}} \\ &= \frac{e^{\varepsilon}-1}{-(e^{-\varepsilon v-\frac{1+\varepsilon}{2}} - e^{\frac{1+\varepsilon}{2}})^2 + e^{1+\varepsilon} - e^{\varepsilon}} \\ &< \frac{e^{\varepsilon}-1}{-(e^{\frac{\varepsilon-1}{2}} - e^{\frac{1+\varepsilon}{2}})^2 + e^{1+\varepsilon} - e^{\varepsilon}} \\ &= \frac{1-e^{-\varepsilon}}{1-e^{-1}}.\end{aligned}$$

Case 4: If $s_y(k-1) < s_{y-1}(k-1) < s_{y-1}(k) < \frac{1}{2}$, then

$$\frac{|I'_y(k)|}{|I_{y-1}(k)|} = \frac{\frac{1}{2}\cdot e^{\varepsilon[\frac{k-\frac{1}{2}}{\varepsilon}-(y-1)]} - \frac{1}{2}\cdot e^{\varepsilon(\frac{k-\frac{1}{2}}{\varepsilon}-y)}}{\frac{1}{2}\cdot e^{\varepsilon[\frac{k+\frac{1}{2}}{\varepsilon}-(y-1)]} - \frac{1}{2}\cdot e^{\varepsilon[\frac{k-\frac{1}{2}}{\varepsilon}-(y-1)]}} = \frac{1-e^{-\varepsilon}}{e-1}.$$

For $\varepsilon \in (0,1)$, we have

$$\frac{1-e^{-\varepsilon}}{e-1} < \frac{1-e^{-\varepsilon}}{1-e^{-1}} = \frac{e-e^{1-\varepsilon}}{e-1} < \frac{e^{\varepsilon}\cdot(e-e^{1-\varepsilon})}{e-1} = \frac{e^{\varepsilon+1}-e}{e-1} < e\cdot\varepsilon.$$

The last inequality holds according to the following three facts: (1) $g_1(x) \stackrel{def}{=} \frac{e^{x+1}-e}{e-1}$ is a convex function; (2) $g_2(x) \stackrel{def}{=} e\cdot x$ is a linear function; (3) $g_1(0) = g_2(0)$ and $g_1(1) = g_2(1)$.

□

Denote $I''_y(k) \stackrel{def}{=} I_{y-1}(k) \setminus I_y(k) = [s_y(k), s_{y-1}(k))$. Similarly, we can get that that there exists a constant C such that $\frac{|I''_y(k)|}{|I_y(k)|} < C\cdot\varepsilon$ for $y,k \in \mathbb{Z}$ and $\varepsilon \in (0,1)$. Due to space limitations, we will omit the corresponding discussion.

Lemma 2. *For all $y,k \in \mathbb{Z}$, we have*

$$|\bar{I}'_y(k)| \le |I'_y(k)|,$$
$$|I_{y-1}(k)| + 2^{-\tau(y-1,k/\varepsilon)} \ge |\bar{I}_{y-1}(k)| \ge |I_{y-1}(k)| - 2^{-\tau(y-1,k/\varepsilon)},$$
$$|I_y(k)| + 2^{-\tau(y-1,k/\varepsilon)} \ge |\bar{I}_y(k)| \ge |I_y(k)| - 2^{-\tau(y-1,k/\varepsilon)}.$$

Proof. Since $s_{y-1}(k-1) \ge \bar{s}_{y-1}(k-1)$, and $\bar{s}_y(k-1) \ge s_y(k-1) - 2^{-\tau(y-1,k/\varepsilon)} + 2^{-\tau(y-1,k/\varepsilon)}$, we get $|\bar{I}'_y(k)| \le |I'_y(k)|$.

Since $\bar{s}_{y-1}(k) \ge s_{y-1}(k) - 2^{-\tau(y-1,k/\varepsilon)}$ and $\bar{s}_{y-1}(k-1) \le s_{y-1}(k-1)$, we have $|\bar{I}_{y-1}(k)| \ge |I_{y-1}(k)| - 2^{-\tau(y-1,k/\varepsilon)}$.

Since $s_{y-1}(k) \ge \bar{s}_{y-1}(k)$ and $s_{y-1}(k-1) \le \bar{s}_{y-1}(k-1) + 2^{-\tau(y-1,k/\varepsilon)}$, we have $|I_{y-1}(k)| + 2^{-\tau(y-1,k/\varepsilon)} \ge |\bar{I}_{y-1}(k)|$.

Since $\bar{s}_y(k) \ge s_y(k) - 2^{-\tau(y-1,k/\varepsilon)} + 2^{-\tau(y-1,k/\varepsilon)}$ and $\bar{s}_y(k-1) \le s_y(k-1) + 2^{-\tau(y-1,k/\varepsilon)}$, we have $|\bar{I}_y(k)| \ge |I_y(k)| - 2^{-\tau(y-1,k/\varepsilon)}$.

Since $\bar{s}_y(k) \le s_y(k) + 2^{-\tau(y-1,k/\varepsilon)}$ and $\bar{s}_y(k-1) \ge s_y(k-1)$, we have $|\bar{I}_y(k)| \le |I_y(k)| + 2^{-\tau(y-1,k/\varepsilon)}$.

□

Assume that Y is a (δ,b,n)-BCL distribution, and $S_0 = \{\mathbf{r} \in \{0,1\}^n \mid \Pr[Y=\mathbf{r}] \neq 0\}$. Denote $\text{STR}\ (I,n) \stackrel{def}{=} \{\mathbf{r} \in \{0,1\}^n \mid \text{REAL}(\mathbf{r}) \in I\}$ as the the set of all n-bit strings whose real representation lies in I. Let $T_1 = \text{STR}\ (\bar{I}_y(k),n) \cap S_0$ and $T_2 = \text{STR}\ (\bar{I}_{y-1}(k),n) \cap S_0$. Then $T_1 \setminus T_2 = \text{STR}\ (\bar{I}'_y(k),n) \cap S_0$.

Lemma 3.

$$2^{n-b} \leq |S_0| \leq 2^n.$$

Proof. We can easily prove it by induction.

□

Lemma 4. *Denote* $\tau(y-1,k/\varepsilon) \stackrel{def}{=} \log \frac{1}{|I_{y-1}(k)|} + \log(2^b+1)$. *For all* $y,k \in \mathbb{Z}$, *we have*

$$|STR\ (\bar{I'}_y(k), \tau(y-1,k/\varepsilon)) \cap S_0| \leq (2^{-b}+1)\cdot e \cdot \varepsilon,\ |STR\ (\bar{I}_{y-1}(k), \tau(y-1,k/\varepsilon)) \cap S_0| \geq 1.$$

Proof. Let $n \stackrel{def}{=} \tau(y-1,k/\varepsilon)$ for shorthand. Consider $|\bar{I'}_y(k)|$ as the probability of sampling a sequence $\mathbf{r}$ from U_{S_0} such that $\mathbf{r} \in \text{STR}\ (\bar{I'}_y(k), n) \cap S_0$, where $2^{n-b} \leq |S_0| \leq 2^n$. Hence,

$$|\bar{I'}_y(k)| = \sum_{\mathbf{r} \in STR(\bar{I'}_y(k),n) \cap S_0} \frac{1}{|S_0|} \geq \sum_{\mathbf{r} \in STR(\bar{I'}_y(k),n) \cap S_0} \frac{1}{2^n} = |STR(\bar{I'}_y(k), n) \cap S_0| \cdot \frac{1}{2^n}.$$

Therefore,

$$\begin{aligned} &|\text{STR}\ (\bar{I'}_y(k), n) \cap S_0| \leq 2^n \cdot |\bar{I'}_y(k)| \leq 2^n \cdot |I'_y(k)| \\ &= (2^b+1) \cdot \frac{|I'_y(k)|}{|I_{y-1}(k)|} \leq (2^b+1)\cdot e \cdot \varepsilon. \end{aligned}$$

Since

$$|\bar{I}_{y-1}(k)| = \sum_{\mathbf{r} \in \text{STR}\ (\bar{I}_{y-1}(k),n) \cap S_0} \frac{1}{|S_0|} \leq \sum_{\mathbf{r} \in \text{STR}\ (\bar{I}_{y-1}(k),n) \cap S_0} (\frac{1}{2})^{n-b} = |\text{STR}\ (\bar{I}_{y-1}(k), n) \cap S_0| \cdot (\frac{1}{2})^{n-b},$$

$$|\bar{I}_y(k)| = \sum_{\mathbf{r} \in \text{STR}\ (\bar{I}_{y-1}(k),n) \cap S_0} \frac{1}{|S_0|} \leq \sum_{\mathbf{r} \in \text{STR}\ (\bar{I}_y(k),n) \cap S_0} (\frac{1}{2})^{n-b} = |\text{STR}\ (\bar{I}_y(k), n) \cap S_0| \cdot (\frac{1}{2})^{n-b},$$

we get

$$|\text{STR}\ (\bar{I}_{y-1}(k), n) \cap S_0| \geq 2^{n-b} \cdot |\bar{I}_{y-1}(k)| \geq 2^{n-b} \cdot (|I_{y-1}(k)| - 2^{-n}) = 2^{-b} + 1 - 2^{-b} = 1.$$

□

Remark 4. We can guarantee that n is legal, in the sense that the modification of the endpoints in $I_{y-1}(k)$ and $I_y(k)$ with respect to n does not cause intervals to "disappear" or for consecutive intervals to "overlap".

Theorem 2. *Denote* $\tau(y-1,k/\varepsilon) \stackrel{def}{=} \log \frac{1}{|I_{y-1}(k)|} + \log(2^b+1)$. *For all* $y,k \in \mathbb{Z}$, *we have*

$$\frac{|STR\ (\bar{I'}_y(k), \tau(y-1,k/\varepsilon)) \cap S_0|}{|STR\ (\bar{I}_{y-1}(k), \tau(y-1,k/\varepsilon)) \cap S_0|} \leq (2^b+1)\cdot e \cdot \varepsilon.$$

Proof. It's straightforward from Lemma 4.

□

Theorem 3. *Denote* $\tau(y-1,k/\varepsilon) \stackrel{def}{=} \log \frac{1}{|I_{y-1}(k)|} + \log(2^b+1)$. *Let* $\boldsymbol{u}$ *be the longest common prefix of all strings in* $\bar{I} \stackrel{def}{=} \bar{I}_y(k) \cup \bar{I}_{y-1}(k)$. *Then*

$$|SUFFIX(\mathbf{u}, \tau(y-1,k/\varepsilon)) \cap S_0| \leq \frac{e \cdot (2^b+1)}{1-e^{-1}}.$$

Proof. For simplicity, let $n \stackrel{def}{=} \tau(y-1, k/\varepsilon)$. Let $\mathbf{u}'$ be the longest common prefix of all strings in $I \stackrel{def}{=} I_y(k) \cup I_{y-1}(k)$. Then we have $|\text{SUFFIX}(\mathbf{u}, n)| \le |\text{SUFFIX}(\mathbf{u}', n)|$. We bound $|\text{SUFFIX}(\mathbf{u}, n)|$ by bounding the number of n-bit strings to the left or right of $\bar{I}$ (depending on which endpoint of the interval [0, 1] is closer to I).

Now we calculate the size of the interval $[s_y(k-1), 1]$ (resp. $[0, s_{y-1}(k)]$), which is an approximation of the size of $[\bar{s}_y(k-1), 1]$ (resp. $[0, \bar{s}_{y-1}(k)]$). Then we can upper bound how many n-bit strings there are in the interval $[\bar{s}_y(k-1), 1]$ (resp. $[0, \bar{s}_{y-1}(k)]$). Let $S \stackrel{def}{=} [s_y(k-1), 1]$.

Recall that $s_y(k) \stackrel{def}{=} \mathsf{CDF}^{\mathsf{Lap}}_{y,\frac{1}{\varepsilon}}(\frac{k+\frac{1}{2}}{\varepsilon})$ for all $k \in \mathbb{Z}$ and

$$\mathsf{CDF}^{\mathsf{Lap}}_{y,\frac{1}{\varepsilon}}(x) = \begin{cases} \frac{1}{2} \cdot e^{\varepsilon(x-y)}, & if\ x < y; \\ 1 - \frac{1}{2} \cdot e^{-\varepsilon(x-y)}, & if\ x \ge y. \end{cases}$$

Note that if $x < y$, $\mathsf{CDF}^{\mathsf{Lap}}_{y,\frac{1}{\varepsilon}}(x) < \frac{1}{2}$; otherwise, $\mathsf{CDF}^{\mathsf{Lap}}_{y,\frac{1}{\varepsilon}}(x) \ge \frac{1}{2}$.
$I'_y(k) = [s_y(k-1), s_{y-1}(k-1))$ and $I'_{y+1}(k) = [s_{y+1}(k-1), s_y(k-1))$.
For simplicity, denote $v \stackrel{def}{=} \frac{k-\frac{1}{2}}{\varepsilon} - y$. We consider four cases.
Case 1: Assume that $\frac{1}{2} \le s_{y+1}(k-1) < s_y(k-1) < s_{y-1}(k-1)$. Then $v \ge 1$.

$$\frac{|I'_y(k)|}{|I'_{y+1}(k)|} = \frac{1 - \frac{1}{2} \cdot e^{-\varepsilon[\frac{k-\frac{1}{2}}{\varepsilon} - (y-1)]} - 1 + \frac{1}{2} \cdot e^{-\varepsilon(\frac{k-\frac{1}{2}}{\varepsilon} - y)}}{1 - \frac{1}{2} \cdot e^{-\varepsilon(\frac{k-\frac{1}{2}}{\varepsilon} - y)} - 1 + \frac{1}{2} \cdot e^{-\varepsilon[\frac{k-\frac{1}{2}}{\varepsilon} - (y+1)]}} = \frac{1}{e^{\varepsilon}}.$$

Case 2: Assume that $s_{y+1}(k-1) < \frac{1}{2} \le s_y(k-1) < s_{y-1}(k-1)$. Then $0 \le v < 1$.

$$\frac{|I'_y(k)|}{|I'_{y+1}(k)|} = \frac{e^{-\varepsilon v} - e^{-\varepsilon(v+1)}}{2 - e^{-\varepsilon v} - e^{\varepsilon(v-1)}} = \frac{1 - e^{-\varepsilon}}{-e^{-\varepsilon}(e^{\varepsilon v} - e^{\varepsilon})^2 + e^{\varepsilon} - 1}.$$

Hence,

$$\frac{1}{e^{\varepsilon}} < \frac{|I'_y(k)|}{|I'_{y+1}(k)|} \le 1.$$

Case 3: Assume that $s_{y+1}(k-1) < s_y(k-1) < \frac{1}{2} \le s_{y-1}(k-1)$. Then $-1 \le v < 0$.

$$\begin{aligned} \frac{|I'_y(k)|}{|I'_{y+1}(k)|} &= \frac{1 - \frac{1}{2} \cdot e^{-\varepsilon[\frac{k-\frac{1}{2}}{\varepsilon} - (y-1)]} - \frac{1}{2} \cdot e^{\varepsilon(\frac{k-\frac{1}{2}}{\varepsilon} - y)}}{\frac{1}{2} \cdot e^{\varepsilon(\frac{k-\frac{1}{2}}{\varepsilon} - y)} - \frac{1}{2} \cdot e^{\varepsilon[\frac{k-\frac{1}{2}}{\varepsilon} - (y+1)]}} \\ &= \frac{1 - \frac{1}{2} \cdot e^{-\varepsilon(v+1)} - \frac{1}{2} \cdot e^{\varepsilon v}}{\frac{1}{2} \cdot e^{\varepsilon v} - \frac{1}{2} \cdot e^{\varepsilon(v-1)}} \\ &= \frac{-(e^{-\varepsilon v - \frac{\varepsilon}{2}} - e^{\frac{\varepsilon}{2}})^2 + e^{\varepsilon} - 1}{1 - e^{-\varepsilon}}. \end{aligned}$$

Therefore,

$$1 < \frac{|I'_y(k)|}{|I'_{y+1}(k)|} \le e^{\varepsilon}.$$

Case 4: Assume that $s_{y+1}(k-1) < s_y(k-1) < s_{y-1}(k-1) < \frac{1}{2}$. Then $v < -1$.

$$\frac{|I'_y(k)|}{|I'_{y+1}(k)|} = \frac{\frac{1}{2} \cdot e^{\varepsilon[\frac{k-\frac{1}{2}}{\varepsilon} - (y-1)]} - \frac{1}{2} \cdot e^{\varepsilon(\frac{k-\frac{1}{2}}{\varepsilon} - y)}}{\frac{1}{2} \cdot e^{\varepsilon(\frac{k-\frac{1}{2}}{\varepsilon} - y)} - \frac{1}{2} \cdot e^{\varepsilon[\frac{k-\frac{1}{2}}{\varepsilon} - (y+1)]}} = \frac{\frac{1}{2} \cdot e^{\varepsilon(v+1)} - \frac{1}{2} \cdot e^{\varepsilon v}}{\frac{1}{2} \cdot e^{\varepsilon v} - \frac{1}{2} \cdot e^{\varepsilon(v-1)}} = e^{\varepsilon}.$$

We only analyze Case 1, the other cases are analogous.

Since $I'_y(k)$ and $I'_{y+1}(k)$ are consecutive intervals for all $y \in \mathbb{Z}$, we have

$$|S| = \sum_{j=-\infty}^{y} |I'_j(k)| \leq \sum_{j=-\infty}^{y} |I'_y(k)|(e^{-\varepsilon})^{y-j} = |I'_y(k)| \sum_{i=0}^{\infty} (e^{-\varepsilon})^i = \frac{|I'_y(k)|}{1-e^{-\varepsilon}} \leq \frac{|I'_y(k)|}{(1-\frac{1}{e})\cdot\varepsilon}.$$

The last inequality holds from the facts: (1) $g_1(x) \overset{def}{=} 1-e^{-x}$ is a concave function; (2) $g_2(x) \overset{def}{=} (1-\frac{1}{e})\cdot x$ is a linear function; (3) $g_1(0) = g_2(0)$ and $g_1(1) = g_2(1)$.

Let $\bar{S} \overset{def}{=} [\bar{s}_y(k-1), 1]$. Then $|\bar{S}| \leq |S| \leq \frac{|I'_y(k)|}{(1-\frac{1}{e})\cdot\varepsilon}$.

On the other hand, $|\bar{S}|$ can be considered as the probability of sampling a sequence $\mathbf{r}$ from the uniform distribution U_{S_0} such that $\mathbf{r} \in \text{STR}\ (\bar{S}, n) \cap S_0$ and $2^{n-b} \leq |S_0| \leq 2^n$.

Therefore,

$$|\bar{S}| = \sum_{\mathbf{r}\in \text{STR}\ (\bar{S},n)\cap S_0} \frac{1}{|S_0|} \geq \sum_{\mathbf{r}\in \text{STR}\ (\bar{S},n)\cap S_0} (\frac{1}{2})^n = |\text{STR}\ (\bar{S}, n) \cap S_0| \cdot (\frac{1}{2})^n.$$

Correspondingly,

$$|\text{STR}\ (\bar{S}, n) \cap S_0| \leq 2^n \cdot |\bar{S}| \leq 2^n \cdot \frac{|I'_y(k)|}{(1-\frac{1}{e})\cdot\varepsilon} = (2^b+1)\cdot\frac{|I'_y(k)|}{|I_{y-1}(k)|}\cdot\frac{1}{(1-\frac{1}{e})\cdot\varepsilon} \leq \frac{e\cdot(2^b+1)}{1-e^{-1}}.$$

Hence,

$$|\text{SUFFIX}(\mathbf{u}, n) \cap S_0| \leq |\text{STR}\ (\bar{S}, n) \cap S_0| \leq \frac{e\cdot(2^b+1)}{1-e^{-1}}.$$

□

Theorem 4. *Mechanism $\overline{M}_{\varepsilon}^{\mathsf{CBCLCS}}$ is a compact $((2^b+1)\cdot e\cdot\varepsilon, \log(\frac{e\cdot(2^b+1)}{1-e^{-1}}))$-BCL-consistent sampling mechanism for (δ, b)-BCL sources. Therefore, $\overline{M}_{\varepsilon}^{\mathsf{CBCLCS}}$ is $(\mathcal{U}, 2e\cdot\varepsilon)$-differentially private and $(\mathcal{BCL}(\delta, b), \xi)$-differentially private for $\xi = (\frac{1+\delta}{1-\delta})^{\log(\frac{e\cdot(2^b+1)}{1-e^{-1}})} \cdot (\frac{1+\delta}{2})^{-b} \cdot (2^b+1)\cdot e\cdot\varepsilon$.*

Proof. The result follows from Theorems 1, 2, and 3.

□

Theorem 5. *Mechanism $\overline{M}_{\varepsilon}^{\mathsf{CBCLCS}}$ has $(\mathcal{BCL}(\delta, b), O(\frac{1}{\varepsilon}\cdot\frac{1}{1-\delta}))$-utility and $(\mathcal{U}, O(\frac{1}{\varepsilon}))$-utility.*

Proof. We only need to prove that for all neighboring databases $D_1, D_2 \in \mathcal{D}$, all queries $f \in \mathcal{F}$, and all distributions $BCL(\delta, b) \in \mathcal{BCL}(\delta, b)$, $\mathbb{E}_{\mathbf{r}\leftarrow BCL(\delta,b)}[|\overline{M}_{\varepsilon}^{\mathsf{CBCLCS}}(D_1, f; \mathbf{r}) - f(D_1)|]$ and $\mathbb{E}_{\mathbf{r}\leftarrow BCL(\delta,b)}[|\overline{M}_{\varepsilon}^{\mathsf{CBCLCS}}(D_2, f; \mathbf{r}) - f(D_2)|]$ are both upper bounded by $O(\frac{1}{\varepsilon}\cdot\frac{1}{1-\delta})$. Without loss of generality, assume that $f(D_1) = y$ and $f(D_2) = y-1$. Then

$$\begin{aligned}
&\mathbb{E}_{\mathbf{r}\leftarrow BCL(\delta,b)}[|\overline{M}_{\varepsilon}^{\mathsf{CBCLCS}}(D_1, f; \mathbf{r}) - y|] \\
&= \sum_{k=-\infty}^{\infty} \Pr_{\mathbf{r}\leftarrow BCL(\delta,b)}[\overline{M}_{\varepsilon}^{\mathsf{CBCLCS}}(D_1, f; \mathbf{r}) = \frac{k}{\varepsilon}] \cdot |\frac{k}{\varepsilon} - y|.
\end{aligned}$$

Let $\mathbf{a}$ be the longest common prefix of all strings in STR $(\bar{I}_y(k), \tau(y-1, k/\varepsilon))$. Denote $I_0 \overset{def}{=} \text{SUFFIX}(\mathbf{a}0, \tau(y-1, k/\varepsilon)) \cap \text{STR}(\bar{I}_y(k), \tau(y-1, k/\varepsilon))$ and $I_1 \overset{def}{=} \text{SUFFIX}(\mathbf{a}1, \tau(y-1, k/\varepsilon)) \cap \text{STR}(\bar{I}_y(k), \tau(y-1, k/\varepsilon))$. Thus, $I_0 \cup I_1 = \text{STR}(\bar{I}_y(k), \tau(y-1, k/\varepsilon))$. Correspondingly, we have

$$\Pr_{\mathbf{r}\leftarrow BCL(\delta,b)}[\overline{M}_{\varepsilon}^{\mathsf{CBCLCS}}(D_1, f; \mathbf{r}) = \frac{k}{\varepsilon}] \leq (\frac{1+\delta}{2})^{|\mathbf{a}0|} + (\frac{1+\delta}{2})^{|\mathbf{a}1|} \leq 2\cdot(\frac{1+\delta}{2})^{\log(\frac{1}{|\bar{I}_y(k)|})}.$$

Similarly, we can conclude that

$$\Pr_{\mathbf{r}\leftarrow BCL(\delta,b)}[\overline{M}_{\varepsilon}^{\mathsf{CBCLCS}}(D_2, f; \mathbf{r}) = \frac{k}{\varepsilon}] \leq 2\cdot(\frac{1+\delta}{2})^{\log(\frac{1}{|\bar{I}_{y-1}(k)|})}.$$

Claim. For all $y, k \in \mathbb{Z}$, we have $|I_y(k)| \le \frac{1}{2} \cdot e^{-\frac{1}{2}} \cdot (e-1) \cdot e^{-|k-\varepsilon y|}$.

Proof. We consider three cases.

Case 1: Assume that $\frac{k-\frac{1}{2}}{\varepsilon} - y \ge 0$ and $\frac{k+\frac{1}{2}}{\varepsilon} - y \ge 0$. Then

$$|I_y(k)| = 1 - \frac{1}{2} \cdot e^{-\varepsilon(\frac{k+\frac{1}{2}}{\varepsilon} - y)} - [1 - \frac{1}{2} \cdot e^{-\varepsilon(\frac{k-\frac{1}{2}}{\varepsilon} - y)}] = \frac{1}{2} \cdot e^{-\frac{1}{2}} \cdot (e-1) \cdot e^{-|k-\varepsilon y|}.$$

Case 2: Assume that $\frac{k-\frac{1}{2}}{\varepsilon} - y < 0$ and $\frac{k+\frac{1}{2}}{\varepsilon} - y \ge 0$. From the fact that $1 - \frac{1}{2}x \le \frac{1}{2} \cdot \frac{1}{x}$ for all $x > 0$, we obtain

$$|I_y(k)| = 1 - \frac{1}{2} \cdot e^{-\varepsilon(\frac{k+\frac{1}{2}}{\varepsilon} - y)} - \frac{1}{2} \cdot e^{\varepsilon(\frac{k-\frac{1}{2}}{\varepsilon} - y)} \le \frac{1}{2} \cdot e^{-\frac{1}{2}} \cdot (e-1) \cdot e^{-|k-\varepsilon y|}.$$

Case 3: Assume that $\frac{k-\frac{1}{2}}{\varepsilon} - y < 0$ and $\frac{k+\frac{1}{2}}{\varepsilon} - y < 0$. Then $|I_y(k)| = \frac{1}{2} \cdot e^{-\frac{1}{2}} \cdot (e-1) \cdot e^{-|k-\varepsilon y|}$.

□

By Lemma 2, $|\bar{I}_y(k)| \le |I_y(k)| + 2^{-\tau(k-1,y)} = |I_y(k)| + \frac{1}{2^b+1}|I_{y-1}(k)|$. Hence,

$$\log(\frac{1}{|\bar{I}_y(k)|}) \ge \log \frac{1}{\frac{1}{2}e^{-\frac{1}{2}}(e-1)(1+\frac{1}{2^b+1})} + \log(e^{\min\{|k-\varepsilon y|, |k-\varepsilon y+\varepsilon|\}}) \ge \min\{|k-\varepsilon y|, |k-\varepsilon y+\varepsilon|\} \ge |k-\varepsilon y| - 1.$$

Similarly, $\log(\frac{1}{|\bar{I}_{y-1}(k)|}) \ge |k - \varepsilon y| - 1$. Therefore,

$$\begin{aligned}
&\sum_{k=-\infty}^{\infty} \Pr_{\mathbf{r} \leftarrow \mathsf{BCL}(\delta,b)}[\overline{M}_{\varepsilon}^{\mathsf{CBCLCS}}(D_1, f; \mathbf{r}) = \frac{k}{\varepsilon}] \cdot |\frac{k}{\varepsilon} - y| \\
&\le \sum_{k=-\infty}^{0} 2 \cdot (\frac{1+\delta}{2})^{|\varepsilon y - k|-1} \cdot |y - \frac{k}{\varepsilon}| + \sum_{k=1}^{\infty} 2 \cdot (\frac{1+\delta}{2})^{|k-\varepsilon y|-1} \cdot |\frac{k}{\varepsilon} - y| \\
&\le \frac{2}{\varepsilon} \cdot (\frac{1+\delta}{2})^{-1} \cdot [\sum_{k=1}^{\infty} (\frac{1+\delta}{2})^{k-1} \cdot k + \sum_{k=-\infty}^{0} (\frac{1+\delta}{2})^{-k} \cdot (-k+1)] \\
&= (\frac{1+\delta}{2})^{-1} \cdot \frac{4}{\varepsilon} \cdot \frac{1}{1-(\frac{1+\delta}{2})^2} = O(\frac{1}{\varepsilon} \cdot \frac{1}{1-\delta}).
\end{aligned}$$

Similarly, we get that

$$\sum_{k=-\infty}^{\infty} \Pr_{\mathbf{r} \leftarrow \mathsf{BCL}(\delta,b)}[\overline{M}_{\varepsilon}^{\mathsf{CBCLCS}}(D_2, f; \mathbf{r}) = \frac{k}{\varepsilon}] \cdot |\frac{k}{\varepsilon} - (y-1)| \le O(\frac{1}{\varepsilon} \cdot \frac{1}{1-\delta}).$$

When $\delta = 0$ and $b = 0$, the BCL source degenerates into the uniform source.

Therefore, the mechanism $\overline{M}_{\varepsilon}^{\mathsf{CBCLCS}}$ has $(\mathcal{BCL}(\delta, b), O(\frac{1}{\varepsilon} \cdot \frac{1}{1-\delta}))$-utility and $(\mathcal{U}, O(\frac{1}{\varepsilon}))$-utility.

□

From Theorems 4 and 5, we get that

Theorem 6. *There exists an explict $(\mathcal{BCL}(\delta, b), \xi)-$differentially private and $(\mathcal{U}, \rho)$-accurate mechanism M for the Hammimg weight queries where*

$$\rho = \frac{2^{b \cdot \log(1+\delta)-9}}{\xi} \cdot (\frac{2}{1+\delta})^{b+1} \cdot \frac{2^b+1}{(1+\delta)^b} \cdot (\frac{1+\delta}{1-\delta})^{\log \frac{(2^b+1)e}{1-e^{-1}}} \cdot \frac{2^{11}}{1-(\frac{1+\delta}{2})^2} \cdot e > \frac{2^{b \cdot \log(1+\delta)-9}}{\xi}.$$

One the other hand, recall that Dodis and Yao [DY14] obtained the following result.

Theorem 7. *If* $b \geq \frac{\log(\xi\rho)+9}{\log(1+\delta)} = \Omega(\frac{\log(\xi\rho)+1}{\delta})$, *then no* $(\mathcal{BCL}(\delta, b), \xi)-$*differentially private and* $(\mathcal{U}, \rho)$*-accurate mechanism for the Hammimg weight queries exists.*

Therefore, we conclude that

Corollary 2. *Assume that the mechanism M is* $(\mathcal{BCL}(\delta, b), \xi)-$*differentially private and* $(\mathcal{U}, \rho)$*-accurate for the Hammimg weight queries, then* $\rho > \frac{2^{b\cdot\log(1+\delta)-9}}{\xi}$.

Corollary 2 implies that it's possible to construct a $(\mathcal{BCL}(\delta, b), \xi)-$differentially private and $(\mathcal{U}, \rho)$-accurate mechanism for Hammimg weight queries, where $\rho > \frac{2^{b\cdot\log(1+\delta)-9}}{\xi}$. In this paper, we show an explicit construction of such mechanisms.

Remark 5. If we replace the rounding method in [DLMV12] with the one in Step 2 of this paper, we can prove that the modified mechanism in [DLMV12] satisfies the compact $(\zeta', O(1))$-SV-consistent sampling. Therefore, the resulting mechanism is differentially private as well. We can also prove that it's accurate. The proofs are similar to Theorems 4 and 5. Due to space limitations, we will omit the analysis.

Acknowledgments. We would like to thank Yevgeniy Dodis, Adriana López-Alt, and Frank Mcsherry for helpful discussions. In particular, we are very grateful to Prof. Yevgeniy Dodis for proposing the project "Do differential privacy with BCL sources for reasonably high b". This work is supported by the Natural Science Foundation of China (60973105, 61370126, and 61170189), the Fund for the Doctoral Program of Higher Education of China (20111102130003), the Fund of the State Key Laboratory of Software Development Environment (SKLSDE-2013ZX-19, SKLSDE-2012ZX-11), the Innovation Foundation of Beihang University for Ph.D. Graduates under Grant No. 2011106014, the Fund of the Scholarship Award for Excellent Doctoral Student granted by Ministry of Education (400618), and the Fund for CSC Scholarship Programme (201206020063).

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] Manuel 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.

[BDMN05] Avrim Blum, Cynthia Dwork, Frank McSherry, and Kobbi Nissim. Practical privacy: the sulq framework. In Chen Li, editor, *PODS*, pages 128-138. ACM, 2005.

[BGMZ97] Andrei Z. Broder, Steven C. Glassman, Mark S. Manasse, and Geoffrey Zweig. Syntactic clustering of the web. *Computer Networks*, 29(8-13): 1157-1166, 1997.

[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] Benny Chor, Oded Goldreich, Johan Håstad, Joel Friedman, Steven Rudich, and Roman 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.

[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.

[DN03] Irit Dinur and Kobbi Nissim. Revealing information while preserving privacy. In Frank Neven, Catriel Beeri, and Tova Milo, editors, *PODS*, pages 202-210. ACM, 2003.

[DN04] Cynthia Dwork and Kobbi Nissim. Privacy-preserving datamining on vertically partitioned databases. In Matthew K. Franklin, editor, *CRYPTO*, volume 3152 of *LNCS*, pages 528-544. Springer, 2004.

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

[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.

[Dwo08] Cynthia Dwork. Differential Privacy: A Survey of Results. *TAMC 2008*, pages 1-19.

[DY14] Yevgeniy Dodis and Yanqing Yao.Privacy and Imperfect Randomness.IACR Cryptology ePrint Archive 2014: 623 (2014).

[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.

[GRS09] Arpita Ghosh, Tim Roughgarden, and Mukund Sundararajan. Universally utilitymaximizing privacy mechanisms. In Michael Mitzenmacher, editor, *STOC*, pages 351-360. ACM, 2009.

[Hol07] Thomas Holenstein. Parallel repetition: simplifications and the no-signaling case. In David S. Johnson and Uriel Feige, editors, *STOC*, pages 411-419. ACM, 2007.

[HT10] Moritz Hardt and Kunal Talwar. On the geometry of differential privacy. In Leonard J. Schulman, editor, *STOC*, pages 705-714. ACM, 2010.

[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.

[Man94] Udi Manber. Finding similar files in a large file system. In *Proceedings of the USENIX Winter 1994 Technical Conference on USENIX Winter 1994 Technical Conference*, pages 2-2, Berkeley, CA, USA, 1994. USENIX Association.

[MMP+10] Andrew McGregor, Ilya Mironov, Toniann Pitassi, Omer Reingold, Kunal Talwar, and Salil P. Vadhan. The limits of two-party differential privacy. In *FOCS*, pages 81-90. IEEE Computer Society, 2010.

[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.

Recall that Lemma A.1. of [DLMV12] and partial proof are as follows.

Lemma 5. *For all* $y, k \in \mathbb{Z}$, $\frac{|I'_y(k)|}{|I_{y-1}(k)|} \leq 6\varepsilon$.

Proof.

...

Case 3: If $s_y(k-1) < s_{y-1}(k-1) < \frac{1}{2} \leq s_{y-1}(k-1)$, then $\frac{|I'_y(k)|}{|I_{y-1}(k)|} \leq \frac{1-e^{-\varepsilon}}{2(e-1)}$.

...

□

There are several points to be noted.

1. Compared with the above lemma, ours is much better. In fact, our upper bound is tight.

2. It's obvious that " $s_{y-1}(k-1) < \frac{1}{2} \leq s_{y-1}(k-1)$" never holds. It must be a typo.

3. "$\frac{|I'_y(k)|}{|I_{y-1}(k)|} \leq \frac{1-e^{-\varepsilon}}{2(e-1)}$" is wrong! Since $-1 - \frac{1}{\varepsilon} \leq v < -1$, without loss of generality, assume that $\frac{1}{\varepsilon}$ is an even integer and $v = -1 - \frac{1}{2\varepsilon}$. Then

$$\frac{|I'_y(k)|}{|I_{y-1}(k)|} = \frac{e^{\varepsilon} - 1}{2 \cdot e^{-\varepsilon v} - e^{-2\varepsilon v - \varepsilon - 1} - e^{\varepsilon}} = \frac{1 - e^{-\varepsilon}}{2(e^{\frac{1}{2}} - 1)} > \frac{1 - e^{-\varepsilon}}{2(e-1)},$$

which stands in contradiction to the inequality $\frac{|I'_y(k)|}{|I_{y-1}(k)|} \leq \frac{1-e^{-\varepsilon}}{2(e-1)}$.