Two-Source Extractors for Leaky Sources

Yu Yu*†, Xiangxue Li†‡ and Haifeng Qian†
*Institute for Interdisciplinary Information Sciences, Tsinghua University, Beijing, 100084, P. R. China
† Department of Computer Science and Technology, East China Normal University, Shanghai, P. R. China
‡Correspondence author: xxli@cs.ecnu.edu.cn

***Abstract*—A (worst-case) 2-source extractor is a deterministic algorithm that transforms pairwise independent weak random sources into almost uniform random strings. Despite non-constructive proofs that such objects exist with almost optimal parameters, it has been a longstanding open problem to construct 'explicit' (aka efficient) functions for sources of 'small' constant entropy rate. In particular, best known constructions either require entropy rate of at least 0.4999 (due to Bourgain), or one source must remain with constant entropy rate above half (due to Raz).**

Motivated by cryptographic applications, we observe that if one source is a leaky source (or it contains a few deterministically extractable entropy), then we will be able to efficiently extract almost all entropy from both sources with nearly optimal entropy loss. Further, our extractor (for leaky sources) does not suffer from the half entropy rate barrier, and it works for all linear (and even sub-linear) entropy sources. The extractor is constructed using the technique of alternating extraction by Dziembowski and Pietrzak (FOCS 2007). Finally, we show that the extractor is almost a worse-case extractor (for the same parameters) in the sense that it only fails for a negligible fraction of sources.

I. Introduction

The research of randomness extraction mainly focuses on construction of efficient functions, called (explicit) randomness extractors, such that when applied on any distributions with nontrivial amount of min-entropy one obtains output distribution statistically close to uniform. It has wide application in many areas of computer science, such as cryptography and derandomization.

Seeded Extractors. For a general weak source, randomness can be extracted efficiently using (necessarily) a short random seed [1]. Shaltiel's survey [2], [3] gave an informative overview of some recent developments in this area, and Lu et al. [4] and Guruswami et al. [5] presented the current state-of-art constructions. It is possible to remove the need of random seed by considering various types of structured sources, such as bit-fixing- (more generally, affine- and polynomial-) sources ([6], [7], [8], [9] and see [3] for more references) and sources samplable by small circuits [10] or generated in small space [11].

Two-Source Extractors and Their Applications. Another line of research focused on deterministic extraction from several independent sources, which originated in the work of von Neumann [12], and got renewed interests recently in [13], [14], [15], [16], [17]. In this paper, we focus on the (most challenging) case of two-source extractors. Chor and Goldreich [18] used inner product as 2-source extractor for equal-length sources of entropy rate above $1/2$, with some improvements made in [19]. Raz [15] showed how to extract almost all entropy where one source has constant entropy rate more than $1/2$ and the other can be of only logarithmic min-entropy. Bourgain [20] gave a breakthrough construction for sources of entropy rate 0.4999. Therefore, known results are far from reaching the existential bounds proven using non-constructive proofs (e.g. counting argument, probabilistic method). We mention also several *conditional* constructions for arbitrary linear entropy sources: the one by Chor and Goldreich [18] based on a conjecture on the Paley Graph, the construction by Kalai, Li and Rao [21] assuming one-way permutations with exponential hardness, and the more recent construction by Zewi and Ben-Sasson [22] based on the Approximate Duality Conjecture. Two-source extractors have found useful cryptographic applications in many recent works, such as leakage-resilient schemes and protocols for secret sharing and storage [23], [24], [25], public-key encryption [26], and distributed computation [24]. However, as no unconditional 2-source extractors are known for entropy rate below 0.4999, none of the aforementioned schemes tolerate leakages of portion beyond 0.5001.

Motivating Scenario. In this paper, we investigate 2-source extractors for two conditional sources (X,Z_X) and (Y,Z_Y), where X is independent of (Y,Z_Y), Y has arbitrary linear min-entropy conditioned on Z_Y, and (X,Z_X) is a leaky source, namely X is uniformly random (e.g. X is a secret key) and it remains with linear entropy given Z_X, which can be correlated to both X and Z_Y. We note that we impose no restriction on (Y,Z_Y), which can be any linear entropy source, e.g., Y can be of min-entropy rate 0.0001 with an empty Z_Y, or Y can be a leaky source with any 0.9999 portion leaked through Z_Y. The only restriction is that X must be uniform by itself (and we will show this condition can be relaxed).

Overview of Our Results. In this paper, we provide an efficient 2-source extractor for the above problem, using the technique of alternating extraction by Dziembowski and Pietrzak [27]. The technique is simple (without deep techniques such as the sum-product theorem) and the results are nearly optimal in terms of entropy loss and the amount of entropy extracted. We now sketch the simplified version of the main results, and it already explains the main idea. For concreteness let $\widetilde{\mathbf{H}}_\infty(X|Z_X) = k_1$ and $\widetilde{\mathbf{H}}_\infty(Y|Z_Y) = k_2$ (definition deferred to Section II-A). As X is uniformly random, by applying a strong extractor on Y (using X as seed) one obtains $k_2 - 2\log(1/\varepsilon) - O(1)$ bits that are ε-close

978-1-4673-0223-4/12/$31.00 ©2012 IEEE

to uniform conditioned on Z_Y, X, and Z_X (which is implied by X). Instantiating the strong extractor with the state-of-art construction [5] (see Theorem 3.1), we proceed to another seeded extraction (using the extracted $k_2 - 2\log(1/\varepsilon)$ bits as a seed) to get further $k_1 - 2\log(1/\varepsilon) - O(1)$ bits from X. In summary, one extracts $k_1+k_2-4\log(1/\varepsilon)-O(1)$ bits that are 2ε-close to uniform (conditioned on Z_X and Z_Y), provided that the first extracted $k_2 - 2\log(1/\varepsilon) - O(1)$ bits must provide a long enough seed for the second extraction, which is easily satisfied for linear entropy sources. In summary, the 2-source extractor extracts nearly all entropy from the sources with (asymptotically) optimal entropy loss. We can further generalize the above results. For example, the condition that (X,Z_X) must be a leaky source is not necessary, it is sufficient to have that X contains a poly-logarithmic amount of randomness for which efficient deterministic extractors exist. In addition, the two sources do not have to be of the same length, they can be linearly or even polynomially related.

CONNECTION TO WORST-CASE 2-SOURCE EXTRACTORS. While explicit *worst-case* two-source extractors for arbitrary linear sources remain unknown, we show a connection to this problem: our extractor is a worst-case two-source extractor (for almost the same parameters) for *all* Y's of min-entropy k_2, and *almost all* (an overwhelming portion of) X's of min-entropy k_1. Otherwise said, the pathological cases for which the extractor fails are of negligible fraction.

II. BACKGROUND

A. Preliminaries

NOTATION AND DEFINITION. Formally, a source X of length n, denoted by $|X| = n$, is a random variable over $\{0,1\}^n$. We let U_n denote the uniform distribution over $\{0,1\}^n$. We write $x \leftarrow X$ to denote the operation of sampling a random x according to X. We use $X \sim Y$ to denote identically distributed X and Y. For a randomized function f, we write $f(x;r)$ to denote the output of f on input x with random coin r. The *min-entropy* of X is defined as $\mathbf{H}_\infty(X) \stackrel{\text{def}}{=} -\log(\max_x \Pr[X = x])$, and X with k bits of min-entropy is referred to as (n,k) source. The min-entropy rate of an (n,k) source is defined as ratio k/n. We say that X is ε-close to Y if their *statistical distance*, defined by

$$\mathsf{SD}(X,Y) \stackrel{\text{def}}{=} \tfrac{1}{2}\textstyle\sum_x |\Pr[X=x] - \Pr[Y=x]|$$

$$= \max_D \ | \ \Pr[D(X)=1] - \Pr[D(Y)=1] \ |$$

is upper bounded by ε, where the maximum is taken over all (including computationally unbounded) distinguishers D. In the conditional case, we write $\mathsf{SD}(X,Y|Z)$ as shorthand for $\mathsf{SD}((X,Z),(Y,Z))$.

Definition 2.1 (Worst-Case Extractor): We say that an efficient function $\mathsf{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ is a strong worst-case (n,k,ε)-*seeded extractor* (for space $\{0,1\}^n$), if for any X with $\mathbf{H}_\infty(X) \geq k$, and for $S \sim U_d$, we get

$$\mathsf{SD}(\ \mathsf{Ext}(X;S)\ ,\ U_m \mid S) \leq \varepsilon$$

where S denotes the coins of Ext (called the *seed*), the value d is called the *seed length* of Ext, and value $L = k - m$ is called the *entropy loss* of Ext.

SEED LENGTH AND ENTROPY LOSS When one is concerned with randomness extraction from a general weak source, a random seed is necessary and entropy loss is inevitable. We already know lower bounds results from [28] that any non-trivial seeded extractor must satisfy seed length $d = \log(n-k)+2\log(1/\varepsilon)-O(1)$ and entropy loss $L = 2\log(1/\varepsilon)-O(1)$. The bounds are tight as there are explicit constructions [4], [5] that match both bounds simultaneously.

LEAKY SOURCES AND AVERAGE-CASE EXTRACTORS A leaky source is a joint distribution (X,Z) where X is uniformly distributed on its own (without Z), and X has some average min-entropy left conditioned on Z. This is defined as

$$\widetilde{\mathbf{H}}_\infty(X|Z) \stackrel{\text{def}}{=} -\log\left(\mathbb{E}_{z\leftarrow Z}[\ \max_x \Pr[X=x|Z=z]\ ]\right)$$
$$= -\log\left(\mathbb{E}_{z\leftarrow Z}\left[2^{-\mathbf{H}_\infty(X|Z=z)}\right]\right)$$

where $\mathbb{E}_{z\leftarrow Z}$ denotes the expected value over $z \leftarrow Z$, and measures the maximal predictability of X by an adversary that may observe a correlated variable Z.

Definition 2.2 (Average-Case Extractor): We say that an efficient function $\mathsf{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ is a strong average-case (n,k,ε)-*seeded extractor* (for space $\{0,1\}^n$), if for $S \sim U_d$, and for all (X,Z) such that X is distributed over $\{0,1\}^n$ and $\widetilde{\mathbf{H}}_\infty(X|Z) \geq k$, we get

$$\mathsf{SD}(\ \mathsf{Ext}(X;S)\ ,\ U_m \mid Z,S) \leq \varepsilon\ .$$

It is easy to see the equivalence between worst-case- and average-case- extractors. On the one hand, an average-case-extractor is also an worst-case one for the same parameters (by considering empty side information); on the other hand, a worst-case extractor is also an average-case extractor for slightly worse parameters using Markov's inequality (see also [29] for tighter parameters). Therefore, there seems no easy way that we could construct efficient average-case 2-source extractors for leaky sources to beat the worst-case 2-source counterparts. However, notice that we have a useful restriction (by definition of leaky source) that X is uniformly distributed, and we exploit this condition to get more efficient constructions in the average-case.

TWO-SOURCE EXTRACTORS. Analogously, we define worst-case- and average-case- exactors below:

Definition 2.3 (Two-Source Extractor): We say that an efficient function $\mathsf{2Ext} : \{0,1\}^{n_1} \times \{0,1\}^{n_2} \to \{0,1\}^m$ is a **worst-case** $(n_1,n_2,k_1,k_2,\varepsilon)$-*2-source extractor* (for space $\{0,1\}^{n_1} \times \{0,1\}^{n_2}$), if for all independent (n_1,k_1)-source X and (n_2,k_2)-source Y, we get

$$\mathsf{SD}(\ \mathsf{2Ext}(X,Y)\ ,\ U_m) \leq \varepsilon$$

Similarly, 2Ext is an **average-case** $(n_1,n_2,k_1,k_2,\varepsilon)$-*2-source extractor* (for space $\{0,1\}^{n_1} \times \{0,1\}^{n_2}$) if for all independent (X,Z_X) and (Y,Z_Y) with $\widetilde{\mathbf{H}}_\infty(X|Z_X) \geq k_1$ and $\widetilde{\mathbf{H}}_\infty(Y|Z_Y) \geq k_2$ respectively, we have

$$\mathsf{SD}(\ \mathsf{2Ext}(X,Y)\ ,\ U_m \mid Z_X, Z_Y) \leq \varepsilon$$

where in both cases the *entropy loss* of 2Ext is the value $L = k_1 + k_2 - m$.

B. Related Work

Dziembowski and Pietrzak [27] studied the problem of *alternating extraction*, where randomness is extracted alternatingly from two independent leaky sources against adaptively chosen leakages. Our work can be viewed as a special (two-round, deterministic) case of theirs by removing the use of random seed (in [27] a public random seed is provided at the beginning). The work of [30], [31] considered practically efficient constructions of key derivation functions (KDFs) based on the *extract-then-expand* approach, where the extractor is built from cryptographic primitives (rather than using existing combinatorial extractors). Their work is incomparable to ours as their extractors also work for computationally unpredictable sources (without any min-entropy) by relying on idealized assumptions such as modeling SHA-1 as random oracles. It is also observed by Bourgain [20] that although the inner product does not work for sources of entropy rate below $1/2$ in general, there are essentially very few counterexamples for which it fails. Thus, the construction of [20] proceeds by first encoding each source in some way and then applying the inner product, such that the constructed extractor works for sources of entropy rate 0.4999. In Section III-D (where we show our extractor is also worst-case extractor for almost all sources), we take a complementary approach. That is, we compromise a bit of generality to make the extractor distill nearly all entropy from almost (an overwhelming portion of) all pairwise independent low-entropy sources with optimal entropy loss.

C. Basic Facts and Lemmas

We recall the following facts and lemmas about statistical distance and min-entropy, which will be used in our proofs.

Fact 2.1: For all random variables X and Y over the same set $\mathcal{S}$, and for any function f for space $\mathcal{S}$, it holds that

$$\mathsf{SD}(f(X), f(Y)) \le \mathsf{SD}(X, Y) \ .$$

Fact 2.2 (Triangle Inequality): For all random variables X, Y, and Z over the same set $\mathcal{S}$, we have

$$\mathsf{SD}(X, Y) \le \mathsf{SD}(X, Z) + \mathsf{SD}(Y, Z) \ .$$

Lemma 2.1 ([32]): For joint random variables (X, Z_X) where Z_X has at most 2^λ possible values, we have

$$\widetilde{\mathbf{H}}_\infty(X|Z_X) \ \ge \ \mathbf{H}_\infty(\ (X, Z_X)\ ) - \lambda \ \ge \ \mathbf{H}_\infty(X) - \lambda \ .$$

III. Main Results

A. Basic Tools

The main technical tool we use are strong seeded extractors, and we will instantiate them with a construction from [5] (as we will see, such instantiation is crucial for the resulted 2-source extractor to have desirable entropy loss and low entropy threshold at the same time). As mentioned, the strong worst-case extractor presented in [5] achieved asymptotically optimal seed length and entropy loss. In fact, the extractor is already average-case for the same parameters, which we stated as Theorem 3.1 below. For completeness, we reproduce its simple proof for completeness.

Theorem 3.1 (The GUV Extractor [5]): For all integers $n \ge k > m > 0$, and for any $\varepsilon > 0$ such that $m \le k - 2\log(1/\varepsilon) - O(1)$, there is an explicit construction of a strong average-case $(k,\ \varepsilon)$-**seeded extractor** $\mathsf{Guv} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$, where $d = \log n + O(\log k \cdot \log(k/\varepsilon))$.

Proof: Let $X_z \stackrel{\mathsf{def}}{=} (X|Z = z)$, then

$$\begin{aligned}
\mathsf{SD} \quad & ((\mathsf{Guv}(X; S), S, Z), (U_m, S, Z)) \\
= \quad & \mathbb{E}_{z \leftarrow Z}\left[\mathsf{SD}((\mathsf{Guv}(X_z; S), S), (U_m, S))\right] \\
\le \quad & \mathbb{E}_{z \leftarrow Z}\left[\ \sqrt{2^{-\mathbf{H}_\infty(X_z)+O(1)} \cdot 2^m}\ \right] \\
\le \quad & \sqrt{\mathbb{E}_{z \leftarrow Z}[2^{-\mathbf{H}_\infty(X_z)+O(1)} \cdot 2^m]} \ = \ 2^{\frac{-k+m+O(1)}{2}} \ \le \ \varepsilon,
\end{aligned}$$

where the first inequality follows from the fact that Guv is a strong worst-case extractor for the same parameters as stated above (see [5]), and the second inequality follows from Jensen's inequality, i.e., $\mathbb{E}\left[\sqrt{T}\right] \le \sqrt{\mathbb{E}\,[T]}$. ∎

Another tool we use are deterministic extractors. Recall that the simplifying statement of our results (see Section I) assumes that (X,Z_X) is a leaky source so that the uniformly random X can be used as a seed for randomness extraction from Y. We can relax this requirement to that X is nearly uniform on some known (or even unknown) d-dimensional subspace (see Theorem 3.1 for the value of d), where d just needs to be logarithmic in the length of X. For better generalization, we assume that efficient deterministic functions exist for X to get sufficient amount of nearly uniform randomness. We refer the readers to [3] for a survey of the rich literature of deterministic extractors.

B. Main Theorem

The main results of this paper are stated as Theorem 3.2 below:

Theorem 3.2 (The Main Results): For all integers n_1, n_2, k_1, k_2, d, m, and for any $\varepsilon > 0$ such that

1) $n_1 \ge d = \log n_2 + O(\log k_2 \cdot \log(k_2/\varepsilon))$,
2) $n_1 \ge k_1 \ge 2\log(1/\varepsilon) + O(1)$,
3) $n_2 \ge k_2 \ge 2\log(1/\varepsilon) + \log n_1 + O(\log k_1 \cdot \log(k_1/\varepsilon))$,
4) $m = k_1 + k_2 - 4\log(1/\varepsilon) - O(1)$.

there is an explicit construction of an average-case (n_1, n_2, k_1, k_2, ε)- 2-source extractor for all pairwise-independent joint distributions (X,Z_X) and (Y,Z_Y) with $\widetilde{\mathbf{H}}_\infty(X|Z_X) \ge k_1$, $\widetilde{\mathbf{H}}_\infty(Y|Z_Y) \ge k_2$, provided that efficient deterministic function $\mathsf{DExt} : \{0,1\}^{n_1} \to \{0,1\}^d$ exists such that $\mathsf{SD}(\mathsf{DExt}(X), U_d) \le \varepsilon$. .

CONSTRUCTIVE PROOF. We give an explicit construction:

$$\begin{aligned}
\mathsf{avg2Ext} : \{0,1\}^{n_1} \times \{0,1\}^{n_2} &\to \{0,1\}^m \\
(x, y) \ &\mapsto (s_2, s_3), \\
\text{where } \ s_2 := \mathsf{Guv}(y; \mathsf{DExt}(x)) \ , \ \ s_3 &:= \mathsf{Guv}(x; s_2) \ ,
\end{aligned}$$

and show that avg2Ext is an average-case (n_1,n_2,k_1,k_2, ε)-2-source extractor. First, by assumption

$$\mathsf{SD}(S_1 \stackrel{\mathsf{def}}{=} \mathsf{DExt}(X), U_d) \le \varepsilon \ . \tag{1}$$

we use S_1 as seed for randomness extraction from Y to get:

$$\mathsf{SD}((S_1, \mathsf{Guv}(Y;S_1), Z_Y), (S_1, U_{|S_2|}, Z_Y)) \ \le \ \underbrace{\mathsf{SD}(\ (S_1, \mathsf{Guv}(Y;S_1), Z_Y)\ ,\ (U_d, \mathsf{Guv}(Y;U_d), Z_Y)\ )}_{\le\varepsilon \text{ by Equation (1)}}$$
$$+ \ \underbrace{\mathsf{SD}(\ (U_d, \mathsf{Guv}(Y;U_d), Z_Y)\ ,\ (U_d, U_{|S_2|}, Z_Y)\ )}_{\le\varepsilon \text{ by Theorem 3.1}} \ + \ \underbrace{\mathsf{SD}(\ (U_d, U_{|S_2|}, Z_Y)\ ,\ (S_1, U_{|S_2|}, Z_Y)\ )}_{\le\varepsilon \text{ by Equation (1)}} \le 3\varepsilon \ .$$

Define a (not necessarily efficient) sampler samp that on input s_1 produces (x, z_x) as output, where the output is sampled from the conditional distribution $(\ (X, Z_X) \mid \mathsf{DExt}(X) = s_1\ )$. Applying samp to the S_1 in the above inequality, we get:

$$\mathsf{SD}(\ (X, Z_X, \mathsf{Guv}(Y;S_1), Z_Y)\ ,\ (X, Z_X, U_{|S_2|}, Z_Y)\ ) \ \le \ 3\varepsilon \tag{2}$$

Write $S_2 \stackrel{\mathsf{def}}{=} \mathsf{Guv}(Y;S_1)$, and use it as seed for extraction from X, we have

$$\mathsf{SD}(\ (S_2, S_3 \stackrel{\mathsf{def}}{=} \mathsf{Guv}(X;S_2)), (U_{|S_2|}, U_{|S_3|}) \ \mid \ Z_X, Z_Y)\ )$$
$$\le \ \underbrace{\mathsf{SD}(\ (S_2, \mathsf{Guv}(X;S_2)), (U_{|S_2|}, \mathsf{Guv}(X;U_{|S_2|}) \mid \ Z_X, Z_Y)\ )}_{\le 3\varepsilon \text{ by Equation (2)}}$$
$$+ \ \underbrace{\mathsf{SD}(\ (U_{|S_2|}, \mathsf{Guv}(X;U_{|S_2|})), (U_{|S_2|}, U_{|S_3|}) \mid \ Z_X, Z_Y\ )}_{\le\varepsilon \text{ by Theorem 3.1}}$$
$$\le \ 4\varepsilon \ ,$$

which completes the proof for the error bound [1]. By Theorem 3.1, we have $|S_1| = d$, $|S_2| = k_2 - 2\log(1/\varepsilon) - O(1)$, and $|S_3| = k_1 - 2\log(1/\varepsilon) - O(1)$, and the parameter conditions #1 and #3 follow from that each of S_1 and S_2 must provide a long enough seed for the Guv extractor respectively, and conditions #2 and #4 are due to $|S_3| \ge 0$ and $m = |S_2| + |S_3|$.

C. Leakage-Resilient Cryptographic Applications

Theorem 3.3: For all integers n_1, n_2, k_1, k_2, m, and for any $\varepsilon > 0$ as stated in Theorem 3.2, there is an efficient function $\mathsf{2Ext} : \{0,1\}^{n_1} \times \{0,1\}^{n_2} \to \{0,1\}^m$ such that for $X \sim U_{n_1}$ and $Y \sim U_{n_2}$, and for all functions $f_1 : \{0,1\}^{n_1} \times \{0,1\}^{n_2-k_2} \to \{0,1\}^{n_1-k_1}$, and $f_2 : \{0,1\}^{n_2} \to \{0,1\}^{n_2-k_2}$ we have

$$\mathsf{SD}(\ \mathsf{2Ext}(X,Y)\ ,\ U_m \mid f_1(X, f_2(Y)),\ f_2(Y)\ ) \le \varepsilon \ .$$

Proof: By Lemma 2.1, we have $\widetilde{\mathbf{H}}_\infty(Y \mid f_2(Y)) \ge k_2$, $\widetilde{\mathbf{H}}_\infty(X \mid f_1(X, f_2(Y))) \ge k_1$, and thus the conclusion immediately follows from Theorem 3.2. Note that here deterministic extractor is trivial as X is uniform by itself. ∎

[1] We can omit the factor 4 from the derived bound by letting $\epsilon = 4\varepsilon$, and hence $\log(1/\epsilon)$=$\log(1/\varepsilon) - 2$, where the additive factor 2 will be absorbed by the big-O notations.

$(1-o(1))$-FRACTION LEAKAGE TOLERANCE. The parameter constraints (see $\#1-\#4$ in Theorem 3.2) are quite loose, and the extractor is actually secure against a $(1-o(1))$-fraction of arbitrary leakages from X and Y (of full entropy), provided that they leak independently. This already covers leakages of any constant fraction, or even arbitrarily close to 1, e.g. $1 - \frac{1}{\sqrt{n}}$. To see this, we let $n_1 = n_2 = n$, $k_1 = k_2 = \sqrt{n}$, and thus we set $\varepsilon = 2^{-t}$ for $t \in \Omega(\sqrt{n}/\log n)$ satisfying all parameters constraints. Moreover, it extracts almost all randomness, namely the entropy loss $4\log(1/\varepsilon) + O(1)$ is optimal (up to factor 2).

D. Almost a Worst-Case Two-Source Extractor

As we additionally require that X is uniform (or deterministically extractable), the above extractor may not be a worst-case 2-source extractor in general. However, we show that as a worst-case extractor it only fails with a negligible fraction of flat sources, for which we call it *almost extractors*[2] defined as the following:

Definition 3.1 (Almost a Worst-Case 2-Source Extractor): We say that an efficient function $\mathsf{2Ext} : \{0,1\}^{n_1} \times \{0,1\}^{n_2} \to \{0,1\}^m$ is a $(1-\varepsilon_1)$-almost worst-case $(n_1, n_2, k_1, k_2, \varepsilon_2)$-2-source extractor (for space $\{0,1\}^{n_1} \times \{0,1\}^{n_2}$), if for all Y with $\mathbf{H}_\infty(Y) \ge k_2$, and independently for at least a $(1-\varepsilon_1)$-fraction of $X \in \{X_i \ \mid \ X_i \in \{0,1\}^{n_1}, \mathbf{H}_\infty(X_i) \ge k_1\}$, we have

$$\mathsf{SD}(\ \mathsf{2Ext}(X,Y)\ ,\ U_m\ ) \le \varepsilon_2 \ .$$

Corollary 3.1: For all integers n_1, n_2, k_1, k_2, m, and for any $\varepsilon > 0$ as stated in Theorem 3.2, there is an explicit construction of a $(1 - \sqrt{\varepsilon})$-almost worst-case (n_1,n_2,k_1,k_2, $\sqrt{\varepsilon}$)- 2-source extractor.

Proof: Consider two joint distributions (X, I) and (Y, Z_Y). For (Y,Z_Y), we let Y to be arbitrary source with $\mathbf{H}_\infty(Y) = k_2$, and let Z_Y be empty (and thus be omitted). Let $\{X_i \ \mid \ X_i$ is uniform over $\mathcal{S}_i \subset \{0,1\}^{n_1}, |\mathcal{S}_i| = 2^{k_1}\}$ be the finite set of all 'flat' sources of min-entropy k_1, indexed by i. Define (X, I) such that $i \leftarrow I$ is uniformly selected from the indices of the above set, and let X be the selected source (i.e.

[2] The same terminology is used in [33] to refer to a somewhat different object, namely a condenser whose output has almost full entropy.

X_i). Then, we have $\mathbf{H}_\infty(X_i) = k_1$ for every i, and $X \sim U_{n_1}$ by Claim 3.1. It follows from Theorem 3.2 that there is an efficient function 2Ext such that

$$\mathsf{SD}(\ \mathsf{2Ext}(X,Y),\ U_m\ \mid\ I) \leq \varepsilon\ .$$

By Markov's inequality, there is a set $\mathcal{S}$ such that $\Pr[I \in \mathcal{S}] \geq 1 - \sqrt{\varepsilon}$, and for every $i \in \mathcal{S}$ $\mathsf{SD}(\ \mathsf{2Ext}(X,Y),\ U_m\ \mid\ I = i) \leq \sqrt{\varepsilon}$. This completes the proof.

Claim 3.1: For all integers $0 \leq k_1 < n_1$, let X be an equal convex combination of all sources from $\{X_i \mid X_i \text{ is uniform over } \mathcal{S}_i \subset \{0,1\}^{n_1}, |\mathcal{S}_i| = 2^{k_1}\}$, then X is uniformly distributed over $\{0,1\}^{n_1}$.

Proof of Claim 3.1. Every string $x \in \{0,1\}^{n_1}$ contributes equally to the combined X, and thus $\Pr[X = x] = 2^{-n_1}$. □

■

Therefore, for the same parameters (except that ε drops to $\sqrt{\varepsilon}$), there is an efficient function which deviates from a perfect worse-case two-source extractor on at most a $\sqrt{\varepsilon}$-fraction of all sources.

ACKNOWLEDGMENT

This work was supported in part by the National Basic Research Program of China Grant 2011CBA00300, 2011CBA00301, the National Natural Science Foundation of China Grant 61033001, 61061130540, 61073174, 61103221, 61172085, 11061130539, 61021004, and 61133014.

REFERENCES

[1] N. Nisan and D. Zuckerman, "Randomness is linear in space," *Journal of Computer and System Sciences*, vol. 52, no. 1, pp. 43–53, 1996.

[2] R. Shaltiel, "Recent developments in explicit constructions of extractors." *Bulletin of the EATCS*, vol. 77, pp. 67–95, 2002.

[3] ——, "An introduction to randomness extractors," in *ICALP (2)*, 2011, pp. 21–41.

[4] C.-J. Lu, O. Reingold, S. P. Vadhan, and A. Wigderson, "Extractors: optimal up to constant factors," in *Proceedings of the Thirty-Fifth Annual ACM Symposium on Theory of Computing*. San Diego, Ca, USA: ACM, 9–11 Jun. 2003, pp. 602–611.

[5] V. Guruswami, C. Umans, and S. P. Vadhan, "Unbalanced expanders and randomness extractors from parvaresh–vardy codes," *J. ACM*, vol. 56, no. 4, 2009.

[6] B. Chor, O. Goldreich, J. Håstad, J. Friedman, S. Rudich, and R. Smolensky, "The bit extraction problem or t-resilient functions," in *Proceedings of the 26th IEEE Symposium on Foundation of Computer Science*, 1985, pp. 396–407.

[7] J. Kamp and D. Zuckerman, "Deterministic extractors for bit-fixing sources and exposure-resilient cryptography," in *44th Annual Symposium on Foundations of Computer Science*. Cambridge, Massachusetts: IEEE, Oct. 2003, pp. 92–101.

[8] J. Bourgain, "On the construction of affine extractors," *Geometric and Functional Analysis*, vol. 17, no. 1, pp. 33–57, 2007.

[9] Z. Dvir, A. Gabizon, and A. Wigderson, "Extractors and rank extractors for polynomial sources," *Computational Complexity*, vol. 18, no. 1, pp. 1–58, 2009.

[10] L. Trevisan and S. Vadhan, "Extracting randomness from samplable distributions," in *41st Annual Symposium on Foundations of Computer Science*. Redondo Beach, California: IEEE, Nov. 2000, pp. 32–42.

[11] J. Kamp, A. Rao, S. P. Vadhan, and D. Zuckerman, "Deterministic extractors for small-space sources," in *Proceedings of the 38th ACM Symposium on the Theory of Computing*.

[12] J. von Neumann, "Various techniques used in connection with random digits," *Monte Carlo Method, U.S. National Bureau of Stadndarts, Applied Mathematics Series*, vol. 12, pp. 36–38, 1951.

[13] B. Barak, R. Impagliazzo, and A. Wigderson, "Extracting randomness using few independent sources," in *45th Symposium on Foundations of Computer Science*. Rome, Italy: IEEE, Oct. 17–19 2004, pp. 384–393.

[14] B. Barak, G. Kindler, R. Shaltiel, B. Sudakov, and A. Wigderson, "Simulating independence: new constructions of condensers, ramsey graphs, dispersers, and extractors," in *Proceedings of the 37th ACM Symposium on the Theory of Computing*, 2005, pp. 1–10.

[15] R. Raz, "Extractors with weak random seeds," in *Proceedings of the 37th ACM Symposium on the Theory of Computing*, 2005, pp. 11–20.

[16] A. Rao, "Extractors for a constant number of polynomially small min-entropy independent sources," in *Proceedings of the 38th ACM Symposium on the Theory of Computing*, 2006, pp. 497–506.

[17] B. Barak, A. Rao, R. Shaltiel, and A. Wigderson, "2-source dispersers for sub-polynomial entropy and ramsey graphs beating the frankl-wilson construction," in *Proceedings of the 38th ACM Symposium on the Theory of Computing*, 2006, pp. 671–680.

[18] B. Chor and O. Goldreich, "Unbiased bits from sources of weak randomness and probabilistic communication complexity," *SIAM Journal on Computing*, vol. 17, no. 2, pp. 230–261, 1988.

[19] Y. Dodis, A. Elbaz, R. Oliveira, and R. Raz, "Improved randomness extraction from two independent sources," in *APPROX-RANDOM*, 2005, pp. 334–344.

[20] J. Bourgain, "More on the sum-product phenomenon in prime fields and its applications," *International Journal of Number Theory*, vol. 1, pp. 1–32, 2005.

[21] Y. T. Kalai, X. Li, and A. Rao, "2-source extractors under computational assumptions and cryptography with defective randomness," in *Proceedings of the 50th IEEE Symposium on Foundation of Computer Science*, 2009, pp. 617–626.

[22] N. Zewi and E. Ben-Sasson, "From affine to two-source extractors via approximate duality," in *STOC*, 2011, pp. 177–186.

[23] F. Davì, S. Dziembowski, and D. Venturi, "Leakage-resilient storage," in *Proceedings of the 7th International Conference, Security and Cryptography for Networks(SCN)*, 2010, pp. 121–137.

[24] E. Boyle, S. Goldwasser, and Y. T. Kalai, "Leakage-resilient coin tossing," in *DISC*, 2011, pp. 181–196.

[25] Y. Dodis, A. Lewko, B. Waters, and D. Wichs, "Storing secrets on continually leaky devices," in *Proceedings of the 52nd IEEE Symposium on Foundation of Computer Science*, 2011, pp. xxx–xxx(to appear).

[26] S. Halevi and H. Lin, "After-the-fact leakage in public-key encryption," in *TCC*, 2011, pp. 107–124.

[27] S. Dziembowski and K. Pietrzak, "Intrusion-resilient secret sharing," in *Proceedings of the 48th IEEE Symposium on Foundation of Computer Science*, 2007, pp. 227–237.

[28] J. Radhakrishnan and A. Ta-Shma, "Bounds for dispersers, extractors, and depth-two superconcentrators." *SIAM Journal on Computing*, vol. 13, no. 1, pp. 2–24, 2000.

[29] S. Vadhan, "Pseudorandomness," to appear in Foundations and Trends in Theoretical Computer Science, 2011, http://people.seas.harvard.edu/ salil/pseudorandomness/.

[30] Y. Dodis, R. Gennaro, J. Håstad, H. Krawczyk, and T. Rabin, "Randomness extraction and key derivation using the cbc, cascade and hmac modes," in *Advances in Cryptology—CRYPTO 2004*, ser. LNCS, M. Franklin, Ed., vol. 3152. Springer-Verlag, 15–19 Aug. 2004, pp. 494–510.

[31] H. Krawczyk, "Cryptographic Extraction and Key Derivation: The HKDF Scheme," in *Advances in Cryptology - CRYPTO 2010*, ser. LNCS, T. Rabin, Ed., vol. 6223. Springer-Verlag, 2010, pp. 631–648.

[32] Y. Dodis, R. Ostrovsky, L. Reyzin, and A. Smith, "Fuzzy extractors: How to generate strong keys from biometrics and other noisy data," *SIAM Journal on Computing*, vol. 38, no. 1, pp. 97–139, 2008.

[33] A. Rao, "A 2-source almost-extractor for linear entropy," in *APPROX-RANDOM*, 2008, pp. 549–556.