CONJURA · OPEN PROBLEM

Is Every Seeded Extractor a Multi-Instance Extractor?

Extract from many correlated blocks with independent seeds; a good fraction of the outputs should look uniform, but only special extractors are known to do it

Status. Statement: AI-written from Jiaxin Guan, Daniel Wichs and Mark Zhandry, *Multi-Instance Randomness Extraction and Security against Bounded-Storage Mass Surveillance*, IACR ePrint 2023/409, TCC 2023. Not yet checked by a human. The source proves the property for extractors built from Hadamard and Reed-Muller codes and leaves the general question open in both directions, recording in a footnote that it tried hard to prove it and failed, as did several experts it consulted.

Category. *Randomness extraction; information theory.*

Conjecture (informal). *A seeded extractor turns one weak source into uniform bits. Now run it t times over t blocks that may be arbitrarily correlated, each with its own fresh independent seed, and suppose only that the blocks* jointly *carry an α fraction of full entropy. No single block need have any entropy at all, so no fixed output can be claimed uniform; but one hopes that some α-fraction of the outputs — chosen after the fact, depending on the source — is jointly indistinguishable from uniform, even to someone holding all the seeds and all the other outputs. The source proves this for two specific code-based extractors. Whether* every *seeded extractor has the property, or some extractor is a counterexample, is open.*

1 The setting

Write $\mathsf{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ for a function taking an n-bit source and a d-bit seed. Recall the standard notion: Ext is a (k,ε) *strong average min-entropy extractor* if for all jointly distributed (X,Y) with X over $\{0,1\}^n$ and $\mathbf{H}_\infty(X \mid Y) \geq k$, the triple $(U_d, \mathsf{Ext}(X;U_d), Y)$ is ε-close to (U_d, U_m, Y).

The multi-instance notion applies Ext separately to each of t blocks, with independent seeds, and asks for a guarantee that is necessarily about a *set* of indices rather than any fixed index. The reason it has to be a set, and a random one, is a two-line example from the source: let $i^* \leftarrow [t]$ be uniform, set X_{i^*} to all zeros and every other block uniform. The joint entropy rate is nearly 1, yet for each fixed i the min-entropy of X_i is at most $\log t$, far too little to extract even one nearly-unbiased bit. So no fixed coordinate can be claimed uniform, and the set of good coordinates must be allowed to depend on the source.

Definition 1 (Multi-instance extraction, [GWZ23, Def. 3.1]). $\mathsf{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ is $(t, \alpha, \beta, \varepsilon)$-*multi-instance extracting* if the following holds. Let $X = (X_1, \ldots, X_t)$ be any random variable with blocks $X_i \in \{0,1\}^n$ and $\mathbf{H}_\infty(X) \geq \alpha \cdot tn$. Then there is a random variable I_X, jointly distributed with X and supported on sets $I \subseteq [t]$ with $|I| \geq \beta \cdot t$, such that

$$(S_1, \ldots, S_t, \mathsf{Ext}(X_1; S_1), \ldots, \mathsf{Ext}(X_t; S_t)) \approx_\varepsilon (S_1, \ldots, S_t, Z_1, \ldots, Z_t),$$

where the $S_i \in \{0,1\}^d$ are independent uniform seeds, and $Z_i \in \{0,1\}^m$ is independent uniform for $i \in I_X$ while $Z_i = \mathsf{Ext}(X_i; S_i)$ for $i \notin I_X$.

2 Notation and parameters

- t is the number of blocks, n the length of each block, d the seed length, m the output length.
- α is the joint min-entropy rate of the whole source, β the guaranteed fraction of good coordinates, ε the statistical error.
- The interesting regime, and the one the source's application needs, is t huge — much larger than the size or the entropy of any one block.
- "Loss" below means $\alpha - \beta$: how far the fraction of uniform outputs falls short of the source's entropy rate.

3 The conjecture

Conjecture 1 ([GWZ23, §1.3]). *Every standard seeded extractor is also a multi-instance randomness extractor. Concretely: there is a function $L(n,m,d,t,\varepsilon)$ such that every (k,ε') strong average min-entropy extractor* $\mathsf{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ *is $(t,\alpha,\beta,\varepsilon)$-multi-instance extracting with $\beta \geq \alpha - L/n$, for a loss L that is polynomial in m,* $\log t$ *and* $\log(1/\varepsilon)$ *and independent of the extractor.*

Remark 1 (How the source states it, and what is being supplied here). The source poses it as a question with both answers live, page 7: "It remains as a fascinating open problem whether every standard seeded extractor is also a multi-instance randomness extractor or if there is some counterexample." It does not attach parameters to the general question.

The quantitative form above is therefore this statement's reading, chosen to match what the source *proves* for its own extractors — see Theorem 1, where the loss has exactly that shape. Two weaker readings are also worth settling and should be reported as such: that every seeded extractor is multi-instance extracting with *some* $\beta = \alpha - o(1)$, without a uniform loss function; and the qualitative form, that no extractor is a counterexample at any non-trivial β. A counterexample refutes all three at once, and is the outcome the source's phrasing keeps open.

Theorem 1 (What is known, [GWZ23, Cor. 3.4]). *For any $n,m,t,\varepsilon > 0, \alpha > 0$ there exist extractors* $\mathsf{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ *that are $(t,\alpha,\beta,\varepsilon)$-multi-instance extracting with either seed length $d = n$ and $\beta = \alpha - O(m + \log t + \log(1/\varepsilon))/n$, or seed length $d = O((\log n)(m + \log\log n + \log t + \log(1/\varepsilon)))$ and $\beta = \alpha - O(d)/n$.*

The two constructions are the Hadamard extractor $\mathsf{Ext}(x;s) = \langle x,s\rangle$ and a Reed-Muller composition with it. Both are standard; what is new is the analysis, which leans on list-decodability together with a property the source isolates and calls *hinting*: the values of $\mathsf{Ext}(x;S_i)$ on a particular exponentially large family of pairwise independent seeds can be compressed into a single hint much shorter than x. Hinting is a feature of local list-decoding algorithms for these two codes, and it is exactly the feature an arbitrary seeded extractor is not known to have.

Remark 2 (The obstruction, named). The natural route is to reduce to the ordinary extractor guarantee: find a large set I_X of blocks that individually have high min-entropy, then apply the extractor blockwise. The first step is available — this is the "block-entropy lemma" of Dodis, Quach and Wichs — and even in the stronger form that X_i for $i \in I_X$ has high min-entropy conditioned on *all past blocks* $X_1, \ldots, X_{i-1}$.

It is the conditioning on the future that fails, and it fails for a concrete reason: no such statement can hold conditioned on all other blocks past and future, as the source's example of X uniform subject to $\bigoplus_{i=1}^{t} X_i = 0$ shows. Since a multi-instance distinguisher sees extracted outputs from *all* the blocks, that is precisely what would be needed. In the source's words: "Unfortunately, this prevents us for using the block-entropy lemma to analyze multi-instance extraction, where the adversary sees some extracted outputs from all the blocks."

A weak version does survive the route, but only when t is small enough that one can afford to lose t bits of entropy from each block — and the regime that matters has t far larger than the entropy of a block.

Remark 3 (Evidence, on both sides). For the conjecture: it is true for two natural extractor families, and concurrent independent work of Dinur, Stemmer, Woodruff and Zhou — which arrives at the same notion from differential-privacy lower bounds — proves it for universal hash functions, by a completely different argument. Three families, two techniques.

Against it: the source records a failed serious attempt, in a footnote worth reading before starting. "We were initially convinced that the general result does hold and invested much effort trying to prove it via some variant of the above approach without success. We also mentioned the problem to several experts in the field who had a similar initial reaction, but were not able to come up with a proof."

The three families that do work — Hadamard, Reed-Muller, universal hashing — are all linear, which is a hint about where a counterexample would have to be looked for.

Remark 4 (Why the answer matters beyond extractor theory). Multi-instance extraction is a drop-in replacement for ordinary ex-

traction in constructions of incompressible encryption, and it is what lifts them from a single user to many: if an adversary stores an α fraction of the total size of all ciphertexts, it learns something about at most a $\beta \approx \alpha$ fraction of the messages. A proof of Conjecture 1 would make every such construction work with any extractor, including short-seed ones with no list-decoding structure. A counterexample would say that these applications depend on a property of the extractor, not on extraction as such, and it would be the more informative outcome.

4 Bibliography

[GWZ23] Jiaxin Guan, Daniel Wichs and Mark Zhandry. *Multi-instance randomness extraction and security against bounded-storage mass surveillance.* IACR ePrint 2023/409; TCC 2023. The source. The open problem and the footnote about the failed attempt are on page 7; the definition of multi-instance extraction is Definition 3.1; the constructions and their parameters are Corollary 3.4; the hinting property is Definition 3.2.

[DQW22] Yevgeniy Dodis, Willy Quach and Daniel Wichs. *Authentication in the bounded storage model.* EUROCRYPT 2022, Part III, LNCS 13277, pages 737–766, Springer, May/June 2022. The block-entropy lemma the failed approach starts from.

[DSWZ23] Itai Dinur, Uri Stemmer, David P. Woodruff and Samson Zhou. *On differential privacy and adaptive data analysis with bounded space.* IACR ePrint 2023/171. Concurrent independent work that reaches an equivalent notion from a different direction and proves it for universal hash functions.

[GWZ22] Jiaxin Guan, Daniel Wichs and Mark Zhandry. *Incompressible cryptography.* EUROCRYPT 2022, Part I, LNCS 13275, pages 700–730, Springer, May/June 2022. The single-user incompressible encryption schemes that multi-instance extraction lifts to the multi-user setting.

[Dzi06] Stefan Dziembowski. *On forward-secure storage.* CRYPTO 2006, LNCS 4117, pages 251–270, Springer, August 2006. The information-theoretic symmetric-key scheme whose extractor the source replaces to obtain multi-user security.

[Nis90] Noam Nisan. *Pseudorandom generators for space-bounded computation.* 22nd ACM STOC, pages 204–212, ACM Press, May 1990. Cited by the source for the definition of a seeded extractor that Conjecture 1 quantifies over.