

Different Flavours of HILL Pseudoentropy and Yao Incompressibility Entropy

Pihla Karanko

Aalto University, Finland

Abstract. There are two popular ways to measure computational entropy in cryptography: (HILL) pseudoentropy and (Yao) incompressibility entropy. Both of these computational entropy notions are based on a natural intuition.

- A random variable X has k bits of *pseudoentropy* if there exists a random variable Y that has k bits ‘real’ entropy and Y is computationally indistinguishable from X .
- A random variable X has k bits of *incompressibility* entropy if X cannot be efficiently compressed to less than k bits.

It is also intuitive, that if a random variable has high pseudoentropy, then it should also have high incompressibility entropy, because a high-entropy distribution cannot be compressed.

However, the above intuitions are not precise. Does ‘real entropy’ refer to Shannon entropy or min-entropy? What kind of correctness do we require from the compressor algorithm? Different papers use slightly different variations of both pseudoentropy and incompressibility entropy.

In this note we study these subtle differences and see how they affect the parameters in the implication that pseudoentropy implies incompressibility.

1 Preliminaries - Shannon Entropy and Min-Entropy

The most common definition for (information theoretic) entropy is *Shannon entropy*, that measures the average amount of uncertainty of a random variable in bits.

Definition 1 (Shannon entropy). *Let X be a random variable and S the support of X . Shannon entropy of X is defined as*

$$H(X) := - \sum_{x \in S} \Pr[X = x] \log_2(\Pr[X = x]).$$

Often in cryptography, the *average* amount of uncertainty is not so interesting, but we are interested in the worst-case scenario. For example, if an adversary tries to guess the value of a random variable X , we do not care about how difficult this guessing is on average, but rather, if the adversary guesses the most likely value of X , how much uncertainty does this guess have. This ‘guessing entropy’ is best captured by min-entropy.

Definition 2 (Min-entropy). Let X be a random variable and S the support of X . Min-entropy of X is defined as

$$H_\infty(X) := -\log_2(\max_{x \in S}(\Pr[X = x])).$$

2 Pseudoentropy

Pseudoentropy was originally defined by Håstad, Impagliazzo, Levin and Luby [HILL99], they say that a variable has pseudoentropy k if it is computationally indistinguishable from a distribution with Shannon entropy k . Later, Barak, Shaltiel and Wigderson [BSW03] study computational analogues of min-entropy, and in that context they define a *min-entropy* based version of HILL pseudoentropy, which they call *HILL-type pseudoentropy*. Subsequent papers (e.g. [HLR07], [Rey11], [HW15], [HMS22]) sometimes refer to both of these pseudoentropy definitions as ‘HILL entropy’ or ‘HILL pseudoentropy’. In this note, we want to highlight the difference, and hence, we refer to them as **HILL avg-pseudoentropy** and **BSW min-pseudoentropy** respectively. The following definition shows both versions and marks the difference with colors.

Definition 3 (HILL avg-pseudoentropy, BSW min-pseudoentropy). Let X_λ be a random variable of length λ . X_λ has **HILL avg-pseudoentropy** or **BSW min-pseudoentropy** $(k_\lambda)_{\lambda \in \mathbb{N}}$ if for every λ , there exists a random variable Y_λ such that

Indistinguishability X_λ and Y_λ are computationally indistinguishable¹ and
Entropy (*HILL avg*) the Shannon entropy $H(Y_\lambda) \geq k_\lambda$.
(BSW min) the min-entropy $H_\infty(Y_\lambda) \geq k_\lambda$.

HILL avg additionally requires that all the random variables are polynomial time sampleable.

Remark 1. Since min-entropy \leq Shannon entropy (by definition of the entropies), we get that if a distribution has **BSW min-pseudoentropy** k , then it has **HILL avg-pseudoentropy** k . The converse is not always true.

3 Incompressibility entropy

Yao [Yao82] introduced the idea to measure computational entropy by how much a variable can be compressed. Since the work of Yao, there have been several different ways to define the incompressibility entropy ([TVZ05], [BSW03], [HLR07], [HW15]), all of them following the same intuition: a variable X has k bits incompressibility entropy, if there is no compressor-decompressor pair where the

¹ both HILL and BSW leave the precise parameter open s.t. one can define indistinguishability differently per use case (e.g. small constant, fixed polynomial, or negligible computational distance).

compressor can compress $x \in X$ to less than k bits and the decompressor can recover x . We now review the slight differences in the formalizations of Yao's incompressibility entropy (sometimes also referred to as Yao pseudoentropy).

Trevisan, Vadhan and Zuckerman [TVZ05], TVZ, consider a compressor, that is perfectly correct, but the output length of the compressor might vary. In turn, Barak, Shaltiel and Wigderson [BSW03], BSW, and Hsiao, Lu and Reyzin [HLR07] consider compressor with fixed output length, but the output of the compressor does not always need to be correctly decompressible, more precisely, the compressor is allowed to make more mistakes, the shorter the output. In turn, Hubacek and Wichs [HW15], HW, use a weaker and simpler version of [HLR07], since it is enough for their use-case, namely, they consider a *fixed* success probability for the compressor regardless of its output length.

In Definition 4, we compare the definitions of incompressibility by TVZ, BSW and HW. Analogously to conditional probability, we can also consider *conditional* (computational) entropy, that is, how much (computational) entropy a variable has conditioned on knowing some other variable. *Conditional incompressibility*² was first defined by HLR [HLR07], HLR take the natural approach of simply giving the compressor and decompressor the conditional information in addition to their original input. HLR base their definition on the BSW version of incompressibility, but the same idea can easily be applied to any of the incompressibility definitions.

Definition 4 (Conditional³ Incompressibility Entropy). *Let X and Z be efficiently samplable random variables (not necessarily independent). The variable $(X|Z)$ is k_λ -incompressible, if for every polynomial size circuit family pair of a compression function $\text{Cmpr}_\lambda(\cdot, \cdot)$ and a decompression function $\text{Decmpr}_\lambda(\cdot, \cdot)$ it holds that if (TVZ: expected) output length $l(\lambda)$ of Cmpr is $< k_\lambda$ (TVZ: $\leq k_\lambda$) then*

Correctness *there exists a negligible function ϵ s.t.*

$$(HW) \Pr[\text{Decmpr}(\text{Cmpr}(X, Z), Z) = X] \leq 1/2 + \epsilon(\lambda)$$

$$(BSW) \Pr[\text{Decmpr}(\text{Cmpr}(X, Z), Z) = X] \leq 2^{l-k} + \epsilon(\lambda)$$

$$(TVZ) \exists x, z \in \text{Supp}(X, Z) \text{ s.t. } \text{Decmpr}(\text{Cmpr}(x, z), z) \neq x$$

Remark 2. Definition 4 directly implies that if X has k bits of BSW incompressibility entropy, then X has k bits of HW incompressibility entropy (but

² Defining conditional *pseudoentropy* is a bit more involved as it often depends on the application whether we want to consider average-case or worst case conditioning and this choice affects what conditional entropy notion to choose in the definition. The subtleties related to conditional BSW min-pseudoentropy are very nicely discussed in [HLR07]. Similar considerations can be done for HILL avg-pseudoentropy too, while the simplest choice, used by [HILL99] among others, is to define conditional pseudoentropy via simply using *conditional* Shannon entropy in Definition 3.

³ We present here the conditional version for full generality. The unconditional version can be obtained by considering empty Z .

the converse is not necessarily true). This is because $2^{l-k} \leq 1/2$ for all $l < k$, assuming k is an integer⁴.

Next, we show that TVZ incompressibility implies BSW incompressibility, but with quite big loss. It is not clear whether the other direction holds for any non-trivial choice of parameters.

Claim 1 (TVZ Incompressibility Implies BSW Incompressibility). *Let X be a random variable of length λ . If X is k_λ TVZ incompressible, then X is $k'_\lambda = k_\lambda + 1 - \lambda/2$ BSW incompressible.*

Proof. WLOG $\lambda \geq k, k'$. Assume towards contradiction that X is not k'_λ BSW incompressible, that is, there exists a compressor-decompressor pair, where compressor's output length l is $< k'$, and there exists a polynomial p s.t. for infinitely many λ

$$\Pr[\text{Decmpr}(\text{Cmpr}(X, Z), Z) = X] > 1/2^{l-k'} + 1/p(\lambda).$$

Now consider the new compressor:

```

Cmprnew(x)
-----
x̃ ← Cmpr(x)
if Decmpr(x̃) ≠ x
    return x
return x̃

```

Now the new compressor is perfectly correct, since the new decompressor can just check the length of \tilde{x} , and if it is λ , return \tilde{x} as is and else return $\text{Decmpr}(\tilde{x})$. Now the expected output length of the new compressor is (for infinitely many λ)

$$\begin{aligned} \mathbb{E}[|\text{Cmpr}_{\text{new}}(X)|] &= \Pr[|\text{Cmpr}_{\text{new}}(X)| = l]l + \Pr[|\text{Cmpr}_{\text{new}}(X)| = \lambda]\lambda \\ &< l + \left(1 - 1/2^{l-k'} - 1/p(\lambda)\right) \lambda & (1) \\ &\leq (k' - 1) + \left(1 - 1/2^{(k'-1)-k'} - 1/p(\lambda)\right) \lambda \\ &< k' - 1 + \lambda/2 = k \end{aligned}$$

where the value of (1) is maximized when l is maximal, that is, $l = k' - 1$. The last line is a contradiction with X being k TVZ incompressible. \square

⁴ This is explicitly assumed only in the HW definition. However, limiting k to integers makes sense in the BSW definition too.

4 Pseudoentropy Implies Incompressibility

Now we present the different versions of the claim that pseudoentropy implies incompressibility. Firstly, the following claim was proven by [Wee04] (see [Wee04] for the precise parameters in the \mathcal{O}).

Claim 2 (High HILL avg-pseudoentropy implies high TVZ incompressibility entropy [Wee04]). *If a random variable X has k bits HILL pseudoentropy, then X has $k - \mathcal{O}(\log \lambda)$ bits (TVZ) incompressibility entropy, where λ is the length of X .*

The following claim was proven by [HLR07], they show a conditional version, and the following is a special case of that.

Claim 3 (High BSW min-pseudoentropy implies high BSW incompressibility entropy). *If random variable X has k bits BSW pseudoentropy, then X has k bits BSW incompressibility entropy.*

Claim 3 is easy to prove also for HW incompressibility entropy. Below we show a proof for both versions of Claim 3.

Proof. Assume towards contradiction that X_λ has $< k_\lambda$ bits incompressibility entropy. Now $\exists \text{Cmpr, Decmpr}$ and exists a polynomial p s.t. for infinitely many λ :

$$\begin{aligned} \text{(HW)} \quad & \Pr_{X_\lambda} [\text{Decmpr}(\text{Cmpr}(X_\lambda)) = X_\lambda] \geq 1/2 + 1/p(\lambda) \\ \text{(BSW)} \quad & \Pr_{X_\lambda} [\text{Decmpr}(\text{Cmpr}(X_\lambda)) = X_\lambda] \geq 2^{l_\lambda - k_\lambda} + 1/p(\lambda) \end{aligned}$$

and output length l_λ of Cmpr is $< k_\lambda$.

Now since X_λ has k_λ bits BSW pseudoentropy, there is a random variable Y_λ with k_λ -bits min-entropy, s.t. X_λ is computationally indistinguishable from Y_λ .

However, consider the following distinguisher for Y_λ and X_λ :

```


$$\frac{\mathcal{A}(x)}{\tilde{x} \leftarrow \text{Cmpr}(x)}$$

if  $|\tilde{x}| \geq k_\lambda$ 
  return 1
 $x' \leftarrow \text{Decmpr}(\tilde{x})$ 
if  $x' = x$ 
  return 0
return 1

```

Now for all λ

$$\Pr[\mathcal{A}(Y_\lambda) = 1] \geq \Pr[\text{Decmpr}(\tilde{x}) \neq x] \geq 1 - 2^{l_\lambda - k_\lambda}$$

where the last inequality follows from the fact that Y_λ has k_λ bits min-entropy while $|\tilde{x}| = l_\lambda$ (so decompressor needs to 'guess' at least $k_\lambda - l_\lambda$ bits).

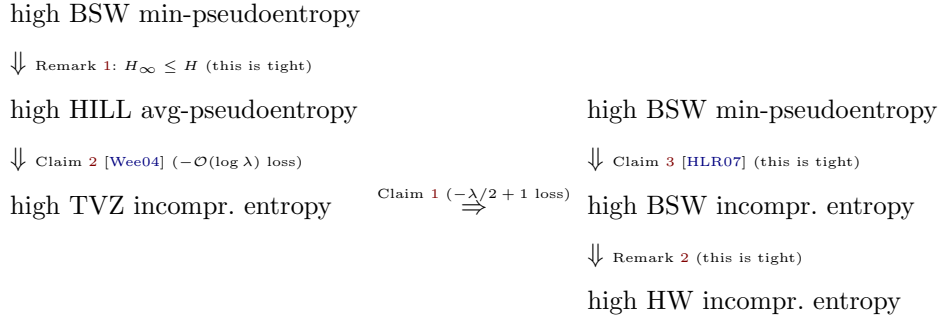
On the other hand

$$\begin{aligned} \Pr[\mathcal{A}(X_\lambda) = 0] &= \Pr[\text{Decmpr}(\tilde{x}) = x] \geq 1/2 + 1/p(\lambda) \quad \left| \text{by HW contradiction assumption} \right. \\ &\geq 2^{l_\lambda - k_\lambda} + 1/p(\lambda) \quad \left| \text{by BSW contradiction assumption} \right. \end{aligned}$$

for infinitely many λ , which is a contradiction, because \mathcal{A} distinguishes X_λ and Y_λ with non-negligible probability. \square

Note that the proof of Claim 3 does not work for HILL avg-pseudoentropy, because high Shannon entropy does not directly imply hard-to-guess⁵, hence, Claim 3 is more tight than Claim 2.

Conclusion of all the results covered in this note:



5 Acknowledgements

Thank you Chris Brzuska for valuable feedback on drafts of this note.

References

- BSW03. Boaz Barak, Ronen Shaltiel, and Avi Wigderson. Computational analogues of entropy. In *APPROX 2003 and 7th International Workshop on Randomization and Approximation Techniques in Computer Science, RANDOM 2003*, volume 2764 of *Lecture Notes in Computer Science*, pages 200–215. Springer, 2003.
- HILL99. Johan Håstad, Russell Impagliazzo, Leonid A. Levin, and Michael Luby. A pseudorandom generator from any one-way function. *SIAM Journal on Computing*, 28(4):1364–1396, 1999.

⁵ Consider for example a random variable Y which is 0^λ with probability $1/2$ and $1||S_{\lambda-1}$ otherwise, where $S_{\lambda-1}$ is a uniformly random string of length $\lambda - 1$. Now the Shannon-entropy of Y is $H(Y) = \frac{1}{2} \cdot 1 + \frac{1}{2}(\lambda - 1) = \lambda/2$ while the min-entropy is $H_\infty(Y) = 1$, that is, the value of Y can be guessed with constant probability ($1/2$).

- HLR07. Chun-Yuan Hsiao, Chi-Jen Lu, and Leonid Reyzin. Conditional computational entropy, or toward separating pseudoentropy from compressibility. In Moni Naor, editor, *Advances in Cryptology – EUROCRYPT 2007*, volume 4515 of *Lecture Notes in Computer Science*, pages 169–186. Springer, Heidelberg, May 2007.
- HMS22. Iftach Haitner, Noam Mazor, and Jad Silbak. Incompressibility and next-block pseudoentropy. Cryptology ePrint Archive, Report 2022/278, 2022. <https://eprint.iacr.org/2022/278>.
- HW15. Pavel Hubacek and Daniel Wichs. On the communication complexity of secure function evaluation with long output. In Tim Roughgarden, editor, *ITCS 2015: 6th Conference on Innovations in Theoretical Computer Science*, pages 163–172. Association for Computing Machinery, January 2015.
- Rey11. Leonid Reyzin. Some notions of entropy for cryptography - (invited talk). In Serge Fehr, editor, *ICITS 11: 5th International Conference on Information Theoretic Security*, volume 6673 of *Lecture Notes in Computer Science*, pages 138–142. Springer, Heidelberg, May 2011.
- TVZ05. Luca Trevisan, Salil Vadhan, and David Zuckerman. Compression of samplable sources. *Comput. Complex.*, 14(3):186–227, December 2005.
- Wee04. Hoeteck Wee. On pseudoentropy versus compressibility. In *Proceedings. 19th IEEE Annual Conference on Computational Complexity, 2004.*, pages 29–41, 2004.
- Yao82. Andrew Chi-Chih Yao. Theory and applications of trapdoor functions (extended abstract). In *23rd Annual Symposium on Foundations of Computer Science*, pages 80–91. IEEE Computer Society Press, November 1982.