

Entropy Loss is Maximal for Uniform Inputs

Leonid Reyzin*

Technical Report BUCS-TR-2007-011

September 20, 2007

Abstract

A secure sketch (defined by Dodis et al.) is an algorithm that on an input w produces an output s such that w can be reconstructed given its noisy version w' and s . Security is defined in terms of two parameters m and \tilde{m} : if w comes from a distribution of entropy m , then a secure sketch guarantees that the distribution of w conditioned on s has entropy \tilde{m} , where $\lambda = m - \tilde{m}$ is called the *entropy loss*. In this note we show that the entropy loss of any secure sketch (or, more generally, any randomized algorithm) on any distribution is no more than it is on the uniform distribution.

1 Introduction

Generalizing the definition of min-entropy, $\mathbf{H}_\infty(A) \stackrel{\text{def}}{=} -\log(\max_a \Pr[A = a])$, Dodis et al. [DORS07] define average min-entropy

$$\tilde{\mathbf{H}}_\infty(A | B) \stackrel{\text{def}}{=} -\log \left(\mathbb{E}_{b \leftarrow B} \left[\max_a \Pr[A = a | B = b] \right] \right) = -\log \left(\mathbb{E}_{b \leftarrow B} \left[2^{-\mathbf{H}_\infty(A|B=b)} \right] \right).$$

For a metric space \mathcal{M} , an $(\mathcal{M}, m, \tilde{m}, t)$ secure sketch [DORS07] is a pair of (possibly randomized) algorithms SS and Rec , such that SS takes an element $w \in \mathcal{M}$ and outputs $s = \text{SS}(w)$, and Rec takes s and an element $w' \in \mathcal{M}$ of distance at most t from w , and produces w . Of interest in this note is the security property: for any random variables W over \mathcal{M} and I over $\{0, 1\}^*$ such that $\tilde{\mathbf{H}}_\infty(W | I) \geq m$, the sketching algorithm guarantees that $\tilde{\mathbf{H}}_\infty(W | \text{SS}(W), I) \geq \tilde{m}$ (the variable I here models external information that may be available about w regardless of the sketching algorithm). The value $\lambda = m - \tilde{m}$ is called the *entropy loss* of a secure sketch.

We show here that if $|\mathcal{M}| = 2^n$ and (SS, Rec) is an $(\mathcal{M}, n, n - \lambda, t)$ -secure sketch, then it is also an $(\mathcal{M}, m, m - \lambda, t)$ -secure sketch for any m . In other words, a secure sketch needs to be analyzed only for the highest-entropy distribution—i.e., the uniform distribution over \mathcal{M} —and will have the same or smaller entropy loss for lower-entropy distributions. We should note, of course, that we do not preclude the possibility that a smaller entropy loss can be obtained by a direct analysis for lower-entropy inputs (although we do not know of any nondegenerate examples of this, i.e., when the entropy loss is not trivially bounded by the input entropy).

*<http://www.cs.bu.edu/~reyzin>. Boston University, Department of Computer Science, 111 Cummington St., Boston MA 02215 USA. This work was supported in part by the National Science Foundation under Grant Nos. CCR-0311485, CCF-0515100, CNS-0546614, and CNS-0202067.

Nothing in our argument uses the fact that Rec exists—in fact, our result applies to any randomized function f , not just SS.

Lemma. *Let f be a probabilistic function on a domain \mathcal{M} of size 2^n (note that n need not be an integer): each $w \in \mathcal{M}$ defines a random variable, which we denote by $f(w)$ (we assume the coins used to compute f are fresh and independent of any distribution on \mathcal{M}). Let U be a uniformly distributed random variable on \mathcal{M} . Suppose for some λ , we have $\tilde{\mathbf{H}}_\infty(U | f(U)) \geq n - \lambda$. Then for any random variable W on \mathcal{M} and (possibly correlated to W) random variable I , we have $\tilde{\mathbf{H}}_\infty(W | f(W), I) \geq \tilde{\mathbf{H}}_\infty(W | I) - \lambda$.*

Proof. First we express λ in terms of properties of f :

$$\begin{aligned}
\tilde{\mathbf{H}}_\infty(U | f(U)) &= -\log \mathbb{E}_y \left[\max_w \Pr[U = w | f(U) = y] \right] \\
&= -\log \sum_y \max_w \Pr[U = w \wedge f(U) = y] \\
&= -\log \sum_y \max_w \Pr[f(w) = y \wedge U = w] \\
&= -\log \frac{1}{|\mathcal{M}|} \sum_y \max_w \Pr[f(w) = y] \\
&= n - \log \sum_y \max_w \Pr[f(w) = y] \\
&= n - \lambda.
\end{aligned}$$

Let W_i be the distribution W conditioned on $I = i$. Let $h_i = \mathbf{H}_\infty(W_i)$, i.e., for any $w \in \mathcal{M}$, $\Pr[W_i = w] \leq 2^{-h_i}$. Note that $\tilde{\mathbf{H}}_\infty(W | I) = -\log \mathbb{E}_i [2^{-h_i}]$.

$$\begin{aligned}
\tilde{\mathbf{H}}_\infty(W | f(W), I) &= -\log \mathbb{E}_{y,i} \left[\max_w \Pr[W = w | f(W) = y \wedge I = i] \right] \\
&= -\log \sum_{y,i} \max_w \Pr[W = w \wedge f(W) = y \wedge I = i] \\
&= -\log \sum_{y,i} \max_w \Pr[f(w) = y \wedge W = w \wedge I = i] \\
&= -\log \sum_{y,i} \max_w \Pr[f(w) = y] \Pr[W_i = w] \Pr[I = i] \\
&\geq -\log \sum_{y,i} \max_w \Pr[f(w) = y] 2^{-h_i} \Pr[I = i] \\
&= -\log \left(\sum_i 2^{-h_i} \Pr[I = i] \right) \left(\sum_y \max_w \Pr[f(w) = y] \right) \\
&= -\log \left(\mathbb{E}_i [2^{-h_i}] \right) - \log \left(\sum_y \max_w \Pr[f(w) = y] \right) \\
&= \tilde{\mathbf{H}}_\infty(W | I) - \lambda.
\end{aligned}$$

□

References

- [DORS07] Yevgeniy Dodis, Rafail Ostrovsky, Leonid Reyzin, and Adam Smith. Fuzzy extractors: How to generate strong keys from biometrics and other noisy data. Technical Report 2003/235, Cryptology ePrint archive, <http://eprint.iacr.org>, 2007. Previous version appeared at *EUROCRYPT 2004*.