Uniform Compression Functions Can Fail to Preserve "Full" Entropy

Daniel R. L. Brown*

November 23, 2012

Abstract

To have "full" entropy has been defined in a draft NIST standard to be to have min-entropy very close, proportionally, to the min-entropy of a uniform distribution. A function is uniform if all its preimages have the same size. This report proves that the output of any uniform compression function can fail to have full entropy, even when the input has full entropy.

1 Introduction

This report involves the notion of *full entropy* from draft standards ANSI X9.82-4 and NIST Special Publication 800-90C. These draft standards specify methods for random bit generation.

A random variable with full entropy has a probability distribution that is very close to be uniform on its domain. Specifically, its min-entropy is within a factor of the min-entropy of the uniform distribution. See Definition 1 of this report for a formal definition of full entropy.

If the output of the random bit generation has full entropy, then it may be quite secure for use as a cryptographic key.

The draft standards include some constructions, *non-deterministic random bit generators* (NRBG), that aim to produce full entropy outputs. The NRBG takes an input seed value, and compresses it with deterministic functions. If the input seed has a uniform distribution, then it plausible that the

*Certicom Research

resulting output has full entropy. If the input seed has a *non-uniform* but high-entropy distribution, then it is still plausible that the output has full entropy, because the constructions use considerable compression. Indeed, the draft standards go a little further and state the following *presumption*: If the input has full entropy, then so does the output. This is used as a possible justification for why the NRBG construction is secure, provided it is supplied with a full entropy seed.

The main observation of this report is that this presumption is likely false, and thus too strong. Specifically, if a compression function f is uniform in the sense of having preimages of constant size (see also Definition 2), then there exists a pathological distribution X (depending on f) with full entropy such that $f(X)$ does not have full entropy. (Exceptions to the result occur if the range is trivial.)

The compression functions used in the NRBG construction of the draft standards are not uniform, so the result of the report does not apply in a strict sense. Intuitively, however, it would seem that non-uniform compression compression would fare worse at preserving a measure of closeness to uniformity such as full entropy. Indeed, it ought to be straightforward to adapt the proof used in this report to the case of non-uniform compression functions.

1.1 Previous Work

It seems quite likely that the main result of this report is just a special case of a more general result that was previously published.

2 Definitions

Consider a random variable X taking values in a set S of size $N = 2^n$. For example, the values of X could be bit strings of length n. Let $P(X = x)$ denote the probability that variable X takes value of $x \in S$.

The min-entropy $H_\infty(X)$ of random variable X is defined to be

$$H_\infty(X) = \min_{x \in S} \left(-\log_2(P(X = x)) \right). \tag{1}$$

Equivalently, $H_\infty(X) = -\log_2 \max_{x \in S} P(X = x)$.

Definition 1 (Full Entropy)**.** *The random variable X taking values in a set S of size $N = 2^n$ is said to have* full entropy *for S, of* fullness level *g, if*

$$H_\infty(X) \geq n(1 - 2^{-g}). \tag{2}$$

If the fullness level is not stated, then full entropy is deemed to be full entropy with fullness level of 64 bits.

In the draft standards the fullness level in full entropy is always assumed to be 64 bits.

Definition 2 (Uniform)**.** *Let A be a set of size 2^m, and B a set of size 2^n. A function $f : A \to B$ is* uniform *if all its preimages have the same size.*

3 Main Result

The following lemma shows that a uniform compression function $f : A \to B$ generally does not always preserve full entropy, by finding a pathological input full entropy distribution upon application of f yields a distribution with a strictly smaller level of fullness.

Lemma 1. *Let $f : A \to B$ be a uniform function, as in Definition 2 and its notation. Suppose $n > 0$, so f has a non-trivial range. A variable X taking values in A exists such that*

- *X has full entropy for A, of fullness level $g \geq \log_2(m/n)$, and*
- *$Y = f(X)$ does not have full entropy for B of fullness level G, for any $G > g - \log_2(m/n)$.*

In particular, if $m > n$, so f is a compression function, then Y does not have full entropy of fullness level g.

Proof. Let $y_0 \in B$. Let random variable X with domain A have distribution given by

$$P(X = x) = \begin{cases} 2^{-m(1-2^{-g})} & \text{if } f(x) = y_0, \text{ and} \\ (2^{-m})\left(\frac{1-2^{(-n+m2^{-g})}}{1-2^{-n}}\right) & \text{otherwise.} \end{cases} \tag{3}$$

First, it must be verified that such a random variable X exists. To show this, it must be shown that the probabilities above are all non-negative, and that they also sum to 1 when x is summed over A.

If $f(x) = y_0$, then $P(X = x) > 0$ because the $P(X = x) = 2^r$ for a real number r, and $2^r > 0$. Otherwise $P(X = x)$ is given by the second case above. Multiplying by positive values 2^m and then by value $1 - 2^{-n}$, which is positive because $n > 0$, it suffices to show that $1 - 2^{(-n+m2^{-g})} \geq 0$. The latter is equivalent to $1 \geq 2^{(-n+m2^{-g})}$, which is equivalent to $0 \geq -n + m2^{-g}$, which is equivalent to $n \geq m2^{-g}$, which is equivalent to $2^g/m \geq 1/n$, which is equivalent to $2^g \geq m/n$, which is equivalent to $g \geq \log_2(m/n)$, which is a hypothesized condition of the lemma.

To sum the probabilities, note that the number of $x \in A$ with $f(x) = y_0$ is 2^{m-n} because f is uniform. So, the number of remaining x is $2^m - 2^{m-n}$. Therefore,

$$\begin{aligned}\sum_{x \in A} P(X = x) &= \left(2^{m-n} 2^{-m(1-2^{-g})}\right) \\ &\quad + \left(\left(2^m - 2^{m-n}\right)\left(2^{-m}\right)\left(\frac{1 - 2^{(-n+m2^{-g})}}{1 - 2^{-n}}\right)\right) \qquad (4) \\ &= \left(2^{(-n+m2^{-g})}\right) + \left(1 - 2^{(-n+m2^{-g})}\right) \\ &= 1\end{aligned}$$

It remains to show that the distribution X has full entropy for A, with fullness level g, and that $f(X)$ lacks the full entropy of the requisite fullness levels.

To show that X has full entropy, of fullness level g, it suffices to show that $P(X = x) \leq 2^{-m(1-2^{-g})}$ for all X. This holds by definition of X for x such that $f(x) = y_0$, where the upper bound is met with equality. Therefore, it suffices to show that the second probability from (3) falls under the same upper bound: which is to say that it must be shown that:

$$\left(2^{-m}\right)\left(\frac{1 - 2^{(-n+m2^{-g})}}{1 - 2^{-n}}\right) \leq 2^{-m(1-2^{-g})}$$

Multiply both sides by the positive number $2^m(1-2^{-n})$, to get the equivalent inequality:

$$1 - 2^{(-n+m2^{-g})} \leq 2^{m2^{-g}}(1 - 2^{-n}).$$

Add $2^{-n+m2^{-g}}$ to both sides, take base-two logarithms, to get the equivalent inequality $0 \leq m2^{-g}$, which is true because $m > 0$.

Lastly, it must be shown that $f(X)$ does not have full entropy of fullness level G. To see this, $P(f(X) = y_0) = 2^{m-n}2^{-m(1-2^{-g})} = 2^{-n(1-(m/n)2^{-g})} = 2^{-n(1-2^{-(g-\log_2(m/n))})} > 2^{-n(1-2^{-G})}$. Hence $H_\infty(f(X)) < n(1-2^{-G})$. □

4 Discussion

This section elaborates further.

4.1 Exceptional Cases: Falsity for Trivial Ranges

Lemma 1 requires $n > 0$, so that $|B| \geq 2$. This is not merely for the convenience of the proof, because when $|B| = 1$, the random variable $f(X)$ is constant in the range, and thus uniform, regardless of the distribution of X.

4.2 Exceptional Cases: Extension to Larger Domain

Lemma 1 requires fullness level $g \geq \log_2(m/n)$ for distribution X. This condition is used in the proof to ensure the constructed distribution X exists. A natural question is what happens in the exceptional case of $g < \log_2(m/n)$.

Taking the default fullness level $g = 64$, this implies that $m > 2^{64}n$. Let n take smallest value allowed in the lemma: $n = 1$. Then $|B| = 2^n$, so we can assume that $B = \{0, 1\}$. In the exceptional case at hand, we would have $m > 2^{64}$. For concreteness, suppose that $m = 2^{65}$, and that the set A be could consist of all bit strings of length at most 2^{64}, together with a null value.

In context of the NRBG construction from the draft standards, this size of m is too large to be relevant. So the discussion here is only for completeness, not for relevance to the draft standards.

Because $\log_2(m/n)$ exceeds 64, the construction of X in the proof of Lemma 1 cannot be used. The following alternative construction of a distribution X on A with full entropy of fullness level 64 bits can be used instead. Every element $x \in A$ has probability at most $2^{-m(1-2^{-64})} = 2^{-(2^{65}-2)}$. Choose the distribution X such that $2^{2^{65}-2}$ elements, in a set A', have this maximal probability under X, and the rest have probability 0. In other words, X has

a uniform distribution on a subset A' of A, which is a quarter of the size of A.

Since $f : A \to B$ is uniform, the preimage of any $y \in B$, has $2^{m-1} = 2^{2^{65}-1}$ elements: half the elements of A. Arrange X so that the A' lies entirely in one preimage, say $f^{-1}(1)$. In this case, $f(X) = 1$ with probability 1, and $f(X)$ has min-entropy zero. Therefore $f(X)$ does not have full entropy for any level of fullness.

More generally, if m is sufficiently large, then there exists an full entropy distribution which maps under a given compression function to a zero entropy output. If $n > 1$, this output distribution cannot have full entropy.

4.3 Pathological Distributions

This report does not prove that the NRBG constructions in the draft standards *always* produce outputs with less than the claimed full entropy. Indeed, the NRBG constructions may possibly produce full entropy outputs given non-pathological full entropy inputs.

All that this report proves is that *pathological* full entropy distributions always exist for each deterministic compression function, such that if the pathological input is fed into the function, then the output has less than full entropy.

This result only implies that an actual proof of security for the NRBG construction from the draft standards cannot be based solely on the uniform NRBG constructions always preserving full entropy.

4.4 Construction of the Pathological Distribution

A computationally limited adversary can feasibly construct the pathological distribution X from Lemma 1. In other words, the pathological distribution X is constructable, not just existential. To construct the distribution, the adversary first chooses x uniformly at random from A. Then the adversary computes $f(x)$. If $f(x) \neq y_0$, then the adversary, with a small probability p will go back and re-generate a new x uniformly at random from A instead. (Just to clarify: the adversary need not iterate the test and re-generation: only one test and re-generate suffices.) The value of p can be chosen such that the final result has the probability distribution from the proof of Lemma 1.

Although the random variable X is constructable, it is not clear how an adversary could supply such a variable to its victim. But if it could, then

the adversary could cause the output to have less than full entropy.

4.5 The Aim for Information-Theoretic Security

The draft standards use the presumption of full entropy preservation (by certain specific compression functions) to justify further security claims about the NRBG constructions: the NRBG construction provide *information-theoretic security*. Although the draft standards do not define this exactly, other than in terms of full-entropy, a reasonable interpretation is that the term implies that the NRBG security does not rely on the hardness of a computational problem.

For example, an NRBG constructed from SHA-1 ought to be able to produce a 256-bit key in such a way that an adversary cannot exploit the size of SHA-1 to launch a 2^{160} step attack to determine this key. To do this, of course, requires the NRBG to fed an input seed of full entropy with considerably more than 256 bits.

But, at least intuitively to the author, the NRBG constructions may possibly provide the some security objective similar in spirit to this. For the NRBG to have this property, the underlying components would necessarily need to have some at least mild security properties, such as some kind of pseudorandomness. For example, if the underlying components were the constant zero function, then the NRBG constructions would be insecure.

4.6 Extension to Non-Uniform Compression Functions

Lemma 1 only addresses the simple case of uniform compression functions. The compression functions used in the draft standards are almost certainly non-uniform. The author's intuition is that such non-uniform compression functions drastically worsen the preservation of full entropy. The informal reasoning for this intuition has two parts. Firstly, full entropy is a measure of closeness to uniformity, and uniform compressions would seem to have an advantage over non-uniform in producing nearly uniform output. Secondly, non-uniform compression function by definition introduce non-uniformity even when supplied with uniform inputs.

It seems that some careful adjustments of the proof of the lemma could extend the result to the case of non-uniform function. The critical aspect to make this extension that the y_0 in the pre-image would have to have a pre-image of size at least 2^{m-n}. (Since f is no longer uniform now, one cannot

presume that 2^{m-n} is an integer.) Then one could use almost the distribution X as described in the proof except that the instead of the first case covering all x in the preimage it instead covers a set of size very close to 2^{m-n}, which is a subset of the preimage. One might have to adjust the probabilities in the second case to accommodate this slight deviation.

This would also entail finding y with the requisite larger than average pre-image. Such a y certainly exists. Taking x uniformly distributed gives a $y = f(x)$ with preimage probably larger than average. To find such a y provably may involve a hard computational problem related to the function f. In the information-theoretic setting, one refuses to rely on the hardness of such problems, and one assumes that such a y can be found easily from the description of f.

4.7 Independence of the Compression Function

It can be argued intuitively that, because the compression function is independent of any naturally occurring bias in the full entropy input seed, the main result of this report in non-applicable. This is yet another strong heuristic argument for the security of the NRBG.

The underlying intuition to this argument involves *independence*. If one is to raise the argument from heuristic to rigorous, one must formalize a notion of independence. The two main formal notions of independence (that the author is aware of) are probabilistic and algebraic (including linear). It is unclear how algebraic independence could be applied here, but perhaps it could be.

Applying probabilistic independence would seem to entail assigning the probabilities to the compression function. A potential formal problem with this approach is that the NRBG compression function is public, so the probabilities would no longer represent the adversary's lack of information. One may view the compression function parameters as random variable that were subsequently leaked to the adversary. One could invoke a computational assumption to say that the parameters of NRBG construction reveal nothing useful to the adversary (much like a public key does not reveal the private key). In this case, one really can use a notion of independence, but one would be relying on some kind of computational problem, and would lose some degree of information-theoretic security.

4.8 Comparison to Key Derivation Functions

One interpretation of the results of this report is that if the output of a deterministic compression function is to claim some of kind information-theoretic security, then its input has to be even closer to uniform.

In the context of the NRBG construction, it may take input of m bits (perhaps streamed from some raw entropy source), and output n bits. If the initial m input bits are already closer to uniform than the output n bits, then why bother at all with the NRBG construction.

Indeed, it seems that the purpose of the NRBG is similar to one of the purposes of a the key derivation function. Loosely speaking, this purpose is to take a distribution that has some significant bias, perhaps with only half the maximal amount of entropy, and then product output with the full entropy.

It should take input with full entropy of low fullness, say of 1 as in the example above, and produce output of higher full entropy. In theory, this report shows that to be impossible. But it may be in practice possible to do this. And some heuristic arguments may support this.

Nevertheless, the role served by the NRBG in this instance is the pretty much the same as the key derivation function. The only distinction would be that the NRBG can taking input as a stream, and output a stream, with a lower rate. The NRBG can also maintain a state. The state helps the NRBG revert to a DRBG if the state has entropy but the input stream stops providing entropy.

So, it would seem best to consider the NRBG not as a construction for providing information-theoretic security, rather as a construction for streamed key derivation with a safe fallback to DRBG when the input stream falls into a low entropy condition.