Comments on Cryptographic Entropy Measurement

Anna M. Johnston
Juniper Networks
amj at juniper dot net

October 30, 2019

One accurate measurement
is worth a thousand expert opinions.

– Grace M. Hopper

Abstract

Random data and the entropy contained within is a critical component of information security. Minimum entropy (min entropy, H_∞) is the base measurement defined by NIST and what many of their entropy tests are based on. However minimum entropy does not satisfy the basic requirements for an entropy measurement in either Shannon's original document [7] or in Rényi's generalization of entropy [6]. This document suggests a different way forward with a reclassification of entropy into two classes. With this differentiation, measurement tools are simplified and allow for more accurate assessments of entropy.

1 Information Security and Entropy

Random data and the entropy it contains is a critical component of information security. Predictable – i.e., poor cryptographic random – can make an otherwise secure system weak. Three NIST[1] documents detail standards on random data:

SP800-90B [9] Specifics on obtaining 'raw' or 'true' random data;

[1] US National Institute of Standards and Technology

SP800-90A [**2**] Specifics on the oxymoronically named (yet ubiquitous) *deterministic random bit generation*, or DRBG;

SP800-90C [**1**] Specifics on combining 90A and 90B for cryptographic use.

While these documents implicitly differentiate between raw (90B) and cryptographic (90A,C) random, the measurement of entropy within them is similar.

This indifferent treatment of random data types makes the testing process messy and unreliable. Cryptographic applications tend to favor more conservative views: assume the worst to maximize security. This idea has been taken to heart in the NIST SP800-90 documents, using the most conservative measure, min-entropy (H_∞). While this may make sense for cryptographic random data [2], it makes no sense when measuring entropy in raw random data. Accurate measuring tools, used conservatively, give a much better picture of what ever is being measured. No one would use a conservative ruler that was always a centimeter too short, nor would they use a conservative thermometer that was always a degree higher.

This paper differentiates the types of random data and their properties. It is hoped that this clarification will lead to simpler, more accurate tests and also help with the design of new entropy extenders (i.e., DRBG).

2 What is Entropy and How do we measure it.

In information theory, entropy in data is defined by two equivalent definitions:

1. The measure of randomness in data;

[2] There is a serious issue if min and mean entropy are not equivalent in cryptographic random data.

2. The amount of information contained in data.

Note that entropy is relative. It is not a solid, physical entity. Entropy depends on perspective or what is known and unknown about the data to a given entity. Once viewed, all information in the data is known to the viewer (zero entropy in the viewers perspective), but the data still contains entropy to non-viewers. The belief that entropy is something that has a classical, fixed measure is false and causes many interpretation issues.

Knowledge of underlying entropy is represented in a probability distribution. Assume that there are r possible states $\{x_j \mid 0 < j \leq r\}$, with state j having probability $pr(x_j) = p_j$ of occurring, with

$$\mathcal{P} = \left\{ p_j \;\middle|\; 0 < j \leq r;\ 0 \leq p_j \leq 1;\ \sum_{j=1}^{r} p_j = 1 \right\}. \tag{1}$$

Shannons[3] are the most commonly used unit[5]. A Shannon (Sh) is the maximal entropy which can be contained in a single bit. Alternatively, a Shannon is the entropy of a two state system with equally probable states. One bit can store at most one Shannon, n-bits can store at most n-Shannons, and in general, if there are k possible states, maximal entropy in the system is

$$log_2(k) = lg(k)\,.$$

Entropy of an individual event is inversely proportional to the probability of it occurring: the more unlikely an event, the higher the entropy from its occurrence. For example, in English we get more information about a hidden

[3]The term 'bits' is often used instead of Shannons. Multitasking the word 'bit' for both an element of the set $\{0, 1\}$ and an information measure can be misleading and confusing. While use of base 2 logarithms is most common, there are other entropy measurement units based off varying bases: hartleys (bans or dits) use base 10 and nat (natural units, nit) use base e.

word if the letter Z is revealed than if the letter E is revealed. The entropy of the event j (for the given probability distribution (equation 1)) is given by:

$$\mathfrak{ent}(j) = -lg(p_j)\,. \tag{2}$$

3 In a Perfect World: Raw vs Cryptographic Entropy

A distinction needs to be made between raw (true) and processed, cryptographically usable random data and the entropy contained within each type. In a perfect world, data with maximal entropy (1-Shannon per bit) from a true random source would be used in each cryptographic application requiring it. Our world is not perfect. Data from a raw random source rarely holds maximal entropy, and even if it did, would be too costly (mostly in terms of time) to generate for every application.

3.1 Our Imperfect World

In our imperfect world, we need a two part system with two different types of entropy:

1. **Raw entropy**: Collected from a 'true' random source. Raw random generally has a fairly low Shannon per bit ratio and is time and/or computationally expensive. Raw entropy is contained in raw random data.

2. **Processed/Cryptographic entropy**: Processed entropy is obtained from an entropy extender, seeded with raw entropy. The device collects, concentrates, and extends the generally poor Shannon-per-bit raw random. Once enough raw entropy has been collected, entities

without knowledge of the seed or state (i.e., observers) of the entropy extender would see output data with full processed entropy (i.e., one Shannon per bit). Furthermore, if n raw Shannons seeded the device, an observer would measure much more than n processed Shannons from the resulting output. No deterministic device generates entropy, but each raw Shannon can produce many processed Shannons.

Cryptographic random data and entropy (output from a cryptographic entropy extender) has the added property that no information on previous or future output can be determined from current output.

By definition, there are no cryptographic concerns for raw entropy. if raw random data contains n Shannons, then due to its true random nature, an observer of any collection of previous output would have only a 2^{-n} probability of guessing the next output.

It is critical is to insure that:

1. raw entropy is **accurately measured**;
2. the entropy extender is **cryptographically sound**;
3. **enough entropy** is collected in the entropy extender before any cryptographic random data is generated;
4. The entropy extender is **judiciously used**.

Insuring good cryptographic entropy implies a careful assessment of what the bold face terms above mean. The following sections briefly outline some of the goals for these terms. Note that this is not intended to be an all inclusive list, only a rough outline.

3.1.1 Accurately Measuring Raw Entropy

It is difficult, if not impossible, to determine with any certainty a precise measure of entropy in raw data. First, entropy is relative, dependent on an entities knowledge of the underlying information. If the source is truly random, its behavior is not well known, can change, and the output is (by definition) not well defined. Second, even the existence of random data can be questioned. However this kind of thinking quickly leads down a deep philosophical rabbit hole, and does us no practical good. This section reviews some of the very basic foundations for entropy measurement, upon which more elaborate working tests should be based.

Shannon [7] described the requirements for an accurate entropy measure H. Three critical properties (pgs 10-11) of this measure were laid out in this text, as well as their implications. The text also contains a proof that the only measure satisfying these properties is a mean (theorem on page 11, proof on page 28).

Rényi [6] lists five initial postulates for an entropy measure which roughly mirror Shannon's properties. Postulate five (pg 551 [6]) correlates to Shannon's third (pg 10 [7]). It is this constraint which leads to the mean measure.

Replacing Rényi's postulate five (arithmetic mean) with a modified constraint (postulate $5'$, the generalized mean, on pg 552 [6]), allows for a whole suite of alternate entropy measures:

$$H_\alpha[\mathcal{P}] = (1-\alpha)^{-1} log_2 \left[\left(\sum_k p_k^\alpha \right) \left(\sum_k p_k \right)^{-1} \right] \text{ with } \alpha > 1.$$

This formula can be extended to include $\alpha = 1$ and $\alpha = \infty$ by taking the limit of H_α as α goes to one or ∞ respectively.

This modified postulate requires a continuous, invertible function $g_\alpha(x)$

which is used in the generalized mean. For mean entropy ($\alpha = 1$), $g_\alpha(x) = ax + b$ ($a = 1, b = 0$ is the most common). For other entropy measures the function becomes:

$$g_\alpha(x) = 2^{(1-\alpha)x} \tag{3}$$

(see equation 2.13 [6] and pg 14 [4]; note that Cachin has a corrected g).

Rényi does not discuss H_∞, minimum entropy measure in [6]. One issue is the fact that the function, g_α, is a constant and not invertible for $\alpha = \infty$. This implies that H_∞ (min-entropy) does not satisfy even the relaxed, generalized properties of a good entropy measure. The central limit theorem does not hold and confidence intervals are not well defined for H_∞.

If minimum entropy does not satisfy the requirements set out by either Shannon or Rényi, why is it used as the foundation for measuring entropy in cryptographic applications? Some have cited the conservative nature of min-entropy ($H_\infty \leq H_\alpha$ for all α); others have cited the accepted 'folklore' [8]. I strongly believe the usage of H_∞ as the foundation of cryptographic entropy measurement should be re-evaluated with the separation of cryptographic and entropic concerns. In this light, the conservative argument fails.

3.1.2 Cryptographically Sound Entropy Extender

Cryptographically sound random data is quite different from raw random. It is expected that the quality – i.e., the Shannons per bit – is maximal within a very small confidence interval. Except for the unsuitable nature of using H_∞ in many statistical tests, H_∞ should equal H_1. It should also be relatively inexpensive in terms of time and computation, and be cryptographically secure.

Cryptographically sound entropy extender should have minimal raw entropy leakage. Leakage of raw entropy through the system implies one or

more of the following:

- Information on raw entropy (seed) can be determined from output;
- Information on previous or future output can be determined from a set of output;

Entropy Extenders are very similar to cryptographic stream ciphers. Just as the only perfectly secure cipher is the one-time pad with a maximal Shannon-per-bit, the only perfect entropy extender is one which only outputs as many Shannons as it receives and can prove no Shannons were lost in the concentration process. Using concentrated raw entropy, while giving the most security, violates the 'inexpensive' requirement of an entropy extender.

No entropy extender can guarantee zero leakage. Cryptanalysis should be performed on an entropy extender to find some estimate on leakage.

3.1.3 What is 'Enough' Entropy

How much raw entropy must be collected in a entropy extender before it may output cryptographic random data? This question depends on the application and the effects of having poor entropy on the results. A minimum of m raw Shannons must be collected, stored and used by the entropy extender when:

1. The entropy extender outputs m-bits of random data;
2. The output data is used in a system requiring an m-bit random block;

Note that these are bare minimums.

3.1.4 Judicious Use of Entropy Extender

Abuse of an entropy extender can lead to security issues. Generating too many cryptographic Shannons per raw Shannon is one issue. This would be equivalent to overuse of a cryptovariable (secret key) in a conventional cryptosystem. Proper analysis of the entropy extender and prior experience should give a rough estimate of an appropriate ratio of raw to cryptographic Shannons.

Another abuse, also covered in the previous section, is using more Shannons in a single application than the entropy extender has collected and stored. For example, if the entropy extender has only 128-bits of storage for raw random, then it should not be used to create a cryptovariable for a 256-bit cryptosystem.

4 Summary

Entropy is not a concrete entity that is easily measured. Obtaining entropy for cryptographic purposes imposes other, slightly less nebulous requirements. Ignoring the differences between raw and cryptographic entropy confuses and complicates the measurement process.

Separating cryptographic entropy generation into the following distinct processes:

1. generating and measuring entropy in raw random data,
2. and concentrating and extending this entropy to create suitable cryptographic random data

simplifies and improves both the measurement and creation process.

The very flawed measure, min-entropy or H_∞, seems to be used due to this confusion. It is conservative, helping to prop up security arguments. This is unnecessary if raw entropy generation is examined separately from cryptographic entropy generation. Instead, the more accurate and usable mean entropy should be used. It removes much of the questionable mathematics involved in many of the existing tests and would give much more accurate estimates of entropy.

> *The most dangerous phrase in the language is,*
> *"We've always done it this way."*
>
> – Grace M. Hopper

A Comments on NIST SP800-90B[9]

A.1 Confidence Intervals

Confidence intervals are used in many places in SP800-90B [9]:

- 6.3.2: Most Common Value Estimate
- 6.3.5: t-Tuple Estimate
- 6.3.6: Longest Repeated Substring Estimate
- 6.3.7: Multi Most Common in Window Prediction Estimate

Min-entropy can not be used to compute confidence intervals as it does not satisfy the central limit theorem. These tests attempt to legitimatize min-entropy usage by converting the distribution from a set of k possible outputs to a simple binary distribution. If the event set is $A = \{x_1, x_2, \ldots, x_k\}$

and the full probability distribution is $\mathcal{P} = \{p_j = pr(x_j) \mid 0 < j \leq r\}$, the reduced binary distribution to allow min-entropy to be used is $\hat{\mathcal{P}} = \left\{p_t, \sum_{j \neq t} p_j\right\}$, where x_t is the most likely event.

There are several problems with this modification

1. The binomial distribution does not play well with the central limit theorem. Standard (Wald) confidence intervals are used, and it can act very erratically with the binomial distribution [3];

2. Reducing the distribution to a binomial throws away a wealth of information about the distribution and thus reduces the accuracy of any statistics arising from the test;

3. These tests (and H_∞ in general) do not deal with multiple sources of entropy well. Getting entropy from more than source is common. Improving one of the sources should improve the entropy output.

   This does not happen with min-entropy. It does hold for H_α for every finite $\alpha > 0$.

   Minimum entropy (H_∞) fails because it does not satisfy postulate 5' in [6] (equation 2.12, pg 552).

A.2 6.4: Reducing the Symbol Space

This section details how to concentrate entropy in a given word of random data. It assumes that in a given word, certain bits occur more frequently than other bits. While this may be true in certain cases, such as when random data is collected from a clock, there are many cases when it would not.

Concentration of entropy is critical, but it is a processing concern, not an entropy source concern. Handling a single type or case of concentration

in a document which focuses on the collection of raw entropy seems out of place. It should be considered for more generalized cases in SP800-90A or 90C, which deal with the processing of raw random data.

A.3 Appendix D: Min-entropy and optimum Guessing Attack Cost

This section supports the use of H_∞ as a foundational entropy measure based off the conservative argument: $H_\infty \leq H_\alpha$ for all α, and the supposition that cryptographic attacks exist against the raw entropy source.

Cryptographic attacks do not exist against raw (true) entropy sources. A guess can not be checked without revealing the value itself. An attacker who sees the raw random data (assuming IID[4]) has no advantage on the next output from the raw source. In other words, if each output word contains k Shannons, an attacker would have only a 2^{-k} probability of randomly guessing any future or previous words.

Cryptographic attacks are a concern only once raw entropy is in a entropy extender and cryptographic random data is output.

References

[1] Elaine Barker and John Kelsey, *Sp800-90c: Recommendation for random bit generator constructions*, Special Publication 800-90C, National Institute of Standards, National Institute of Standards and Technology Computer Security Division, Information Technology Laboratory 100 Bureau Drive (Mail Stop 8930) Gaithersburg, MD 20899-8930, June 2012, `https://csrc.nist.gov/CSRC/media/Publications/sp/800-90c/draft/documents/sp800_90c_second_draft.pdf`.

[4] Independent, Identically Distributed

[2] ———, *Sp800-90a: Recommendation for random number generation using deterministic random bit generator*, Special Publication 800-90A, National Institute of Standards, National Institute of Standards and Technology Computer Security Division, Information Technology Laboratory 100 Bureau Drive (Mail Stop 8930) Gaithersburg, MD 20899-8930, June 2015, `http://dx.doi.org/10.6028/NIST.SP.800-90Ar1`.

[3] Lawrence D. Brown, T. Tony Cai, and Anirban DasGupta, *Interval estimation for a binomial proportion*, Statistical Science **16** (2001), no. 2, 101–103.

[4] Christian Cachin, *Entropy measures and unconditional security in cryptography*, Ph.D. thesis, Swiss Federal Institute of Technology, 1997.

[5] ISO, *IEC 80000-13:2008: Quantities and units – part 13: Information science and technology*, Standards document 13, International Organization for Standardization (ISO), Geneva, Switzerland, March 2008.

[6] Alfréd Rényi, *On measures of entropy and information*, Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics (Berkeley, California), University of California Press, 1961, `http://digitalassets.lib.berkeley.edu/math/ucb/text/math_s4_v1_article-27.pdf`, pp. 547–561.

[7] C. E. Shannon, *A mathematical theory of communication*, The Bell System Technical Journal **27** (1948), 379–423, 623–656.

[8] Maciej Skorksi, *Shannon entropy versus rényi entropy from a cryptographic viewpoint*, IMA International Conference on Cryptography and

Coding, IMACC 2015: Cryptography and Coding, Lecture Notes in Computer Science, vol. 9496, Springer-Link, 2015, pp. 257–274.

[9] Meltem Sönmez Turan, Elaine Barker, John Kelsey, Kerry A. McKay, Mary L. Baish, and Mike Boyle, *SP800-90B: Recommendation for the entropy sources used for random bit generation*, Special Publication 800-90B, National Institute of Standards, National Institute of Standards and Technology Computer Security Division, Information Technology Laboratory 100 Bureau Drive (Mail Stop 8930) Gaithersburg, MD 20899-8930, January 2018, `https://doi.org/10.6028/NIST.SP.800-90B`.