Entropy Estimation on the Basis of a Stochastic Model

Werner Schindler

Bundesamt für Sicherheit in der Informationstechnik (BSI)
Godesberger Allee 185–189
53175 Bonn, Germany
Werner.Schindler@bsi.bund.de

Abstract

The core of any physical RNG is its noise source. In [3], Subsection 3.2.2, among other requirements it is demanded that the documentation shall provide an explicit statement of the expected entropy rate and a technical argument, which supports this claim. This presentation considers entropy estimation on the basis of a stochastic model.

The (optional) conditioning is typically realised by a cryptographic postprocessing algorithm, which is too complicated to allow direct entropy analysis of the output data. Instead, we focus on the data after digitization, denoted by 'digitized data' in the following, or on the 'raw data' (terminology as in [3]). We point out that in the literature other definitions are also widespread, e.g. 'das bits' (das = digitized analog signal) or 'raw random numbers' for 'digitized data' and 'internal random numbers' for 'raw data'.) We assume that the digitized data and thus also the raw data are realizations of (i.e., values taken on by) random variables.

The stochastic model specifies of a family of probability distributions, which contains the true (but unknown) probability distribution of the digitized data. Usually, this family of distributions can be parametrized by one or several parameters. The distributions of different copies of this RNG (often components of smart cards) shall be contained in this family, maybe represented by different parameter(s). The variation of parameters may be caused by tolerances of components, fluctuations in production, ageing effects, or varying environmental conditions. The stochastic model should be justified by physical / technical arguments and supported by empirical experiments. If the stochastic model has successfully been verified the estimation of the entropy reduces to the estimation of one or several parameters.

Based on the analysis of the stochastic model the distribution of the raw data or at least a lower entropy bound is derived under consideration of the (non-cryptographic) post-processing algorithm. In exceptional cases the stochastic model considers the raw data (i.e., the data after post-processing) directly. In the presentation the concept of stochastic models will be addressed and explained by examples.

The stochastic model also allows to develop effective health tests (online tests) for the corresponding RNG design since the distribution of the digitized data or the raw data, respectively, are contained in the specified family of distributions, even in the case when their quality goes down. Ideally, this is also the case for accidential total failures of the noise source (and maybe even for fault attacks).

The presentation closes with brief remarks on the German certification scheme according to the Common Criteria where stochastic models have proven successful for many years [1, 2].

References

1. Bundesamt für Sicherheit in der Informationstecnik (BSI): Anwendungshinweise und Interpretationen zum Schema (AIS) AIS 31. Version 3, 15.05.2013; https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Zertifizierung /Interpretationen/AIS_31_pdf.html
2. W. Killmann, W. Schindler: A Proposal for: Functionality Classes for Random Number Generators. Mathematical-Technical Reference to [1], Version 2, 18.09.2011; https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Zertifizierung/ Interpretationen/AIS_31_Functionality_classes_for_random_number_generators _e.pdf?__blob=publicationFile
3. NIST Special Publication 800-90B (Second Draft): Recommendation for the Entropy Sources Used for Random Bit Generation.