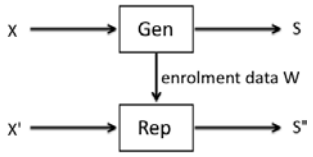
An efficient fuzzy extractor for limited noise

Boris Škorić and Pim Tuyls

A fuzzy extractor is a security primitive that allows for reproducible extraction of an almost uniform key from a non-uniform noisy source. We analyze a fuzzy extractor scheme that uses universal hash functions for both information reconciliation and privacy amplification. This is a useful scheme when the number of error patterns likely to occur is limited, regardless of the error probabilities. We derive a sharp bound on the uniformity of the extracted key, making use of the concatenation property of universal hash functions and a recent tight formulation of the leftover hash lemma.

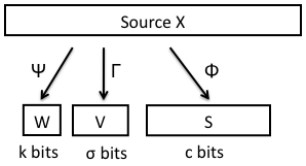
Many security applications require input bitstrings to be uniformly distributed and exactly reproducible. Physical sources of randomness, such as biometrics and Physical Unclonable Functions (PUFs), however, are neither uniform nor noise-free. For security/privacy reasons it is often necessary to apply a one-way hash function to the biometric/PUF measurement, in analogy with the `/etc/passwd` file in UNIX. The storage of biometric data is assumed to be public; the hashing step hides the measurement data. However, as measurements are noisy, it is not possible to directly hash; a single bit error in the input causes roughly 50% of the output bits to flip. Hence, an error-correction step is required first. This is not trivial, since the redundancy data has to be stored publicly and may reveal too much sensitive information. Similarly, a key derived from a PUF should be thoroughly noise-corrected first. Here, too, it is crucial that the publicly stored redundancy data does not reveal secrets. After information reconciliation, the step of *privacy amplification* is applied, mapping a non-uniform random string to a shorter, almost uniform string. The requirement of uniformity is obvious in the case of key extraction. Interestingly, extracting uniform bitstrings is also desirable in biometric systems and PUF-based anti-counterfeiting, applications where the identifiers are *not* considered to be secret. A uniform string is the most efficient way of storing the entropy present in a measurement. Furthermore, database search speed is improved. The concept of a *Fuzzy Extractor* was introduced as a primitive that achieves both information reconciliation and privacy amplification. The publicly stored enrolment data W suffices to reproducibly reconstruct a string S from noisy measurements X', yet leaks only a negligible amount of information about the extracted key S.



One of the nontrivial aspects of the information reconciliation step is the 'shape' of the noise. Noise patterns are not always nicely compatible with a binary

Dagstuhl Seminar Proceedings 09282
Foundations for Forgery-Resilient Cryptographic Hardware
http://drops.dagstuhl.de/opus/volltexte/2010/2409

representation. Error-correcting codes (ECCs) work best under the condition that the likely to occur error patterns are completely random. For biometrics and PUFs, this assumption on the error patterns does not hold, due to various properties of the sources and the applied discretization methods. Typical ECCs are not able to capitalize on the low entropy of the errors, since they must be able to correct the 'worst case'; consequently a large part of the entropy present in the source gets wasted. Furthermore, ECCs can approach the Shannon bound only when the code words are very long. The challenge is to construct a practical error correction method that, in the case of very non-uniform noise probabilities, extracts more information than typical ECCs.



We analyze a fuzzy extractor based on a specific form of Slepian-Wolf coding which employs (almost-)universal hash functions. Three such hash functions are applied to the source X, to produce the secret string S, the helper data W, and a MAC key V which is used to authenticate the helper data. W and S can be considered to be part of the same big hash value. If this is taken literally, then it can be said that the scheme performs information reconciliation and privacy amplification *at the same time* or even *in the opposite order* compared to other schemes. The scheme is efficient if the number of likely-to-occur error patterns is limited; computation of short universal hashes is very fast.
We derive a sharp bound on the uniformity of the extracted key. If the string lengths c, k, σ satisfy

$$c \leq \max_{\rho} \left[\mathsf{H}_2^{\rho}(X) + 2 - \log \frac{1}{\varepsilon(\varepsilon - \rho) - \delta/4} \right] - k - \sigma$$

then the statistical distance between the key S and the uniform distribution (given that the attacker has seen W and the MAC) is smaller than ε. In the above equation, ρ is a 'smoothing' parameter, $\mathsf{H}_2^{\rho}(X)$ is the *smooth Rényi entropy of order 2*, and δ is a parameter derived from the 'almost-ness' of the employed almost-universal hash functions.
Finally we argue that the scheme can be implemented efficiently.