Key extraction from general non-discrete signals

E. Verbitskiy[1,3], P. Tuyls[1], C. Obi[2], B. Schoenmakers[2], B. Škorić[1,2]

September 2008

Abstract

We address the problem of designing optimal schemes for the generation of secure cryptographic keys from continuous noisy data. We argue that, contrary to the discrete case, a universal fuzzy extractor does not exist. This implies that in the continuous case, key extraction schemes have to be designed for particular probability distributions. We extend the known definitions of the correctness and security properties of fuzzy extractors. Our definitions apply to continuous as well as discrete variables.

We propose a generic construction for fuzzy extractors from noisy continuous sources, using independent partitions. The extra freedom in the choice of discretisation, which does not exist in the discrete case, is advantageously used to give the extracted key a uniform distribution. We analyze the privacy properties of the scheme and the error probabilities in a one-dimensional toy model with simplified noise.

Finally, we study the security implications of incomplete knowledge of the source's probability distribution $\mathbb{P}$. We derive a bound on the min-entropy of the extracted key under the worst case assumption, where the attacker knows $\mathbb{P}$ exactly.

1 Introduction

1.1 Fuzzy Extractors

Extraction of secure cryptographic keys from noisy measurements is a problem that received a lot of attention in recent years [9, 5, 2, 4, 8]. The main motivations originate from the area of biometrics and Physical Unclonable Functions (PUFs) [6]. Within the field of biometrics key extraction plays a role in protecting the privacy of stored biometric templates and in the formation of new applications such as file access based on biometric data. PUFs on the other hand are used for anti-counterfeiting (e.g. making devices such as RFID tags unclonable) and secure storage of cryptographic keys [7].

Most of the research up to now has focused on the extraction of secure keys from *discrete* noisy sources [2]. The basic primitive that resulted from this work is the fuzzy extractor. A fuzzy extractor is a general primitive that allows one to extract a secure cryptographic key from a noisy source. It consists basically of two phases. In a first phase the source is challenged; a secure bit string (the key) as well as *helper data* are extracted by means of a probabilistic procedure. The helper string has to be considered as publicly available data and hence can be observed by an attacker. In the second phase, the key has to be reconstructed from a fresh measurement of the noisy source. This measurement will in general differ slightly from the first one. The reconstruction procedure takes as input the fresh measurement and the helper data to reconstruct the original key.

Many sources, however, produce *continuous* rather than discrete data. Fingerprint templates for instance are represented by sequences of points in a continuous domain such as $\mathbb{R}^n$. Speckle patterns originating from optical PUFs and capacitance measurements of coating PUFs [7] are

[1] Philips Research Laboratories, Eindhoven, The Netherlands
[2] Department of Mathematics and Computer Science, Eindhoven University of Technology, The Netherlands
[3] Department of Mathematics, University of Groningen, The Netherlands

typically continuous data. In order to use these responses for cryptographic purposes, some discretisation step has to be performed. The choice of this discretisation procedure is relevant since it determines the quality of the input of all the discrete procedures that follow. A bad choice can lead to large entropy loss. An important consideration here is the noise reduction. A carefully chosen discretisation allows for extraction of most of the entropy available in the measurement while reducing the noise to a point where an efficient error correcting code can be deployed.
There is only a small amount of literature [4, 1] addressing the extraction of secure keys from noisy continuous sources; only particular quantization schemes have been considered.
The notion of a discrete fuzzy extractor can not be immediately generalized to this situation. This is primarily because several entropy notions (min-entropy and Renyi entropy) used in the definition of the discrete fuzzy extractor are not well defined for the continuous case.

1.2 Contributions and organisation of this paper

In Section 2 we extend the known definitions of the correctness and security properties of fuzzy extractors. Our definitions apply to continuous as well as discrete variables. We argue that the universality property of fuzzy extractors for discrete distributions does not directly apply to continuous distributions, since concepts such as min-entropy are undefined if the discretisation is left unspecified.
Then, in Section 3.1, we present a geometric approach to the problem of key extraction from noisy continuous sources. The extraction scheme is based on an independent pair $(\mathcal{A}, \mathcal{B})$ of partitions of the measurement space. The extra freedom in the choice of discretisation, which does not exist in the discrete case, is advantageously used to give the extracted key a uniform distribution. In Section 3.2 the privacy aspects of the scheme are addressed. In section 3.3 various types of errors are introduced. In Section 3.4 we introduce a one-dimensional toy model with simplified noise, for which we analyze the error probabilities.
In practice, the designer of a key extraction scheme does not exactly know the true probability distribution ρ of the source. He only has an estimate $\tilde{\rho}$ based on a finite set of experimental data. The mismatch between the empirical distribution $\tilde{\rho}$ and the real distribution ρ affects the security of the extracted key. In Section 4 we compute a lower bound on the min-entropy of the extracted key, taking this mismatch into account. This is done under the worst case assumption, i.e. we assume that the attacker has perfect knowledge of ρ. Finally, to illustrate our approach we analyze the case of a source with normal distribution.

2 Preliminaries

Throughout this paper, except when otherwise stated, log is taken to base 2. When an algorithm or a function f is randomized, we use the semicolon when we wish to make the randomization explicit: i.e. we denote by $f(x; r)$, the outcome of computing f on input x with randomness r. Pair $(\mathcal{M}, d)$ will represent a discrete metric space and $(\mathcal{X}, d)$ an uncountable ("continuous") metric space [1].
Let $\mathbb{P}$ and $\mathbb{Q}$ be probability measures on $\mathcal{M}$. The *total variation* between $\mathbb{P}$ and $\mathbb{Q}$ is defined as[2] $\Delta(\mathbb{P}, \mathbb{Q}) = \frac{1}{2} \sum_{x \in \mathcal{M}} |\mathbb{P}(x) - \mathbb{Q}(x)|$. The min-entropy of $\mathbb{P}$ is given by $H_\infty(\mathbb{P}) = -\log \max_{x \in \mathcal{M}} \mathbb{P}(x)$. For a random variable X with distribution $\mathbb{P}$ (denoted by $X \sim \mathbb{P}$), we often write $H_\infty(X)$ instead of $H_\infty(\mathbb{P})$. For a joint distribution $(X, Y) \sim \mathbb{P}$ the marginal distribution of Y is denoted as $\mathbb{P}^{(2)}(y) = \sum_x \mathbb{P}(x, y)$. The *conditional min-entropy* of X given Y is defined as

$$H_\infty(X|Y) = -\log \max_{x,y} \mathbb{P}[X = x | Y = y] = -\log \max_{x,y} \frac{\mathbb{P}(x, y)}{\mathbb{P}^{(2)}(y)}.$$

[1] We abuse notation by setting the same notation d for the metric in both $\mathcal{M}$ and $\mathcal{X}$. The precise meaning will however be clear from the context.
[2] For $X \sim \mathbb{P}$, $X' \sim \mathbb{Q}$ we often write $\Delta(X, X')$ or $\Delta(P_X, P_{X'})$ instead of $\Delta(\mathbb{P}, \mathbb{Q})$.

Definition 2.1. *([2]) For $(X, Y) \sim \mathbb{P}$, the average min-entropy of X given Y is defined as*

$$\widetilde{H}_\infty(X|Y) = -\log \mathbb{E}_{\mathbb{P}^{(2)}}\left(2^{-H_\infty(X|Y=y)}\right) = -\log\left(\sum_y \max_x \mathbb{P}(x,y)\right). \quad (1)$$

We recall and extend the definition of a fuzzy extractor as introduced by Dodis et al. in [2] (Definitions 2.2, 2.3 below).

Definition 2.2. *A fuzzy extractor of length ℓ on a discrete metric space $(\mathcal{M}, d)$ is a pair of (randomized) procedures **generate** (Gen) and **reproduce** (Rep) with*

$$\mathsf{Gen} : \mathcal{M} \to \{0,1\}^\ell \times \{0,1\}^* : x \mapsto (k, w),$$
$$\mathsf{Rep} : \mathcal{M} \times \{0,1\}^* \to \{0,1\}^\ell : (x', w) \mapsto \hat{k},$$

We write $(k_x, w_x) = \mathsf{Gen}(x)$ for values of the Gen function at point $x \in \mathcal{M}$. If X is a random variable with values in $\mathcal{M}$, then (K, W), given by $(K, W) = \mathsf{Gen}(X)$, are random variables with values in $\{0,1\}^\ell$ and $\{0,1\}^*$ respectively, with the joint distribution induced by Gen from $\mathbb{P}$.

Naturally, fuzzy extractors must be resilient with respect to small measurement errors and should produce sufficiently random keys. The following conditions are the reformulation of the corresponding conditions of the definition in [2].

Definition 2.3 ([2])**.** *Let $(\mathsf{Gen}, \mathsf{Rep})$ be a fuzzy extractor of length ℓ on a metric space $(\mathcal{M}, d)$. Let $(k_x, w_x) = \mathsf{Gen}(x)$. We say that $(\mathsf{Gen}, \mathsf{Rep})$ is*

($\mathtt{C}_1$) *t-correct for $t > 0$, if for all $x, x' \in \mathcal{M}$ with $d(x, x') < t$ one has $\mathsf{Rep}(x', w_x) = k_x$.*

($\mathtt{S}_1$) *(m, δ)-secure for some $m > 0$ and $\delta \in [0, 1)$, if for any probability distribution $\mathbb{P}$ on $\mathcal{M}$ with min-entropy at least m, the random variables $(K, W) = \mathsf{Gen}(X)$, with $X \sim \mathbb{P}$, satisfy*

$$\Delta\Big((K, W), (U, W)\Big) \leq \delta,$$

where U is a uniformly distributed random variable on $\{0,1\}^\ell$.

By definition, the entropy loss of an (m, δ)-secure fuzzy extractor of length ℓ is $m - \ell$.

Fuzzy extractors aim at correcting for noise corrupting the data. However, in the definition above the distribution of the noise is not taken into account. One can argue that the correctness property of the fuzzy extractor should be stated in terms of the noise distribution.
To illustrate this argument further, we give a number of conditions that a "good" fuzzy extractor could satisfy. We modify our notations slightly, for better understanding. The noisy measurement X' of X will be modeled by means of of a family $\{\mathbf{P}_x\}_{x \in \mathcal{M}}$ of probability measures $\mathbf{P}_x$, which define the distribution of X' given $X = x$.

Definition 2.4 (***Correctness***)**.** *Suppose that X is a random variable with values in $\mathcal{M}$ distributed according to a probability law $\mathbb{P}$; X' is a noisy measurement of X for the noise $\mathbf{P} = \{\mathbf{P}_x\}_{x \in \mathcal{M}}$, i.e., $X' \sim \mathbf{P}_x$, if $X = x$. We say that a fuzzy extractor $(\mathsf{Gen}, \mathsf{Rep})$ is*

($\mathtt{C}_2$) *on average ϵ-stochastically resilient for noise $\{\mathbf{P}_x\}$, if*

$$\int \mathbf{P}_x(\mathsf{Rep}(X', w_x) = k_x)\ \mathbb{P}(dx) \geq 1 - \epsilon.$$

($\mathtt{C}_3$) *worst case ϵ-stochastically resilient for noise $\{\mathbf{P}_x\}$, if for any $x \in \mathcal{X}$ one has*

$$\mathbf{P}_x(\mathsf{Rep}(X', w_x) = k_x) \geq 1 - \epsilon.$$

Conditions $(\mathtt{C}_2)$ and $(\mathtt{C}_3)$ are meant to generalize $(\mathtt{C}_1)$, by taking into account the effect of the measurement noise. For the same value of the parameter ϵ, $(\mathtt{C}_3)$ implies $(\mathtt{C}_2)$. Condition $(\mathtt{C}_1)$ can be reformulated in probabilistic terms as well: a fuzzy extractor $(\mathsf{Gen}, \mathsf{Rep})$ is t-resilient for some $t > 0$, if and only if $(\mathsf{Gen}, \mathsf{Rep})$ is worst case 0-stochastically resilient for any noise $\{\mathbf{P}_x\}$ such that

$$\mathbf{P}_x(d(x, X') < t) = 1.$$

Definition 2.5 (***Security***)**.** *Suppose X is a random variable with values in $\mathcal{X}$ distributed according to a probability law $\mathbb{P}$. Fuzzy extractor $(\mathsf{Gen}, \mathsf{Rep})$ is called*

$(\mathtt{S}_2)$ *on average $\tilde{m}$-secure if*

$$\tilde{H}_\infty(K|W) \geq \tilde{m};$$

$(\mathtt{S}_3)$ *worst-case $\tilde{m}$-secure if for all $w \in \{0,1\}^*$ such that $\mathbb{P}^{(2)}(W = w) > 0$,*

$$H_\infty(K|W = w) \geq \tilde{m}.$$

For the same value of parameter $\tilde{m}$, $(\mathtt{S}_3)$ is stronger than $(\mathtt{S}_2)$. We argue that $(\mathtt{S}_3)$ is too strict: it takes the worst case helper data, while for predictability by an adversary, the average case suffices. Averaging over all possible helper data as in $(\mathtt{S}_2)$ suffices because the helper data is not under attack by the adversary. The advantage the adversary has depends on how much control he has in creating the worst case. The attacker has no influence on the key generating procedure Gen so the best he can do is hope that the helper data is the worst possible (i.e. the helper data reveals a lot of information about the key). However, the worst case occurs with negligible probability. A proof of this fact is given in Appendix A. We prove that if $(\mathtt{S}_2)$ holds, then the probability of $w \in W$ occurring such that $H_\infty(K|W = w) \leq \xi \ll \tilde{m}$ is bounded from above by $2^{-(\tilde{m}-\xi)}$. Since this bad class of helper data occurs with negligible probability, describing the security of a fuzzy extractor in terms of property $(\mathtt{S}_2)$ is valid.
Moreover, one can also compare the security conditions $(\mathtt{S}_1)$ and $(\mathtt{S}_2)$. Namely, suppose $(\mathsf{Gen}, \mathsf{Rep})$ is (m, δ)-secure in the sense of $(\mathtt{S}_1)$. Then for $(K, W) = \mathsf{Gen}(X)$, one has

$$\Delta\Big((K, W), (U, W)\Big) = \sum_w \mathbb{P}(W = w)\left[\frac{1}{2}\sum_k \Big|\mathbb{P}(K = k|W = w) - \frac{1}{2^\ell}\Big|\right] = \sum_w \mathbb{P}(W = w)\Delta_w \leq \delta,$$

where

$$\Delta_w = \frac{1}{2}\sum_k \Big|\mathbb{P}(K = k|W = w) - \frac{1}{2^\ell}\Big|.$$

Therefore, for any w

$$\max_k \mathbb{P}(K = k|W = w) \leq \frac{1}{2^\ell} + \Delta_w,$$

and hence

$$\begin{aligned}
\tilde{H}_\infty(K|W) &= -\log\left(\sum_w \mathbb{P}(W = w)\max_k \mathbb{P}(K = k|W = w)\right) \\
&\geq -\log\Big(\frac{1}{2^\ell} + \sum_w \mathbb{P}(W = w)\Delta_w\Big) \\
&\geq -\log\Big(\frac{1}{2^\ell} + \delta\Big).
\end{aligned}$$

Together, with an obvious bound $\tilde{H}_\infty(K|W) \leq \ell$, this provides a relation between $\tilde{m}$ and δ

$$\tilde{m} = \ell - \log\Big(1 + 2^\ell\delta\Big).$$

Since the above estimates are rather sharp, one has to realize than for large ℓ, seemingly small values δ, can result in low entropy values. Therefore, we argue that the condition $(\mathtt{S}_2)$ is more suitable for ensuring the security of the derived keys.

A practical fuzzy extractor must have at least one of the correctness properties and one of the security properties. Fuzzy extractors satisfying properties $(\mathtt{C}_1)$ and $(\mathtt{S}_1)$ are precisely those introduced in [2].
It is expected that the helper data might reveal some amount of information about the original measurement and/or the bit string derived from the measurement. In the discrete case, this can be quantified in terms of the universal loss $\mathcal{L}$ (see [4]),

$$\mathcal{L} \geq H_\infty(K) - H_\infty(K|W).$$

The universality reffers to the fact that the inequality should be valid for **all** or at least a large class of probability distributions on the measurement space.
The fuzzy extractor allows one to extract an ϵ-secure bit string of length $m - 2\log(1/\epsilon)$ from **all** discrete noisy sources that have min-entropy at least m. This is called the universality property for fuzzy extractors for discrete distributions.
For continuous distributions, a quantization scheme $\mathcal{Q}$ is applied to transform the continuous domain to a discrete domain. A fuzzy extractor for discrete domains is then applied. During reconstruction, the discretized version is reconstructed instead of the original x in the continuous domain. Quantization $\mathcal{Q}(X)$ is treated as the "discrete original" [4]. The entropy loss in this phase of the construction is given by $H_\infty(\mathcal{Q}(X)) - H_\infty(\mathcal{Q}(X)|W)$. The second term is called the left-over entropy [4]. In the continuous case, we aim at maximizing left-over entropy because it is the "source entropy" for the strong extractor phase. A strong extractor can now be applied to $\mathcal{Q}(X)$ to extract a secure bit string. The total entropy loss of the fuzzy extraction scheme using this construction is

$$[H_\infty(\mathcal{Q}(X)) - H_\infty(\mathcal{Q}(X)|W)] + [H_\infty(\mathcal{Q}(X)|W) - \ell] = H_\infty(\mathcal{Q}(X)) - \ell,$$

where ℓ is the length of extracted string.
It is interesting to find out how much information the helper data reveals about the extracted string. Unlike in the discrete case, for **any** quantization scheme on $\mathcal{X}$ and any fixed positive $\mathcal{L}$, one can find a probability distribution on $\mathcal{X}$ such that $H_\infty(K)$ is smaller than $\mathcal{L}$, hence making the above estimate useless. This distribution can be made to have a large value of appropriately defined entropy. Since for the continuous case the quantization scheme is part of the fuzzy extractor, one therefore concludes that there are no 'universally' good fuzzy extractors for continuous spaces. Hence, contrary to the discrete case, one must construct good quantization schemes for a **fixed** distribution, since schemes with reasonable properties do not exist for any sufficiently large class of distributions, e.g., the class of distributions with reasonably large entropy.
The scheme presented in the next section is derived for a fixed distribution and has the property that $H_\infty(\mathcal{Q}(X)) = H_\infty(\mathcal{Q}(X)|W)$, i.e. the entropy loss due to the helper data is zero. Furthermore, the output after quantization is uniform, so we do not require a strong extractor. Consequently, we have

$$H_\infty(K|W) = H_\infty(\mathcal{Q}(X)|W) = H_\infty(\mathcal{Q}(X)),$$

i.e. the helper data does not reveal any information about the key.

3 Fuzzy extractors for continuous distributions

3.1 Construction based on independent partitions

There are many ways to partition a measurable space $\mathcal{X}$. In fact, there is one-to-one correspondence between measurable function and partitions: For example, any measurable function $h : \mathcal{X} \to \Sigma$, where Σ is a finite set, gives rise to a finite partition into level sets:

$$\mathcal{A} = \{A_\sigma | \sigma \in \Sigma\},$$

where $A_\sigma = \{x \in \mathcal{X} : h(x) = \sigma\} = h^{-1}(\sigma)$. Similarly, any partition $\mathcal{A} = \{A_\sigma | \sigma \in \Sigma\}$ gives rise to a function $h : \mathcal{X} \to \Sigma$, given by $h(x) = \sigma$ if $x \in A_\sigma$. Moreover, by considering functions from $\mathcal{X} \times \mathcal{R}$, we can associate partitions to randomized functions h on $\mathcal{X}$. To simplify notation we only consider deterministic functions and corresponding partitioning schemes.
Consider a surjective function $\mathcal{Q}$ (the 'quantizer') on $\mathcal{X}$. $\mathcal{Q} : \mathcal{X} \to [n]$. Here $[n]$ denotes the set $\{1, \ldots, n\}$; $i = \mathcal{Q}(x)$ will be referred to a the key extracted from x. As mentioned above, $\mathcal{Q}$ induces a natural partitioning $\mathcal{A}$ of $\mathcal{X}$, which consists of n subsets $\{A_1, \ldots, A_n\}$,

$$A_i = \{x \in \mathcal{X} : \mathcal{Q}(x) = i\},$$

satisfying $A_i \cap A_k = \emptyset$ for $i \neq k$ and $\cup_{i=1}^n A_i = \mathcal{X}$. To explicitly state the relationship between the quantizer $\mathcal{Q}$ and the partitioning $\mathcal{A}$, we write $\mathcal{Q}_\mathcal{A} : \mathcal{X} \to [n]$, $\mathcal{Q}_\mathcal{A}(x) = i$ if and only if $x \in A_i$. If $\mathbb{P}$ is a probability distribution on $\mathcal{X}$, $\mathcal{Q}_\mathcal{A}$ induces a discrete probability distribution $P_\mathcal{A}$ on $[n]$, $P_\mathcal{A} = (\mathbb{P}(A_1), \ldots, \mathbb{P}(A_n))$.
Now we are going to incorporate a noise correction mechanism, because the measurement can be corrupted by noise. This is done by means of another partition $\mathcal{B}$ of size m of $\mathcal{X}$ as follows. First of all, for two partitions $\mathcal{A} = \{A_i\}_{i=1}^n$ and $\mathcal{B} = \{B_j\}_{j=1}^m$ of $\mathcal{X}$, the refinement of $\mathcal{A}$ and $\mathcal{B}$ is a partition consisting of the sets $\{A_i \cap B_j : i \in [n],\, j \in [m]\}$. The corresponding quantizer, $\mathcal{Q}_{(\mathcal{A},\mathcal{B})}$ is given as $\mathcal{Q}_{(\mathcal{A},\mathcal{B})}(x) = (i, j)$ if and only if $x \in A_i \cap B_j$. In this case, we say that i is the extracted key and j is the helper data. On input of a noisy observation $x' = x + e$ and helper data j, the key is recovered as follows:

$$\hat{\imath} = \operatorname*{argmin}_k \operatorname{dist}(x', A_k \cap B_j) = \operatorname*{argmin}_k \min_{\hat{x} \in A_k \cap B_j} d(x', \hat{x}),$$

where $\operatorname{dist}(U, V) = \min_{u \in U, v \in V} d(u, v)$. To ensure that the key is recovered correctly, we demand that a sufficient **gap** should exist between each pair of sets $A_i \cap B_j$, $A_k \cap B_j$, with $i, k \in [n]$ and $i \neq k$.

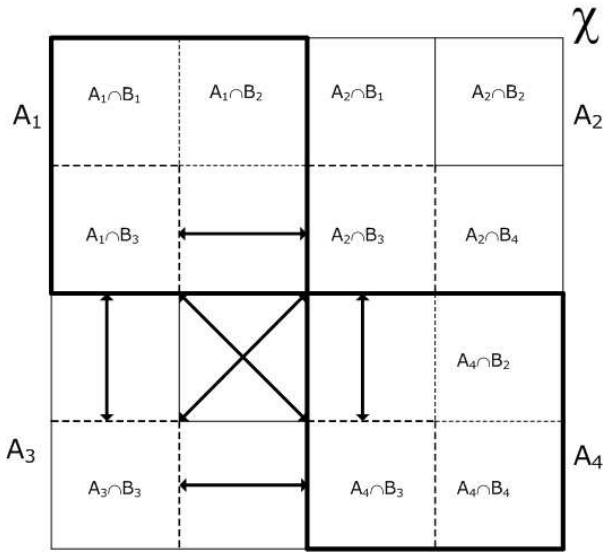

Figure 1: *For each fixed j, large gaps exist between each pair of subsets in $\{A_i \cap B_j\}$. This allows for efficient error correction.*

It is easy to see that in the case of additive noise, i.e., $x' = x + e$, where the noise e is such that $2e \leq d_{\min} = \min_{i,j,k:\, i \neq k} \operatorname{dist}(A_i \cap B_j, A_k \cap B_j)$, then the noise cannot cause a reconstruction error, and we will always have $\hat{\imath} = i$.

It is desirable that the helper data j reveals the least information possible about the extracted key i. To accomplish this, we demand that $\mathcal{A}$ and $\mathcal{B}$ are independent, i.e. $P(A_i \cap B_j) = P(A_i)P(B_j) \forall i, j$. This implies that the helper data reveals no information about the extracted key. So $H_\infty(I|J) = H_\infty(I)$.
To optimize our construction, we require that the probability distribution induced on $[n] \times [m]$ by the quantizer $\mathcal{Q}_{(\mathcal{A},\mathcal{B})}$ is uniform[3], i.e.

$$P(A_i) = \frac{1}{n}, \quad P(B_j) = \frac{1}{m}$$

for all $i \in [n]$, $j \in [m]$. Equivalently, $H_\infty(P_\mathcal{A}) = \log n$ and $H_\infty(P_\mathcal{B}) = \log m$. In line with the notation in [2], we give a formal construction of a continuous space fuzzy extractor as follows:

Construction 3.1. *Let $(\mathcal{X}, d)$ be a continuous metric space. Let X be a random variable on $\mathcal{X}$. Let $\mathcal{A}$ and $\mathcal{B}$ be independent partitions of $\mathcal{X}$ of sizes n and m respectively, such that $\mathcal{Q}_{(\mathcal{A},\mathcal{B})}$ induces a uniform distribution on $[n] \times [m]$. We define the generation procedure Gen: $\mathcal{X} \to [n] \times [m]$ and the reproduction procedure Rep: $\mathcal{X} \times [m] \to [n]$ as follows:*

- *Generation: $\mathsf{Gen}_{(\mathcal{A},\mathcal{B})}(x) = \mathcal{Q}_{(\mathcal{A},\mathcal{B})}(x) = (i, j)$. Here i is the key and j is the helper data.*
- *Reproduction: $\hat{\imath} = \mathsf{Rep}_{(\mathcal{A},\mathcal{B})}(x', j) = \operatorname{argmin}_k \inf_{\hat{x} \in A_k \cap B_j} d(x', \hat{x})$.*

By construction this scheme yields perfect security in the sense of $(\mathtt{C}_1)$: $\Delta\big((I, J), (U, J)\big) = 0$, where $(I, J) = \mathsf{Gen}(X)$, $X \sim \mathbb{P}$, and U is uniform on $[n]$. Moreover, the scheme has zero entropy loss incurred from publishing the helper data, $H_\infty(I|J) = H_\infty(I) = \log n$.

3.2 Privacy

In the case X represent the biometric data, the privacy of the original measurement has to be preserved. Therefore it is interesting to investigate how much information is gained on the measurement X from the value of helper data J. It was shown in [5, 2] that in the discrete case, the helper data always leaks information on the measurement X. A natural measure for determining the amount of leakage is the mutual information between the measurement and the helper data. In the case of our construction 3.1, the amount of information an adversary learns about X is given by

$$I(X; J) = H(J) - H(J|X) = H(J) = \log m.$$

Here we have used $H(J|X) = 0$, which holds because because given $x \in \mathcal{X}$, an adversary can easily compute $\mathcal{Q}_{(\mathcal{A},\mathcal{B})}(x) = (i, j)$. In order to reveal as little information as possible about X, m has to be chosen as small as possible. However, the choice of m (together with n) has an impact on the amount of noise the scheme can tolerate.

3.3 Error analysis

A properly designed fuzzy extractor scheme extracts a long key in a robust fashion. Yet, errors are inevitable. Therefore, error analysis of our proposed scheme is essential. Let X' be a noisy version of X. A reconstruction error occurs when $\mathsf{Rep}(X', J) \neq I$. We examine three ways of evaluating the probability of such an erroneous re-extraction. First of all, for a fixed x, X' is a random variable with distribution $X' \sim P_x = P(\cdot|X = x)$. Hence, we might be interested in evaluating the **x-dependent error probability**:

$$p(x) = \mathbf{P}_x[\mathsf{Rep}(X', j_x) \neq i_x].$$

[3]Demanding the partitions to be uniform implies that they are independent. However in our construction, independence of the partitions is a more important property, because it ensures that the helper data leaks no information about the extracted key. We treat uniformity as an additional requirement to achieve optimality.

Secondly, we can evaluate the **average error probability** $\bar{p}$. Let $X \sim \mathbb{P}$, then

$$\bar{p} = \int_{x \in \mathcal{X}} p(x)\, \mathbb{P}(dx).$$

Finally, one might be interested in the **maximal error probability**:

$$p^{\max} = \sup_{x \in \mathcal{X}} p(x).$$

3.4 Toy example: Uniform distribution on $[0, 1)$ with simplified noise

In this section we consider the simplest example. The continuous space $\mathcal{X}$ is the unit circle $[0, 1)$. The random variable X is uniformly distributed[4]. The noisy measurement X' is written as $X' = X + Z \mod 1$. The noise Z is assumed to be uniformly distributed on the interval $[-\delta, \delta]$. This example is by no means realistic, but it is susceptible to complete analysis and the geometric idea – introduction of gaps to obtain robustness – is transparent.

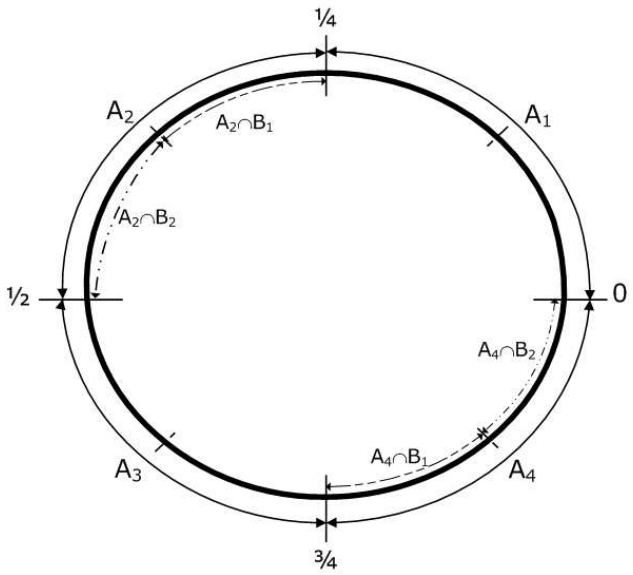

Figure 2: *Partitioning the unit circle. Note that for each fixed j, large **gaps** exist between each pair $A_i \cap B_j$, $A_k \cap B_j$ with $k \neq i$.*

Construction 3.2. *Let $\mathcal{X}$ be the unit interval $[0, 1]$. We construct the partitions $\mathcal{A}$ and $\mathcal{B}$ as follows.*

$$A_i = \left[\frac{i-1}{n}, \frac{i}{n}\right), \quad i \in [n],$$
$$A_i \cap B_j = \left[\frac{i-1}{n} + \frac{j-1}{mn}, \frac{i-1}{n} + \frac{j}{mn}\right), \quad j \in [m].$$

A_i is an interval of length $1/n$, and B_j is a union of n intervals of length $1/mn$. To complete the partition scheme we have to specify the distance function d. We let d be the usual distance on the circle. The distance between two sets $A_i \cap B_j$ for any fixed j is

$$\operatorname{dist}(A_i \cap B_j, A_{i'} \cap B_j) \geq \frac{1}{n} - \frac{1}{nm} =: d_{mn} \quad \forall i, i', i \neq i',$$

[4]Any one-dimensional continuous density function can be mapped to a uniform distribution on the unit circle. Let $\rho_X(x)$ be the density function of X. Let $y = \int_{-\infty}^{x} \rho_X(t)dt$, then $Y \in [0, 1]$ and $Y \sim U[0, 1]$, where $U[0, 1]$ is the uniform distribution on the unit interval. Identifying the end points 0 and 1 we get the unit circle.

with equality if and only if $|i - i'| = 1 \mod n$, i.e., when A_i and $A_{i'}$ are neighboring intervals. Hence, we say that the **gap size** is d_{mn}. Recall that the key reconstruction is based on the distance from $x' \in \mathcal{X}$ to the nearest $A_i \cap B_j$,

$$\hat{\imath}(X', j) = \operatorname*{argmin}_i d(A_i \cap B_j, X').$$

Hence, for $J = j$ the set of x' such that $\hat{\imath}$ assumes value k is an interval

$$\begin{aligned} T_{k,j} &:= \{x' : \mathsf{Rep}_{(\mathcal{A},\mathcal{B})}(x', j) = k\} = \left(\inf(A_k \cap B_j) - \frac{d_{mn}}{2}, \sup(A_k \cap B_j) + \frac{d_{mn}}{2}\right) \\ &= \left(\frac{k-1}{n} + \frac{j-1}{mn} - \frac{m-1}{2mn}, \frac{k-1}{n} + \frac{j}{mn} + \frac{m-1}{2mn}\right). \end{aligned}$$

The length of this interval is $1/n$. Trivially, all key values are equiprobable not only for noise-free measurements (achieved by design), but for noisy measurements as well.
We will now evaluate the error probabilities. To stress the dependence on the parameters m, n, and on the noise $Z \sim U([-\delta, \delta])$, we denote by $p_{m,n}(\delta, x)$, $p^{\max}_{m,n}(\delta)$ and $\bar{p}_{m,n}(\delta)$, the x-dependent, maximal and average probabilities, respectively.

Theorem 3.3. *Let X be a random variable with uniform distribution on the unit circle. Let the noise Z be uniformly distributed on $[-\delta, \delta]$. Let $\mathcal{A}$, $\mathcal{B}$ be the partitioning according to Construction 3.2. Let $(i, j) = Q_{(\mathcal{A},\mathcal{B})}(x)$, and $q = x - [\frac{i-1}{n} + \frac{j-1}{mn} + \frac{1}{2mn}]$, i.e. $|q|$ is the distance between x and the middle point of the interval $A_i \cap B_j$.*

(i) The x-dependent error probability is given by

$$p_{m,n}(\delta, x) = \begin{cases} 0, & \text{if } \delta \in \left[0, \frac{1}{2n} - |q|\right] \\ \frac{1}{2}\left(1 - \frac{1}{2\delta n}\right) + \frac{|q|}{2\delta}, & \delta \in \left(\frac{1}{2n} - |q|, \frac{1}{2n} + |q|\right) \\ 1 - \frac{1}{2\delta n}, & \text{if } \delta \in \left[\frac{1}{2n} + |q|, \frac{1}{2}\right]. \end{cases} \tag{2}$$

(ii) The maximal error probability $p^{\max}_{m,n}(\delta) = \max_x p_{m,n}(\delta, x)$ is attained at the endpoints of $A_i \cap B_j$, and is given by

$$p^{\max}_{m,n}(\delta) = \begin{cases} 0, & \delta \in [0, \frac{m-1}{2mn}), \\ \frac{1}{2} - \frac{m-1}{4\delta mn}, & \delta \in [\frac{m-1}{2mn}, \frac{m+1}{2mn}], \\ 1 - \frac{1}{2\delta n}, & \delta \in (\frac{m+1}{2mn}, \frac{1}{2}]. \end{cases} \tag{3}$$

(iii) The average error-probability is given by

$$\bar{p}_{m,n}(\delta) = \begin{cases} 0, & \delta \in [0, \frac{m-1}{2mn}), \\ \frac{mn}{2\delta}\left(\delta - \frac{m-1}{2mn}\right)^2, & \delta \in [\frac{m-1}{2mn}, \frac{m+1}{2mn}], \\ 1 - \frac{1}{2\delta n}, & \delta \in (\frac{m+1}{2mn}, \frac{1}{2}]. \end{cases} \tag{4}$$

The proof is straightforward, but laborious (see Appendix B). The result is useful for the optimization of design parameters.
In terms of Definitions 2.3, 2.4, and 2.5, Construction 3.2 for noise $Z \sim U([-\delta, \delta])$ produces a fuzzy extractor which is $(\mathtt{C}_1)$ $\frac{m-1}{2mn}$-correct; $(\mathtt{C}_2)$ on average $\bar{p}_{m,n}(\delta)$-stochastically resilient; $(\mathtt{C}_3)$ worst case $p^{\max}_{m,n}(\delta)$-stochastically resilient, with $\bar{p}_{m,n}(\delta)$ and $p^{\max}_{m,n}(\delta)$ as given in Theorem 3.3. With respect to security conditions, the fuzzy extractor is $(\mathtt{S}_2)$ on average $\log n$-secure; $(\mathtt{S}_3)$ worst case $\log n$-secure. Recall that the security condition $(\mathtt{S}_1)$cannot be applied in the continuous case.

Example 3.4. Suppose $n = 2$, i.e., we want to extract 1 bit. If $m = 1$, i.e., no helper data is used, then

$$p^{\max}_{1,2}(\delta) = \frac{1}{2} \quad \text{for all } \delta \in (0, 1/2].$$

On the other hand, if $\delta < 1/8$, then the partition scheme with $m = 2$ gives no error at all.

More generally, the following proposition aids the optimal selection of parameter m. Let $\lfloor x \rfloor$ denote the integer part of x for $x \geq 0$.

Proposition 3.5. *For $n \geq 2$ and $\delta \in [0, 1/2]$, let $\gamma = 2\delta n$. Then*

(a) If $\gamma < 1$, then there exists an optimal m such that $p^{\max}_{m,n}(\delta) = 0$.

(b) If $\gamma = 1$, then for every $\varepsilon > 0$, there exists an m such that $p^{\max}_{m,n}(\delta) < \varepsilon$.

(c) If $\gamma \in (1, 1.5]$, then there exists an optimal m such that

$$p^{\max}_{m,n}(\delta) = \frac{1}{2}\left(1 - \frac{1}{\gamma\lfloor \frac{1}{\gamma - 1} \rfloor}\right).$$

(d) If $\gamma > 1.5$, then independently of m, $p^{\max}_{m,n}(\delta) = 1 - \frac{1}{\gamma}$.

A proof is given in Appendix C.

Remark 3.6. If $\gamma = 2\delta n > 3/2$, then we can say that the noise level is too large for the desired number of key values n, and the introduction of helper data does not improve the performance. Moreover, for any δ we can find a partition scheme which allows errorless extraction of up to $n = \lfloor (2\delta)^{-1} \rfloor$. Moreover, if we are prepared to tolerate some moderate error levels, we can extract up to $n = \lfloor 1.5(2\delta)^{-1} \rfloor$, which is roughly an increase of 50% for small δ's.

4 Imperfect knowledge of the source

Up to this point, we have investigated the problem of extracting a secret key from a continuous source distributed according to $\mathbb{P}$. In general, we can not assume that $\mathbb{P}$ is known precisely. This is due to the fact that in practice one often has to learn the distribution $\mathbb{P}$ empirically. Hence, one obtains only an estimate $\tilde{\mathbb{P}}$ of the true distribution $\mathbb{P}$. From a security point of view we have to assume that an attacker has put more effort in learning the distribution and therefore has a more accurate knowledge of $\mathbb{P}$ than the designer. This implies of course that the schemes that we have described before do not a priori guarantee complete security of the extracted key in this situation. We give a security derivation assuming the worst case scenario: the attacker knows $\mathbb{P}$ exactly. We prove a bound on the min-entropy of the extracted key, taking into account the mismatch between the empirical $\tilde{\mathbb{P}}$ and the true $\mathbb{P}$.

The designer of the extraction scheme $(\mathcal{A}, \mathcal{B})$ bases his design on $\tilde{\rho}$. Therefore, the induced probabilities on elements of the partition are $\tilde{P}_{ij} := \tilde{\mathbb{P}}(A_i \cap B_j) = 1/(nm)$, i.e. uniform by construction. However, according to the true distribution, $Q_{(\mathcal{A},\mathcal{B})}$ induces an an unknown probability distribution $\{\mathbb{P}_{ij}\}_{i\in[n],j\in[m]}$, which is different from uniform. If the empirical estimate $\tilde{\mathbb{P}}$ is good, then it will be close to uniform.

In slight abuse of notation we write $(I, J) \sim \mathbb{P}$ and $(\tilde{I}, \tilde{J}) \sim \tilde{\mathbb{P}} = U_{[n]\times[m]}$. By construction we have that $H_\infty(\tilde{I}|\tilde{J}) = \log n$. We now give a lower bound on the average min-entropy $\tilde{H}_\infty(I|J)$, which tells us how secure the key is in the case of imperfect knowledge of the true distribution.

Lemma 4.1. *Suppose $\tilde{\mathbb{P}}$ is a probability distribution on $\mathcal{X}$, and $Q_{(\mathcal{A},\mathcal{B})} : \mathcal{X} \to [n] \times [m]$ is the key extraction scheme according to Construction 3.1 with respect to $\tilde{\mathbb{P}}$. Suppose X is a random variable on $\mathcal{X}$, $X \sim \mathbb{P}$, and put $(I, J) = Q_{(\mathcal{A},\mathcal{B})}(X)$. Then*

$$\tilde{H}_\infty(I|J) \geq \log n - \log\bigl(1 + nm\Delta\bigr),$$

where

$$\Delta = \frac{1}{2}\sum_{i\in[n]}\sum_{j\in[m]}\Bigl|\mathbb{P}(A_i \cap B_j) - \tilde{\mathbb{P}}(A_i \cap B_j)\Bigr| = \frac{1}{2}\sum_{i\in[n]}\sum_{j\in[m]}\Bigl|\mathbb{P}(A_i \cap B_j) - \frac{1}{mn}\Bigr|.$$

Proof. Using

$$\max_{i,j} \mathbb{P}(A_i \cap B_j) \leq \frac{1}{mn} + \Delta,$$

and Definition 2.1, we have

$$\begin{aligned} \tilde{H}_\infty(I|J) &= -\log \sum_{j\in[m]} \max_{i\in[n]} \mathbb{P}(A_i \cap B_j) \geq -\log \left(m \max_{i\in[n], j\in[m]} \mathbb{P}(A_i \cap B_j) \right) \\ &\geq -\log \left(m \left[\frac{1}{mn} + \Delta \right] \right) = \log n - \log(1 + nm\Delta). \end{aligned}$$

□

For a good practical fuzzy extractor, $m\Delta$ should be of order $\mathcal{O}(\frac{1}{n})$ implying that $\tilde{H}_\infty(I|J)$ is of order $\mathcal{O}(\log n)$.
If we assume that both $\mathbb{P}$ and $\tilde{\mathbb{P}}$ are absolutely continuous probability distributions on $\mathcal{X}$ with respect to some reference measure λ, and if ρ, $\tilde{\rho}$ are the corresponding densities, i.e.,

$$\frac{d\mathbb{P}}{d\lambda} = \rho, \quad \frac{d\tilde{\mathbb{P}}}{d\lambda} = \tilde{\rho},$$

then the quantity Δ in the above lemma satisfies the following inequalities

$$\Delta \leq \frac{1}{2} \int_{\mathcal{X}} \left| \rho(x) - \tilde{\rho}(x) \right| \lambda(dx), \quad \Delta \leq \sqrt{\frac{1}{2} \int_{\mathcal{X}} \rho(x) \log \frac{\rho(x)}{\tilde{\rho}(x)} \lambda(dx)}.$$

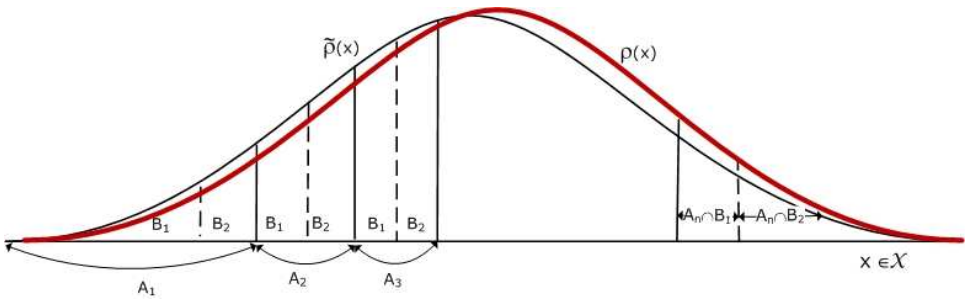

Figure 3: *Effect of partitioning scheme designed using known density $\tilde{\rho}(x)$ instead of unknown density $\rho(x)$. Partitions $\mathcal{A}$ and $\mathcal{B}$ are sizes n and $m = 2$ respectively.*

We conclude this section with a simple example, namely $\mathbb{P}$ being the Gaussian distribution $N(\mu, \sigma^2)$, and $\tilde{\mathbb{P}}$ being the maximal likelihood estimate of $\mathbb{P}$ based on sample of size N. More specifically, $\tilde{\mathbb{P}}$ is a Gaussian distribution with parameters $(\hat{\mu}, \hat{\sigma}^2)$, where

$$\hat{\mu} = \frac{1}{N} \sum_{i=1}^{N} X_i, \quad \hat{\sigma}^2 = \frac{1}{N} \sum_{i=1}^{N} (X_i - \hat{\mu})^2.$$

It is well known that $\hat{\mu} \to \mu$, $\hat{\sigma}^2 \to \sigma^2$ as $N \to \infty$, and the error is of order $N^{-\frac{1}{2}}$. We begin with the following simple lemma, which gives an estimate of parameter Δ in Lemma 4.1 in case of two Guassian distributions in terms of the parameters of these distributions.

Lemma 4.2. *Let $\mathbb{P}$, $\tilde{\mathbb{P}}$ be normal distributions with parameters (μ, σ^2), $(\tilde{\mu}, \tilde{\sigma}^2)$ respectively. Then for any key extraction scheme $Q_{(\mathcal{A},\mathcal{B})} : \mathcal{X} \to [n] \times [m]$ given by Construction 3.1 with respect to the empirical distribution $\tilde{\mathbb{P}}$, one has*

$$\Delta \leq \frac{1}{\min(\sigma, \tilde{\sigma})} \sqrt{(\sigma - \tilde{\sigma})^2 + (\mu - \tilde{\mu})^2},$$

The proof is given in appendix E. Hence, applying this lemma, for sufficiently large N, one can conclude that Δ will be sufficiently small as well: $\Delta = \mathcal{O}(N^{-\frac{1}{2}})$. In fact, the estimate above can be used to derive bounds on N sufficient to ensure accurate estimate of $\mathbb{P}$, and hence small Δ. Moreover, statistics [3] provides even stronger results: Since $\hat{\mu}$, $\hat{\sigma}$ are random variables, it can happen with low probability that these estimates lead to large Δ. Nevertheless, if an independent sample of size N has been used to estimate the unknown parameters of distribution $\mathbb{P}$, then for some constants $c_1, c_2, c_3, c_4 > 0$ one has

$$\mathbb{P}\Big(\Delta > c_1 \frac{(\log N)^{c_2}}{\sqrt{N}}\Big) \leq c_3 e^{-c_4 (\log N)^2},$$

hence any deviation from the order $N^{-\frac{1}{2}}$ is quite improbable.

5 Summary

We have extended the known definitions of the correctness and security properties of fuzzy extractors. Our definitions apply to continuous as well as discrete variables. We have introduced a generic construction for fuzzy extractors for noisy continuous sources using independent partitions. The extra freedom in the choice of discretisation, which does not exist in the discrete case, is advantageously used to give the extracted key a uniform distribution. We have analyzed the privacy properties of the scheme and the error probabilities in a one-dimensional toy model with simplified noise.
We have studied the security implications of incomplete knowledge of the source's probability distribution $\mathbb{P}$. We have derived a bound on the min-entropy of the extracted key under the worst case assumption, where the attacker knows $\mathbb{P}$ exactly. We have worked this out for the case of a normal-distributed variable.
Finally, we conclude by observing that the proposed approach is easily extended to produce schemes which allow for extraction of multiple keys from the same source X as well with the same security, robustness and privacy properties. This is relevant for biometric applications.

References

[1] I.R. Buhan, J.M. Doumen, P.H. Hartel, and R.N.J. Veldhuis. Fuzzy extractors for continuous distributions. In *Proceedings of the 2nd ACM Symposium on Information, Computer and Communications Security (ASIACCS), 20-22 March 2007*, pages 353–355, 2007.

[2] Y. Dodis, M. Reyzin, and A. Smith. Fuzzy extractors: How to generate strong keys from biometrics and other noisy data. In C. Cachin and J. Camenisch, editors, *Advances in Cryptology — Eurocrypt 2004*, volume 3027 of *LNCS*, pages 523–540. Springer-Verlag, 2004.

[3] S. Ghosal and A.W. van der Vaart. Entropies and rates of convergence for maximum likelihood and Bayes estimation for mixtures of normal densities. *Ann. Statist.*, 29(5), 2001.

[4] Q. Li, Y. Sutcu, and N. Memon. Secure sketch for biometric templates. In X. Lai and K. Chen, editors, *ASIACRYPT*, volume 4284 of *Lecture Notes in Computer Science*, pages 99–113. Springer, 2006.

[5] J.-P.M.G. Linnartz and P. Tuyls. New shielding functions to enhance privacy and prevent misuse of biometric templates. In J. Kittler and M. Nixon, editors, *Conference on Audio and Video Based Person Authentication*, volume 2688 of *LNCS*, pages 238–250. Springer-Verlag, 2003.

[6] R. Pappu, B. Recht, J. Taylor, and N. Gershenfeld. Physical one-way functions. *Science*, 297:2026–2030, Sept. 2002.

[7] P. Tuyls, G.J. Schrijen, B. Škorić, J. van Geloven, R. Verhaegh, and R. Wolters. Read-proof hardware from protective coatings. In L. Goubin and M. Matsui, editors, *Cryptographic Hardware and Embedded Systems — CHES 2006*, volume 4249 of *LNCS*, pages 369–383. Springer-Verlag, 2006.

[8] P. Tuyls, B. Škorić, and T. Kevenaar. *Security with Noisy Data: Private Biometrics, Secure Key Storage and Anti-Counterfeiting*. Springer, London, 2007.

[9] E. Verbitskiy, P. Tuyls, D. Denteneer, and J.-P.M.G. Linnartz. Reliable biometric authentication with privacy protection. In *Proceedings 24th Benelux Symposium on Information Theory*, 2003.

A Motivation for using security property S_2

Let $\tilde{m}, \xi \geq 0$. Let $(K, W) \sim \mathbb{P}$. We prove that if $\tilde{H}_\infty(K|W) \geq \tilde{m}$, then the probability is negligible of $w \in \mathcal{W}$ occurring such that $H_\infty(K|W=w) \leq \xi \ll \tilde{m}$. Let Ω_ξ be the set of 'bad' helper data values, $\Omega_\xi = \{w \in \mathcal{W} : \max_k \mathbb{P}^{(1|2)}(k,w) \geq 2^{-\xi}\}$.

$$\begin{aligned}
\mathbb{P}[w \in \Omega_\xi] &= \sum_{w\in\Omega_\xi} \mathbb{P}^{(2)}(w) \leq \sum_{w\in\Omega_\xi} \mathbb{P}^{(2)}(w) \cdot 2^\xi \max_k \mathbb{P}^{(1|2)}(k,w) && (5)\\
&\leq \sum_{w\in\mathcal{W}} \mathbb{P}^{(2)}(w) \cdot 2^\xi \max_k \mathbb{P}^{(1|2)}(k,w) = 2^\xi \sum_{w\in\mathcal{W}} \max_k \mathbb{P}(k,w) = 2^{\xi - \tilde{H}_\infty(K|W)} && (6)\\
&\leq 2^{-(\tilde{m}-\xi)}. && (7)
\end{aligned}$$

In (5) we used the defining property of w in Ω_ξ. In (6) we made use of the fact that $\sum_{w\in\Omega_\xi}(\cdots) \leq \sum_{w\in\mathcal{W}}(\cdots)$. Finally the inequality in (6) follows from the given property $\tilde{H}_\infty(K|W) \geq \tilde{m}$.

B Proof of Theorem 3.3

Let $\delta \leq 1/2$. Let $Y = X + Z$ be the noisy measurement, with $Z \sim U[-\delta, \delta]$. For $x \in A_i \cap B_j$, the error probability is $\Pr[Y \notin T_{i,j}]$. From the uniformity of X and the independence of Z from X it follows that the error probabilities are independent of i, j. Without loss of generality we consider $(i,j) = (1,1)$. Let $q = x - 1/(nm)$. Let a^+ denote $\max(a, 0)$.
(i) x-dependent error probability:

$$\begin{aligned}
p_{mn}(\delta, x) &= P_Z(x + Z \notin T_{1,1})\\
&= P_Z(Z \in [-\delta, -x - \frac{1}{2}d_{mn})) + P_Z(Z \in (-x + \frac{1}{2}d_{mn} + \frac{1}{nm}, \delta])\\
&= \frac{1}{2\delta}(\delta - x - \frac{1}{2}d_{mn})^+ + \frac{1}{2\delta}(\delta + x - \frac{1}{2}d_{mn} - \frac{1}{nm})^+\\
&= \frac{1}{2\delta}(\delta - q - \frac{1}{2n})^+ + \frac{1}{2\delta}(\delta + q - \frac{1}{2n})^+. && (8)
\end{aligned}$$

Separate evaluation of this expression for different ranges of q and δ yields (2).
(ii) Maximal error probability: We note that $p_{mn}(\delta, x)$ is an increasing function of $|q|$ and that $q \in (-\frac{1}{2nm}, \frac{1}{2nm})$. Hence the maximum is obtained at $q = \pm\frac{1}{2nm}$. Substitution into (8) gives

$$p_{mn}^{\max}(\delta) = \frac{1}{2\delta}\left(\delta - \frac{m-1}{2mn}\right)^+ + \frac{1}{2\delta}\left(\delta - \frac{m+1}{2mn}\right)^+.$$

Separate evaluation for different ranges of δ gives (3).

(iii) Average error probability. We are now going to compute the average error probability $\bar{p}_{m,n}(\delta) = P_{X,Z}(X + Z \notin D_{1,1})$ where $X \sim U[0, \frac{1}{2mn}]$ and $Z \sim U[-\delta, \delta]$. One possible approach is to integrate with respect to x the expression for $p_{m,n}(\delta, x)$ obtained in (2). However, it is easier to follow a different approach. We observe that

$$\begin{aligned}\bar{p}_{m,n}(\delta) &= P_{X,Z}(X + Z \notin D_{1,1}) \\ &= P_{X,Z}\left(Z + \left\{X - \frac{1}{2mn}\right\} \le -\frac{1}{2n}\right) \\ &\quad + P_{X,Z}\left(Z + \left\{X - \frac{1}{2mn}\right\} \ge \frac{1}{2n}\right) \\ &= F_{(X-\frac{1}{2mn})+Z}\left(-\frac{1}{2n}\right) + \left(1 - F_{(X-\frac{1}{2mn})+Z}\left(\frac{1}{2n}\right)\right),\end{aligned}$$

where $F_{(X-\frac{1}{2mn})+Z}(t)$ is the distribution function of a random variable $(X - \frac{1}{2mn}) + Z$. Note that $X - \frac{1}{2mn} \sim U(-\frac{1}{2mn}, \frac{1}{2mn})$, and hence the distribution of $X - \frac{1}{2mn} + Z$ is also symmetric, and therefore

$$\bar{p}_{m,n}(\delta) = 2F_{(X-\frac{1}{2mn})+Z}\left(-\frac{1}{2n}\right).$$

First of all, as was noted above, $\bar{p}_{m,n}(\delta) = 0$ for

$$-\frac{1}{2n} \le -\delta - \frac{1}{2mn} \iff \delta \le \frac{m-1}{2mn}.$$

Suppose now $\delta > \frac{m-1}{2mn}$. If $m \ge 2$, then $\delta \ge \frac{1}{2mn}$. We now can use the result of Lemma D.1 with $a = \delta$, $b = \frac{1}{2mn}$ and $t = -\frac{1}{2n}$, and conclude that for $m \ge 2$

$$\bar{p}_{m,n}(\delta) = \begin{cases} 0, & \delta \in [0, \frac{m-1}{2mn}), \\ \frac{mn}{2\delta}\left(\delta - \frac{m-1}{2mn}\right)^2, & \delta \in [\frac{m-1}{2mn}, \frac{m+1}{2mn}], \\ 1 - \frac{1}{2\delta n}, & \delta \in (\frac{m+1}{2mn}, \frac{1}{2}]. \end{cases} \tag{9}$$

C Proof of Proposition 3.5

Proof of Proposition 3.5. If $\gamma < 1$, then since $\lim_{m\to\infty}(m-1)/m = 1$, for sufficiently large m we have

$$\frac{m-1}{m} > \gamma.$$

But then $\delta < (m-1)/(2mn)$, and, by (3), the partition scheme with parameters (m, n) allows errorless extraction of n key values. Suppose $\gamma > 1$, but $\gamma < 3/2$. Then the set

$$M_\gamma := \left\{m \ge 2 : \frac{m+1}{m} > \gamma\right\}$$

is not empty: M_γ contains at least one element $m = 2$. Since $\gamma > 1$, M_γ is bounded. For all $m \in M_\gamma$, $\delta < (m+1)/(2mn)$, and hence

$$p^{\max}_{m,n} = \frac{1}{2} - \frac{(m-1)}{4\delta mn} = \frac{1}{2} - \frac{m-1}{2m\gamma} = \frac{1}{2} - \frac{1}{2\gamma}\left(1 - \frac{1}{m}\right).$$

The optimal choice for m is clearly the maximal element in M_γ:

$$m^* = \sup M_\gamma = \left\lfloor \frac{1}{\gamma - 1} \right\rfloor,$$

where $\lfloor \cdot \rfloor$ is the integer part, and hence

$$p^{\max}_{m^*,n} = \frac{1}{2}\left(1 - \frac{1}{\gamma \left\lfloor \frac{1}{\gamma - 1} \right\rfloor}\right).$$

If $\gamma = 2\delta n = 1$, then the set M_γ coincides with $\mathbb{N}$ and hence is not bounded. Thus we are able to choose arbitrarily large m's. Therefore,

$$\inf_m p^{\max}_{m,n}(\delta) = 0,$$

but the infimum is not achieved. Hence, by choosing m appropriately we can only make $p^{\max}_{m,n}$ less than any fixed positive number. Finally, if $\gamma > 1.5$, then for any $m \in \mathbb{N}$, one has $\delta > \frac{m+1}{2mn}$, and by (4), we can only achieve $p^{max}_{m,n} = 1 - 1/\gamma$. □

D Distribution of the sum of two independent uniform random variables

Lemma D.1. *Suppose x_1, x_2, are independent random variables, and $x_1 \sim U[-a, a]$, $x_2 \sim U[-b, b]$. Assume that $a \geq b$. Then $x_1 + x_2$ takes values in $[-a-b, a+b]$, and the density is given by*

$$p_{x_1+x_2}(t) = \begin{cases} 0, & t < -a-b \\ \frac{1}{4ab}(a+b+t), & t \in [-a-b, -a+b] \\ \frac{1}{2a}, & t \in (-a+b, a-b) \\ \frac{1}{4ab}(a+b-t), & t \in [a-b, a+b] \\ 0, & t > a+b \end{cases}. \tag{10}$$

Moreover, the distribution function $F_{x_1+x_2}(t) = \int_{-\infty}^{t} p_{x_1+x_2}(\tau)d\tau$ is given by

$$F_{x_1+x_2}(t) = \begin{cases} 0, & t < -a-b \\ \frac{(a+b+t)^2}{8ab}, & t \in [-a-b, -a+b] \\ \frac{t+a}{2a}, & t \in (-a+b, a-b) \\ 1 - \frac{(a+b-t)^2}{8ab}, & t \in [a-b, a+b] \\ 1, & t > a+b \end{cases}. \tag{11}$$

Proof. The densities of x_1, x_2 are given by

$$p_{x_1}(z) = \begin{cases} \frac{1}{2a}, & |z| \leq a \\ 0, & |z| > a \end{cases}, \qquad p_{x_2}(z) = \begin{cases} \frac{1}{2b}, & |z| \leq b \\ 0, & |z| > b \end{cases}.$$

The distribution of $x_1 + x_2$ has the density

$$\begin{aligned} p_{x_1+x_2}(t) &= \int_{-\infty}^{\infty} p_{x_1}(t-z)p_{x_2}(z)dz = \frac{1}{2b}\int_{-b}^{b} p_{x_1}(t-z)dz \\ &= \frac{1}{2b}\int_{-b}^{b} \frac{1}{2a}\mathbf{I}\left[-a \leq t - z \leq a\right] dz \\ &= \frac{1}{4ab}\int_{-b}^{b} \mathbf{I}\left[t-a \leq z \leq t+a\right] dz \\ &= \frac{1}{4ab}\big|[-b, b] \cap [t-a, t+a]\big|. \end{aligned}$$

Note that we assumed $a \geq b$.
Case 1. If $t \in [-a+b, a-b]$, then $t-a \leq -b$ and $b \leq t+a$. Hence $[-b,b] \subset [t-a, t+a]$. Therefore $p_{x_1+x_2}(t) = (2a)^{-1}$.
Case 2. If $t \in [-a-b, -a+b]$, then $t+a \in [-b,b]$, $t-a < -b$, and hence $[-b,b] \cap [t-a,t+a] = [-b, t+a]$. Hence $p_{x_1+x_2}(t) = (4ab)^{-1}(a+b+t)$.
Case 3. If $t \in [a-b, a+b]$, then $t-a \in (-b,b)$, $t+a > b$, and hence $[-b,b] \cap [t-a,t+a] = [t-a,b]$. Hence $p_{x_1+x_2}(t) = (4ab)^{-1}(a+b-t)$.
Finally, integrating density (10) we obtain expression for the distribution function (11). □

E Proof of Lemma 4.2

Proof. For two normal distributions $\mathbb{P} = \mathcal{N}(\mu, \sigma^2)$ and $\tilde{\mathbb{P}} = \mathcal{N}(\tilde{\mu}, \tilde{\sigma}^2)$, by the Pinsker and the log-sum inequalities we have

$$\begin{aligned}\Delta &= \frac{1}{2}\sum_{i\in[n]}\sum_{j\in[m]}\Big|\mathbb{P}(A_i \cap B_j) - \tilde{\mathbb{P}}(A_i \cap B_j)\Big| \\ &\leq \min\left\{\sqrt{\frac{1}{2}\int_{-\infty}^{+\infty}\rho(x)\log\frac{\rho(x)}{\tilde{\rho}(x)}dx}, \sqrt{\frac{1}{2}\int_{-\infty}^{+\infty}\rho(x)\log\frac{\rho(x)}{\tilde{\rho}(x)}dx}\right\},\end{aligned}$$

where the Kullback-Leibler divergences are given by

$$\begin{aligned}\int_{-\infty}^{+\infty}\rho(x)\log\frac{\rho(x)}{\tilde{\rho}(x)}dx &= \frac{1}{2}\left(\log\frac{\tilde{\sigma}^2}{\sigma^2} + \frac{\sigma^2}{\tilde{\sigma}^2} + \Big[\frac{\tilde{\mu}-\mu}{\tilde{\sigma}}\Big]^2 - 1\right), \\ \int_{-\infty}^{+\infty}\tilde{\rho}(x)\log\frac{\tilde{\rho}(x)}{\rho(x)}dx &= \frac{1}{2}\left(\log\frac{\sigma^2}{\tilde{\sigma}^2} + \frac{\tilde{\sigma}^2}{\sigma^2} + \Big[\frac{\tilde{\mu}-\mu}{\sigma}\Big]^2 - 1\right).\end{aligned}$$

Since $\min\{a,b\} \leq \sqrt{\frac{1}{2}(a^2+b^2)}$, one concludes that

$$\begin{aligned}\Delta &\leq \sqrt{\frac{1}{4}\Big((\sigma^2-\tilde{\sigma}^2)^2\frac{1}{\sigma^2\tilde{\sigma}^2} + (\mu-\tilde{\mu})^2\Big[\frac{1}{\sigma^2}+\frac{1}{\tilde{\sigma}^2}\Big]\Big)} \\ &= \frac{1}{2\sigma\tilde{\sigma}}\sqrt{(\sigma^2-\tilde{\sigma}^2)^2 + (\mu-\tilde{\mu})^2(\sigma^2+\tilde{\sigma}^2)} \leq \frac{\sigma+\tilde{\sigma}}{2\sigma\tilde{\sigma}}\sqrt{(\sigma-\tilde{\sigma})^2+(\mu-\tilde{\mu})^2} \\ &\leq \frac{1}{\min(\sigma,\tilde{\sigma})}\sqrt{(\sigma-\tilde{\sigma})^2+(\mu-\tilde{\mu})^2}.\end{aligned}$$

□