ssdnm
środowiskowe
studia doktoranckie
z nauk matematycznych

Karol Cwalina

Uniwersytet Warszawski

Explicit construction of a non-malleable code: an analysis of a recent approach by D., K. and O.

Praca semestralna nr 3

(semestr letni 2010/11)

Opiekun pracy: Stefan Dziembowski

KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY

EXPLICIT CONSTRUCTION OF A NON-MALLEABLE CODE: AN ANALYSIS OF A RECENT APPROACH BY D., K. AND O.

KAROL CWALINA

Abstract. In this note we analyze the approach of DKO [1] towards an explicit construction of non-malleable codes. This is a new primitive introduced recently by Dziembowski, Pietrzak and Wichs [2], which weakens the notion of error detection codes. While its existence is easily provable by a fairly standard probabilistic argument, no explicit construction exists up-to-date in any quasi-general case.

This note, aiming at being relatively self-contained, presents also a relevant piece of theory of extractors.

1. Introduction

Let us follow the approach of Dziembowski, Pietrzak and Wichs [2] and let us consider the following three-step process, called a *tampering experiment*:

(1) A *source message* $s \in \{0,1\}^k$ is encoded via a (possibly randomized) procedure $\mathsf{Enc} : \{0,1\}^k \to \{0,1\}^n$, yielding a *codeword* $c = \mathsf{Enc}(s)$.
(2) The codeword is modified under a *tampering function* $f : \{0,1\}^n \to \{0,1\}^n$ to an *erroneous codeword* $\tilde{c} = f(c)$.
(3) The erroneous codeword $\tilde{c}$ is decoded using a (deterministic) procedure $\mathsf{Dec} : \{0,1\}^n \to \{0,1\}^k \cup \{\perp\}$, resulting in the decoded message $\tilde{s} = \mathsf{Dec}(\tilde{c})$. Here, $\perp$ represents the case when the codeword $\tilde{c}$ is not a valid code for any message in $\{0,1\}^k$.

The tampering experiment serves us as a universal model of a possible attack against a cryptographic system $(\mathsf{Enc}, \mathsf{Dec})$.

With this model in mind, Dziembowski, Pietrzak and Wichs [2] introduce a new primitive called non-malleable codes. They call a system $(\mathsf{Enc}, \mathsf{Dec})$ non-malleable with respect to (w.r.t.) a family $\mathcal{F}$ of tampering functions if for all $f \in \mathcal{F}$ the distribution of a result of the tampering experiment is independent of the source message, with possibly extra probability of leaving the message unaffected by the experiment.

Depending on the interpretation of the last clause, two distinct notions of non-malleability can be defined: (weak) non-malleability, when we impose no restrictions on the reason for preservation of the message being promoted, and strong non-malleability, when we only allow explicit promotion of preservation of the message, i.e. the one corresponding to the event $\tilde{c} = f(c)$.

We remark here that whenever we claim a distribution to be independent of the source message, this is only valid up to some error. It is a standard practice

Date: September, 2011.

2000 *Mathematics Subject Classification.* 94A60 (cryptography).

Key words and phrases. non-malleable codes, extractors.

in the field to require the error to be negligible w.r.t. n in the metric given by the *statistical distance*. As a consequence, we shall have to consider parametrized families of codes.

Dziembowski, Pietrzak and Wichs prove in their paper [2] that there exist (strong) codes that are non-malleable w.r.t. limited in size families of tampering functions. The proof is non-existential, however, as it crucially depends on a probabilistic argument. In the same paper the authors construct a code non-malleable with respect to the family $\mathcal{F}_{\mathsf{BIT}}$ of bit-wise independent tampering functions.

This lack of explicitness, suffered by the former result, is a motivation behind DKO's[1] attempt at providing a construction, in a setting limited to $k = 1$, of a code non-malleable w.r.t. a much larger family $\mathcal{F}_{\frac{1}{2},\frac{1}{2}}$ of tampering functions acting independently on two distinct halves of a codeword.

A presentation and an analysis of their approach is the main goal of this paper and it is organized as follows. In the next section necessary definitions are introduced. Then, in Sections 3 to 5, we present the construction proposed by DKO. Section 6 is devoted to presenting some part of theory of randomness extraction. Some issues relating it to problems of our interest are sketched there and are further developed in the following section. Section 8 concludes the paper suggesting some possible directions of a further research.

2. Preliminaries and definitions

Let us begin with some standard definitions.

Definition 2.1 (negligible functions)**.** We define the family of *negligible functions* as all those $f : \mathbb{N} \to \mathbb{R}_+$ that are smaller, in the limit, than the reciprocals of all polynomials.

Definition 2.2 (statistical distance)**.** For any two distributions X, Y (possibly implicitly given by random variables X, Y) on a common set S we define their *statistical distance* $\mathsf{SD}(X, Y)$ to be

$$\mathsf{SD}(X,Y) = \frac{1}{2}\sum_{s\in S} |X(s) - Y(s)| = \sup_{A\subset S} X(A) - Y(A),$$

where we treat X and Y as functions, mapping any $A \subset S$ to the probability of the event A w.r.t. X and Y, respectively, and abusing this notations for singletons, i.e. $X(s) = X(\{s\})$.

If, for some ε, we have $\mathsf{SD}(X, Y) < \varepsilon$, then we say that X is ε-close to Y and write it $X =_\varepsilon Y$. In the particular case of parametrized families of distributions we omit the subscript if the corresponding error function is negligible.

We now proceed towards introducing the notion of non-malleable codes.

Definition 2.3 (coding scheme)**.** A *coding scheme* consists of two functions: a randomized encoding function $\mathsf{Enc} : \{0,1\}^k \to \{0,1\}^n$ and a deterministic decoding function $\mathsf{Dec} : \{0,1\}^n \to \{0,1\}^k \cup \{\perp\}$ such that $\mathsf{Dec}(\mathsf{Enc}(s)) = s$, for $s \in \{0,1\}^k$.

[1] As we shall proceed with the analysis, some serious flaws of the approach will become clear. For this reason, we shall not disclose the identities of the authors of the presented construction, constantly referring to DKO instead. If the reader tries to uncover it on his own, please, let be aware of the fact that the word DKO is an anagram of the word *kod* which stands for *code* in Polish.

In our presentation we shall use a different yet equivalent perspective on coding schemes.

Definition 2.4 (coding scheme, an equivalent definition)**.** A coding scheme consists of a deterministic decoding function $\mathsf{Dec} : \{0,1\}^n \to \{0,1\}^k \cup \{\perp\}$ and a family $\{C_s\}_{s\in\{0,1\}^k}$ of random variables with values in $\{0,1\}^n$, the codewords, such that $\mathsf{Dec}(C_s) = s$, for each $s \in \{0,1\}^k$.

Usually we shall write S instead of $\mathsf{Dec}(C)$ and $(C|S = s)$ rather than C_s, keeping it consistent with a case particularly interesting to us, when all the codewords come from the same random variable, conditioned on different source messages being encoded. If C is uniformly distributed we shall simplify the notation even further, treating a deterministic decoding function Dec as a code.

For a random variable $X = X(C)$, deterministically dependent on C, and a tampering function f we shall write $\tilde{X}$ to denote $X(f(C))$. The tampering function will always be clear from context.

Definition 2.5 (non-malleability)**.** Let $\mathcal{F}$ be a family of tampering functions and $\varepsilon > 0$ be some real. We call a cryptographic system *non-malleable* w.r.t. the family $\mathcal{F}$ and with the security parameter ε if for each $f \in \mathcal{F}$ there exists a probability measure μ_f on $\{0,1\}^k \cup \{\perp, \mathsf{same}*\}$ such that

$$\begin{aligned} &\mathbb{P}(\tilde{S} = \tilde{s}|S = s) =_\varepsilon \mu_f(\tilde{s}) + \mu_f(\mathsf{same}*) && \text{for } \tilde{s} = s; \\ &\mathbb{P}(\tilde{S} = \tilde{s}|S = s) =_\varepsilon \mu_f(\tilde{s}) && \text{otherwise, also for } \tilde{s} = \perp. \end{aligned}$$

Just to mark the contrast, we follow with the definition of strong non-malleable codes.

Definition 2.6 (strong non-malleability)**.** Let $\mathcal{F}$ be a family of tampering functions and $\varepsilon > 0$ be some real. We call a cryptographic system *strong non-malleable* w.r.t. the family $\mathcal{F}$ and with the security parameter ε if for each $f \in \mathcal{F}$ there exists a probability measure μ_f on $\{0,1\}^k \cup \{\perp, \mathsf{same}*\}$ such that

$$\begin{aligned} &\mathbb{P}(\tilde{S} = \tilde{s} \wedge \tilde{c} \neq c|S = s) =_\varepsilon \mu_f(\tilde{s}) && \text{for all } \tilde{s} \in \{0,1\}^k \cup \{\perp\}; \\ &\mathbb{P}(\tilde{c} = c|S = s) =_\varepsilon \mu_f(\mathsf{same}*) \end{aligned}$$

It is straightforward to prove that strong non-malleable codes are non-malleable in the weak sense.

We know that strong non-malleable codes exist for families $\mathcal{F}$ limited in size as states the following theorem of Dziembowski, Pietrzak and Wichs.

Theorem 2.7 ([2, Theorem 5.1])**.** *Let $\mathcal{F}$ be any family consisting of functions $f : \{0,1\}^n \to \{0,1\}^n$. Let $\varepsilon, \rho > 0$ be arbitrary values and $k, n > 0$be integers such that*

$$n > \log\log|\mathcal{F}| + 3k + \log k + 2\log(1/\varepsilon) + \log\log(1/\rho) + 9.$$

Then there exists a strong non-malleable code w.r.t. $\mathcal{F}$, with k-bit source messages, n-bit codewords and with the security parameter ε. Moreover, a randomly chosen decoding function $\mathsf{Dec} : \{0,1\}^n \to \{0,1\}^k$ gives rise to such a code with probability $\geq 1 - \rho$.

Note that for the family $\mathcal{F}_{\mathsf{all}}$ of all tampering functions on $\{0,1\}^n$ we have $\log\log|\mathcal{F}_{\mathsf{all}}| = n + \log n$. Since no non-malleable code can exist w.r.t. $\mathcal{F}_{\mathsf{all}}$, this theorem is, up to minor additive terms, as strong as possible.

However, as we already said in the introduction, the proof of the theorem heavily relies on probabilistic method and, hence, gives no clues about how to construct such codes. This was the motivation behind the work of DKO.

Before presenting the construction, we shall now define the family of tampering functions that we shall be interested in. It naturally arises from considerations about the security of codes split between two parties, which share a secret, and stored independently.

Definition 2.8 (family $\mathcal{F}_{\frac{1}{2},\frac{1}{2}}$)**.** Let $n \in \mathbb{N}$ be even. Then we define the family $\mathcal{F}_{\frac{1}{2},\frac{1}{2}}$ as consisting of all those functions $f : \{0,1\}^n \to \{0,1\}^n$ which, when considered as $f : \{0,1\}^{n/2} \times \{0,1\}^{n/2} \to \{0,1\}^{n/2} \times \{0,1\}^{n/2}$, can be represented, for some functions $f^{\mathsf{L}}, f^{\mathsf{R}} : \{0,1\}^{n/2} \to \{0,1\}^{n/2}$, as $f(x^{\mathsf{L}}, x^{\mathsf{R}}) = (f^{\mathsf{L}}(x^{\mathsf{L}}), f^{\mathsf{R}}(x^{\mathsf{R}}))$.

3. DKO's construction

We shall begin with the following observation, crucially dependent on DKO's choice of the limited case $k = 1$.

Theorem 3.1 ([1, Definition 2.])**.** *In the setting of $k = 1$, if for each $f \in \mathcal{F}$*

$$\mathbb{P}(S = 0) = \mathbb{P}(S = 1) = \frac{1}{2},$$

$$\mathbb{P}(\tilde{S} = \perp | S = 0) = \mathbb{P}(\tilde{S} = \perp | S = 1)$$

and

$$\mathbb{P}(\tilde{S} = S) \geq \mathbb{P}(\tilde{S} = \neg S),$$

then the coding scheme is non-malleable w.r.t. the family $\mathcal{F}$.

Proof. It suffices to construct a requested probability measure μ_f. Let us take

$$\begin{aligned}
\mu_f(0) &= \mathbb{P}(\tilde{S} = 0 | S = 1)\\
\mu_f(1) &= \mathbb{P}(\tilde{S} = 1 | S = 0)\\
\mu_f(\perp) &= \mathbb{P}(\tilde{S} = \perp | S = 0) = \mathbb{P}(\tilde{S} = \perp | S = 1)\\
\mu_f(\mathsf{same}*) &= \mathbb{P}(\tilde{S} = 0 | S = 0) - \mathbb{P}(\tilde{S} = 0 | S = 1)\\
&= \mathbb{P}(\tilde{S} = 1 | S = 1) - \mathbb{P}(\tilde{S} = 1 | S = 0) \geq 0,
\end{aligned}$$

where the last inequality stems from the condition $\mathbb{P}(\tilde{S} = S) \geq \mathbb{P}(\tilde{S} = \neg S)$.

It is a straightforward computation to check that

$$\mu_f(0) + \mu_f(1) + \mu_f(\perp) + \mu_f(\mathsf{same}*) = 1.$$

□

A fundamental profit from the theorem is this: we can work with one distribution C, which will be uniform in our approach, hence simplifying the setting even further, instead of considering several distributions $(C|S = s)$, which, obviously, cannot be all uniform.

DKO propose an essentially two-step construction. More precisely, for some l' between k and n, they define two functions

$$\mathsf{BE} : \{0,1\}^n \to \{0,1\}^{l'}$$

and

$$\mathsf{FE} : \{0,1\}^{l'} \to \{0,1\}^k \cup \{\perp\},$$

which we shall call the *back-end* and the *front-end*, respectively, and define Dec as the composition of both, i.e.

$$\mathsf{Dec} = \mathsf{FE} \circ \mathsf{BE} : \{0,1\}^n \to \{0,1\}^k \cup \{\perp\}.$$

Together with a random variable C uniformly distributed on $\{0,1\}^n$ this gives rise to a legitimate coding scheme.

Following the convention of DKO [1], we shall write $Z = BE(C)$.

The idea behind this two-layer construction lies in separating two distinct goals, which can therefore be achieved by quite different means. Ideally, we would like to prove a kind of reduction from the case of the family $\mathcal{F}_{\frac{1}{2},\frac{1}{2}}$, applied to codewords $c \in \{0,1\}^n$, to some well structured and intelligible family $\mathcal{H}$ of functions acting on $z \in \{0,1\}^{l'}$. In the second step, we would exploit structural properties of functions in the class $\mathcal{H}$ to prove non-malleability w.r.t. $\mathcal{H}$ of the code derived from FE and, hence, of the code resulting from Dec too.

4. The back-end

In this section we aim at presenting a closer look at the back-end proposed by DKO. We shall also briefly analyze this approach.

The fundamental component in the back-end design by DKO is Bourgain's extractor $\mathsf{Ext}_{\mathsf{Bou}} : \{0,1\}^{n/2l'} \times \{0,1\}^{n/2l'} \to \{0,1\}$ which is applied independently to all l' pairs of blocks of codeword's bits, every block being $(n/2l')$-bits long. More formally, for $l = n/2l'$ and $x_{a,b}$ denoting the concatenation of all bits of $x \in \{0,1\}^{n/2}$ from a-th to b-th, the back-end is defined as

$$\mathsf{BE}(x^{\mathsf{L}}, x^{\mathsf{R}}) = (\mathsf{Ext}_{\mathsf{Bou}}(x^{\mathsf{L}}_{1,l}, x^{\mathsf{R}}_{1,l}), \ldots, \mathsf{Ext}_{\mathsf{Bou}}(x^{\mathsf{L}}_{(l'-1)l+1,n/2}, x^{\mathsf{R}}_{(l'-1)l+1,n/2})).$$

The motivation behind this kind of construction, especially behind employing the Bourgain extractor, lies in the ability of extractors to behave uniformly randomly as long as they are fed with sufficiently random, but possibly far from uniform, random variables. This is ensured by the following theorem.

Theorem 4.1 ([3, Theorem 3.8] after [4])**.** *There exists a universal constant $\gamma > 0$ and a polynomial time computable function $\mathsf{Ext}_{\mathsf{Bou}} : (\{0,1\}^l)^2 \to \{0,1\}^r$ such that if X, Y are two independent random variables with values in $\{0,1\}^l$ and their min-entropies (see: Definition 6.1) are greater than $(\frac{1}{2} - \gamma)l$, then*

$$\mathbb{E}_{y \leftarrow Y} \mathsf{SD}(\mathsf{Ext}_{\mathsf{Bou}}(X, y), \mathsf{U}_r) < \varepsilon,$$

with $\varepsilon = 2^{-\Omega(l)}$ and $r = \Omega(l)$. The same holds for the expectations w.r.t. $x \leftarrow X$.

This function is called the Bourgain extractor*.*

Actually, in the setting of our interest, $r = 1 = o(l)$, this theorem can be slightly strengthened. Following the idea of Barack [3, Section 5] and applying Chebyshev's inequality we can prove that Bourgain's extractor is strong.

Theorem 4.2. *Under the assumptions of Theorem 4.1, Bourgain's extractor is strong for $r = o(l)$, i.e.*

$$\mathbb{P}_{y \leftarrow Y}\left[\mathsf{SD}(\mathsf{Ext}_{\mathsf{Bou}}(X, y), \mathsf{U}_r) > \varepsilon\right] < \varepsilon,$$

with $\varepsilon = 2^{-\Omega(l)}$. The same holds w.r.t. $x \leftarrow X$.

In particular, the postulated uniformity of codewords C implies the same for Z, up to a negligible error.

Having introduced the tools, we shall now present a short argument behind the plausible usefulness of this particular back-end. For the sake of comprehensibility of exposition, we shall develop the argument for $l' = 2$.

Consider some particular block of a tampered codeword, e.g. $\tilde{C}^{\mathsf{L}}_{a,b} = f^{\mathsf{L}}(C^{\mathsf{L}})_{a,b}$. Then we partition $\{0,1\}^{n/2}$, the domain of C^{L}, w.r.t. the tampered block $\tilde{C}^{\mathsf{L}}_{a,b}$:

$$\mathsf{Dom}_{\tilde{c}_{a,b}} = \Big\{c \in \{0,1\}^{n/2} : \ f^{\mathsf{L}}(c)_{a,b} = \tilde{c}_{a,b}\Big\}$$

and, further, we split $\mathsf{Dom}_{\tilde{c}_{a,b}}$ in three parts

$$\begin{aligned}
\mathsf{Dom}^{A}_{\tilde{c}_{a,b}} &= \Big\{c \in \mathsf{Dom}_{\tilde{c}_{a,b}} : \ |\mathsf{Dom}_{\tilde{c}_{a,b}} \cap \{c' \in \{0,1\}^{n/2} : \ c'_{1,l} = c_{1,l}\}| \geq 2^{l/2}\Big\}\\
\mathsf{Dom}^{B}_{\tilde{c}_{a,b}} &= \Big\{c \in \mathsf{Dom}_{\tilde{c}_{a,b}} \backslash \mathsf{Dom}^{A}_{\tilde{c}_{a,b}} :\\
&\qquad |(\mathsf{Dom}_{\tilde{c}_{a,b}} \backslash \mathsf{Dom}^{A}_{\tilde{c}_{a,b}}) \cap \{c' \in \{0,1\}^{n/2} : \ c'_{l+1,2l} = c_{l+1,2l}\}| \geq 2^{l/2}\Big\}\\
\mathsf{Dom}^{C}_{\tilde{c}_{a,b}} &= \mathsf{Dom}_{\tilde{c}_{a,b}} \backslash (\mathsf{Dom}^{A}_{\tilde{c}_{a,b}} \cup \mathsf{Dom}^{B}_{\tilde{c}_{a,b}}).
\end{aligned}$$

Intuitively, $\mathsf{Dom}^{A}_{\tilde{c}_{a,b}}$ and $\mathsf{Dom}^{B}_{\tilde{c}_{a,b}}$ denote long "rows" and "columns" of $\mathsf{Dom}_{\tilde{c}_{a,b}}$.

Let Hint be any of the sets $\mathsf{Dom}^{?}_{\tilde{c}_{a,b}}$ we have partitioned $\{0,1\}^{n/2}$ into.

If $\mathsf{Hint} = \mathsf{Dom}^{A}_{\tilde{c}_{a,b}}$ or $\mathsf{Hint} = \mathsf{Dom}^{B}_{\tilde{c}_{a,b}}$, for some $\tilde{c}_{a,b}$, then

$$H_\infty(C^{L}_{1,l} | C^{L} \in \mathsf{Hint}) \geq l/2$$

or

$$H_\infty(C^{L}_{l+,2l} | C^{L} \in \mathsf{Hint}) \geq l/2,$$

respectively. Also, if $\mathsf{Hint} = \mathsf{Dom}^{C}_{\tilde{c}_{a,b}}$ and $|\mathsf{Hint}| \geq 2^{l(1-\gamma)}$ then the above inequalities hold in an almost equally strong form:

$$\begin{aligned}
\mathbb{P}(C^{L}_{1,l} = c_{1,l} | C^{L} \in \mathsf{Hint}) &= \frac{|\mathsf{Dom}_{\tilde{c}_{a,b}} \cap \{c' \in \{0,1\}^{n/2} : \ c'_{1,l} = c_{1,l}\}|}{|\mathsf{Hint}|}\\
&\leq \frac{2^{l/2}}{2^{l(1-\gamma)}} \leq 2^{-l(\frac{1}{2}-\gamma)},
\end{aligned}$$

hence

$$H_\infty(C^{L}_{1,l} | C^{L} \in \mathsf{Hint}) \geq l(\frac{1}{2} - \gamma)$$

and, analogously,

$$H_\infty(C^{L}_{l+1,2l} | C^{L} \in \mathsf{Hint}) \geq l(\frac{1}{2} - \gamma).$$

If $C^{L} \in \mathsf{Dom}^{C}_{\tilde{c}_{a,b}}$ and $|\mathsf{Dom}^{C}_{\tilde{c}_{a,b}}| < 2^{l(1-\gamma)}$ we cannot prove a good bound on the min-entropy. This is, however, a rare event:

$$\begin{aligned}
\mathbb{P}(C^{L} &\in \mathsf{Dom}^{C}_{\tilde{c}_{a,b}} \text{ and } |\mathsf{Dom}^{C}_{\tilde{c}_{a,b}}| < 2^{l(1-\gamma)} \text{ for some } \tilde{c}_{a,b})\\
&= \sum_{\tilde{c}_{a,b}} \mathbb{P}(C^{L} \in \mathsf{Dom}^{C}_{\tilde{c}_{a,b}} \text{ and } |\mathsf{Dom}^{C}_{\tilde{c}_{a,b}}| < 2^{l(1-\gamma)})\\
&\leq \sum_{\tilde{c}_{a,b}} \frac{2^{l(1-\gamma)}}{2^{2l}} = 2^{l} \cdot \frac{2^{l(1-\gamma)}}{2^{2l}}\\
&= 2^{-\gamma l},
\end{aligned}$$

which is negligible.

Therefore, up to a negligible probability,

$$H_\infty(C^L_{1,l}|C^L \in \mathsf{Hint}) \geq l(\frac{1}{2} - \gamma)$$

or

$$H_\infty(C^L_{l+,2l}|C^L \in \mathsf{Hint}) \geq l(\frac{1}{2} - \gamma)$$

and the same applies for C^{R}.

Let us now apply Theorem 4.2 in order to exploit the fact that Bourgain's extractor is strong. This results, again up to a negligible probability, in

$$(Z_1|\mathsf{Hint}_{\tilde{c}^{\mathsf{L}}_{a,b}}, \mathsf{Hint}_{\tilde{c}^{\mathsf{R}}_{a,b}}) = \mathsf{U}_1$$

or

$$(Z_2|\mathsf{Hint}_{\tilde{c}^{\mathsf{L}}_{a,b}}, \mathsf{Hint}_{\tilde{c}^{\mathsf{R}}_{a,b}}) = \mathsf{U}_1$$

which reduces further to $(Z_1|\tilde{Z}_j = \tilde{z}_j) = \mathsf{U}_1$ or $(Z_2|\tilde{Z}_j = \tilde{z}_j) = \mathsf{U}_1$, for some j and $\tilde{z}_j$.

The above reasoning can be generalized to the case of arbitrarily many blocks, i.e. arbitrary l'.

Now, it is tempting to claim that uniformity of all Z_i $(i = 1, \ldots, l)$ but one $(Z_{i*}$ say) w.r.t. all $\tilde{Z}_j$ $(j = 1, \ldots, l)$ but one (Z_{j*}) is an argument for dependence of $\tilde{Z}_{j*}$ on Z_{i*} only. In particular, DKO aim at showing that we can model a random mapping $Z \to \tilde{Z}$ by a function of the shape

$$\tilde{Z}_i = h_i(Z_j) \text{ for some } j,$$

where $h_i : \{0,1\} \to \{0,1\}$ denotes some deterministic function, or by the evaluation at Z of a randomly (and independently of C) chosen function of the above-specified form. The formerly-defined class of such deterministic bit-modifying functions will be denoted by $\mathcal{H}$.

This can be, however, easily shown to be incorrect and the error is founded on the same kind of subtlety as the distinction between independence of random variables and their pairwise independence. To make the thing more explicit, we shall now define a distribution $(\tilde{Z}|Z = z)$ such that any distribution possibly suspected of being uniform in general case is such but, nonetheless, $\tilde{Z}$ does not depend on Z in a manner foreseen by DKO.

To this end, let us consider the following distribution: if all bits of Z sum in $\mathbb{Z}_2$ to 0 then $\tilde{Z}$ is uniformly distributed over the whole domain; otherwise, it is uniformly distributed on the part of its domain where the sum of bits equals 1 only.

One can check that the only functions $h \in \mathcal{H}$ respecting those conditions can be either those fixing their output to some value with an odd sum of its bits, or those swapping an even number of bits and, possibly, permuting them. Those functions, however, cannot be combined in such a way that $(\tilde{Z}|Z = 0)$ is uniformly distributed. On the other hand, Z is completely uniformly distributed provided not all bits of $\tilde{Z}$ are known.

While this proves the approach incorrect, we shall discuss some issues related to the existence of codes non-malleable w.r.t. the family $\mathcal{H}$ in the next section.

5. The front-end: codes non-malleable w.r.t. $\mathcal{H}$

In this section we shall briefly sketch a front-end function proposed by DKO. Later on we shall consider to what extent trivial Hadamard's extractor fulfills the requirements of being non-malleable w.r.t. $\mathcal{H}$.

In DKO's construction the l'-bits long input is split in m' equally long blocks and an extra bit β. Then, every block $Z_{a,b}$ encodes 0, 1 or $\perp$ and we shall denote it by $\mathsf{FE}_{\mathsf{help}}(Z_{a,b})$. The definition of $\mathsf{FE}(Z)$ is now straightforward: it is the value represented by the first "valid" block, i.e. the first one such that $\mathsf{FE}_{\mathsf{help}}(Z_{a,b}) \neq \perp$. If no such block exists then we define $\mathsf{FE}(Z)$ as equal to the extra bit β.

$\mathsf{FE}_{\mathsf{help}}$, for a suitably chosen m, looks at its argument as $m/2$ numbers in the range $\{0, m-1\}$, each one $\log m$-bits long, followed by m bits. We shall call the set of $m/2$ numbers the *mask* and m bits – the *face*. Then $\mathsf{FE}_{\mathsf{help}}$ is 0 or 1 if all $m/2$ numbers in the mask are distinct and the bits of the face equal to 0 (or 1, respectively) are exactly those whose positions are listed in the mask. Otherwise we define $\mathsf{FE}_{\mathsf{help}}$ as equal to $\perp$.

A few words of remark would fit now. First of all, the way we define $\mathsf{FE}_{\mathsf{help}}$ assures that a uniformly random argument will be mapped to $\perp$ with high probability. On the other hand, suitably chosen values of m' and m can make the value of $\mathsf{FE}(Z)$ independent of the bit β, up to a negligible error. Finally, at the intuitive level at least, the only way of negating $\mathsf{FE}(Z)$ is to negate the first valid block which, ultimately, boils down to the ability of successfully negating any bit encoded with help of $\mathsf{FE}_{\mathsf{help}}$. The latter can be shown, however, to be impossible.

We shall not develop on this construction any more.

It is worth noting here, that codes non-malleable w.r.t. an even tinier family of functions have already been successfully constructed in the foundational paper of Dziembowski, Pietrzak and Wichs: see [2, Theorem 4.1]. The family of interest there was the family $\mathcal{F}_{\mathsf{bit}}$ of functions deterministically but independently modifying every bit of the input.

In view of powerful properties exhibited by extractors (see the discussion in this and previous section), it is of independent interest to find if extractors are good candidates for being non-malleable codes. While, to our knowledge, this seems to be intractable at the time, we shall discuss non-malleability w.r.t. $\mathcal{H}$ of Hadamard's extractor, the archetype of Bourgain's extractor. This will feature next two sections.

6. Randomness extractors

Definition 6.1 (min-entropy)**.** For a probability distribution X, defined on some set S, we define its *min-entropy* $H_\infty(X)$ to be

$$H_\infty(X) = -\log \sup_{x \in S} X(x).$$

The following folklore observation is worth noting.

Theorem 6.2. *Let $k \in \mathbb{N}$ and X be a random variable of the min-entropy k. Then X is a convex combination of random variables uniformly distributed on a supporting set of size 2^k.*

By an immediate argument, the min-entropy $H_\infty(X)$ is bounded from above by the well-known Shannon entropy $H(X)$. It is quite surprising, however, to learn that it is $H_\infty(X)$, and not $H(X)$, the one which is, from an extractor's point of view, the right measure of quantity of randomness carried by a random variable.

We shall now introduce some issues directly related to the topic of this section, i.e. to randomness extractors.

An unattainable goal of randomness extractors is to produce an almost uniform output, i.e. uniform up to a negligible error, when given any source, possibly very far from being completely random but, nonetheless, sufficiently random. Of course, by definition, considered extractors must be deterministic, the only randomness coming from the random argument. As we shall now prove, no such extractor exists, even if we require it to extract only one uniformly random bit from any source of min-entropy (and hence entropy, too) at least $n-1$. This theorem is also a folklore in the field.

Theorem 6.3. *There is no function* $\mathsf{Ext} : \{0,1\}^n \to \{0,1\}$ *such that for every random variable* X *on* $\{0,1\}^n$ *with* $H_\infty(X) \geq n-1$ *the distribution* $\mathsf{Ext}(X)$ *is (almost) uniformly 0 or 1.*

Proof. Suppose that such a function exists. Then it outputs 0 or 1 for at least 2^{n-1} distinct arguments. The random variable X uniformly distributed on this set satisfies $H_\infty(X) \geq n-1$ but $\mathsf{Ext}(X)$ is fixed – a contradiction. □

In the light of this result a search for extractors splits between two main lines of interest: seeded extractors, which are additionally given some truly random (i.e. uniformly distributed) bits from a perfect source, or multisource extractors, which are given some fixed number of independent samples from sufficiently random sources. The 2-source extractor of our interest, Bourgain's extractor, is an example of this kind. In both cases, explicit construction are far from the limits established by probabilistic method. For more information on a state-of-the-art development in the area of seeded and unseeded extractors see [6, 7]. In Nisan [5] one can find general discussion of historical motivations behind development of a concept of extractors.

We shall now present some considerations behind the construction of the currently most powerful 2-source extractor, i.e. the aforementioned Bourgain's extractor, and its archetype: Hadamard's extractor.

Let us consider the function $\mathsf{Ext}_{\mathsf{Had}} : (\{0,1\}^n)^2 \to \{0,1\}$ interpreted as the scalar product of two n-bits long vectors over $\mathbb{Z}_2$. Then, it is an extractor for any pair of two n-bit sources provided that their min-entropies sum to $n + \omega(\log n)$. The proof of this statement is our goal in this first of the section.

Theorem 6.4. *The function* $\mathsf{Ext}_{\mathsf{Had}}$*, is a 2-source extractor for any pair of two* n*-bit sources whose min-entropies sum to* $n + \omega(\log n)$*.*$\mathsf{Ext}_{\mathsf{Had}}$ *is called Hadamard's extractor.*

Proof. We shall sketch the proof as it is presented in [3, Subsection 3.1]. Intentionally we shall prove it in the setting of our interest but the proof remains unchanged in the case of any field.

For the non-trivial character ψ of $\mathbb{Z}_2$ we consider the random variable $\mathsf{bias}_\psi(X,Y)$ defined as

$$\mathsf{bias}_\psi(X,Y) = |\mathbb{E}\psi(\mathsf{Ext}_{\mathsf{Had}}(X,Y))|$$

and we denote by $e_x(\cdot)$ the character $y \mapsto \psi(xy)$.

Below we shall use Fourier analysis w.r.t. characters given by e and the definitions of appropriate Fourier-analytic notions will be assumed to be deducible from context.

Now, note that

$$\mathsf{bias}_\psi(X,Y) = \left|\sum_{y\in\mathbb{Z}_2^n} Y(y) \sum_{x\in\mathbb{Z}_2^n} X(x)\psi(xy)\right|$$

and, because $\sum_{x\in\mathbb{Z}_2^n} X(x)\psi(xy) = 2^n\overline{\widehat{X}(y)}$, we have

$$\begin{aligned}\mathsf{bias}_\psi(X,Y)^2 &= 2^{2n}\left|\sum_{y\in\mathbb{Z}_2^n} Y(y)\overline{\widehat{X}(y)}\right|^2\\ &\leq 2^{2n}||Y||_2^2||\widehat{X}||_2^2\\ &= 2^n||Y||_2^2||X||_2^2\\ &\leq 2^{n-H_\infty(Y)-H_\infty(X)}.\end{aligned}$$

It follows from the above estimate that

$$||\mathsf{bias}_\psi\widehat{(X,Y)} - \mathsf{U}_1||_\infty < 2^{\frac{1}{2}(n-H_\infty(X)-H_\infty(Y))}$$

and, as a result,

$$\mathsf{SD}(\mathsf{bias}_\psi(X,Y),\mathsf{U}_1) = \frac{1}{2}||\mathsf{bias}_\psi(X,Y) - \mathsf{U}_1||_1 < 2^{\frac{1}{2}(n+1-H_\infty(X)-H_\infty(Y))}.$$

The last step of this reasoning is interesting in its own right and is known as the XOR lemma [3, Lemma 4.1]. □

While the proof shows that the bound for applicability of the theorem has to be at least n for the sum of the min-entropies, the following example suggests why it is so. To this end, consider two random variables X, Y which are uniformly distributed on the set of those elements of $\{0,1\}^n$ whose first $n/2$ bits (last bits, respectively) are 0. In this case, $\mathsf{Ext}_{\mathsf{Had}}$ always outputs 0, which is very non-uniform.

As the above example suggests, it is a particular structure of the random variables that caused this high non-uniformity. Actually, what matters is the fact that $H_\infty(X)$ does not differ from $H_\infty(2X)$, since, in the Minkowskian sense, $2X = X$. This intuition is formalized in the following observation.

Theorem 6.5. *Let X, Y be independent random variables such that for some $c_1, c_2 \in \mathbb{N}$ the sources $2^{c_1}X - 2^{c_1}X$ and $2^{c_2}Y - 2^{c_2}Y$ have min-entropies k_1 and k_2, respectively. Then, keeping the definition of $\mathsf{bias}_\psi(\cdot,\cdot)$ as given in the proof of Theorem 6.4,*

$$\mathsf{bias}_\psi(X,Y) \leq (2^n 2^{-(k_1+k_2)})^{1/2^{c_1+c_2+2}},$$

for every non-trivial character ψ of $\mathbb{Z}_2$.

Proof. Once again, we shall follow Rao's exposition [3, Subsection 3.2.1] of Bourgain's coonstruction.

A key observation is the inequality $\mathsf{bias}_\psi(X,Y)^2 \leq \mathsf{bias}_\psi(X-X,Y)$ which constitutes the inductive step for the proof.

$$\begin{aligned}
\mathsf{bias}_\psi(X,Y) &= \left|\sum_{y\in\mathbb{Z}_2^n} Y(y) \sum_{x\in\mathbb{Z}_2^n} X(x)\psi(xy)\right| \\
&\leq \sum_{y\in\mathbb{Z}_2^n} Y(y) \left|\sum_{x\in\mathbb{Z}_2^n} X(x)\psi(xy)\right|
\end{aligned}$$

Then, by convexity,

$$\begin{aligned}
\mathsf{bias}_\psi(X,Y)^2 &\leq \sum_{y\in\mathbb{Z}_2^n} Y(y) \left|\sum_{x\in\mathbb{Z}_2^n} X(x)\psi(xy)\right|^2 \\
&= \left|\sum_{y\in\mathbb{Z}_2^n} Y(y) \sum_{x_1,x_2\in\mathbb{Z}_2^n} X(x_1)X(x_2)\psi(x_1y)\psi(-x_2y)\right| \\
&= \left|\sum_{y\in\mathbb{Z}_2^n} Y(y) \sum_{x_1,x_2\in\mathbb{Z}_2^n} X(x_1)X(x_2)\psi((x_1-x_2)y)\right| \\
&= \left|\sum_{y\in\mathbb{Z}_2^n} Y(y) \sum_{x\in\mathbb{Z}_2^n} (X-X)(x)\psi(xy)\right| \\
&= \mathsf{bias}_\psi(X-X,Y).
\end{aligned}$$

□

This theorem is a key for successfully weakening the restriction on the min-entropy, imposed by the proof method used above. In the work of Bourgain [4], the "additive grow-up" of a random variable's min-entropy is assured by subtle combinatorial properties which feature the field of so-called sum-product estimates. At the surface, however, everything boils down to some particular yet simple application of Theorem 6.4.

Definition 6.6 (Bourgain's extractor)**.** Let $\mathbb{F}$ be the field of cardinality 2^n. Then $\mathsf{Ext}_{\mathsf{Bou}} : (\{0,1\}^n)^2 \to \{0,1\}$ is defined as

$$\mathsf{Ext}_{\mathsf{Bou}}(x,y) = \mathsf{Ext}_{\mathsf{Had}}((x,x^2),(y,y^2)),$$

where x, y are treated as elements of $\mathbb{F}$, when x^2 and y^2 are considered, and as elements of a vector space over $\mathbb{Z}_2$ elsewhere.

For more details on this subject consult the original paper [4] or Rao's exposition [3]. Here, we shall only make one word of comment on the apparent differences between a clear approach present in Rao's article and followed in our definition of Bourgain's extractor, and Bourgain's original exposition [4, Section 3]. This difference is mainly due to the necessity of interpreting extractor's arguments as elements of a field and squaring the arguments interpreted this way. Our approach deliberately leaves it implicit.

It is worth noting that the same construction is valid over any final field.

Having presented Bourgain's extractor and an archetypal Hadamard's extractor, we shall now pass to a discussion of their possible non-malleability w.r.t. $\mathcal{H}$.

7. Extractors as non-malleable codes

As we have already mentioned in Section 5, Dziembowski, Pietrzak and Wichs [2, Theorem 4.1] establish an explicit code non-malleable w.r.t. the family $\mathcal{F}_{\mathsf{bit}}$ of functions independently modifying every bit of its argument. While the result is worth appreciation, it was obtained by fairly involved means. In this section we shall discuss reasons why extractors presented in the previous section are or are not good candidates for codes non-malleable w.r.t. a larger family $\mathcal{H}$.

Let us consider the code stemming from Hadamard's extractor $\mathsf{Ext}_{\mathsf{Had}}$.

First, let a tampering function $f_1(x^{\mathsf{L}}, x^{\mathsf{R}})$ set the first bits of its arguments to 1 (or set the first bit of x^{L} to 1 and negate the first bit of x^{R}). It is clear that, in the setting established by Theorem 3.1, $\mathbb{P}(\tilde{S} = \neg S) = \frac{3}{4}$ and the code is not non-malleable w.r.t. $\mathcal{F}_{\mathsf{bit}}$, not to mention $\mathcal{H}$.

For another considerations, let $f_2(x^{\mathsf{L}}, x^{\mathsf{R}})$ negate the first two bits of x^{L} and both negate and swap the first two bits of x^{R}. It requires a little check to verify that, in this case, $\mathbb{P}(\tilde{S} = \neg S) = \frac{3}{4}$. Again, this proves non-malleability of the code w.r.t. $\mathcal{H}$.

In the next paragraphs we shall show that these two tampering functions represent all possible obstacles to non-malleability of $\mathsf{Ext}_{\mathsf{Had}}$ w.r.t. the family $\mathcal{H}$.

Fix a tampering function $f \in \mathcal{H}$ and let us suppose that it does not set deterministically any bit. For the reason of convenience we shall also require f to depend on all bits of its argument.

Let us consider a graph G whose vertices correspond to distinct bits of f's input: we shall call them $z_1, z_2, \ldots, z_{2n}$. By convention $\mathsf{Ext}_{\mathsf{Had}}(Z) = z_1 z_{n+1} + \ldots + z_n z_{2n}$. Then, we connect two vertices with an edge whenever corresponding bits, or their deterministic functions, are multiplied together, either in the formula for $\mathsf{Ext}_{\mathsf{Had}}(Z)$ or $\mathsf{Ext}_{\mathsf{Had}}(\tilde{Z})$.

By the extra assumption concerning f, we can conclude that G is bipartite and every vertex has either one or two neighbours: in the first case the vertex is also its neighbour's only neighbour. We can check that whenever there exists any such pair and at least one of its "vertices" is modified by h, then $\mathbb{P}(\tilde{S} = S) = \frac{1}{2} = \mathbb{P}(\tilde{S} = \neg S)$. If this is not the case G splits in cycles of even length and some pairs independent of the events $\tilde{S} = S$ and $\tilde{S} = \neg S$.

We shall now analyze one such cycle and denote by $\mathsf{Odd}/\mathsf{Even}$ the two distinct maximal sets of its verices containing every second vertex. Let a_1, b, a_2 be three consecutive elements of the cycle, $a_1, a_2 \in \mathsf{Odd}$ and $b \in \mathsf{Even}$. The crucial observation now is that for every possible $a_1 \in \{0,1\}$ there is exactly one $a_2 \in \{0,1\}$ such that exactly one of $(Z|a_1, a_2)$ and $(\tilde{Z}|a_1, a_2)$ depends on b. We can check that in such a case

$$\mathbb{P}(\tilde{S} = S|a_1, a_2 \text{ and other bits in } \mathsf{Odd} \text{ fixed}) = \frac{1}{2} = \mathbb{P}(\tilde{S} = \neg S|a_1, a_2 \text{ and ...})$$

up to a negligible error. Moreover, there are only two valuations of bits belonging to Odd such that the condition does not hold for any pair a_1, a_2 of two "consecutive" bits in the set Odd.

Now, denoting by $\bigcup \mathsf{Odd}$ the union of all Odd-sets, we conclude that if there are at least $\omega(\log n)$ vertices in cycles altogether, then every valuation, up to a negligible error, of the bits in $\bigcup \mathsf{Odd}$ results in some pair a_1, a_2 satisfying the condition mentioned above, hence resulting in

$$\mathbb{P}(\tilde{S} = S) = \frac{1}{2} = \mathbb{P}(\tilde{S} = \neg S).$$

8. Concluding remarks

In this final section we shall discuss some points related to the topics presented in this paper. In particular, we shall address limitations of the proposed methods.

A very weakness of the approach presented in this paper is its ineffectiveness, meaning not non-constructivity, but still worse, its fundamental flaws and incorrectness at the current stage of development. It is the back-end's design and deemed reduction from $\mathcal{F}_{\frac{1}{2},\frac{1}{2}}$ to $\mathcal{H}$ which experience this problem.

While we can imagine this construction being correct it is currently out of our range to successfully prove it. One, quite natural approach to this would be to exclude any possibility of a "hidden conspiracy" between distinct bits of $\tilde{Z}$ and Z. A probabilistic argument proves that almost every extractor design behaves well w.r.t. those hidden conspiracies when not too many bits of Z are considered.

On the other hand, let us consider the code originating from Hadamard's extractor. If a tampering function permutes bits of both halves of a codeword in the same manner, then distinct bits of $\tilde{Z}$ and Z are independent but conspiracy is still present as the sum of all bits of Z remains unchanged. This example is quite embarrassing as the design of Bourgain's extractor is strongly influenced by that of Hadamard's extractor.

The second issue where some improvement would be appreciated is the limited setting $k = 1$ that we work in. It is an intrinsic feature of the approach supported by Theorem 3.1. A natural question is whether two definitions of non-malleability: Definition 2.5 and one somehow imitating that originating from Theorem 3.1 are equivalent in the general case.

In the case of this question it is even unclear what the latter should be. One immediate candidate should require that whatever a tampering function is, it is more probable to leave the source message unchanged than to change it according to some other bijective mapping.

Finally, there remains wide open a question of designing strong non-malleable codes, which has not been mentioned at all in this paper.

Acknowledgements

The author is supported by the National PhD Programme in Mathematical Sciences at the University of Warsaw.

References

[1] D., K., O., *Non-malleable codes*, in preparation

[2] S. Dziembowski, K. Pietrzak, D. Wichs, *Non-Malleable Codes*, Innovations in Computer Science (ICS) 2010

[3] A. Rao, *An Exposition of Bourgain's 2-Source Extractor*

[4] J. Bourgain, *More on the sum-product phenomenon in prime fields and its applications*, International Journal of Number Theory, Vol. 1, No. 1 (2005) 1-32

[5] N. Nisan, *Extracting Randomness: How and Why. A survey*

[6] R. Shaltiel, *Recent developments in explicit constructions of extractors*

[7] B. Barak, R. Impagliazzo, A. Wigderson, *Extracting Randomness Using Few Independent Sources*

Faculty of Mathematics, Informatics and Mechanics, University of Warsaw, Banacha 2, 02-097 Warszawa, Poland
E-mail address: cwalina@mimuw.edu.pl