Elastyczne ekstraktory dwuźródłowe i ich zastosowania

autoreferat rozprawy doktorskiej

Maciej Obremski
Wydział Matematyki, Informatyki i Mechaniki, Uniwersytet Warszawski

1 Wstęp

Prezentujemy nowe pojęcie *elastycznego* extraktora dwuźródłowego. W przeciwieństwie do standardowych dwuźródłowych ekstraktorów, które wymagają by każde ze źródeł osobno miało pewną entropię, *elastyczny* ekstraktor wymaga by sumaryczna entropia źródeł przekraczała daną wartość. Wyróżniamy słabe i silne *elastyczne* ekstraktory i podobnie jak w przypadku słabych i silnych ekstraktorów dwuźródłowych dowodzimy, że każdy słaby *ekstraktor* jest też silny kosztem nieznacznego pogorszenia jego parametrów. Ponadto dowodzimy, że dwa z powszechnie znanych i używanych ekstraktorów są *elastyczne* co znacząco wzmacnia tezę Leftover Hash Lemma dla tych ekstraktorów. Pojęcia *elastycznych* ekstraktorów używamy we wspólnej pracy ze Stefanem Dziembowskim i Tomaszem Kazaną "Non-Malleable Codes from Two-Source Extractors", praca ta została wysłana na międzynarodową konferencję. Konstruujemy w niej wydajny, teorio-informacyjnie bezpieczny kod niekowalny w modelu z przepołowioną pamiecią dla wiadomości jednobitowych. Pojęcie kodów niekowalnych zostało wprowadzone przez S.Dziembowskiego, K.Pietrzaka i D.Wichsa (ICS 2010), jako narzędzie do składowania danych na urządzeniu, które może być poddane działaniu przeciwnika modyfikującego dane. Nieformalnie ujmując, schemat $(\mathsf{Enc} : \mathcal{M} \to \mathcal{L} \times \mathcal{R}, \mathsf{Dec} : \mathcal{L} \times \mathcal{R} \to \mathcal{M})$ jest *kodem niekowalnym w modelu z przepołowioną pamięcią* jeśli wspomniany przeciwnik manipulujący *niezależnie* L i R (gdzie (L, R) koduje pewną wiadomość m) nie może otrzymac, kodu wiadomości m', która byłaby różna od m ale z nią "skorelowana"(np. $m' = m + 1$). Do teraz efektywna konstrukcja informacyjnie bezpiecznego kodu o takiej wlasności pozostawala nieznana nawet dla wiadomości ze zbioru $\{0, 1\}$. Nasza konstrukcja rozwiązuje ten problem. Ponadto dowodzimy jej odporności na wycieki, w następującym sensie: przeciwnik zanim wybierze dwie funkcje manipulujące (jedną na L, drugą na R) może poznać dowolną, ustaloną wcześniej funkcję wycieku z (L, R). Formalnie, dla każdego $\xi < 1/4$ potrafimy podać efektywną konstrukcje kodu niekowalnego taką, że przeciwnik przed wyborem funkcji manipulacji pozna wartości

wybranych przez siebie adaptywnie funkcji $F_1(L), F_2(R), F_3(L), F_4(R)...$ byle tylko sumaryczna dlugość wyjścia tych funkcji nie przekraczała $\xi \cdot (|L| + |R|)$. Konstrukcja naszego kodu jest oparta na iloczynie skalarnym nad ciałami skończonymi, ale pokazujemy jak zbudowac kod z dowolnego innego dwuźródłowego ekstraktora, który jest *elastyczny (flexible)*. Poza tym pokazujemy, ze definicja kodów niekowalnych w przypadku wiadomości jednobitowych ma równoważną, prostszą charakteryzację mianowicie: jeśli wybierzemy wiadomość m jednostajnie z $\{0,1\}$ wtedy prawdopodobieństwo, że przeciwnik bedzie w stanie uzyskac (w sposób opisany powyżej) wiadomość przeciwną do m jest niewiększe niż $1/2 + \epsilon$.

2 Ekstraktory

Zacznijmy od podstawowych definicji.

Definicja (dystans statystyczny). Dystansem statystycznym (albo po prostu odległością) między dwiema zmiennymi losowymi A i B określonymi na zbiorze $\mathcal{A}$ będziemy nazywać

$$\Delta(A;B) = \frac{1}{2} \sum_{a \in \mathcal{A}} |P(A = a) - P(B = a)|.$$

Dla zmiennej U o rozkładzie jednostajnym na zbiorze $\mathcal{A}$ będziemy mówili, że $\Delta(A, U)$ jest odległością zmiennej losowej A od rozkładu jednostajnego bedziemy to oznaczać po prostu przez $d(A)$.

Definicja (min-entropia). Przez pojęcie *min-entropii* będziemy rozumieć:

$$\mathbf{H}_\infty(X) = \log\left(\frac{1}{\max_{x \in \mathcal{X}} X(x)}\right) = -\log\left(\max_{x \in \mathcal{X}} X(x)\right)$$

gdzie $X(x) = P(X = x)$.

Definicja (ekstraktor). Mówimy, że $\mathsf{ext} : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ jest (k, ϵ)-dwuźródłowym ekstraktorem jeśli dla niezależnych zmiennych losowych X na $\mathcal{X}$ i Y na $\mathcal{Y}$, takich, że $\mathbf{H}_\infty(X) \geqslant k$ i $\mathbf{H}_\infty(Y) \geqslant k$, wynik $\mathsf{ext}(X, Y)$ jest zmienną losową, której odległość od rozkładu jednostajnego na $\mathcal{Z}$ jest niewiększa niż ϵ. Formalnie $\Delta(\mathsf{ext}(X, Y); U_{\mathcal{Z}}) \leqslant \epsilon$, gdzie $U_{\mathcal{Z}}$ jest rozłożona jednostajnie na $\mathcal{Z}$.

Definicja (silny ekstraktor). Mówimy, że $\mathsf{ext} : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ jest silnym (k, ϵ)-dwuźródłowym ekstraktorem, jeśli dla niezależnych zmiennych losowych X na $\mathcal{X}$ i Y na $\mathcal{Y}$ takich, że $\mathbf{H}_\infty(X) \geqslant k$ i $\mathbf{H}_\infty(Y) \geqslant k$, zachodzi $\Delta((\mathsf{ext}(X, Y), X); (U_{\mathcal{Z}}, X)) \leqslant \epsilon$. Co nieformalnie oznacza, że nawet jeśli przeciwnik zna jedno ze źródeł (X lub Y) wciąż nie jest on w stanie odróżnić wyniku ekstraktora od uczciwie jednostajnej zmiennej losowej z prawdopodobieństwem istotnie wyższym niż $\frac{1}{2}$. W [43] (Twierdzenie 5.1) znajduje się twierdzenie

przypisane Boaz Barakowi, że każdy 2-źródłowy ekstraktor jest też silnym 2-źródłowym ekstraktorem tylko z nieco gorszymi parametrami.

Definicja (słaby elastyczny ekstraktor). Funkcję $\mathsf{ext} : \mathcal{L} \times \mathcal{R} \to \mathcal{C}$ nazwiemy *słabym elastycznym* (k, ϵ)*-2-źródłowym ekstraktorem* jeśli dla każdego $L \in \mathcal{L}$ oraz $R \in \mathcal{R}$ takich, że $\mathbf{H}_\infty(L) + \mathbf{H}_\infty(R) \geqslant k$ zachodzi $d(\mathsf{ext}(L, R)) \leqslant \epsilon$.

Definicja (silny elastyczny ekstraktor). Funkcję $\mathsf{ext} : \mathcal{L} \times \mathcal{R} \to \mathcal{C}$ nazwiemy *silnym elastycznym* (k, ϵ)*-2-źródłowym ekstraktorem* (lub po prostu : *elastycznym* (k, ϵ)*-ekstraktorem*) jeśli dla każdego $L \in \mathcal{L}$ oraz $R \in \mathcal{R}$ takich, że $\mathbf{H}_\infty(L)+\mathbf{H}_\infty(R) \geqslant k$ zachodzi $d(\mathsf{ext}(L, R)|L) \leqslant \epsilon$ i $d(\mathsf{ext}(L, R)|R) \leqslant \epsilon$.

Dwie ostatnie definicje są nowymi pojęciami wprowadzonymi przez nas.

Wyniki.

Twierdzenie 1. *Niech* $\mathsf{ext} : (\{0,1\}^N)^2 \to \{0,1\}^M$ *będzie* słabym *elastycznym* (K, ϵ)*-ekstraktorem, dla* $K \geqslant N$*. Wtedy dla każdego* $K' \geqslant K$ *mamy, że* ext *jest* silnym *elastycznym* (K', ϵ')*-ekstraktorem, gdzie* $\epsilon' = 2^M(\epsilon + 2^{K-K'})$.

To twierdzenie daje nam narzędzie do dowodzenia, że ekstraktor jest silny. Korzystamy z niego w dowodzie następującego twierdzenia.

Twierdzenie 2. *Dla każdego skończonego ciała* $\mathbb{F}$ *i dowolnego* n *mamy, że* $\mathsf{ext} : \mathbb{F}^n \times \mathbb{F}^n \to \mathbb{F}$ *zdefiniowany jako* $\mathsf{ext}^n_{\mathbb{F}}(L, R) = \langle L, R\rangle$ *jest elastycznym* (k, ϵ)*-ekstraktorem dla dowolnego* k *oraz* ϵ *takiego, że*

$$\log(1/\epsilon) = \frac{k - (n+4)\log|\mathbb{F}|}{3} - 1.$$

Elastyczność iloczynu skalarnego jest kluczowa przy tworzeniu kodu niekowalnego, wykorzystujemy ją zarówno w dowodzie twierdzenia 9, jak i w przypadku kodu niekowalnego odpornego na wycieki tj. twierdzenie 10. Własność elastyczności w przypadku iloczynu skalarnego uogólnia wynik Leftover Hash Lemma [3].
Kolejne twierdzenie dotyczy innego znanego ekstraktora wprowadzonego przez Holensteina w [33]. W owej pracy dowodzi, że poniższy ekstraktor jest silnym ekstraktorem z ziarnem co jest słabszym wynikiem od uzyskanego przez nas. Niech $GF(2^n)$ oznacza ciało Galois.

Twierdzenie 3. *Ekstraktor* $\mathsf{ext} : GF(2^n) \times GF(2^n) \to GF(2^\lambda)$*, zdefiniowany jako* $\mathsf{ext}(X, Y) = (X \cdot Y)_\lambda$ *jest* $(k, 2^{\frac{n-k+2\lambda-2}{2}})$ *słabym elastycznym 2-źródłowym ekstraktorem.*

Gdzie $(X)_\lambda$ oznacza obcięcie ciagu bitów do λ najbardziej istotnych. Stosując tutaj Twierdzenie 1 możemy uzyskać, że powyższy ekstraktor jest ekstraktorem elastycznym.

3 Wycieki i Leftover Hash Lemma

3.1 Leftover Hash Lemma dla iloczynu skalarnego

Rodzinę $\mathcal{H}$ deterministycznych funkcji $h : \mathcal{X} \to \{0,1\}^v$ nazywamy *p-uniwersalną rodziną haszującą* (na przestrzeni $\mathcal{X}$), jeśli dla każdej pary $x_1 \neq x_2 \in \mathcal{X}$ zachodzi $P_{h \leftarrow \mathcal{H}}[h(x_1) = h(x_2)] \leqslant p$. Jeśli $p = \frac{1}{2^v}$, mówimy, że $\mathcal{H}$ jest *uniwersalna*. Przejdzmy do sformułowania *Leftover Hash Lemma (LHL).*

Twierdzenie 4. *(Leftover-Hash Lemma) Niech rodzina $\mathcal{H}$ funkcji deterministycznych $h : \mathcal{X} \to \{0,1\}^v$ będzie $\frac{1+\gamma}{2^v}$-uniwersalną rodziną haszującą. Wtedy ekstraktor $\mathsf{ext}(x;h) = h(x)$, gdzie h jest wybierane jednostajnie z $\mathcal{H}$, jest (m,ϵ)-ekstraktorem, gdzie $\epsilon = \frac{1}{2}\sqrt{\gamma + \frac{1}{2^{m-v}}}$.*

Dowód tego twierdzenia znajduje się w [3]. LHL zastosowany wprost do iloczynu skalarnego daje następujące twierdzenie:

Twierdzenie 5. *Dla X i Y niezależnych zmiennych losowych na $\mathbb{F}^n$ takich, że Y jest jednostajnia oraz $\mathbf{H}_\infty(X) \geqslant m$ zzachodzi*

$$\Delta[(\langle X, Y\rangle, Y); (U_{\mathbb{F}}, Y)] \leqslant \frac{1}{2}\sqrt{\frac{1}{2^{m-\log|\mathbb{F}|}}}$$

Korzystając z faktu, że iloczyn skalarny jest elastycznym $(k, 2^{-\left[\frac{k-(n+4)\log|\mathbb{F}|}{3}-1\right]})-$ekstraktorem możemy pominąć założenie o jednostajnym rozkładzie Y otrzymując twierdzenie w ogólnej wersji:

Twierdzenie 6. *Dla dowolnych niezależnych zmiennych losowych X i Y na $\mathbb{F}^n$ takich, że $\mathbf{H}_\infty(Y) \geqslant k - m$ i dla dowolnej funkcji $f : \mathbb{F}^n \to \mathcal{G}$ takiej, że $P_{x \leftarrow f(U_{\mathbb{F}^n})}(\mathbf{H}_\infty(X|f(X) = x) \geqslant m) \geqslant 1 - \epsilon$. otrzymujemy:*

$$\Delta[(\langle X, Y\rangle, f(X), Y); (U_{\mathbb{F}}, f(X), Y)] \leqslant 2^{-\left[\frac{k-(n+4)\log|\mathbb{F}|}{3}-1\right]} + \epsilon$$

Następujący lemat podaje klasę funkcji spełniających powyższe założenia

Lemat 1. *Dla każdej zmiennej losowej X na $\mathcal{X}$ takiej, że $\mathbf{H}_\infty(X) = k$ i dla dowolnej $f : \mathcal{X} \to \{0,1\}^c$ zachodzi*

$$P_{y \leftarrow f(U_{\mathcal{X}})}\left(\mathbf{H}_\infty(X|f(X) = y) \leqslant m\right) \leqslant 2^{-k+c+m}.$$

3.2 Wycieki adaptywne

Twierdzenie 7. *Niech $\mathsf{ext} : \mathcal{X} \times \mathcal{X} \to \mathcal{G}$ będzie elastycznym $(k,\epsilon)-$ekstraktorem. Niech X i Y będą niezależnymi zmiennymi losowymi takimi, że $\mathbf{H}_\infty(X) = k_x$ i $\mathbf{H}_\infty(Y) = k_y$. Dla każdej adaptywnej sekwencji funkcji $f_i : \mathcal{X} \to \{0,1\}^{a_i}$ i $g_i : \mathcal{X} \to \{0,1\}^{b_i}$ (gdzie wybór $i-$tej funkcji może zależeć od wyników $i-1$ wcześniejszych f_i oraz g_i), takiej, że $\lambda_x + \lambda_y \leqslant \lambda$ gdzie λ jest parametrem zaś $\sum_i a_i = \lambda_x$ i $\sum_i b_i = \lambda_y$ zachodzi:*

$$d(\mathsf{ext}(X,Y)|\mathit{view}_{f,g}) \leqslant \epsilon + 2^{-k_x - k_y + k + \lambda}$$

gdzie $\mathit{view}_{f,g} = (f_1(X), g_1(Y), f_2(X), g_2(Y)...)$.

3.3 Odporne na wycieki schematy przechowywania danych w modelu split-state

W [18] wprowadzono pojęcie *odpornych na wycieki schematów przechowywania danych*, na potrzeby tej rozprawy generalizujemy niektóre pojęcia z tej pracy. Niech $\mathsf{ext} : \mathcal{X} \times \mathcal{X} \to \mathbb{F}$ będzie (k, ϵ)- elastycznym ekstraktorem. Niech L i R będą niezależnymi zmiennymi losowymi takimi, że $\mathbf{H}_\infty(X) + \mathbf{H}_\infty(Y) > k$. Zdefiniujmy $((k, \epsilon) - \mathsf{ext}, L, R)$-schemat jako parę funkcji $\mathsf{Enc} : \mathbb{F} \to \mathcal{X} \times \mathcal{X}$ i $\mathsf{Dec} : \mathcal{X} \times \mathcal{X} \to \mathbb{F}$ zdefiniowanych następująco:

$$\mathsf{Dec}(l, r) = \mathsf{ext}(l, r)$$
$$\mathsf{Enc}(m) = (l, r) \text{ takie, że } l \leftarrow L, r \leftarrow R \text{ i } \mathsf{ext}(l, r) = m.$$

Zdefiniujmy (λ, t)-*przeciwnika w model split-state* (w skrócie (λ, t)- SSM przeciwnika) podobnie jak w pracy [18]. Niech pamięć urządzenia będzie rozdzielona na dwie części L, R, t-krotnie powtarzamy nasępującą procedurę: dla $i = 1, 2, ..., t$ przeciwnik wybiera $f_i : \mathcal{X} \to \{0,1\}^{a_i}$ i $g_i : \mathcal{X} \to \{0,1\}^{b_i}$ i poznaje $f_i(L)$ oraz $g_i(R)$. Przeciwnik może wybrać dowolne a_i i b_i takie, że $\sum_i a_i + b_i < \lambda$. Wektor $(f_1(L), g_1(R), f_2(L), ..., g_t(R))$ wyników poznanych przez przeciwnika $\mathcal{A}$ oznaczamy jako $\mathsf{view}_{\mathcal{A}}(L, R)$.

Mówimy, że $((k, \epsilon) - \mathsf{ext}, L, R)$-schemat jest (λ, t, δ)-słabo bezpieczny w modelu split-state jeśli dla każdego (λ, t)-SSM przeciwnika $\mathcal{A}$ zachodzi:

$$d(\mathsf{ext}(L, R) | \mathsf{view}_{\mathcal{A}}(L, R)) \leqslant \delta.$$

Przywołajmy definicję z [18] i przepiszmy ją w języku modelu split-state. Mówimy, że schemat $(\mathsf{Enc}, \mathsf{Dec})$ jest (λ, t, δ)-bezpieczny jeśli dla każdych dwóch wiadomości m_0 i m_1 oraz dla każdego (λ, t)-SSM przeciwnika zachodzi:

$$\Delta\left(\mathsf{view}_{\mathcal{A}}(\mathsf{Enc}(m_0)); \mathsf{view}_{\mathcal{A}}(\mathsf{Enc}(m_1))\right) \leqslant \delta$$

Twierdzenie 8. *Jeśli $((k, \epsilon) - \mathsf{ext}, L, R)$-schemat jest (λ, t, δ)-słabo bezpieczny wtedy jest też $(\lambda, t, 4|\mathbb{F}| \cdot \delta)-$bezpieczny.*

W Twierdzeniu 7 udowodniliśmy, że $((k, \epsilon)-\mathsf{ext}, L, R)$- schemat jest $(\lambda, \infty, \epsilon + 2^{-\mathbf{H}_\infty(L) - \mathbf{H}_\infty(R) + k + \lambda})$- słabo bezpieczny stąd dzięki Twierdzeniu 8 otrzymujemy, że każdy schemat oparty na ekstraktorze elastycznym jest $(\lambda, \infty, 4|\mathbb{F}|(\epsilon + 2^{-\mathbf{H}_\infty(L) - \mathbf{H}_\infty(R) + k + \lambda}))$- bezpieczny.

4 Kody niekowalne

4.1 Definicje

W ten części podamy definicję kodu niekowalnego z [26]. Formalnie, niech $(\mathsf{Enc} : \mathcal{M} \to \mathcal{X}, \mathsf{Dec} : \mathcal{X} \to \mathcal{M} \cup \{\perp\})$ będzie schematem kodowania. Dla $F : \mathcal{X} \to \mathcal{X}$ i dla dowolnego

$m \in \mathcal{M}$ definiujemy eksperyment Tamper_m^F jako:

$$\mathsf{Tamper}_m^F = \left\{ \begin{array}{c} X \leftarrow \mathsf{Enc}(m), \\ X' := F(X), \\ m' := \mathsf{Dec}(X') \\ \text{output: } m' \end{array} \right\}$$

Niech $\mathcal{F}$ będzie rodziną funkcji z $\mathcal{X}$ w $\mathcal{X}$. Mówimy, że schemat kodowania $(\mathsf{Enc}, \mathsf{Dec})$ jest *ϵ-niekowalny względem rodziny $\mathcal{F}$* jeśli dla każdej funkcji $F \in \mathcal{F}$ istnieje rozkład D^F na $\mathcal{M} \cup \{\mathsf{same}^*, \perp\}$ taki, że dla każdej wiadomości $m \in \mathcal{M}$ mamy

$$\mathsf{Tamper}_m^F \approx_\epsilon \left\{ \begin{array}{c} d \leftarrow D^F \\ \text{if } d = \mathsf{same}^* \text{ then output } m \\ \text{otherwise output } d. \end{array} \right\}$$

Nieformalnie symbol "$\perp$" oznacza błąd odkodowania, to znaczy slowo kodowe było niepoprawne. Nasza konstrukcja nie wymaga jednak tego symbolu więc dalej będziemy go pomijać przyjmując, że $\mathsf{Dec} : \mathcal{X} \to \mathcal{M}$. Pokazujemy też własną, nieco prostszą definicję kodu niekowalnego w przypadku wiadomości $m \in \{0,1\}$.

Lemat 2. *Niech $\mathcal{M} = \{0,1\}$. Niech $\mathcal{F}$ będzie rodziną funkcji z $\mathcal{X}$ w $\mathcal{X}$. Schemat kodowania $(\mathsf{Enc} : \mathcal{M} \to \mathcal{X}, \mathsf{Dec} : \mathcal{X} \to \mathcal{M})$ jest ϵ-niekowalny względem rodziny $\mathcal{F}$ wtedy i tylko wtedy gdy dla każdej $F \in \mathcal{F}$ i $B \leftarrow \{0,1\}$ mamy*

$$P\left(\mathsf{Dec}\left(F(\mathsf{Enc}(B))\right) \neq B\right) \leqslant \frac{1}{2} + \epsilon. \tag{1}$$

Definicja (kody split-state). W rozprawie interesują nas kody split-state. Kodem *split-state* nazywamy parę $(\mathsf{Enc} : \mathcal{M} \to \mathcal{L} \times \mathcal{R}, \mathsf{Dec} : \mathcal{L} \times \mathcal{R} \to \mathcal{M})$. Mówimy, że taki kod jest *ϵ-niekowalny* jeśli jest ϵ-niekowalny względem rodziny *wszystkich* funkcji $\mathsf{Mall}^{f,g}$ zdefiniowanych jako $\mathsf{Mall}^{f,g}(L,R) = (f(L), g(R))$.

4.2 Konstrukcja kodu niekowalnego i jego własności.

Zaprezentuję teraz konstrukcję kodu niekowalnego.

Niech $\mathsf{ext} : \mathcal{L} \times \mathcal{R} \to \mathcal{C}$ będzie elastycznym (k, ϵ)-ekstraktorem, dla pewnych k i ϵ, oraz niech $c \in \mathcal{C}$ dowolne, wybrane z góry. Najpierw zdefiniujemy funkcję odkodowującą . Niech $\mathsf{D}_{\mathsf{ext}}^c : \mathcal{L} \times \mathcal{R} \to \{0,1\}$ będzie zdefiniowane następująco

$$\mathsf{D}_{\mathsf{ext}}^c(L,R) = \left\{ \begin{array}{ll} 0 & \text{if } \mathsf{ext}(X) = c \\ 1 & \text{w przeciwnym przypadku.} \end{array} \right.$$

Niech $\mathsf{E}_{\mathsf{ext}}^c : \{0,1\} \to \mathcal{L} \times \mathcal{R}$ będzie funkcją kodującą zdefiniowaną $\mathsf{E}_{\mathsf{ext}}^c(b) := (L,R)$, gdzie (L,R) jest parą wybraną jednostajnie ze zbioru $\{(L,R) : \mathsf{D}_{\mathsf{ext}}^c(L,R) = b\}$. Wymagamy

jeszcze jednego założenia o ext. Żądamy by dla $\tilde{L}$ i $\tilde{R}$ jednostajnych odpowiednio na $\mathcal{L}$ i $\mathcal{R}$ zachodziło, że $\mathsf{ext}(\tilde{L}, \tilde{R})$ jest idealnie jednostajny. Formalnie

$$\text{dla } \tilde{L} \leftarrow \mathcal{L} \text{ i } \tilde{R} \leftarrow \mathcal{R} \text{ mamy } d(\mathsf{ext}(\tilde{L}, \tilde{R})) = 0. \tag{2}$$

Łatwo zauważyć, że każdy ekstraktor można zmodyfikować tak, by spełniał (2). [1].

Twierdzenie 9. *Niech $\mathcal{L}'$ i $\mathcal{R}'$ będą podzbiorami odpowiednio $\mathcal{L}$ i $\mathcal{R}$. Niech $\mathsf{ext} : \mathcal{L}\times\mathcal{R} \to \mathcal{C}$ będzie elastycznym (k, ϵ)-ekstraktorem spełniającym (2), takim, że dla parametru δ mamy:*

$$k = \frac{2}{3} \cdot (\log|\mathcal{L}'| + \log|\mathcal{R}'|) - \frac{2}{3} \cdot \log(1/\delta).$$

Weźmy dowolne funkcje $f : \mathcal{L} \to \mathcal{L}$ i $g : \mathcal{R} \to \mathcal{R}$, niech B będzie wybrane jednostajnie z $\{0,1\}$ oraz niech $(L, R) \leftarrow \mathsf{E}^c_{\mathsf{ext}}(B)$. Wtedy

$$P\left(\mathsf{D}^c_{\mathsf{ext}}(f(L), g(R)) \neq B \mid (L, R) \in (\mathcal{L}', \mathcal{R}')\right) \leqslant \frac{1}{2} + |\mathcal{C}|^{-1} + 2|\mathcal{C}|^2\epsilon + \delta/(|\mathcal{C}|^{-1} - \epsilon),$$

co w szczególności oznacza, że $(\mathsf{E}^c_{\mathsf{ext}}, \mathsf{D}^c_{\mathsf{ext}})$ jest $\left(|\mathcal{C}|^{-1} + 2|\mathcal{C}|^2\epsilon + \delta/(|\mathcal{C}|^{-1} - \epsilon)\right)$-niekowalny.

4.3 Odporność na wycieki.

Nasza konstrukcja jest też odporna na wycieki informacji. Zacznijmy od zdefiniowana bezpieczeństwa w przypadku dodatkowego wycieku. Przyjmiemy naturalny pomysł by przeciwnik mógł poznać wyciek z L i R przed wyborem funkcji modyfikujących. Dla każdego schematu kodowania w modelu split-state $(\mathsf{E}^c_{\mathsf{ext}} : \mathcal{M} \to \mathcal{L}\times\mathcal{R}, \mathsf{D}^c_{\mathsf{ext}} : \mathcal{L}\times\mathcal{R} \to \mathcal{M})$, rodzimy funkcji $\mathcal{F}$, dowolnego $m \in \mathcal{M}$ i każdego przeciwnika $\mathcal{A}$ definiujemy grę $\mathsf{Tamper}^{\mathcal{A}}_m$ (gdzie λ jest pewnym parametrem) następująco. Najpierw, niech $(L, R) \leftarrow \mathsf{E}^c_{\mathsf{ext}}(m)$. Teraz przeciwnik $\mathcal{A}$ wybiera ciąg funkcji $(v^1, w^1, \ldots, v^t, w^t)$, gdzie każda z v^i jest typu $v^i : \mathcal{L} \to \{0,1\}^{\lambda_i}$ zaś każda z w^i jest typu $w^i : \mathcal{R} \to \{0,1\}^{\rho_i}$ gdzie każdy z parametrów λ_i i ρ_i jest taki, że

$$\lambda_1 + \cdots + \lambda_t + \rho_1 + \cdots \rho_t \leqslant \lambda. \tag{3}$$

Przeciwnik poznaje $\mathsf{Leak}(L, R) = (v^1(L), w^1(R), \ldots, v^t(L), w^t(R))$. Dopuszczamy by ten proces był *adaptywny*, t.j. przeciwnik wybiera i-tą funkcję w ciągu (3) po poznaniu $i-1$ wcześniejszych wartości $\mathsf{Leak}(L, R)$. Na końcu przeciwnik wybiera funkcje $f : \mathcal{L} \to \mathcal{L}$ oraz $g : \mathcal{R} \to \mathcal{R}$. Wynikiem gry jest:

$$\mathsf{Tamper}^{\mathcal{A}}_m := (f(L), g(R)).$$

[1] Iloczyn skalarny spełnia (2) jeśli założymy, że na przykład pierwsza współrzędna $\mathcal{L}$ oraz ostatnia współrzędna $\mathcal{R}$ są niezerowe. W ogólności, jesli $\mathsf{ext} : \mathcal{L}\times\mathcal{R} \to \mathcal{C}$ jest ekstraktorem, wtedy $\mathsf{ext}' : (\mathcal{L}\times\mathcal{C})\times\mathcal{R} \to \mathcal{C}$ zdefiniowany jako $\mathsf{ext}((C, L), R) = ext(L, R) + C$ (zakładając, że $(\mathcal{C}, +)$ jest grupą) spełnia (2).

Mówimy, że schemat kodowania $(\mathsf{E}^c_{\mathsf{ext}}, \mathsf{D}^c_{\mathsf{ext}})$ jest ϵ-*niekowalny z wyciekiem* λ jeśli dla każdego przeciwnika $\mathcal{A}$ istnieje rozkład $D^{\mathcal{A}}$ na $\mathcal{M} \cup \{\mathsf{same}^*\}$ taki, że dla każdej wiadomości $m \in \mathcal{M}$ zachodzi

$$\mathsf{Tamper}^{\mathcal{A}}_m \approx_\epsilon \left\{ \begin{array}{c} d \leftarrow D^{\mathcal{A}} \\ \text{if } d = \mathsf{same}^* \text{ then output } m, \\ \text{otherwise output } d. \end{array} \right\}$$

Twierdzenie 10. *Niech* $\mathsf{ext} : \mathcal{L} \times \mathcal{R} \to \mathcal{C}$ *będzie elastycznym* (k, ϵ)*-ekstraktorem spełniającym (2), niech parametry* δ *i* λ *spełniają*

$$k = \frac{2}{3} \cdot (\log |\mathcal{L}| + \log |\mathcal{R}| - \lambda) - \frac{4}{3} \cdot \log(1/\delta).$$

Wtedy schemat kodowania oparty na tym ekstraktorze jest $\Big(|\mathcal{C}|^{-1} + 2|\mathcal{C}|^2\epsilon + 2\delta/(|\mathcal{C}|^{-1} - \epsilon)\Big)$*-niekowalny przy wycieku* λ*.*

Używając Twierdzenia 10. do iloczynu skalarnego oraz korzystając z Twierdzenia 2. otrzymujemy następujące twierdzenie.

Twierdzenie 11. *Połóżmy dowolne* $\xi \in [0, 1/4)$ *oraz* $\gamma > 0$ *wtedy istnieje efektywny kod split-state* $(\mathsf{Enc} : \{0,1\} \to \{0,1\}^{N/2} \times \{0,1\}^{N/2}, \mathsf{Dec} : \{0,1\}^{N/2} \times \{0,1\}^{N/2} \to \{0,1\})$ *który jest* γ*-niekowalny z wyciekiem* $\lambda := \xi N$ *taki, że* $N = \mathcal{O}(\log(1/\gamma) \cdot (1/4 - \xi)^{-1})$*. Funkcja kodująca i odkodowująca są wyliczalne w czasie* $\mathcal{O}(N \cdot \log^2(\log(1/\gamma)))$ *zaś stała w notacji* $\mathcal{O}$*w tej formule jest około* 100*.*

W dowodzie tego twierdzenia podajemy jak dobrać parametry iloczynu skalarnego by otrzymac odpowiedni kod.

4.4 Odporność na ataki afiniczne.

Schemat kodowania można nieznacznie zmodyfikować i otrzymać odporność na ataki afiniczne w modelu z połączoną pamięcią. Okazuje się że nasz schemat $(\mathsf{E}^c_{\mathsf{ext}}, \mathsf{D}^c_{\mathsf{ext}})$, oparty na iloczynie skalarnym (w tym przypadku niestety konieczne są inne własności iloczynu skalarnego niż tylko elastyczność) w modelu w którym $(L, R) \in \mathbb{F}^n \times \mathbb{F}^n$ może być modyfikowane równocześnie (nie używamy modelu split-model, w którym L i R są atakowane osobno), jednakże wtedy trzeba ograniczyć rodzinę funkcji manipulujących. Zakładamy więc że ową rodziną będą funkcje afiniczne nad $\mathbb{F}^n$, tj. każda z funkcji manipulacji h jest postaci

$$h((L_1, \ldots, L_n), (R_1, \ldots, R_n)) = M \cdot (L_1, \ldots, L_n, R_1, \ldots, R_n)^T + V^T,$$

gdzie M jest macierzą rozmiaru $(2n \times 2n)$ o wyrazach w $\mathbb{F}$ oraz $V \in \mathbb{F}^{2n}$.

References

[1] A. Akavia, S. Goldwasser, and V. Vaikuntanathan. Simultaneous hardcore bits and cryptography against memory attacks. *TCC*, pages 474–495, 2009.

[2] R. Anderson and M. Kuhn. Tamper resistance - a cautionary note. In *The Second USENIX Workshop on Electronic Commerce Proceedings*, November 1996.

[3] B. Barak, Y. Dodis, H. Krawczyk, O. Pereira, K. Pietrzak, F.-X. Standaert, and Y. Yu. Leftover hash lemma, revisited. 2011. `http://eprint.iacr.org/`.

[4] M. Bellare and T. Kohno. A theoretical treatment of related-key attacks: Rka-prps, rka-prfs, and applications. *EUROCRYPT 2003*, pages 647–647, 2003.

[5] E. Biham. New types of cryptanalytic attacks using related keys. *Journal of Cryptology*, 7(4):229–246, 1994.

[6] M. Blum. Independent unbiased coin flips from a correlated biased source: a finite state markov chain. *25th Annual Symposium on Foundations of Computer Science*, pages 425–433.

[7] J. Bourgain. More on the sum-product phenomenon in prime fields and its applications. *International Journal of Number Theory*, 1(01):1–32, 2005.

[8] C.E.Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27, 1948.

[9] H. Chabanne, G. Cohen, J. Flori, and A. Patey. Non-malleable codes from the wire-tap channel. In *Information Theory Workshop (ITW), 2011 IEEE*, pages 55–59. IEEE, 2011.

[10] H. Chabanne, G. Cohen, and A. Patey. Secure network coding and non-malleable codes: Protection against linear tampering. In *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on*, pages 2546–2550, 2012.

[11] S. Choi, A. Kiayias, and T. Malkin. Bitr: built-in tamper resilience. *ASIACRYPT 2011*, pages 740–758, 2011.

[12] B. Chor and O. Goldreich. Unbiased bits from sources of weak randomness and probabilistic communication complexity. *SIAM Journal on Computing*, 17(2):230–261, 1988.

[13] B. Chor, O. Goldreich, J. Hastad, J. Friedman, S. Rudich, and R. Smolensky. The bit extraction problem or t-resilient functions. *26th Annual Symposium on Foundations of Computer Science*, 1985.

[14] S.-Y. Chung, G. D. F. Jr., T. J. Richardson, and R. Urbanke. On the design of low-density parity-check codes within 0.0045 db of the shannon limit. *IEEE Communications Letters*, 5.

[15] A. Cohen and A. Wigderson. Dispersers, deterministic amplification, and weak random sources. *30th Annual IEEE Symposium on Foundations of Computer Science*, 1989.

[16] G. Cohen, R. Raz, and G. Segev. Non-malleable extractors with short seeds and applications to privacy amplification. In *Computational Complexity (CCC)*, pages 298–308, 2012.

[17] D. Dachman-Soled and Y. Kalai. Securing circuits against constant-rate tampering. *CRYPTO 2012*, pages 533–551, 2012.

[18] F. Davì, S. Dziembowski, and D. Venturi. Leakage-resilient storage. *Security and Cryptography for Networks*, pages 121–137, 2010.

[19] Y. Dodis, Y. T. Kalai, and S. Lovett. On cryptography with auxiliary input. pages 621–630, 2009.

[20] Y. Dodis, X. Li, T. Wooley, and D. Zuckerman. Privacy amplification and non-malleable extractors via character sums. In *FOCS 2011*, pages 668–677, 2011.

[21] Y. Dodis and D. Wichs. Non-malleable extractors and symmetric key cryptography from weak secrets. In *STOC*, pages 601–610, 2009.

[22] D. Dolev, C. Dwork, and M. Naor. Nonmalleable cryptography. *SIAM review*, 45(4):727–784, 2003.

[23] S. Dziembowski and S. Faust. Leakage-resilient circuits without computational assumptions. *TCC*, pages 230–247, 2012.

[24] S. Dziembowski and K. Pietrzak. Intrusion-resilient secret sharing. In *FOCS'07*, pages 227–237.

[25] S. Dziembowski and K. Pietrzak. Leakage-resilient cryptography. In *FOCS'08*, pages 293–302. IEEE.

[26] S. Dziembowski, K. Pietrzak, and D. Wichs. Non-malleable codes. *ICS*, pages 434–452, 2010.

[27] S. Faust, T. Rabin, L. Reyzin, E. Tromer, and V. Vaikuntanathan. Protecting circuits from leakage: the computationally-bounded and noisy cases. pages 135–156, 2010.

[28] R. Gennaro, A. Lysyanskaya, T. Malkin, S. Micali, and T. Rabin. Algorithmic tamper-proof (atp) security: Theoretical foundations for security against hardware tampering. *TCC*, pages 258–277, 2004.

[29] O. Goldreich. Modern cryptography, probabilistic proofs and pseudorandomness. *Algorithms and Combinatorics*, 1998.

[30] S. Goldwasser and G. Rothblum. How to compute in the presence of leakage, 2012. accepted to FOCS 2012.

[31] S. Halevi and H. Lin. After-the-fact leakage in public-key encryption. *TCC*, pages 107–124, 2011.

[32] J. HÅstad, R. Impagliazzo, L. Levin, and M. Luby. A pseudorandom generator from any one-way function. *SIAM Journal on Computing*, 28(4):1364–1396, 1999.

[33] T. Holenstein. Pseudorandom generators from one-way functions: A simple construction for any hardness. In *TCC*, pages 443–461, 2006.

[34] Y. Ishai, M. Prabhakaran, A. Sahai, and D. Wagner. Private circuits ii: Keeping secrets in tamperable circuits. *EUROCRYPT*, pages 308–327, 2006.

[35] Y. Ishai, A. Sahai, and D. Wagner. Private circuits: Securing hardware against probing attacks. *CRYPTO*, pages 463–481, 2003.

[36] Y. Kalai, B. Kanukurthi, and A. Sahai. Cryptography with tamperable and leaky memory. *CRYPTO 2011*, pages 373–390, 2011.

[37] F. Liu and A. Lysyanskaya. Tamper and leakage resilience in the split-state model. *CRYPTO 2012*, pages 517–532, 2012.

[38] S. Micali and L. Reyzin. Physically observable cryptography. *TCC*, pages 278–296, 2004.

[39] R. Motwani and P. Raghavan. Randoomized algorithms. *Cambridge University press*, 1995.

[40] M. Naor and G. Segev. Public-key cryptosystems resilient to key leakage. *CRYPTO 2009*, pages 18–35, 2009.

[41] E. N. of Excellence (ECRYPT). Side channel cryptanalysis lounge. `http://www.emsec.rub.de/research/projects/sclounge`.

[42] L. Pontryagin and R. Gamkrelidze. *Topological Groups*. Gordon and Breach Science Publishers.

[43] A. Rao. An exposition of bourgain 2-source extractor. In *Electronic Colloquium on Computational Complexity (ECCC)*, volume 14, page 034, 2007.

[44] R.Dorfman. The detectation of defective members of large population. *Ann.Math.Statist.*, 1943.

[45] M. Santha and U. Vizirani. Generating quasi-random sequences from semi-random sources. *Journal of Computer and System Science*, 1986.

[46] R. Shaltiel. Recent developments in explicit constructions of extractors. *Bulletin of the EATCS*, 77:67–95, 2002.

[47] H. Wee. Public key encryption against related key attacks. *PKC 2012*, pages 262–279, 2012.

[48] H. Yamamoto. Rate-distortion theory for the shannon cipher system. *IEEE Transactions on Information Theory*, 43.