ELASTYCZNE EKSTRAKTORY DWUŹRÓDŁOWE I ICH ZASTOSOWANIA

Maciej Obremski

Wydział Matematyki, Informatyki i Mechaniki,
Uniwersytet Warszawski

Streszczenie. Prezentujemy nowe pojęcie elastycznego ekstraktora dwu-źródłowego. Prezentujemy cały wachlarz metod i twierdzeń uzupełniających wiedzę o przypadkach nierozpatrywanych przez lemat Lidsey'a i Leftover Hash Lemma. Pokazujemy analog twierdzenia Barak'a o silnych i słabych ekstraktorach dla przypadku elastycznego. Na przykładzie odpornych na wycieki schematów składowania danych prezentujemy przykład zastosowania elastycznych ekstraktorów dwu-źródłowych. Otrzymujemy w ten sposób lepsze parametry niż w przypadku standardowej ekstrakcji oraz możliwość prowadzenia adaptywnych wycieków.

Słowa kluczowe: elastyczne ekstraktory dwu-źródłowe, ekstraktory dwu-źródłowe, wycieki, odporne na wycieki schematy składowania danych, twierdzenie Baraka.

1. Wstęp

Prezentujemy nowe pojęcie *elastycznego* extraktora dwuźródłowego. Pojęcie to pojawiło się po raz pierwszy we wspólnej pracy ze Stefanem Dziembowskim i Tomaszem Kazaną „Non-Malleable Codes from Two-Source Extractors", opublikowanej na konferencji CRYPTO 2013. 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.

Ekstraktory Losowości są kluczowym komponentem wielu schematów kryptograficznych na przykład kodów niekowalnych [1, 2, 3, 6] czy schematów składowania danych odpornych na wycieki [5]. Dotychczas stosowane metody tj. standardowa dwu-źródłowa ekstrakcja i Leftover Hash Lemma niestety miały istotną wadę, mianowicie gdy chcemy użyć ekstraktora do składowania danych zazwyczaj bierzemy dane m i losujemy dwa wektory L, R, takie że $\mathrm{ext}(L, R) = m$, następnie składujemy L, R na różnych serwerach. Żeby w tej sytuacji stosować Leftover Hash Lemma musimy założyć,

że L zostało wylosowane idealnie jednostajnie. Jeśli nie mogliśmy wylosować L jednostajnie musimy uciec się do standardowej ekstrakcji, która daje nam drastycznie gorsze parametry. Ten przeskok jakościowy parametrów LHL względem standardowej ekstrakcji zainspirował nas do poszukiwania rozwiązania pośredniego łączącego oba te przypadki.

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:

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

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

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

Definicja (**silny ekstraktor**). Mówimy, że $\mathrm{ext} : \mathcal{X}\times\mathcal{Y} \to \mathcal{Z}$ jest silnym (k,ϵ)-dwuźródłowym ekstraktorem, jeśli dla niezależnych zmiennych losowych X na $\mathcal{X}$ i Y na $\mathcal{Y}$ takich, że $\mathbf{H}_\infty(X) \geq k$ i $\mathbf{H}_\infty(Y) \geq k$, zachodzi $\Delta(\mathrm{ext}(X,Y),X);(U_\mathcal{Z},X) \leq \epsilon$. 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 [8] (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 $\mathrm{ext} : \mathcal{L} \times \mathcal{R} \to \mathcal{C}$ nazwiemy *słabym elastycznym* (k, ϵ)*-2-źródłowym ekstraktorem* jeśli dla każdego $L \in \mathcal{L}$ oraz $R \in \mathcal{R}$ takich, że $\mathbf{H}_\infty(L) + \mathbf{H}_\infty(R) \geq k$ zachodzi $d(\mathrm{ext}(L, R)) \leq \epsilon$.

Definicja (**silny elastyczny ekstraktor**). Funkcję $\mathrm{ext} : \mathcal{L} \times \mathcal{R} \to \mathcal{C}$ nazwiemy *silnym elastycznym* (k, ϵ)*-2-źródłowym ekstraktorem* jeśli dla każdego $L \in \mathcal{L}$ oraz $R \in \mathcal{R}$ takich, że $\mathbf{H}_\infty(L) + \mathbf{H}_\infty(R) \geq k$ zachodzi $d(\mathrm{ext}(L, R)|L) \leq \epsilon$ i $d(\mathrm{ext}(L, R)|R) \leq \epsilon$.
Dwie ostatnie definicje są nowymi pojęciami wprowadzonymi przez nas.

Twierdzenie 1. *Niech* $\mathrm{ext} : (\{0,1\}^N)^2 \to \{0,1\}^M$ *będzie słabym elastycznym* (K, ϵ)*-ekstraktorem, dla* $K \geq N$*. Wtedy dla każdego* $K' \geq K$ *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* $\mathrm{ext} : \mathbb{F}^n \times \mathbb{F}^n \to \mathbb{F}$ *zdefiniowany jako* $\mathrm{ext}^n_{\mathbb{F}}(L, R) = \langle L, R \rangle$ *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 [6] jak i w innych pracach o zbliżonej tematyce [1, 3]. Własność elastyczności w przypadku iloczynu skalarnego uogólnia wynik Leftover Hash Lemma [4].
Kolejne twierdzenie dotyczy innego znanego ekstraktora wprowadzonego przez Holensteina w [7]. 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* $\mathrm{ext} : GF(2^n) \times GF(2^n) \to GF(2^\lambda)$*, zdefiniowany jako* $\mathrm{ext}(X, Y) = (X \cdot Y)_\lambda$ *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 silnym 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 $P_{h \leftarrow \mathcal{H}}[h(x_1) = h(x_2)] \leq p$. 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* $\mathrm{ext}(x; h) = h(x)$*, 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 [4]. 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* $\mathbf{H}_\infty(X) \geq m$ *zachodzi*

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

Korzystając z faktu, że iloczyn skalarny jest silnym 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* $\mathbf{H}_\infty(Y) \geq k - m$ *i dla dowolnej funkcji* $f : \mathbb{F}^n \to \mathcal{G}$ *takiej, że* $P_{x \leftarrow f(U_{\mathbb{F}^n})}(\mathbf{H}_\infty(X|f(X) = x) \geq m) \geq 1 - \epsilon$*. Otrzymujemy:*

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

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*

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

3.2. Wycieki adaptywne

Twierdzenie 7. *Niech* ext $: \mathcal{X} \times \mathcal{X} \to \mathcal{G}$ *będzie silnym elastycznym (k, ϵ)-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 $\lambda_x + \lambda_y \leq \lambda$ gdzie λ jest parametrem zaś $\sum_i a_i = \lambda_x$ i $\sum_i b_i = \lambda_y$ zachodzi:*

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

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

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

W [5] wprowadzono pojęcie *odpornych na wycieki schematów przechowywania danych*, na potrzeby tej rozprawy generalizujemy niektóre pojęcia z tej pracy. Niech ext $: \mathcal{X} \times \mathcal{X} \to \mathbb{F}$ będzie silnym (k, ϵ)- elastycznym ekstraktorem. Niech L i R będą niezależnymi zmiennymi losowymi takimi, że $\mathbf{H}_\infty(X) + \mathbf{H}_\infty(Y) > k$. Zdefiniujmy $((k, \epsilon) - \text{ext}, L, R)$-schemat jako parę funkcji $\text{Enc} : \mathbb{F} \to \mathcal{X} \times \mathcal{X}$ i $\text{Dec} : \mathcal{X} \times \mathcal{X} \to \mathbb{F}$ zdefiniowanych następująco:

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

Zdefiniujmy (λ, t)*-przeciwnika w model split-state* (w skrócie (λ, t)-SSM przeciwnika) podobnie jak w pracy [5]. Niech pamięć urządzenia będzie rozdzielona na dwie części L, R, t-krotnie powtarzamy nasępującą procedurę: dla $i = 1, 2, \ldots, t$ 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 $(f_1(L), g_1(R), f_2(L), \ldots, g_t(R))$ wyników poznanych przez przeciwnika $\mathcal{A}$ oznaczamy jako $\text{view}_\mathcal{A}(L, R)$.

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

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

Przywołajmy definicję z [5] i przepiszmy ją w języku modelu split-state. Mówimy, że schemat (Enc, Dec) jest (λ, t, δ)-bezpieczny jeśli dla każdych dwóch wiadomości m_0 i m_1 oraz dla każdego (λ, t)-SSM przeciwnika zachodzi:

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

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

W Twierdzeniu 7 udowodniliśmy, że $((k,\epsilon)-\text{ext}, L, R)$-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.

Literatura

[1] D. Aggarwal, Y. Dodis, T. Kazan, and M. Obremski, *Reductions and their applications*, unpublished manuscript, 2014.

[2] D. Aggarwal, Y. Dodis, and S. Lovett, *Non-malleable codes from additive combinatorics*, FOCS, 2014.

[3] D. Aggarwal, S. Dziembowski, T. Kazan, and M. Obremski, *Interactive non-malleable codes*, unpublished manuscript, 2014.

[4] 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/.

[5] F. Daví, S. Dziembowski, and D. Venturi, *Leakage-resilient storage*, Security and Cryptography for Networks, 2010, p. 121–137.

[6] S. Dziembowski, T. Kazana, and M. Obremski, *Non-malleable codes from two-source extractors*, CRYPTO, 2013.

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

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

FLEXIBLE TWO-SOURCE EXTRACTORS AND APPLICATIONS

Abstract. We introduce new notion – flexible two-source extractor. We show various methods and theorems that fill in cases not covered by Lidsey's lemma and Leftover Hash Lemma. We present flexible version of Barak's theorem for strong and weak flexible extractors. We also show possible application of flexible extractors to leakage resilient storage schemes. We are able to obtain better parameters then in standard extraction case, moreover we are able to obtain resilience to adaptive leakages.

Keywords: flexible two-source extractors, two-source extractors, leakage, leakage resilient storage, Barak's theorem.