NIEKOWALNE EKSTRAKTORY LOSOWOŚCI

Konrad Durnoga

Instytut Matematyki; Wydział Matematyki, Informatyki i Mechaniki
Uniwersytetu Warszawskiego

Streszczenie. Ekstraktory losowości należą do jednego z głównych nurtów badań współczesnej kryptografii teoretycznej. Zadaniem tych deterministycznych funkcji jest przekształcenie źródeł słabej losowości w takie, których rozkład jest bliski rozkładowi jednostajnemu. W pracy przedstawiona jest teorioliczbowa konstrukcja ekstraktora o pewnych szczególnych własnościach – ekstraktora niekowalnego. Wynik ten stanowi udoskonalenie warunkowego rezultatu Y. Dodisa *i in.* opublikowanego na prestiżowej konferencji FOCS'11.
Słowa kluczowe: ekstraktor losowości, ekstraktor niekowalny, logarytm dyskretny

1. Ekstraktory losowości

We współczesnej informatyce ważne miejsce zajmują algorytmy randomizowane. W wymiarze praktycznym istotną kwestią staje się zapewnienie tym algorytmom źródła losowych bitów „wysokiej jakości". Jest to kluczowy problem zwłaszcza w przypadku zastosowań kryptograficznych, gdzie na założeniu jednostajnej losowości różnych elementów kryptosystemu może opierać się jego bezpieczeństwo. Fizyczne źródła losowości, powszechnie dostępne w komputerach osobistych, mogą nie zapewniać wystarczających statystycznych własności generowanego strumienia bitów. Tu pojawia się potrzeba konstrukcji *ekstraktorów losowości* – deterministycznych funkcji przekształcających niedoskonałe źródła losowości na takie, które są w statystycznym sensie bliskie rozkładom jednostajnym.

Rozwój dziedziny związanej z poprawianiem statystycznych własności rozkładów dyskretnych dokonał się w zasadzie w całości na przestrzeni ostatniego ćwierćwiecza, choć jej początki sięgają lat 50-tych ubiegłego stulecia i pionierskich prac von Neumanna o symulowaniu rzutów symetryczną monetą przy użyciu monety, na której orzeł wypada ze stałym prawdopodobieństwem $\neq 1/2$. Od tego czasu problematyka ta była przedmiotem badań czołowych naukowców, czego owocem są liczne prace na temat ekstraktorów losowości. Sam termin *ekstraktor* został zaproponowany przez Nisana i Zuckermana [23]. Za fundamentalne uznaje się prace Chora i Goldreicha [7], Cohena i Widgersona [8] oraz Zuckermana [27]. Jedne z najwcześniejszych konstrukcji ekstraktorów pochodzą od Chora i Goldreicha [7] oraz Impagliazzo *i in.* [19]. Przełomowym osiągnięciem ostatniej dekady był wynik Bourgain [4], laureata medalu Fieldsa, wskazujący istnienie

ekstraktorów nawet dla źródeł losowości o niskim współczynniku entropii. Rozprawa w znaczącej części zajmuje się analizą aspektów obliczeniowych klasycznego już dziś przykładu ekstraktora Chora-Goldreicha.

Ekstraktor losowości formalnie definiuje się w terminach entropii oraz statystycznej odległości rozkładów prawdopodobieństwa. Wielkością mierzącą stopień losowości dyskretnej zmiennej losowej X jest, znana z teorii informacji, tzw. *min-entropia* $\mathbf{H}_\infty(X)$, którą określamy jako:

$$\mathbf{H}_\infty(X) := \min_x \log_2 \frac{1}{\Pr(X = x)},$$

gdzie x przebiega przez wszystkie elementy zbioru nośnika zmiennej X. Spośród wszystkich rozkładów prawdopodobieństwa na n bitach min-entropia przyjmuje wartość maksymalną dla rozkładu jednostajnego – $\mathbf{H}_\infty(U_{\{0,1\}^n}) = n$. Do określenia odległości dwóch zmiennych losowych X i X' o rozkładzie na zbiorze $\mathcal{X}$ używamy standardowej definicji dystansu statystycznego $\Delta(X, X')$:

$$\begin{aligned}\Delta(X, X') &:= \max_{\mathcal{S}\subseteq\mathcal{X}} \mid \Pr(X \in \mathcal{S}) - \Pr(X' \in \mathcal{S}) \mid \\ &= \frac{1}{2} \sum_{x\in\mathcal{X}} \mid \Pr(X = x) - \Pr(X' = x) \mid\end{aligned}$$

. Zapis $X \approx_\epsilon X'$ oznacza $\Delta(X, X') \leq \epsilon$ dla pewnego $\epsilon \geq 0$. Przy tym typowo wymaga się, by ϵ był *zaniedbywalny* jako funkcja pewnego parametru n, tzn. $\epsilon = \epsilon(n)$ powinien dążyć do 0 szybciej niż odwrotność dowolnego wielomianu w punkcie $n \to \infty$.

Funkcję $\mathrm{Ext}\colon \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ nazywamy (k, ϵ)*–niekowalnym ekstraktorem*, jeśli dla każdej pary niezależnych zmiennych losowych X i Y nad zbiorami, odpowiednio, $\mathcal{X}$ oraz $\mathcal{Y}$, i dowolnej funkcji $A\colon \mathcal{Y} \to \mathcal{Y}$ spełniającej $A(y) \neq y$ dla wszystkich $y \in \mathcal{Y}$, zachodzi:

$$\big(\mathrm{Ext}(X, Y), \mathrm{Ext}(X, A(Y)), Y\big) \approx_\epsilon \big(U_\mathcal{Z}, \mathrm{Ext}(X, A(Y)), Y\big), \qquad (1)$$

o ile $\mathbf{H}_\infty(X) \geq k$ oraz Y jest rozkładem jednostajnym nad $\mathcal{Y}$.

Intuicyjnie, powyższa definicja oznacza, że jeśli argument x wybrany został z dostatecznie losowego rozkładu to żaden, w tym nawet nieograniczony obliczeniowo, adwersarz nie potrafi odróżnić $\mathrm{Ext}(x, y)$ od wartości losowej przy znanym losowym ziarnie y oraz $\mathrm{Ext}\big(x, A(y)\big)$, gdzie A jest dowolnie ustaloną (a priori) przez adwersarza funkcją nie posiadającą punktów stałych. Jest to również znaczące wzmocnienie klasycznej definicji *ekstraktora*, w której warunek (1) można zastąpić przez $\mathrm{Ext}(X, Y) \approx_\epsilon U_\mathcal{Z}$, oraz *silnego ekstraktora*, gdzie przyjmuje on postać: $\big(\mathrm{Ext}(X, Y), Y\big) \approx_\epsilon \big(U_\mathcal{Z}, Y\big)$.

Pojęcie *ekstraktora niekowalnego* wprowadzone zostało przez Dodisa i Wichsa [13], którzy wskazali, że hipotetyczna funkcja o takich własnościach może posłużyć do budowy protokołu tzw. *wzmacniania prywatności* (ang. *privacy amplification*). Ten drugi termin zaproponowany został wcześniej w pracy Bennetta, Brassarda i Roberta [3]. Postawili oni następujący problem: dwóch użytkowników posiada te same dane, które nie są idealnie losowe z punktu widzenia adwersarza; czy istnieje protokół komunikacji między tymi użytkownikami, który umożliwi wzmocnienie prywatności tych danych tak, by po jego wykonaniu nowe dane był nieodróżnialne od losowych? Przy tym przyjmuje się założenie, że kanał komunikacyjny pozostaje niezabezpieczony i adwersarz ma dostęp do przesyłanych informacji oraz, dodatkowom posiada nieograniczoną moc obliczeniową. Mając do dyspozycji silny ekstraktor losowości można skonstruować nieskomplikowany, jednorundowy (tzn. wymagający wysłania tylko jednej wiadomości) protokół realizujący to wymaganie, w sytuacji gdy adwersarz pozostaje pasywny, czyli może poznać treść przesyłanych komunikatów, ale nie może w nie ingerować. Zadanie zaprojektowania bezpiecznego protokołu w przypadku istnienia aktywnego adwersarza, który może dowolnie modyfikować wiadomości, staje się dużo trudniejsze. W literaturze znanych było kilka rozwiązań (np. Maurera i Wolfa [22] czy Dodisa i Wichsa [12]), które jednak wymagają dodatkowych założeń co do poziomu entropii wejściowych danych (czyli wiedzy *a priori* adwersarza o danych) lub są nieoptymalne z punktu widzenia liczby rund protokołu bądź współczynnika utraty entropii między wyjściowymi a wyjściowymi danymi. Wiadomo również, że dla danych o niskiej entropii jednorundowy protokół wzmacniania prywatności przy aktywnym adwersarzu nie może istnieć. Dodis i Wichs [13] jako pierwsi podali optymalny dwurundowy protokół dla tej wersji problemu. Jednocześnie przedstawili oni argument probabilistyczny dowodzący istnienia ekstraktorów niekowalnych dla szerokiego zakresu parametrów k oraz ϵ. Jednak problemem otwartym pozostawała kwestia podania jawnego przykładu konstrukcji tego typu.

2. Ekstraktor Chora-Goldreicha

W literaturze poświęconej teorii ekstraktorów odnaleźć można bogactwo rozmaitych metod – technik analizy fourierowskiej, teorii kodów, kombinatoryki czy teorii liczb. Głębokie rezultaty z tej ostatniej pozwoliły m.in. na uzyskanie przez Bourgaina [4] wspomnianego wyżej ekstraktora dla źródeł o niskiej min-entropii. Teorioliczbowy charakter ma również konstrukcja Chora i Goldreicha [7], oparta na logarytmie dyskretnym, którą krótko przytoczymy poniżej.

Niech p będzie nieparzystą liczbą pierwszą, a g generatorem cyklicznej grupy multiplikatywnej $(\mathbb{Z}/p\mathbb{Z})^*$ ciała $(\mathbb{Z}/pZ)$. Ponadto, niech $M > 1$ oznacza dowolnie ustalony dzielnik rzędu tej grupy, czyli $p-1$. Podstawą ekstraktora Chora-Goldreicha jest następująca funkcja

$$f_g(a) := \log_g a \bmod M, \tag{2}$$

wyznaczona przez logarytm dyskretny przy podstawie g w $(\mathbb{Z}/p\mathbb{Z})^*$, dodatkowo zredukowany modulo M.

Dodis *i in.* [10] wykazali, używając oszacowań Weila dla sum charakterów multiplikatywnych nad ciałem skończonym (zob. np. monografię Schmidta [25]), że f_g zadaje ekstraktor niekowalny.

Twierdzenie 1. (Dodis *i in.* [10], twierdzenie 4.1) *Niech* $\mathrm{Ext}\colon \mathbb{Z}/p\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z} \to \mathbb{Z}/M\mathbb{Z}$ *będzie określona wzorem*

$$\mathrm{Ext}(x,y) := f_g(x+y) = \log_g (x+y) \bmod M\ . \tag{3}$$

Wtedy, dla dowolnego k*, funkcja* Ext *jest* (k,ϵ)*–niekowalnym ekstraktorem przy* $\epsilon = 2Mp^{1/4}2^{-k/2}$.

Twierdzenie to stanowiło rozszerzenie rezultatu z oryginalnej pracy Chora i Goldreicha [7]. Zawarta tam analiza w istocie implikowała, że (3) jest ekstraktorem losowości, choć w momencie ukazania się tego artykułu pojęcie ekstraktora nie było jeszcze znane w literaturze. Dodis i Oliveira [11] wskazali, że ta funkcja spełnia warunki definicji silnego ekstraktora.

Praca Durnogi i Źrałka [15] poświęcona jest zagadnieniu efektywnego obliczania wartości ekstraktora (3) lub, równoważnie, funkcji f_g danej wzorem (2). Należy tu zwrócić uwagę, że, inaczej niż ma to miejsce w typowych zastosowaniach kryptograficznych opierających swoje bezpieczeństwo na założeniu wysokiej złożoności obliczeniowej problemu logarytmu dyskretnego, w tym przypadku wymagamy, by problem ten był „łatwo" rozwiązywalny. Użycie standardowej metody do wyznaczania logarytmu dyskretnego, czyli algorytmu Pohliga-Hellmana [24], z niewielką modyfikacją polegającą na obliczaniu $\log_g z$ modulo każdy dzielnik pierwszy $q \mid M$, pozwala na znalezienie wartości funkcji $f_g(z)$ w czasie proporcjonalnym do $\mathrm{P}^+(M)$ – największego dzielnika pierwszego M. Jest to czas wielomianowy wyłącznie pod warunkiem, że M jest liczbą *gładką*, tzn. ma jedynie małe dzielniki pierwsze. W tym kontekście kluczowego znaczenia nabiera kwestia wydajnego generowania liczb p oraz $M \mid p-1$ z dodatkowym wymaganiem gładkości M.

Dodis *i in.* [10] sugerują następującą procedurę znajdowania p oraz M. Najpierw ustalmy liczbę gładką M, np. wybierając odpowiednio dużą

potęgę 2, dla której $\log_2 M$ odpowiada w przybliżeniu oczekiwanej liczbie bitów wyjścia ekstraktora. Następnie przeglądamy kolejne wyrazy postępu arytmetycznego $\equiv 1 \pmod{M}$, tzn. $M+1$, $2M+1$, $3M+1$, etc., w poszukiwaniu liczby pierwszej p. Sprawdzenie czy dana liczba jest pierwsza może być zrealizowane za pomocą wielomianowego deterministycznego testu pierwszości [20]. Cała procedura jest również efektywna o ile najmniejsze p w tym postępie nie jest zbyt duże. W przypadku ekstraktorów o tzw. krótkich wyjściach, czyli dla niewielkich, na przykład stałych, wartości M istnienie takiego p wynika z oszacowań dla stałej Linnika. Najlepsze znane obecnie ograniczenie bezwarunkowe otrzymane przez Xylourisa [26] zapewnia znalezienie $p = O(M^5)$. Natomiast w przypadku ogólnym, dla dużych M, konieczne jest odwołanie się do następującej, powszechnie uważanej za prawdziwą (zob. artykuł Granville'a i Pomerance'a [17]), hipotezy.

Hipoteza 2. *Dla dowolnych a oraz M takich, że* $\mathrm{NWD}(a, M) = 1$, *najmniejsza liczba pierwsza* $p \equiv a \bmod M$ *spełnia* $p = O(\varphi(M) \ln^2 M)$, *gdzie φ oznacza funkcję Eulera.*

Konstrukcja Dodisa *i in.* [10] jest tym samym wynikiem warunkowym. Przytoczona wyżej procedura zawiera jednak, jak zostało to zauważone przez Durnogę i Źrałka [15], pewną usterkę. Mianowicie nie gwarantuje ona zachowania odpowiednich proporcji między liczbami M i p. W szczególności dla wartości M bliskich p z twierdzenia 1 otrzymujemy błąd ϵ, który nie jest zaniedbywalny, przez co samo twierdzenie traci zupełnie siłę wyrazu. Prosta modyfikacja metody generowania M i p pozwala na uniknięcie tego problemu, jednak za cenę wprowadzenia zależności od hipotezy 2 również w przypadku ekstraktorów o krótkich wyjściach.

Ostatnim parametrem ekstraktora (3), którego wybór wymaga osobnego komentarza jest podstawa logarytmu dyskretnego g. Autorzy oryginalnej pracy [10] nie specyfikują jednak żadnego sposobu konstrukcji takiego g. Istnienie tej luki (również zidentyfikowanej w pracy Durnogi i Źrałka [15]) zostało potwierdzone przez autorów (komunikacja prywatna). W istocie kwestia efektywnego znajdowania generatora grupy multiplikatywnej ciała $(\mathbb{Z}/pZ)$ w ogólnym przypadku pozostaje ważnym problemem otwartym. Naturalnie istnieje szybki, niedeterministyczny algorytm wyznaczający pewien generator, ale jedynie przy znanej pełnej faktoryzacji liczby $p-1$ niezbędnej do weryfikacji czy dany losowy element jest pełnego rzędu. W świetle wyniku Ankeny'ego [2] ten proces może być zderandomizowany przy założeniu uogólnionej Hipotezy Riemanna (ERH). Użycie hipotezy 2 implikuje, że dla znalezionej liczby p rząd grupy $(\mathbb{Z}/p\mathbb{Z})^*$, czyli $p-1$, jest gładki. Deterministyczny algorytm wyznaczania g wymaga jednak dodatkowo zastosowania Hipotezy Riemanna. W rozprawie uzasadniamy, że

zakładając jedynie prawdziwość ERH, ale bez odwoływania się do hipotezy 2, możliwe jest wydajne generowanie odpowiednich wartości p, M oraz g w sposób randomizowany.

W chwili obecnej znanych jest kilka innych przykładów ekstraktora niekowalnego [9, 21]. Są to konstrukcje bezwarunkowe pozwalające na osiągnięcie lepszych parametrów (np. oszacowania na wyraz błędu ϵ czy słabszego warunku na poziom min-entropii źródła k) niż ekstraktor Chora-Goldreicha i używające metod spoza teorii liczb.

Poniżej przedstawimy wyniki analizy problemu efektywnego wyboru parametrów p, M i g w ekstraktorze Chora-Goldreicha w formie przedstawionej przez Dodisa *i in.* [10]. Pochodzą one z pracy Durnogi i Źrałka [15] oraz rozprawy doktorskiej autora [14], a ich rezultat stanowi definicja alternatywnej wersji ekstraktora niekowalnego, która nie zakłada prawdziwości ERH ani hipotezy 2.

3. Generator pseudolosowy online

Pierwszym krokiem do stworzenia bezwarunkowej konstrukcji ekstraktora jest próba budowy algorytmu wyznaczającego wartości funkcji f_g bez danego generatora g. Z pozoru zadanie to wydaje się nie być poprawnie sformułowane ze względu na fakt, że nie jest możliwe obliczanie funkcji, która sama w sobie pozostaje nieznana. Proponowana tu metoda pozwala na obliczenie wartości pewnej funkcji częściowej $f(a_1), \ldots, f(a_\ell)$ dla wielu argumentów $a_1, \ldots, a_\ell$ gwarantując istnienie takiego, nie danego jawnie, generatora g, że $f(a_i) = f_g(a_i)$ dla wszystkich $i = 1, \ldots, \ell$. Metoda ta jest algorytmem typu *online*, tzn. takim, w którym argumenty przetwarzane są sekwencyjnie i całość wejścia może nie być dostępna na początku działania algorytmu. Algorytmy online mają fundamentalne znaczenie w obliczeniach interaktywnych.

Można wykazać następujące twierdzenie:

Twierdzenie 3. *(*Durnoga i Źrałek [15], twierdzenie 8) *Istnieje deterministyczny algorytm online, który dla danego ciągu argumentów $a_1, \ldots, a_\ell \in (\mathbb{Z}/p\mathbb{Z})^*$ oblicza $f_g(a_1), \ldots, f_g(a_\ell)$, gdzie f_g jest epimorfizmem danym przez (2) dla pewnego, a priori nieznanego, generatora g grupy $(\mathbb{Z}/p\mathbb{Z})^*$ zależnego od $a_1, \ldots, a_\ell$. Algorytm znajduje $f_g(a_i)$ dla $i = 1, \ldots, \ell$ przed pobraniem kolejnego argumentu a_{i+1} z wejścia. Pojedyncza wartość $f_g(a_i)$ obliczana jest w czasie $O\big(\mathrm{P}^+(M) \cdot s^{-1} \mathrm{poly}(\log p)\big)$ przy użyciu $O\big(s\,\mathrm{poly}(\log p)\big)$ bitów pamięci, gdzie parametr $0 < s \leq \sqrt{\mathrm{P}^+(M)}$ może być wybrany dowolnie.*

Należy przy tym zaznaczyć, że algorytm z powyższego twierdzenia nie może być bezpośrednio użyty do obliczania wartości ekstraktora niekowalnego (3), który zdefiniowany jest w terminach statystycznej odległości rozkładów. Dla ciągów wartości pojawiających się na wyjściu ekstraktora możliwe jest jednak, analogicznie jak w przypadku ekstraktorów, udowodnienie pewnej własności nieodróżnialności od ciągu wartości losowych.

4. Bezwarunkowa konstrukcja ekstraktora niekowalnego

Osiągnięcie celu bezwarunkowej konstrukcji ekstraktora niekowalnego możliwe jest po zaskakująco prostej modyfikacji oryginalnego rozwiązania pochodzącego od Chora i Goldreicha [7]. Pomysł polega na zastąpieniu logarytmu dyskretnego występującego w definicji (3), który stanowił źródło opisanych wyżej problemów, przez potęgowanie. Operacja ta przekształca otrzymane argumenty na elementy pewnej grupy G zdefiniowanej jako:

$$G := \{a^{(p-1)/M} \mid a \in (\mathbb{Z}/p\mathbb{Z})^*\}, \tag{4}$$

a sam ekstraktor jest naturalnym odwzorowaniem $(\mathbb{Z}/p\mathbb{Z})^*$ w G. Mówi o tym poniższe twierdzenie:

Twierdzenie 4. (Durnoga i Źrałek [15], lemat 10) *Niech p, M, k oraz ϵ będą takie jak w twierdzeniu 1. Wtedy funkcja* $\mathrm{Ext}_G\colon (\mathbb{Z}/pZ)\times(\mathbb{Z}/pZ) \to G$ *dana wzorem*

$$\mathrm{Ext}_G(x, y) := (x + y)^{(p-1)/M}$$

jest efektywnie obliczalnym (k, ϵ)–niekowalnym ekstraktorem.

W oczywisty sposób definicja Ext_G nie wymaga znajomości generatora grupy $(\mathbb{Z}/p\mathbb{Z})^*$. Ponadto wartości tej funkcji, w odróżnieniu od ekstraktora (3), można szybko obliczać również w przypadku, gdy M ma duże dzielniki pierwsze. Ta obserwacja pozwala na efektywne wygenerowanie parametrów p oraz M bez dodatkowego warunku na gładkość liczby M. Opierając się na głębokim wyniku Alforda *i in.* [1] można dowieść następującego wniosku:

Twierdzenie 5. *Dla wszystkich dostatecznie dużych z, z' oraz dowolnego $l > 0$ spełniających $(z + l)^4 < \frac{1}{2}z'$, istnieje co najmniej $\frac{1}{3}z'/(\varphi(M)\log z')$ liczb pierwszych w przedziale $(\frac{1}{2}z', z']$, które należą do postępu arytmetycznego $\equiv 1 \pmod M$ dla każdego modułu M z przedziału $(z, z + l)$, poza co najwyżej $O\big(l/\log(z + l)\big)$ wyjątkowymi wartościami M.*

To twierdzenie natychmiast prowadzi do randomizowanego algorytmu, o oczekiwanej wielomianowej złożoności, generacji parametrów ekstraktora Ext_G.

Z praktycznego punktu widzenia pewną wadą ekstraktora Ext_G z twierdzenia 4 jest fakt, że wartości pojawiające się na jego wyjściu są elementami podgrupy G rzędu M określonej przez (4). Grupa ta, interpretowana jako zbiór ciągów bitów, ma nieregularną strukturę, przez co może nie być odpowiednia w zastosowaniach oczekujących właśnie losowych ciągów bitów. Ciekawym i nietrywialnym problemem staje się więc kwestia przekształcenia elementów G na takie ciągi bitów. To zagadnienie jest ściśle związane ze znanym problemem tzw. ekstrakcji klucza (ekstraktor deterministyczny, por. Fouque *i in.* [16]).

W pracy Durnogi i Źrałka [15] zaproponowany jest algorytm do wyznaczania wartości pewnej bijekcji $G \to \mathbb{Z}/M\mathbb{Z}$. Szczególne znaczenie miałoby skonstruowanie podobnej bijekcji dla grup stosowanych w protokole Diffie-Hellmana, w których problem obliczania logarytmu dyskretnego jest uznawany za trudny. Wspomniany algorytm działa w czasie proporcjonalnym do największego dzielnika pierwszego $\mathrm{P}^+(M)$ liczby M, a więc jest efektywny jedynie dla tych M, które są gładkie. Tym samym zakładamy, że problem znajdowania logarytmów dyskretnych w G jest „łatwy". Mimo, że w takim przypadku istnieje naturalny i prosty do obliczenia izomorfizm między G i $\mathbb{Z}/M\mathbb{Z}$, zadany właśnie przez logarytm dyskretny, to jednak jego konstrukcja wymaga znajomości generatora grupy G. Rozważana metoda obliczania bijekcji nie zakłada, że taki generator jest dany. Ten przypadek nie był, wg wiedzy autora, wcześniej badany w literaturze i może być interesujący również poza kontekstem ekstraktorów.

Można wykazać następujące twierdzenie:

Twierdzenie 6. (Durnoga i Źrałek [15], twierdzenie 14) *Istnieje bijekcja $\sigma\colon G \to \mathbb{Z}/M\mathbb{Z}$ taka, że dla danego $a \in G$ odpowiadającą wartość $\sigma(a)$ można obliczyć w sposób deterministyczny w czasie $O\big(\mathrm{P}^+(M)\mathrm{poly}(\log p)\big)$.*

Jednocześnie konstrukcja może być zmodyfikowana tak, by uzyskać algorym obliczający bijekcję dla wielu argumentów:

Twierdzenie 7. (Durnoga i Źrałek [15], twierdzenie 16) *Istnieje deterministyczny algorytm online obliczający wartości pewnej bijekcji $G \to \mathbb{Z}/M\mathbb{Z}$. Algorytm wykorzystuje $O(s\mathrm{poly}(\log p))$ bitów pamięci, a jego łączny czas działania dla ℓ argumentów to*

$$O\big(\mathrm{P}^+(M)(\ell s^{-1}+1)\mathrm{poly}(\log p)\big),$$

gdzie $0 < s \leq \sqrt{\mathrm{P}^+(M)}$ może być ustalone dowolnie.

Wspomniana w twierdzeniu 6 bijekcja $G \to \mathbb{Z}/M\mathbb{Z}$ pozwala „bezstratne" przekształcenie wyjścia ekstraktora Ext_G w element zbioru $\mathbb{Z}/M\mathbb{Z}$.

ciślej: złożenie $\sigma \circ \mathrm{Ext}_G$ jest ekstraktorem niekowalnym z identycznym wyrazem błędu ϵ jak w przypadku ekstraktora Ext_G. Ograniczeniem obliczeniowym jest jednak tutaj fakt, że wartości σ mogą być efektywnie znalezione dla M będących liczbami gładkimi. Jest to w istocie problem podobnej natury, co ten występujący w oryginalnej konstrukcji Dodisa *i in.* [10] – metoda generacji parametrów p i $M \mid p-1$ z dodatkowym warunkiem na poziom gładkości M jest warunkowa, chyba że M jest niewielkie, tzn. wielomianowe jako funkcja $\log p$. Praca Durnogi [14] proponuje stosowne rozwiązanie dla ekstraktorów o długich wyjściach, a więc takich, których zbiór wartości jest rozmiaru rzędu $p^{\Omega(1)}$. Podstawą tej konstrukcji jest obserwacja, że przekształcając elementy grupy G na ciągi losowych bitów można dopuścić pewną dodatkową stratę entropii. Innymi słowy, σ w złożeniu $\sigma \circ \mathrm{Ext}_G$ nie musi być funkcją różnowartościową. Naturalnym, a jednocześnie niezwykle prostym, pomysłem na takie przekształcenie zmniejszające dziedzinę jest obcięcie reprezentacji bitowej elementu grupy G do najmniej znaczących bitów. To podejście znane już było w literaturze i sprawdziło się w pokrewnych zastosowaniach – przykładem może być tutaj ekstraktor Holensteina [18]. Chevalier *i in.* [6] pokazali, korzystając z oszacowań dla sum wykładniczych, że wynikiem działania funkcji $\mathrm{lsb}_n(a)$, obcinającej reprezentację bitową elementu a wybranego losowo (z rozkładu jednostajnego) z dowolnej, odpowiednio dużej podgrupy $(\mathbb{Z}/p\mathbb{Z})^*$ do n jej najmniej znaczących bitów, jest niemal losowy ciąg zer i jedynek. Funkcja lsb_n jest tym samym tzw. deterministycznym ekstraktorem o tej własności, że przekształca on losowe elementy grupy w (prawie) losowe ciągi bitów. Efektywne obliczeniowo ekstraktory tego typu są użyteczne w praktyce – umożliwiają przeniesienie klucza uzyskanego w realizacji protokołu Diffie-Hellmana (który to klucz, przy odpowiednich założeniach, jest nieodróżnialny od losowego elementu grupy) do świata kryptografii symetrycznej (w którym zwykle wymaga się, by klucze były losowymi ciągami bitów).

Opierając się na wynikach Fouque *i in.* [16], Chevalier *i in.* oraz głębokim rezultacie Bourgain i Konyagina [5] można pokazać następujące twierdzenie:

Twierdzenie 8. (Durnoga [14], twierdzenie 7.10) *Dla dowolnej stałej $\alpha > 0$ istnieje takie $\beta = \beta(\alpha) > 0$, że dla każdej liczby pierwszej p i całkowitego $M \mid p-1$, spełniających $M \geq p^{\alpha}$, oraz dowolnego ograniczenia na min-entropię k i $n > 0$ funkcja $\mathrm{Ext}' \colon (\mathbb{Z}/pZ) \times (\mathbb{Z}/pZ) \to \mathbb{Z}/N\mathbb{Z}$ dana wzorem*

$$\mathrm{Ext}'(x,y) := \mathrm{lsb}_n\big(\mathrm{Ext}_G(x,y)\big) = \mathrm{lsb}_n\big((x+y)^{(p-1)/M}\big)$$

jest efektywnie obliczalnym (k, ϵ')*–niekowalnym ekstraktorem z wyrazem błędu*

$$\epsilon' = 2Mp^{1/4}2^{-k/2} + \frac{1}{2}2^{n/2}p^{-\beta}\log_2^{1/2} p.$$

Literatura

[1] William R. Alford, Andrew Granville, and Carl Pomerance, *There are infinitely many Carmichael numbers*, Annals of Mathematics, 139:703–722, 1994.

[2] Nesmith C. Ankeny, *The least quadratic non residue*, Annals of Mathematics, 55:65–72, 1952.

[3] Charles H. Bennett, Gilles Brassard, and Jean-Marc Robert, *Privacy amplification by public discussion* SIAM J. Comput., 17(2):210–229, April 1988.

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

[5] Jean Bourgain and Sergei Konyagin, *Estimates for the number of sums and products and for exponential sums over subgroups in fields of prime order*, Comptes Rendus Mathematique, 337(2):75–80, 2003.

[6] Céline Chevalier, Pierre-Alain Fouque, David Pointcheval, and Sbastien Zimmer, *Optimal randomness extraction from a Diffie-Hellman element*, In Antoine Joux, editor, *Advances in Cryptology* – Proceedings of EUROCRYPT 09, volume 5479 of Lecture Notes in Computer Science, pages 572–589, Cologne, Germany, 2009. Springer.

[7] Benny Chor and Oded Goldreich, *Unbiased bits from sources of weak randomness and probabilistic communication complexity*, SIAM J. Comput., 17(2):230–261, April 1988.

[8] Aviad Cohen and Avi Wigderson, *Dispersers, deterministic amplification, and weak random sources* (extended abstract), In FOCS, pages 14–19, IEEE Computer Society, 1989.

[9] Gil Cohen, Ran Raz, and Gil Segev, *Non-malleable extractors with short seeds and applications to privacy amplification*, In IEEE Conference on Computational Complexity, pages 298–308, IEEE, 2012.

[10] Yevgeniy Dodis, Xin Li, Trevor D. Wooley, and David Zuckerman, *Privacy amplification and non-malleable extractors via character sums*, In Proceedings of the 2011 IEEE 52nd Annual Symposium on Foundations of Computer Science, FOCS 11, pages 668–677, Washington, DC, USA, 2011, IEEE Computer Society.

[11] Yevgeniy Dodis and Roberto Oliveira, *On extracting private randomness over a public channel* In Sanjeev Arora, Klaus Jansen, Jos D. P. Rolim, and Amit Sahai, editors, RANDOM-APPROX, volume 2764 of Lecture Notes in Computer Science, pages 252–263, Springer, 2003.

[12] Yevgeniy Dodis and Daniel Wichs, *Non-malleable extractors and symmetric key cryptography from weak secrets*, IACR Cryptology ePrint Archive, 2008:503, 2008.

[13] Yevgeniy Dodis and Daniel Wichs, *Non-malleable extractors and symmetric key cryptography from weak secrets*, In Proceedings of the 41st annual ACM symposium on Theory of computing, STOC 09, pages 601610, New York, NY, USA, 2009, ACM.

[14] Konrad Durnoga, *Non-malleable Randomness Extractors*, PhD thesis, University of Warsaw, 2014.

[15] Konrad Durnoga and Bartosz Zrałek, *On randomness extractors and computing discrete logarithms in bulk*, 2013, preprint.

[16] Pierre-Alain Fouque, David Pointcheval, Jacques Stern, and Sbastien Zimmer, *Hardness of distinguishing the MSB or LSB of secret keys in Diffie-Hellman schemes*, In Michele Bugliesi, Bart Preneel, Vladimiro Sassone, and Ingo Wegener, editors, Automata, Languages and Programming, volume 4052 of Lecture Notes in Computer Science, pages 240–251. Springer Berlin Heidelberg, 2006.

[17] Andrew Granville and Carl Pomerance, *On the least prime in certain arithmetic progressions*, Journal of the London Mathematical Society, 2(2):193–200, 1990.

[18] Thomas Holenstein, *Pseudorandom generators from one-way functions: A simple construction for any hardness*, In Shai Halevi and Tal Rabin, editors, In 3rd Theory of Cryptography Conference – (TCC 06), Lecture Notes in Computer Science. Springer-Verlag, 2006.

[19] Russell Impagliazzo, Leonid A. Levin, and Michael Luby, *Pseudo-random generation from one-way functions*, In Proceedings of the twenty-first annual ACM symposium on Theory of computing, STOC 89, pages 12–24, New York, NY, USA, 1989. ACM.

[20] Hendrik W. Lenstra. Primality testing with Gaussian periods, *In Manindra Agrawal and Anil Seth*, editors, FST TCS 2002: Foundations of Software Technology and Theoretical Computer Science, 22nd Conference Kanpur, India, December 12–14, 2002, Proceedings, volume 2556 of Lecture Notes in Computer Science, page 1, Springer, 2002.

[21] Xin Li, *Non-malleable extractors, two-source extractors and privacy amplification*, In FOCS, volume 0, pages 688697, Los Alamitos, CA, USA, 2012, IEEE Computer Society.

[22] Ueli M. Maurer and Stefan Wolf, *Privacy amplification secure against active adversaries* In Burton S. Kaliski Jr., editor, CRYPTO, volume 1294 of Lecture Notes in Computer Science, pages 307–321, Springer, 1997.

[23] Noam Nisan and David Zuckerman, *More deterministic simulation in logspace*, In Proceedings of the twenty-fifth annual ACM symposium on Theory of computing, STOC 93, pages 235–244, New York, NY, USA, 1993, ACM.

[24] Stephen Pohlig and Martin Hellman, *An improved algorithm for computing logarithms over $GF(p)$ and its cryptographic significance*, IEEE Transactions on Information Theory, 24(1):106–110, 1978.

[25] Wolfgang M. Schmidt, *Equations over finite fields: an elementary approach*, Lecture Notes in Mathematics. Springer-Verlag, 1976.

[26] Triantafyllos Xylouris, *Über die Nullstellen der Dirichletschen L-Funktionen und die kleinste Primzahl in einer arithmetischen Progression*, PhD thesis, Mathematisch-Naturwissenschaftliche Fakultät der Universität Bonn, 2011.

[27] David Zuckerman, *General weak random sources*, In FOCS 90, pages 534–543, 1990.

NON-MALLEABLE RANDOMNESS EXTRACTORS

Abstract. We give an unconditional construction of a non-malleable extractor improving the solution from the recent paper Privacy Amplification and Non-Malleable Extractors via Character Sums by Dodis et al. (FOCS'11). There, the authors provide the first explicit example of a non-malleable extractor - a cryptographic primitive that significantly strengthens the notion of a classical randomness extractor. In order to make the extractor robust, so that it runs in polynomial time and outputs a linear number of bits, they rely on a certain conjecture on the least prime in a residue class. In this paper we present a modification of their construction that allows to remove that dependency and address an issue we identified in the original development.

Keywords: randomness extractor, non-malleable extractor, discrete logarithm