KODOWANIE DETERMINISTYCZNE NA KRZYWYCH ELIPTYCZNYCH

Mariusz Skałba

Wydział Matematyki, Informatyki i Mechaniki
Uniwersytet Warszawski
ul. Banacha 2, 02-097 Warszawa, Polska
skalba@mimuw.edu.pl

Streszczenie. W pracy przedstawiono problematykę znajdowania punktów na krzywych eliptycznych określonych nad ciałami skończonymi, ze szczególnym uwzględnieniem algorytmów deterministycznych. Algorytmy takie nie były znane do 2005 roku. Wcześniejsze metody, chociaż dość praktyczne, miały charakter probabilistyczny, a ich efektywność była uwarunkowana hipotezami Riemanna.

Słowa kluczowe: krzywe eliptyczne, ciała skończone, kodowanie deterministyczne

Przed zaszyfrowaniem informacji, na ogół przecież werbalnej, należy ją zakodować za pomocą liczb. I tak na przykład kolejnym literom alfabetu można na stałe przyporządkować kolejne liczby

$$A \longrightarrow 0, \quad B \longrightarrow 1, \quad \ldots, Z \longrightarrow 25, \quad \text{spacja} \longrightarrow 26$$

i zdanie ALA MA KOTA zakodować jako ciąg liczb ze zbioru $\{0, 1, \ldots, 26\}$

$$0\ 11\ 0\ 26\ 12\ 0\ 26\ 10\ 14\ 19\ 0.$$

Dopiero teraz możemy wiadomość zaszyfrować. Potraktujmy powyższe liczby jako reszty modulo 27, czyli elementy pierścienia $\mathbb{Z}_{27}$. Przekształcenie szyfrujące E zadajemy wzorem

$$E(x) = 5x + 3.$$

Powyższa wiadomość zaszyfrowana wygląda tak

$$3\ 4\ 3\ 25\ 9\ 3\ 25\ 26\ 19\ 17\ 3.$$

Przeprowadzając proste rachunki modulo 27 łatwo obliczamy przekształcenie rozszyfrowujące D

$$D(y) = 11y - 6.$$

Rzeczywiście, dla każdej liczby całkowiej x mamy

$$D(E(x)) = 11(5x + 3) - 6 = 55x + 27 \equiv x \pmod{27}.$$

Oczywiście szyfrogram można dla niepoznaki zapisać literami DEDZJDZ TRD. Już na tym elementarnym przykładzie widać, że w istocie szyfrowanie polega na jednoznacznym przyporządkowaniu układowi liczb innego układu liczb, według tajnej reguły (E), która jest odwracalna, w tym sensie, że, ze zmienionego układu można powrócić do wyjściowego za pomocą reguły odwrotnej (D).

Metodę powyższą można wyrafinować pracując w grupie multyplikatywnej reszt, zamiast w grupie addytywnej, jak wyżej. Niech mianowicie p będzie liczbą pierwszą nieparzystą, a liczba $g \in \{2, 3, \ldots, p-1\}$ będzie tak wybrana, że reszty liczb $g, g^2, \ldots, g^{p-1}$ wyczerpują wszystkie reszty modulo p (oczywiście z wyjątkiem reszty 0). Takie liczby g istnieją dla każdej liczby pierwszej p – nazywamy je pierwiastkami pierwotnymo modulo p, lub generatorami grupy $\mathbb{Z}_p^*$. I tak dla $p = 29$ można wziąć $g = 2$. Nasze ulubione zdanie można zakodować teraz za pomocą reszt modulo 29 w następujący sposób

$$2^0\ 2^{11}\ 2^0\ 2^{26}\ 2^{12}\ 2^0\ 2^{26}\ 2^{10}\ 2^{14}\ 2^{19}\ 2^0, \text{czyli}$$

$$1\ 18\ 1\ 22\ 7\ 1\ 22\ 9\ 28\ 26\ 1.$$

Szyfrować i rozszyfrowywać można na przykład za pomocą funkcji

$$E(x) = 8x^5, \text{ oraz } D(y) = (11y)^{17}.$$

Rzeczywiście, mamy następujący ciąg kongruencji mod 29, gdzie w pewnym momencie używamy małego twierdzenia Fermata:

$$D(E(x)) = (11 \cdot 8x^5)^{17} \equiv (x^5)^{17} \equiv (x^{28})^3 \cdot x \equiv x \pmod{29}.$$

Ponieważ oswoiliśmy się nieco z dodawaniem i mnożeniem reszt modulo m warto odnotować najważniejszy fakt: jeżeli p jest liczbą pierwszą to $\mathbb{Z}_p$ jest ciałem. Równania z dwiema niewiadomymi x, y o współczynnikach z ciała opisują krzywe, które można badać metodami algebry liniowej, lub dogłębniej metodami geometrii algebraicznej. Szczególnie interesujący jest przypadek równania typu

$$y^2 = x^3 + Ax + B, \quad \text{gdzie} \quad A, B \in \mathbb{Z}_p \quad \text{oraz} \quad \Delta = 4A^3 - 27B^2 \neq 0.$$

Zakładamy ponadto, dla uproszczenia dalszych wywodów, że $p > 3$. Mówimy, że powyższe równanie zadaje krzywą eliptyczną nad ciałem $\mathbb{Z}_p$. „Krzywa" ta składa się ze skończonej liczby punktów – wszystkich par (x, y), gdzie $x, y \in \mathbb{Z}_p$, jest przecież $p \cdot p$.

Rozważmy jako przykład krzywą E nad ciałem $\mathbb{Z}_{31}$ zadaną równaniem

$$y^2 = x^3 + 7x + 13.$$

Łatwo sprawdzić, że poniższa lista zawiera wszystkie elementy (x, y) zbioru $\mathbb{Z}_{31}^2$ spełniające powyższe równanie:

$$\{(2,2),(2,-2),(5,7),(5,-7),(7,8),(7,-8),(13,10),(13,-10),(16,6),$$
$$(16,-6),(18,9),(18,-9),(20,0),(21,11),(21,-11),(26,15),(26,-15),$$
$$(27,13),(27,-13),(30,6),(30,-6)\}.$$

Podobnie, jak reszty, które można dodawać i mnożyć, punkty (x, y) na krzywej eliptycznej też można dodawać. Jest to pięknie i wyczerpująco opisane w klasycznej książce [4], a my poprzestaniemy tylko na uwadze, że chociaż punktów nie dodajemy „po współrzędnych" to współrzędne sumy dwóch punktów wyrażają się poprzez współrzędne punktów składników za pomocą stosunkowo prostych wzorów. Powstaje w ten sposób skończona grupa abelowa, która w naszym przypadku ma 22 elementy. Na powyższej liście wymienione są wszystkie punkty afiniczne – jest ich 21 – niewymieniony punkt to tzw. punkt w nieskończoności, oznaczany zwykle ∞ lub $\mathcal{O}$. Jest on elementem neutralnym rzeczonego działania grupowego.

Uznajmy, że najbardziej wyrafinowany sposób kodowania z powyżej opisanych polega na zastąpieniu liter (lub kilkuliterowych zbitek liter) przez punkty krzywej eliptycznej określonej nad ciałem skończonym $\mathbb{Z}_p$. I tu pojawia się tytułowy problem: inaczej niż w przypadku kodowania za pomocą reszt modulo m, czy to w wersji addytywnej, czy multyplikatywnej, w przypadku krzywej eliptycznej nie mamy naturalnego sposobu wyboru punktów, które będą kodować litery. Jeśli liczba pierwsza p oraz parametry krzywej eliptycznej A, B są ogromnymi liczbami kilkusetcyfrowymi to nie jest jasne, jak znaleźć $x_0, y_0 \in \mathbb{Z}_p$ spełniające równanie

$$y_0^2 = f(x_0),$$

gdzie oznaczyliśmy $f(x) = x^3 + Ax + B$.

Problem ten składa się z dwóch części:

1. Znaleźć $x_0 \in \mathbb{Z}_p$ takie, że $f(x_0)$ jest kwadratem w $\mathbb{Z}_p$.
2. Znaleźć $y_0 \in \mathbb{Z}_p$ spełniające $y_0^2 = f(x_0)$.

W żargonie podproblem 2 określa się jako *wyciąganie pierwiastka kwadratowego mod p.* Od dawna znany jest efektywny algorytm wyciągania pierwiastka kwadratowego modulo p, działający w czasie $O(\ln^3 p)$, o ile dysponujemy elementem $n \in \mathbb{Z}_p$, który nie jest kwadratem w $\mathbb{Z}_p$ [8]. Algorytm ten należy uznać za probabilistyczny, gdyż póki co nikt nie potrafi znajdować niereszty kwadratowej modulo p w czasie wielomianowym (od $\ln p$). Co prawda, przy założeniu rozszerzonej hipotezy Riemanna E. Bach wykazał w [1] , że najmniejsza niereszta kwadratowa jest mniejsza od $2\ln^2 p$, ale wynik ten jest przez to bardzo warunkowy!

Jeszcze gorzej było z podproblemem 1. Przynajmniej do 2005 roku „rozwiązywano" go w następujący sposób. Za x_0 podstawiano małe liczby naturalne $0, 1, 2, \ldots$ dopóki nie natrafiono na x_0 takie, że $f(x_0)$ jest resztą kwadratową modulo p. To, że powyższe postępowanie jest praktycznie skuteczne nawet dla dużych liczb pierwszych p wynika ze słynnego twierdzenia Hassego [3]:

Niech N będzie liczbą $\mathbb{F}_q$-punktów na krzywej eliptycznej zdefiniowanej nad ciałem $\mathbb{F}_q$. Wtedy

$$|N - (q+1)| \leq 2\sqrt{q}.$$

Ponieważ $\sqrt{p}$ jest bardzo małe w stosunku do p, dla dużych p, więc teza twierdzenia Hassego oznacza, że dla około połowy liczb x_0 ze zbioru $\{0, 1, 2 \ldots, p-1\}$ liczba $f(x_0)$ jest kwadratem modulo p, i każde takie x_0 daje dwa punkty na krzywej $(x_0, y_0), \quad (x_0, -y_0)$. Szansa wylosowania dobrego x_0 wynosi około $1/2$ i dlatego takie poszukiwanie po omacku jest praktycznie bardzo efektywne – mimo tego, przedstawiona powyżej metoda postępowania to typowy algorytm zrandomizowany.

Neal Koblitz, pionier i guru zastosowań krzywych eliptycznych w kryptografii, pisał w swojej książce [5] tak:
Ale główną przeszkodą w skonstruowaniu deterministycznego algorytmu wielomianowego znajdującego punkt na krzywej eliptycznej E wcale nie jest pierwiastkowanie $f(x)$. Problem leży raczej w znalezieniu $x \in \mathbb{Z}_p$, dla którego $f(x)$ jest kwadratem. Chociaż własność tę ma około 50% elementów x, to – z wyjątkiem kilku specjalnych przypadków – nie jest znany żaden efektywny sposób deterministyczny wyznaczenia takiego x. Wszystkie metody deterministyczne wymagają czasu wykładniczego. Jest rzeczą godną uwagi, że nikt nie wie, jak znaleźć na krzywej eliptycznej jakikolwiek punkt (różny od punktu w nieskończoności) w czasie podwykładniczym.

Ten podproblem udało mi się jednak rozwiązać w 2005 roku w pracy [9]. Przy założeniu $A \neq 0$ oraz $p > 3$ podałem wzory wyraźne na niestałe funkcje wymierne $X_1(t), X_2(t), X_3(t), U(t)$ spełniające tożsamość

$$f(X_1(t))f(X_2(t))f(X_3(t)) = U(t)^2.$$

Podstawmy teraz za t taki element $t_0 \in \mathbb{Z}_p$, który nie jest miejscem zerowym żadnego z mianowników funkcji $X_1(t), X_2(t), X_3(t), U(t)$. Z równości w ciele $\mathbb{Z}_p$,

$$f(X_1(t_0))f(X_2(t_0))f(X_3(t_0)) = U(t_0)^2 \tag{1}$$

i faktu, że grupa multyplikatywna $\mathbb{Z}_p^*$ jest cykliczna dostajemy natychmiast wniosek:

istnieje $j \in \{1, 2, 3\}$ *takie, że* $f(X_j(t_0))$ *jest kwadratem w* $\mathbb{Z}_p$.

Jednak najbardziej zaskakujące jest to, że korzystając z (1) można również rozwiązać efektywnie i **bezwarunkowo** podproblem 2! Dokonał tego Christian van de Woestijne w swojej pracy doktorskiej, napisanej pod kierunkiem H.W. Lenstry na Uniwersytecie w Lejdzie. Natomiast A. Shallue podał deterministyczny sposób znajdowania punktu na krzywej eliptycznej nad ciałem skończonym charakterystyki 2. W ten sposób powstała ważna praca wspólna obu autorów [7]. Opiszemy teraz krótko na czym polega rozwiązanie podproblemu 2 zaproponowane przez van de Woestijne. Na podstawie (1) mamy cztery elementy $a, b, c \in \mathbb{Z}_p$ takie, że

$$abc = d^2 \tag{2}$$

i bardzo chcemy wyciągnąć efektywnie pierwiastek kwadratowy z przynajmniej jednego elementu spośród a, b, c. Niech 2^t będzie największą potęgą 2, która dzieli $p-1$. Mamy więc $p-1 = 2^t s$, gdzie liczba s jest nieparzysta. Oznaczmy

$$\alpha = a^s, \quad \beta = b^s, \quad \gamma = c^s, \quad \delta = d^s.$$

Jeśli uda nam się znaleźć μ, takie, że $\alpha = \mu^2$ to przyjmując $f = \mu a^{(1-s)/2}$ dostajemy $a = f^2$. Tak więc wystarczy, jeśli wyciągniemy efektywnie pierwiastek kwadratowy z jednego z elementów α, β lub γ. Możemy napisać, że $\alpha, \beta, \gamma, \delta \in H$, gdzie H jest maksymalną 2-podgrupą grupy $\mathbb{Z}_p^*$. Dla $h \in H$ niech $\mathrm{ord}(h)$ oznacza rząd elementu h. Rozróżniamy teraz dwa przypadki:

- jeśli $\mathrm{ord}(\alpha) < \mathrm{ord}(\beta)$ to istnieje liczba naturalna k taka, że $\alpha_1 = \alpha\beta^{2^k}$ ma rząd mniejszy niż rząd α. Kontynuując to postępowanie uzyskujemy przedstawienie $\alpha = \beta^{2K}$.
- jeżeli natomiast $\mathrm{ord}(\alpha) = \mathrm{ord}(\beta)$ to $\mathrm{ord}(\alpha\beta) < \mathrm{ord}(\beta)$ i jak wyżej uzyskamy efektywnie przedstawienie typu $\alpha\beta = \beta^{2K}$. Ale wówczas $\gamma = (\delta\beta^{-K})^2$.

Podsumowując: metoda van de Woestijne pozwala efektywnie wyciągnąć pierwiastek kwadratowy z przynajmniej jednego spośród elementów

$a, b, c \in \mathbb{Z}_p$, o ile spełniają one równanie (2) – nie potrzebujemy przy tym niereszty kwadratowej, a więc obywamy się bez nieudowodnionych hipotez.

Należy jeszcze dodać, że A. Schinzel i M. Skałba we wspólnej pracy [6] rozpatrzyli przypadek $A = 0$. Ponadto M. Ulas w pracy [10] znacznie uprościł wyprowadzenie tożsamości (1) i, co najważniejsze, jego sposób obejmuje również krzywe hipereliptyczne postaci

$$y^2 = x^n + Ax + B \quad \text{oraz} \quad y^2 = x^n + Ax^2 + Bx, \quad \text{o ile} \quad AB \neq 0.$$

Metody naszkicowane w bieżącej pracy były potem i są dalej rozwijane intensywnie na całym świecie – obecny stan badań i bogatą literaturę można znaleźć w pracy [2].

Literatura

[1] E. Bach, *Explicit bounds for primality testing and related problems*, Math. Comp. 55 (1990), 355–380

[2] R.R. Farashahi, P.A. Fouque, I.E.Shparlinski, M. Tibouchi, J.F. Voloch, *Indifferentiable deterministic hashing to elliptic and hyperelliptic curves*, Math. Comp. 82 (2013), 491–512.

[3] H. Hasse, *Zur Theorie der abstrakten elliptischen Funktionenkrper. I, II & III*, Crelle's Journal 175 (1936).

[4] N. Koblitz, *Wykład z teorii liczb i kryptografii*, Wydawnictwa Naukowo-Techniczne,
Warszawa 2006.

[5] N. Koblitz, *Algebraiczne aspekty kryptografii*, Wydawnictwa Naukowo-Techniczne,
Warszawa 2000.

[6] A. Schinzel, M. Skałba, On equations $y^2 = x^n + k$ in a finite field, Bull. Polish Acad. Sci. Math. 52 (2004), 223–226.

[7] A. Shallue, C. van de Woestijne, *Construction of rational points on elliptic curves over finite fields*, Lecture Notes in Computer Science 4076, Springer 2006, 510–524.

[8] D. Shanks, *Five number-theoretic algorithms*, Congressus Numerantium 7 , Proc. 2 nd Manitoba Conf. on Numerical Math. (University of Manitoba), 1972, 51–70.

[9] M. Skałba, *Points on elliptic curves over finite fields*, Acta Arith. 117 (2005), 293–301.

[10] M. Ulas, *Rational points on certain hyperelliptic curves over finite fields*, Bull. Polish Acad. Sci. Math. 55 (2007), 97–104.

DETERMINISTIC ENCODING ON ELLIPTIC CURVES

Abstract. The methods of finding points on elliptic curves over finite fields are presented with special emphasis on deterministic algorithms. Such algorithms were unknown until 2005. Earlier methods were probabilistic in nature and their efficiency was strongly conditioned on unproved Riemann conjectures.

Keywords: elliptic curves, finite fields, deterministic encoding.