Přínos teorie eliptických křivek k řešení moderních kryptografických systémů

Eliška Ochodková

Katedra informatiky, FEI, VŠB - Technická Univerzita Ostrava,
17. listopadu 15, 708 33, Ostrava-Poruba
eliska.ochodkova@vsb.cz

Abstrakt. Na rozdíl od symetrické kryptografie, kde je bezpečnost a efektivita kryptosystému dána vlastním návrhem algoritmu, hraje u asymetrických kryptosystémů významnou roli algebraická platforma, nad níž je asymetrický kryptografický algoritmus realizován, a to jak pro bezpečnost tak i pro efektivitu výsledné implementace. Cílem tohoto článku je objasnit principy ktyptosystémů na bázi eliptických křivek, uvedení konkrétního kryptosystému ECDSA (Elliptic Curve Digital Signature Algorithm) a nastínit požadavky na výběr tělesa a eliptické křivky pro realizaci bezpečného a efektivního asymetrického kryptosystému (podle některých standardů).

Klíčová slova: asymetrická kryptografie, eliptická křivka, konečné těleso, ECDSA

1 Úvod

Kryptografie se zabývá transformací otevřeného textu na šifru prostřednictvím šifrování a transformací šifry na otevřený text prostřednictvím dešifrování. Ideální kryptosystémy podporují tři nejdůležitější bezpečnostní funkce, a to utajení, autentizaci a integritu dat, a popřípadě nepopiratelnost. Existují dva typy kryptografických algoritmů:

1. asymetrické algoritmy, které používají dvojici klíčů - soukromý a veřejný klíč (např. RSA, ElGamal, viz [10]),
2. aymetrické algoritmy, které používají tajný klíč (např. AES [1], DES [10]). Symetrické algoritmy se dělí na blokové šifry (operují nad blokem dat) a proudové šifry (zpracovávají data bit po bitu (nebo byte po bytu)).

Definice 1. *Kryptografický systém (dále jen kryptosystém) je pětice* $\{M, C, K, E, D\}$*, kde* M *je konečná množina srozumitelných (otevřených) textů (prostor textů (zpráv)),* C *konečná množina možných šifer (prostor šifer),* K *je konečná množina možných klíčů (prostor klíčů),* E *je množina šifrovacích funkcí (pravidel, algoritmů),* D *je množina dešifrovacích funkcí.*

Dnešní systémy s veřejným klíčem vznikly proto, aby byl s jejich pomocí vyřešen problém klíčového hospodářství pro symetrické algoritmy. Mezi N účastníky, kdy chce komunikovat každý účastník s každým utajeně pomocí symetrického algoritmu, je počet tajných klíčů roven $(N * (N-1))/2$ a tyto tajné klíče je třeba distribuovat pouze příslušným dvěma účastníkům. Pokud však účastníci používají asymetrický algoritmus, je celá situace jednodušší. Každý účastník vlastní dvojicí klíčů - veřejný a soukromý. Veřejné klíče jsou vhodným způsobem „zveřejněny“. Každý účastník zašle tajnou zprávu při použití pouze veřejně dostupné informace. Zaslanou zprávu může dešifrovat pouze zamýšlený příjemce neboť jen on je vlastníkem příslušného soukromého klíče.

V současné době jsou asymetrické kryptosystémy používány:

1. pro výměnu tajných klíčů symetrické kryptografie,
2. pro digitální podpisy,
3. pro šifrování.

Mnoho současných asymetrických kryptosystémů je založeno na využití operací nad velkými konečnými matematickými grupami. Kryptografická síla těchto systémů je odvozena ze složitosti řešení úlohy diskrétního logaritmu nad multiplikativní grupou tělesa F_p, kde p je velké prvočíslo, např. algoritmus ElGamal (viz [2]). Bezpečnost těchto systémů závisí na velikosti prvočísla a řádu vybraného prvku grupy. Přitom jsou prvočíselný modul, vybraný prvek i jeho řád uživatelům systému známy, a tak může být použito speciálního tvaru hodnot pro zefektivnění kryptografických operací. Nicméně se tato technika v praxi zpravidla nepoužívá.

Typickým představitelem asymetrických kryptosystémů je dále algoritmus RSA (viz [2]), jehož síla spočívá v řešení úlohy faktorizace velkého přirozeného čísla (na součin dvou prvočísel). K provedení rozkladu dostatečně velkého čísla však potřebujeme výrazně efektivní algoritmy. Takové algoritmy existují, ale mají své výpočetní meze (v současnosti algoritmus Number field sieve viz [10]). K tomu, aby používaný kryptosystém měl z hlediska takovýchto útoků dostatečnou záruku bezpečnosti, je třeba stanovit dostatečnou délku klíče, kterou je dnes > 1024b. Dále je bezpečnost implementace závislá na vhodné volbě jeho faktorů. Pro co nejvyšší rychlost provádění kryptografických operací se používá velmi krátkých veřejných klíčů, zpravidla $e = 3$, a na druhé straně držitel soukromého klíče využívá jeho znalosti k efektivnímu provádění algebraických operací na základě čínské věty o zbytcích (Chinese reminder theorem, viz [10]). Požadavky na volbu modulu jsou motivovány výhradně bezpečnostními hledisky. Volba speciálního tvaru modulu by sice mohla zvýšila efektivitu prováděných operací, na druhé straně by ale mohla významně omezit prostor použitelných hodnot, tím zvýšit predikovatelnost a snížit bezpečnost systému.

Nevýhodou asymetrických kryptosystémů je, že bezpečná velikost klíče je podstatně větší než u symetrických kryptosystémů a asymetrické algoritmy jsou značně pomalejší. V tabulce 1. je uvedena srovnatelná bezpečnost různých kryptosystémů při různých délkách klíčů.

Tabulka 1. Srovnatelná bezpečnost různých kryptosystémů při různých délkách klíčů podle [15].

Blokové šifry	RSA/DL	Eliptické křivky
56	417	105
64	682	120
80	1464	149
86	1881	161
...	...	...
109	4047	206

Matematici proto hledali jiné cesty, nové algoritmy pro asymetrické kryptosystémy. Kryptografie na bázi eliptických křivek (Elliptic Curve Cryptography, dále jen ECC) je moderní směr, který v řadě ukazatelů přináší lepší výsledky než nejrozšířenější používané kryptosystémy. Užití eliptických křivek pro návrh asymetrických kryptosystémů poprvé navrhli v r. 1985 nezávisle pánové Victor Miller [4] a Neal Koblitz [3]. Jedná se vlastně o analogii již existujících systémů s veřejným klíčem, kdy je modulární aritmetika nahrazena aritmetikou budovanou na základě operací s body na eliptické křivce. U asymetrických kryptosystémů definovaných nad eliptickou křivkou (ECC) se hierarchicky volí dva typy algebraických struktur: konečné těleso a eliptická křivka reprezentující grupu bodů, nad níž je vlastní asymetrický algoritmus definován. Volba obou těchto algebraických struktur významně ovlivňuje bezpečnost a efektivitu kryptosystému. Požadavky kladené na tyto dvě struktury spolu vzájemně souvisí.

Bezpečnost eliptických kryptosystémů spočívá v obtížnosti řešení úlohy diskrétního logaritmu pro eliptické křivky. V současné době je tato úloha podstatně obtížněji řešitelná než je úloha klasického diskrétního logaritmu. Dokonce nejsou pro tyto algoritmy známy žádné subexponenciální algoritmy (jako pro klasický diskrétní logaritmus, resp. úlohu faktorizace), nejlepší algoritmy mají plně exponenciální charakter (což platí pro obecné křivky, nikoliv však pro některé speciální pod-

třídy eliptických křivek). V důsledku toho lze konstruovat bezpečné kryptosystémy s výrazně kratší délkou klíče (viz tabulka 1.). To vede mimo jiné k implementacím s menšími nároky na paměť, které jsou současně i výrazně rychlejší ve srovnání např. s kryptosystémy na bázi diskrétního logaritmu. Teoretická konstrukce přitom umožňuje vytvořit systémy zcela analogické klasickým modelům.

2 Matematický background

Definice 2. *Nechť x je reálné číslo. Potom $\lfloor x \rfloor$ je největší celé číslo menší nebo rovno x (tzv. dolní celá část čísla x).*

Definice 3. *Jestliže a je celé číslo a N je kladné celé číslo, pak definujeme $a \bmod N$ jako zbytek po dělení čísla a číslem N. Pro každé a můžeme psát $a = \lfloor a/n \rfloor * n + a \bmod N$.*

Definice 4. *Celé číslo a je kongruentní modulo N s c.č. b tehdy, když je rozdíl $a - b$ dělitelný číslem N. Píšeme $a \equiv b \bmod N$ (tj. platí $a \bmod N = b \bmod N$).*

Definice 5. *Nechť p je prvočíslo. Prvočíselné těleso F_p je tvořeno množinou prvků $\{0, 1, ..., p-1\}$ spolu s následujícími operacemi:*

- *jestliže $a, b \in F_p$, pak výsledkem $a + b$ bude číslo $c \in F_p$, pro které platí $a + b \equiv c \bmod p$,*
- *jestliže $a, b \in F_p$, pak výsledkem $a * b$ bude číslo $c \in F_p$, $0 \leq c \leq p-1$, pro které platí $a * b \equiv c \bmod p$,*
- *jestliže $a \in F_p$, pak multiplikativní inverzní prvek k prvku a je a^{-1}: $a * a^{-1} \equiv 1 \bmod p$. Takovýto inverzní prvek existuje tehdy a jen tehdy když $NSD(a, p) = 1$ (kde NSD znamená největší společný dělitel).*

Definice 6. *Na binární těleso[1] F_{2^m} se můžeme dívat jako na vektorový prostor dimenze m nad tělesem F_2, které se skládá z prvků $0, 1$. Tj. v tělese F_{2^m} existuje m prvků $\alpha_0, \alpha_1, ..., \alpha_{m-1}$ takových, že každý prvek $\alpha \in F_{2^m}$ se dá jednoznačně zapsat ve tvaru: $\alpha = a_0 * \alpha_0 + a_1 * \alpha_1 + ... + a_{m-1} * \alpha_{m-1}$, kde $a_i \in \{0, 1\}$.*

Definice 7. *Množina prvků $\{\alpha_0, \alpha_1, ..., \alpha_{m-1}\}$ se nazývá báze F_{2^m} nad F_2.*

Existuje mnoho různých bází, norma ANSI X9.62 se omezuje na dvě báze: polynomiální a normální bázi viz [14]. Pokud se omezíme na polynomiální bázi, pak prvky tělesa F_{2^m} jsou všechny polynomy stupně $\leq m$ s koeficienty z $\{0, 1\}$. Všechny operace (sčítání polynomů, násobení polynomů a nalezení inverzního plynomu k danému polynomu (viz [2])) jsou prováděny modulo ireducibilní polynom stupně m s koeficienty z $\{0, 1\}$.

Operace nad binárním tělesem F_{2^m} s polynomiální bází se staly také základem nového symetrického standardu, algoritmu AES (viz [1]).

3 Úvod do teorie eliptických křivek

Pod pojmem rovinná křivka rozumíme množinu bodů, které splňují rovnici $F(x, y) = 0$. Nejjednodušší rovinnou křivkou je přímka. Rovnici přímky lze zapsat jako polynom s proměnnými x a y, přitom tyto proměnné se v těchto rovnicích objevují v podobě lineární závislosti. Kuželosečky (parabola, hyperbola, elipsa) lze popsat rovnicemi, kde závislost proměnných je popsána kvadratickou rovnicí (v x a y). Logicky navazují kubické křivky, závislost proměnných je popsána rovnicí třetího stupně. Jejich speciální podtřídou jsou eliptické křivky.

[1] Konečným tělesům se říká také Galoisova pole (Galoisova tělesa) a značí se GF_p resp. GF_{2^m}.

Eliptická křivka je algebraická struktura konstruovaná obecně nad tělesem. V kryptografii se používají eliptické křivky nad konečnými tělesy, která lze algebraicky klasifikovat a každé konečné těleso je pak jednoznačně určeno řádem (počtem svých prvků).

Proto pro specifikaci konkrétního konečného tělesa stačí použít označení F_q, kde $q = p^m$ je počet prvků, p prvočíslo a m přirozené číslo. Velmi zjednodušeně lze definovat eliptickou křivku nad konečným tělesem F_q jako množinu $E = \{[x, y] \in F_q^2 \setminus \{[0, 0]\}, F(x, y) = 0\} \cup \{\mathcal{O}\}$, kde $\mathcal{O}$ je dodefinovaný neutrální prvek eliptické křivky (tzv. bod v nekonečnu) a $F(x, y) = y^2 + a_1 * x * y + a_2 * y - x^3 - a_3 * x^2 - a_4 * x - a_5$ je polynom nad F_q. Na takto vytvořené množině je možno definovat binární operaci tak, že eliptická křivka opatřená touto operací má strukturu komutativní grupy viz [23]. Vztah $F(x, y) = 0$ je obecnou Weierstrassovou rovnicí eliptické křivky. Tu lze zjednodušit pro jednotlivé tvary těles a doplnit podmínky pro hodnoty koeficientů rovnic tak, aby eliptická křivka nebyla singulární:

1. Pro $q = p^m$, kde $p > 3$, $m \geq 1$ a koeficienty $a, b \in F_q$, $4 * a^3 + 27 * b^2 \not\equiv 0$, splňuje bod eliptické křivky $[x, y] \in E \setminus \{\mathcal{O}\}$ následující rovnici nad $F_q : y^2 = x^3 + a * x + b$, kde a, b jsou celá čísla *mod p*.
2. Pro $q = 2^m$, $m \geq 1$ a koeficienty $a, b \in F_q$, $b \neq 0$, splňuje bod eliptické křivky $[x, y] \in E \setminus \{\mathcal{O}\}$ následující rovnici nad $F_q : y^2 + x * y = x^3 + a * x^2 + b$, kde a, b jsou celá čísla *mod q*.
3. Pro $q = 3^m$, $m \geq 1$ a koeficienty $a, b \in F_q$, $b \neq 0$, splňuje bod eliptické křivky $[x, y] \in E \setminus \{\mathcal{O}\}$ následující rovnici nad $F_q : y^2 = x^3 + a * x^2 + b$, kde a, b jsou celá čísla *mod q*.

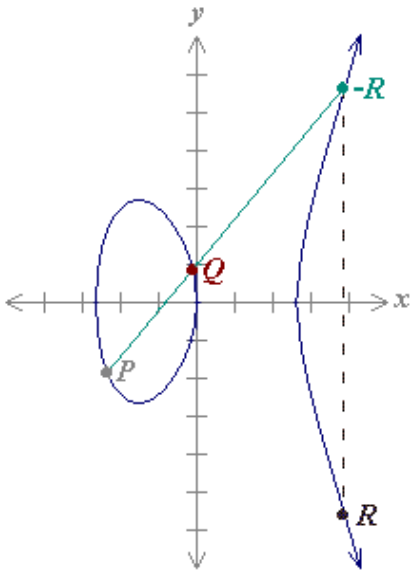
3.1 Eliptické křivky nad F_p

Sčítání: jak na to? Nejprve přibližme graficky operaci sčítání na „reálné“ křivce (v rovině). Na obrázku 1 je znázorněna konkrétní eliptická křivka E daná rovnicí $y^2 = x^3 - 7 * x$.

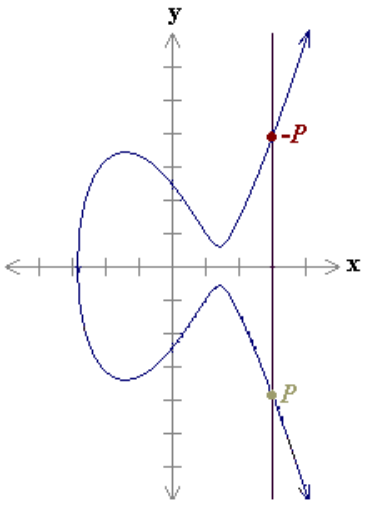
- Výsledkem součtu $P + Q$ dvou různých bodů P, Q, které leží na křivce, bude opět bod R křivky E a vznikne takto (viz obr. 1.): Spojíme body $P = (x_P, y_P)$ a $Q = (x_Q, y_Q)$ přímkou, ta protne křivku v dalším bodě, který označíme $-R$, a výsledkem sčítání je bod R, symetrický k $-R$ podle osy x. Body symetrické podle osy x se nazývají opačné.
- Algebraicky je směrnice přímky, která spojuje body P, Q (předpokládejme, že jsou různé a nikoli opačné), rovna $s = (y_Q - y_P)/(x_Q - x_P)$ a souřadnice bodu $R = (x_R, y_R)$ lze pak odvodit z rovnice křivky jako $x_R = s^2 - x_P - x_Q$ a $y_R = s * (x_P - x_R) - y_P$.
- V případě $P = Q$ přechází jejich spojnice v tečnu ke křivce E a její směrnice je rovna $s = (3 * x_P^2 + a)/(2 * y_P)$.
- Když sčítáme body opačné $P + (-P)$, jejich spojnice (rovnoběžka s osou y) eliptickou křivku E protne jakoby v nekonečnu. Proto matematici definitoricky přidali ke křivce E bod v nekonečnu $\mathcal{O}$ a sčítání dodefinovali i pro body opačné: $P + (-P) = \mathcal{O}$, viz obr. 2.
- Pro bod v nekonečnu $\mathcal{O}$ nadefinujeme pravidla pro sčítání takto: pro každý bod P na křivce E definujeme $P + \mathcal{O} = P$ a také $\mathcal{O} + \mathcal{O} = \mathcal{O}$, přičemž $-\mathcal{O} = \mathcal{O}$.
- Pro přičtení bodu k sobě samému, tj. $P + P$, je vedena tečna ke křivce E z bodu P. Jestliže $y_P \neq 0$, protne tečna křivku v bodě $-R$. Výsledkem sčítání $P + P = 2P = R$ je bod R, symetrický k $-R$ podle osy x, viz obr.3.

Sčítání: eliptická křivka nad F_p. Eliptická křivka E nad tělesem F_p je tedy definována jako bod v nekonečnu $\mathcal{O}$ společně s množinou bodů $P = (x, y)$, kde x a y jsou z tělesa F_p a splňují rovnici $y^2 = x^3 + ax + b$ v F_p, tj. $y^2 \equiv x^3 + a * x + b \ (mod\ p)$. Víme, že koeficienty a, b jsou také prvky tělesa F_p a musí splňovat podmínku $4 * a^3 + 27 * b^2 \not\equiv 0 \ mod\ p$, která zaručuje, že takto definovaná množina bodů tvoří grupu (jinak koeficienty a ab můžeme volit libovolně budou to později veřejné parametry příslušného kryptosystému).

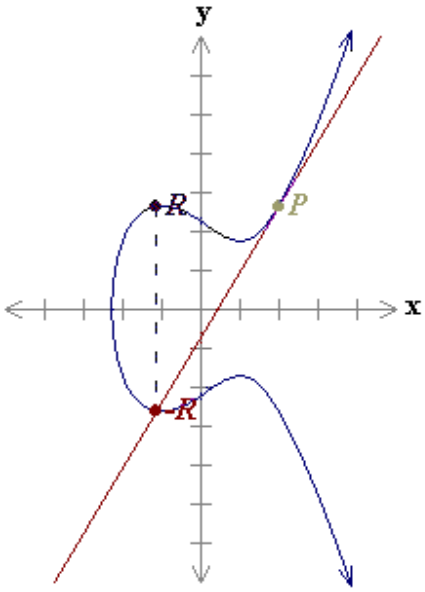
- V této grupě definujeme opačný bod k $\mathcal{O}$ jako $-\mathcal{O} = \mathcal{O}$ a pro ostatní nenulové $P = (x_P, y_P) \in E$ definujeme $-P = (x_P, -y_P \ mod\ p)$, dále pro všechny body $P \in E$ definujeme $P + -P = \mathcal{O}$ a $P + \mathcal{O} = P$. Bod $\mathcal{O}$ nazýváme také nulový bod, vzhledem k jeho roli při sčítání v grupě E.



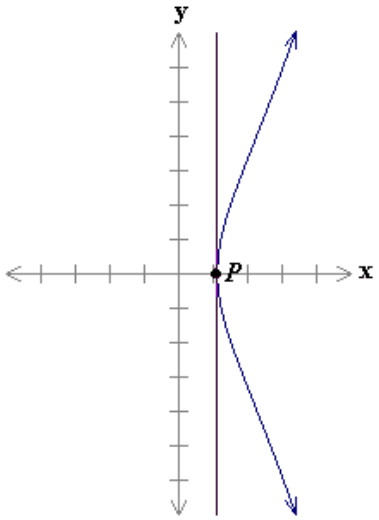
Obr. 1. Sčítání bodů $P + Q = R$ křivky $E : y^2 = x^3 - 7 * x$.



Obr. 2. Sčítání opačných bodů $P + (-P) = \mathcal{O}$ křivky $E : y^2 = x^3 - 6 * x + 6$.



Obr. 3. Přičítání bodu k sobě samému $P + P = 2P$, křivka $E : y^2 = x^3 - 3 * x + 5$.



Obr. 4. Přičítání bodu k sobě samému $P + P = 2P = \mathcal{O}$, pokud je $y_P = 0$, křivka $E : y^2 = x^3 + 5 * x - 7$.

– Sčítání různých nenulových a ne vzájemně opačných bodů $P = (x_P, y_P)$ a $Q = (x_Q, y_Q)$: jestliže P a Q jsou různé body takové, že $P \neq -Q$, pak $P + Q = R$, kde
$s = (y_P - y_Q)/(x_P - x_Q) \ mod \ p^2$,
$x_R = s^2 - x_P - x_Q \ mod \ p$
$y_R = -y_P + s * (x_P - x_R) \ mod \ p$
– Zdvojení bodu P: pokud $y_P \neq 0$, pak $2P = R$, kde
$s = (3 * x_P^2 + a)/(2 * y_P) \ mod \ p$,
$x_R = s^2 - 2 * x_P \ mod \ p$,
$y_R = -y_P + s * (x_P - x_R) \ mod \ p$.

Příklad 1. (Eliptická křivka nad F_p.) Nechť $p = 23$ a předpokládejme křivku $E : y^2 = x^3 + x + 4$ nad F_{23}, tj. $a = 1, b = 4$. Platí $4 * a^3 + 27 * b^2 = 4 + 432 = 436 \equiv 22 \ (mod \ 23)$.

1. Body (viz Obr 5.) této křivky jsou $\mathcal{O}$ a $(0, 2), (0, 21), (1, 11), (1, 12), (4, 7), (4, 16), (7, 3), (7, 20), (8, 8), (8, 15), (9, 11), (9, 12), (10, 5), (10, 18), (11, 9), (11, 14), (13, 11), (13, 12), (14, 5), (14, 18), (15, 6), (15, 17), (17, 9), (17, 14), (18, 9), (18, 14), (22, 5), (22, 19)$.
2. Nechť $P = (4, 7)$ a $Q = (13, 11)$, potom $P + Q = R$ se vypočte takto: $x_R = ((11 - 7)/(13 - 4))^2 - 4 - 13 = 3^2 - 4 - 13 = -8 \equiv 15 \ mod \ 23$ a $y_3 = 3 * (4 - 15) - 7 = -40 \equiv 6 \ mod \ 23$. Tedy $P + Q = (15, 6)$.
3. Nechť $P = (4, 7)$ potom $2P = P + P = R$ se vypočte takto: $x_R = ((3 * 4^2 + 1)/14)^2 - 8 = 15^2 - 8 = 217 \equiv 10 \ mod \ 23$ a $y_3 = 15 * (4 - 10) - 7 = -97 \equiv 18 \ mod \ 23$. Tedy $2P = (10, 18)$.

Podobně jako $2P$ vypočítáme $3P = (P + P) + P = 2P + P, 4P, 5P, ...$ Získáme obecně různé body xP na křivce $\#E$. Protože má křivka konečný počet bodů, musí se po určitém počtu (l) kroků tato posloupnost zacyklit. V bodě zacyklení (lP) tak platí $lP = mP$, kde mP je nějaký dřívější bod. Odtud ale dostáváme $lP - mP = (l - m)P = \mathcal{O}$, čili existuje nějaké $n = l - m, n < l$) takové, že $nP = \mathcal{O}$. Je tedy jasné, že v posloupnosti $P, 2P, 3P, 4P, ...$ se vždy nakonec dostaneme do bodu $\mathcal{O}$ a poté cyklus začíná znovu od bodu P, neboť $(n + 1)P = nP + P = \mathcal{O} + P = P$. Různé body na křivce E mohou mít různý řád. V kryptografické praxi vybíráme takový bod, jehož řád je roven největšímu prvočíslu v rozkladu čísla $\#E$ (pokud je n velké, například řádově 2^{256}, dostaneme velmi dlouhou posloupnost, než se „zacyklí“) nebo jeho násobku (tzv. kofaktoru, kofaktor $h = \#E/n$).

Definice 8. *Řádem eliptické křivky E rozumíme celkový počet bodů této křivky a ozn. jej jako $\#E$.*

Definice 9. *Nechť n je přirozené číslo, bod $P \in E$. Nejmenší takové n, pro něž je $nP = \mathcal{O}$, nazýváme řád bodu P.*

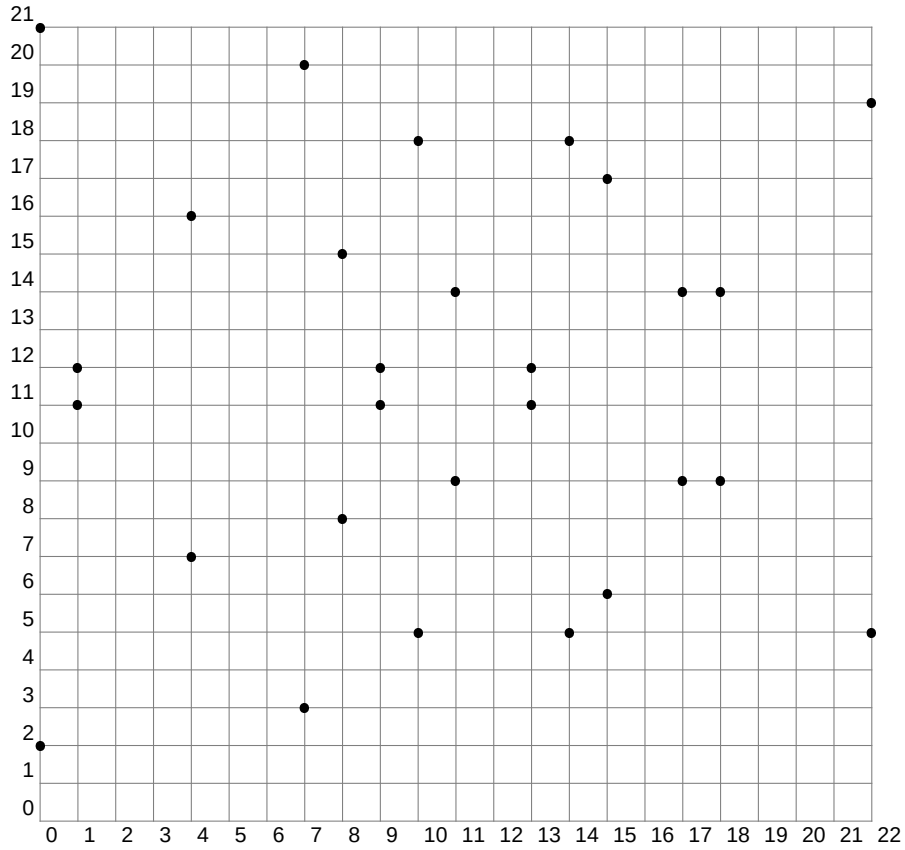
Lze dokázat, že řád bodu dělí řád křivky, viz [11].

Příklad 2. (F_p, $p = 23$, $E : y^2 = x^3 + x + 4$, $P = (0, 2)$) Řád křivky je $\#E = 29$,
$1P = (0, 2), 2P = (13, 12), 3P = (11, 9), 4P = (1, 12), 5P = (7, 20), 6P = (9, 11), 7P = (15, 6), 8P = (14, 5), 9P = (4, 7), 10P = (22, 5), 11P = (10, 5), 12P = (17, 9), 13P = (8, 15), 14P = (18, 9), 15P = (18, 14), 16P = (8, 8), 17P = (17, 14), 18P = (10, 18), 19P = (22, 18), 20P = (4, 16), 21P = (14, 18), 22P = (15, 17), 23P = (9, 12), 24P = (7, 3), 25P = (1, 11), 26P = (11, 14), 27P = (13, 11), 28P = (0, 21), 29P = \mathcal{O}$.

4 ECDSA - Elliptic Curve Digital Signature Algorithm

Problémem diskrétního logaritmu nad eliptickou kkřivkou (ECDLP, eliptic curve discrete logarithm problem) je tato úloha: je dána eliptická křivka E nad konečným tělesem F_q, bod $P \in E$, bod $Q = dP$, kde $1 \leq d \leq n - 1$. Určete číslo d tak, že $Q = dP$. Schéma DSA [5](multiplikativní grupa) se transformuje na eliptickou křivku (aditivní grupa) tak, že operace násobení prvků $g*g*g*g*....$(tj. g^k) se převede na sčítání bodů na křivce $P + P + P + P + ...$ (tj. kP).

[2] Operaci dělení definujeme jako násobení inverzním prvkem, například x/y je $x * y^{-1}$ a přirozeně y^{-1} je ten prvek tělesa F_p, který vynásoben y dává jedničku: $y * y^{-1} = 1$.



Obr. 5. Body křivky $E : y^2 = x^3 + x + 4$ nad F_{23}.

Parametry eliptického kryptosystému pokud $q = p$ je prvočíslo:

- prvočíslo p definující velikost příslušného tělesa $F_p, p > 3$,
- (nepovinně) bitový řetězec SEED délky alespoň 160 bitů, pokud eliptická křivka byla náhodně generována (viz [17], [13]),
- $a, b \in F_p$, které definují rovnici eliptické křivky $E : y^2 = x^3 + a * x + b$,
- $x, y \in F_p$ dva prvky tělesa, které definují bod $P = (x, y)$, tento bod má prvočíselný řád a $P \neq \mathcal{O}$,
- řád n bodu P (musí platit $n > 2^{160}, n > 4\sqrt{q}$), tento řád bodu P je dělitelem řádu křivky E,
- (nepovinně) kofaktor $h = \#E/n$.

Parametry eliptického kryptosystému pokud q je rovno 2^m:

- číslo $q = 2^m$ definující velikost příslušného tělesa F_q, dále jaká báze je použita k vyjádření jednotlivých prvků tohoto tělesa a příslušný redukční polynom řádu m,
- (nepovinně) bitový řetězec SEED délky alespoň 160 bitů, pokud eliptická křivka byla náhodně generována,
- $a, b \in F_p$, které definují rovnici eliptické křivky $E : y^2 + x * y = x^3 + a * x^2 + b$,
- $x, y \in F_p$ dva prvky tělesa, které definují bod $P = (x, y)$, tento bod má prvočíselný řád a $P \neq \mathcal{O}$,
- řád n bodu P (musí platit $n > 2^{160}, n > 4\sqrt{q}$), tento řád bodu P je dělitelem řádu křivky E,
- (nepovinně) kofaktor $h = \#E/n$.

Generování dvojice klíčů pro eliptické kryptosystémy. Daná množina parametrů eliptického kryptosystému je asociována s dvojicí klíčů, která je vytvářena následovně: soukromý klíč d je celé číslo náhodně vygenerované v intervalu $1 < d < n - 1$. Veřejný klíč je bod Q na eliptické křivce vypočtený jako $Q = dP$. Bod Q (i P) můžeme nyní zveřejnit budou součástí veřejného klíče, kterým je čtveřice (E, P, n, Q).

Vytvoření podpisu pomocí schématu ECDSA. Mějme zprávu M.

- vybereme jedinečné a nepredikovatelné číslo $1 \leq k \leq n - 1$,
- vypočteme bod $kP = (x_1, y_1)$ a číslo $r = x_1 \ mod \ n$,
- je-li $r = 0$, pak postup opakujeme od generování čísla k (to je nutné proto, aby v hodnotě s byl obsažen privátní klíč, viz dále),
- vypočteme $k^{-1} \ mod \ n$,
- vypočteme $s = k^{-1}(h(M) + dr) \ mod \ n$, kde h je hashovací funkce SHA-1 (viz literatura),
- je-li $s = 0$, pak opět jdeme na první bod generování nového k (neexistovalo by $s^{-1} \ mod \ n$, viz dále proces ověření),
- podpisem zprávy M je dvojice čísel (r, s).

Ověření podpisu ECDSA. Mějme zprávu M a její podpis (r, s).

- důvěryhodným způsobem získáme veřejný klíč podepisujícího (E, P, n, Q),
- ověříme, že $r, s \in \langle 1, n - 1 \rangle$,
- vypočteme $w = s^{-1} \ mod \ n$ a $h(M)$,
- vypočteme $u_1 = h(M)w \ mod \ n$ a $u_2 = rw \ mod \ n$,
- vypočteme $u_1P + u_2Q = (x_0, y_0)$ a $v = x_0 \ mod \ n$,
- podpis je platný právě tehdy, když $v = r$.

5 Standardy

V původním návrhu využití eliptických křivek pro kryptografii se užívalo výhradně eliptických křivek nad F_p. Z důvodu snadné hardwarové implementace (např. čipové karty bez nutnosti vnitřního speciálního koprocesoru) byly zavedeny i eliptické křivky nad tělesem F_{2^m}. Pro tyto dvě kategorie byla zveřejněna kritéria pro výběr vhodných těles a eliptických křivek nad nimi.

Pro implementaci asymetrického kryptosystému je nutno zvolit bod eliptické křivky s maximálním možným řádem, tj. uvedený prvek generuje pracovní podgrupu maximálního řádu. Jelikož řád podgrupy vždy dělí řád grupy, je podíl obou hodnot celočíselný (kofaktor). Snahou je volit parametry systému tak, aby kofaktor byl co nejnižší.

5.1 Volba bezpečných parametrů

Řešení problému diskrétního logaritmu nad eliptickou křivkou má obecně exponenciální složitost na rozdíl od hledání přirozeného diskrétního logaritmu, pro který byly nalezeny výpočetní metody se subexponenciální složitostí. Pro některé speciální typy eliptických křivek však existují metody řešení diskrétního logaritmu s menší výpočetní složitostí než v obecném případě. Proto se s výjimkou tzv. Koblitzových křivek (viz [3]), které jsou jako speciální tvar používány pro vyšší efektivitu výpočtu, a jednoho pevného parametru eliptické křivky nad prvočíselnými tělesy volí parametry (koeficienty a výchozí bod P) eliptických křivek náhodně s tím, že se takto získaná křivka následně otestuje, splňuje-li bezpečnostní požadavky.

Kritéria jsou následující ($\#E$ je řád eliptické křivky E nad F_q, n řád pracovní podgrupy eliptické křivky generované výchozím bodem P) [23]:

– Křivka nesmí být singulární (pokud toto kritérium není již zahrnuto v generování koeficientů):
  1. pro $q = p^m$, kde $p > 3$ a $m \geq 1$, musí koeficienty rovnice splňovat $4 * a^3 + 27 * b^2 \neq 0$,
  2. pro $q = 2^m$ musí být $b \neq 0$,
  3. pro $q = 3^m$ musí být $b \neq 0$.
– Křivka nesmí být supersingulární (Jinak je možno problém diskrétního logaritmu nad eliptickou křivkou E převést na problém diskrétního logaritmu nad tělesem F_q.):
  1. pro $q = p$, nesmí platit $p = \#E - 1$,
  2. pro $q = p^m$, kde $m > 1$, nesmí být hodnota $|\#E - 1|$ dělitelná prvočíslem p (pro $q = 2^m$ mají supersingulární křivky jinou rovnici, než je zavedeno v tomto textu, proto se v takto vygenerovaných elipticky křivkách supersingulární eliptické křivky nevyskytují).
– Křivka musí splňovat MOV viz [9] podmínku (zobecnění supersingularity): $q^B \not\equiv \mod n$ pro $B \in 1, ..., 20$. (Jinak je možno problém diskrétního logaritmu nad eliptickou křivkou E převést na problém diskrétního logaritmu nad tělesem F_{q^B}. Volba testovacího rozsahu pro hodnotu B je určena tak, aby řešení problému diskrétního logaritmu nad tělesem F_{q^B} bylo výpočetně náročnější než řešení diskrétního logaritmu nad příslušnou eliptickou křivkou.)
– Křivka nesmí být anomální: nesmí platit $\#E = q$. (Jinak je možno najít izomorfizmus eliptické křivky E do aditivní grupy tělesa F_q a tam vyřešit úlohu diskrétního logaritmu nad eliptickou křivkou velice jednoduše.)
– Kofaktor musí mít hodnotu nejvýše čtyři. (Jinak je konstrukce neefektivní a bezpečnost eliptické křivky nižší než odpovídají předpoklady pro volbu tělesa.)

Pro tělesa typu F_{p^m}, kde $m > 1$ je požadováno volit hodnotu parametru m prvočíselnou, v opačném případě je možno provést útok popsaný v [19].

Pro tělesa typu F_p, se doporučuje [17] volit hodnotu koeficientu $a = -3$, což zaručuje hodnotu kofaktoru rovnu 1.

V některých případech muže být žádoucí prokázat, že křivka je vygenerována opravdu náhodně. V tom případě se negenerují přímo parametry ale náhodný řetězec Seed, ze kterého se jednosměrnou funkcí odvodí náhodné parametry eliptické křivky. Potom se Seed stává součástí veřejných parametrů pro případnou verifikaci [23].

5.2 Volba efektivních parametrů

Tato problematika svým rozsahem překračuje rámec tohoto článku, bližší informace lze nalézt např. v [21].

5.3 Křivky doporučené NIST

Doporučená konečná tělesa:

1. prvočíselná tělesa: F_p pro $p = 2^{192} - 2^{64} - 1$, $p = 2^{224} - 2^{96} + 1$, $p = 2^{256} - 2^{224} + 2^{192} + 2^{96} - 1$, $p = 2^{384} - 2^{128} - 2^{96} + 2^{32} - 1$, $p = 2^{521} - 1$,
2. binární tělesa: $F_{2^{163}}$, $F_{2^{233}}$, $F_{2^{283}}$, $F_{2^{409}}$, $F_{2^{571}}$.

Kritéria výběru těles:

1. délka řádu tělesa je dvojnásobkem délky tajného klíče běžného symetrického algoritmu. Dosud nejúčinnější metodou na řešení úlohy ECDLP je totiž Pollardova rho metoda (viz [8]), jejíž složitost je řádově $(\pi * n/2)^{1/2}$ kroků. Pokud je $n = 2^{256}$, dostáváme cca 2^{128} kroků, což je zhruba na úrovni luštitelnosti symetrické blokové šifry se 128-bitovým klíčem a z výpočetního hlediska neřešitelné. Proto říkáme, že příslušná šifra je výpočetně bezpečná. Úlohu lze sice paralelizovat, takže pokud použijeme N procesorů, dostáváme složitost $(\pi * n/2)^{1/2}/N$, ale pro velká n je to stále výpočetně neřešitelná úloha.
2. Pro prvočíselná tělesa jsou prvočísla p tzv. zobecněná Mersenova čísla viz [7].
3. Pro binární tělesa F_{2^m} bylo m vybráno tak, že existuje Koblitzova křivka [3] prvočíselného řádu nad F_{2^m}.

Tabulka 2. Doporučená velikost těles podle FIPS 186-2 [17].

Délka symetrického klíče	Příklad algoritmu	Délka p pro F_p	Velikost m pro F_{2^m}
80	Skipjack	192	163
112	3-DES	224	233
128	AES 128	256	283
192	AES 192	384	409
256	AES 256	521	571

Doporučené křivky:

1. Náhodné eliptické křivky nad F_p.
2. Koblitzovy křivky nad F_{2^m}.
3. Náhodné eliptické křivky nad F_{2^m}.

Příklad 3. Křivka P-192, $E : y^2 = x^3 - 3 * x + b \ (mod\ p)$

- prvočíselný modul $p = 2^{192} - 2^{64} - 1$
  $= 6277101735386680763835789423207666416083908700390324961279$,
- prvočíselný řád křivky $\#E = 6277101735386680763835789423176059013767194773182842284081$,
- kofaktor $= 1$, protože řád křivky je prvočíslo,
- $a = -3$,
- $b = 64210519e59c80e70fa7e9ab72243049feb8deecc146b9b1 (hex.)$,
- (generující) bod P: $x_P = 188da80eb03090f67cbf20eb43a18800f4ff0afd82fff1012$ (hex.),
- $y_P = 07192b95ffc8da78631011ed6b24cdd573f977a11e794811$ (hex.),
- $n = 6277101735386680763835789423176059013760590137671947731828422840 81$.

Doporučení pro výběr parametrů kryptografických systémů obsahuje celá řada standardů (SEC [11], FIPS 186-2, ANSI X9.62 [14], ISO 15946 [18], IEEE P1363 [13]), vzájemně jsou však mezi nimi odchylky. Často citovanou normou pro digitální podpis je FIPS 186-2 [17], která zrovnoprávňuje podpis na bázi RSA, DSA i ECDSA. ECDSA, vychází z normy ANSI X9.62. Ta, stejně jako FIPS 186-2, pak těží z práce skupiny P1363 organizace IEEE, která definuje řadu asymetrických algoritmů, včetně těch na bázi eliptických křivek. ECC se zabývá i ANSI norma X9.63. Další skupinu tvoří různé normy ISO používající ECC: například ISO 14888-3 definuje digitální podpis, ISO/IEC 15946 [18] definuje podpisy, šifrování a výměnu klíče, ISO/IEC 9798-3 autentizaci a ISO/IEC 11770-3 klíčové hospodářství. Dále jsou k dispozici různé internetové standardy IETF, využívající eliptické křivky pro internetové použití [24], standardy WAP fóra pro bezdrátové komunikace, zejména mobilní telefony (například Wireless Transport Layer Security, [25]).

6 Závěr

V současnosti se staly eliptické kryptosystémy alternativou ke klasickým asymetrickým kryptosystémům. Mají své výhody zejména v rychlosti a menší náročnosti na hardware i software. Nasazení eliptických kryptosystémů se zdá být pomalé. Příčinou může být to, že klasické asymetrické kryptosystémy jsou používány, studovány a známy déle. Avšak výhodou kryptosystémů na bázi eliptických křivek je jejich velká kryptografická bezpečnost vzhledem k dané velikosti klíče. Význačně kratší délka klíčů oproti klasickým kryptosystémům vede k menším parametrům systému, a tedy i k větší výpočetní efektivnosti algoritmů. Další výhodou je, že fakticky všechna již známá použití v systémech na bázi diskrétního logaritmu (např. ElGamal) lze převést do systémů na bázi eliptických křivek, což se podařilo zejména při převodu DSA na ECDSA.

Reference

1. http://csrc.nist.gov/CryptoToolkit/aes/rijndael/
2. Koblitz N. *A Course in Number Theory and Cryptography*. Springer-Verlag, 1994, New York. ISBN 0-387-94293-9.
3. Koblitz N. CM-curves with good cryptographic properties. *Advances in Cryptology - Crypto 91, Lecture Notes in Computer Science, 576 (1992), 279-287*, Springer-Verlag, 1992.
4. Miller V. Uses of elliptic curves in cryptography. *Advances in Cryptology Crypto 85, Lecture Notes in Computer Science, 218 (1986) 417-426*, Springer-Verlag, 1986.
5. National Institute of Standards and Technology, Digital Signature Standard, FIPS Publication 186. 1994.
6. Stallings W. *Cryptography and Network Security*. Prentice Hall, 1999, New Jersey. ISBN 0-13-869017-0.
7. Solinas, Jerome A.: Generalized Mersenne Numbers, Technical Reports, CORR 99-39, University of Waterloo. 1999. http://www.cacr.math.uwaterloo.ca/techreports/1999/corr99-39.pdf.
8. Pollard J. Monte Carlo methods for index computation mod c. *Mathematics of Computation, 32 (1978), 918-924*. 1978.
9. Menezes A., Okamoto T. and Vanstone S. Reducing elliptic curve logarithms to logarithms in a finite field. *IEEE Transactions on Information Theory, 39 (1993), 1639-1646*. 1993.
10. Menezes A., van Oorschot P. Vanstone S. *Handbook of Applied Cryptography*. http://cacr.math.uwaterloo.ca/hac/.
11. Standards for Efficient Cryptography Group: SEC 1: Elliptic Curve Cryptography, ver. 1.0. 2000. http://www.secg.org.
12. Standards for Efficient Cryptography Group: SEC 2: Recommended Elliptic Curve Domain Parameters, ver. 1.0. 2000. http://www.secg.org.
13. IEEE P1363: Standard Specifications for Public-Key Cryptography, last draft. 2000. http://www.grouper.ieee.org/groups/1363.
14. ANSI X9.62: Public Key Cryptography for the Financial Services Industry: The Elliptic Curve Digital Signature Algorithm (ECDSA), draft. 1999.
15. Lenstra A., Verheul E. Selecting Cryptographic Key. *Journal of Cryptology*. 2001.
16. ANSI X9.63: Public Key Cryptography for the Financial Services Industry: Elliptic Curve Key Agreement and Key Transport Protocol, draft. 1998.
17. FIPS 186-2: Digital Signature Standard (DSS), Federal Information Processing Standard Publication 186-2. 2000. http://csrc.nist.gov.
18. ISO/IEC 15946-1: Cryptographic Techniques based on Elliptic Curves: Part 1 General, DIN Berlin. 2000. http://www.din.de/ni/sc27.
19. Galbraith, S. D. and Smart, N. P. A cryptographic application of Wail descent. *Cryptography and Codes, pp. 191-200*. 1999.
20. Mersenne Primes: History, Theorems and Lists. http://www.mersenne.org/prime.htm/.
21. Dolník K. Hlediska výběru eliptických křivek pro asymetrické kryptosystémy. *Workshop Velikonoční kryptologie*. 2002.
22. http://www.aec.cz/
23. http://www.certicom.com/.
24. http://www.ietf.org/.
25. http://www.wapforum.org/.