Obsah

I Základní pojmy faktorizace čísel a testů prvočíselnosti, včetně některých jednodušších postupů

I.1 Úvod

Především je třeba zdůraznit, že úloha rozložit číslo na jeho činitele se jeví být podstatně složitější než úloha rozhodnout, zda dané číslo je či není prvočíslo. To je podstata veřejného klíče RSA.

Faktorizace (rozklad) složeného čísla je svou povahou rekurzivní — je třeba nalézt nějakého vlastního dělitele, provést dělení a rozkládat dále nalezené činitele. Lze také říci, že se snažíme dané číslo rozštěpit, takže o nalezení vlastního dělitele se mluví jako *rozštěpení* (splitting) daného čísla.

Nejjednodušší metodou, jak nalézt vlastního dělitele, je *zkusmé dělení* (trial division). I takovýto jednoduchý algoritmus poskýtá prostor pro drobná vylepšení. Doporučuje se začínat jakékoliv pokusy o rozštěpení daného čísla zkusmým dělením prvočísly až do 500000 (těch je 41538).

Další algoritmy se podle svého určení dělí do několika skupin. O *testu složenosti* (compositness test) mluvíme u postupů, které u některých čísel odhalí, že jde o číslo složené, avšak ne u všech. Testy složenosti většinou vycházejí z malé Fermatovy věty. Pokud je dáno číslo n a nalezne se $a < n$, jež nesplňuje $a^{n-1} \equiv 1 \pmod{n}$, tak n nemůže být prvočíslo. Potíž je v tom, že existují složená čísla n, jež splňují $a^{n-1} \equiv 1 \pmod{n}$ pro všechna $a < n$ nesoudělná s n; říká se jim *Carmichaelova*. Nejmenším Carmichaelovým číslem je $561 = 3 \cdot 11 \cdot 17$. Ví se, že pro každé x existuje alespoň x^C Carmichaelových čísel menších než x, kde C je konstanta blízká $1/10$.

Je dobré si uvědomit, že význam malé Fermatovy věty pro testy složenosti je spjat s tím, že mocnění modulo n umíme provádět efektivně. To vynikne srovnáním s Wilsonovou větou, která říká, že $(p-1)! \equiv -1 \pmod{p}$ pro každé prvočíslo p. (Navíc pro n složené, $n \neq 4$, vždy máme $(n-1)! \equiv 0 \pmod{n}$.) Pokud bychom uměli efektivně počítat faktoriály modulo n, byl by k dispozici efektivní algoritmus zjišťující prvočíselnost. Zatím však žádný efektivní algoritmus, který by využíval Wilsonovu větu, znám není.

Pokud testy složenosti vedou ke zjištění, že jde o číslo složené, mohou nastoupit vlastní štěpící algoritmy. Pokud testy složenosti nedají jednoznačný závěr, lze s vysokou mírou pravděpodobnosti (která je u řady testů explicitně známa) předpokládat, že jde o prvočíslo. Algoritmům, které tento předpoklad potvrdí nebo vyvrátí, se říká *testy prvočíselnosti* (primality tests).

I.2 Rabin–Millerův test

Ať $n > 1$ je liché kladné číslo a ať $n - 1 = 2^t q$, ke q je liché. Řekneme, že n je *silné pseudoprvočíslo v bázi* a, jestliže buď $a^q \equiv 1 \pmod{n}$ nebo $a^{2^e q} \equiv -1 \pmod{n}$ pro nějaké e, $0 \leq e < t$.

Pokud $n = p$ je prvočíslo, tak q-tá mocnina každého $a \in \mathbb{Z}_p^* \cong C_{p-1}$ padne do 2^t-prvkové podgrupy $\mathbb{Z}_p^*$. Tato podgrupa má jediný prvek řádu 2, a ten je roven -1. Je-li a^q řádu 2^f, kde $1 \leq f \leq t$, tak je (2^{f-1})-tá mocnina a^q rovna -1.

Vidíme, že prvočíslo p je silné pseudoprvočíslo pro každou bázi $a < p$. Později vyložíme, že pro n složené je bází menších než n vždy méně než $n/4$. Předpokládejme, že n je složené. Pokud zvolíme $a < n$ náhodně, je pravděpodobnost, že jde o bázi, menší než $1/4$. Pokud vybíráme $a < n$ náhodně c-krát za sebou, a vždy vybereme bázi, můžeme se s pravděpodobností $< 4^{-c}$ domnívat, že se jedná o prvočíslo.

Uvedený postup se nazývá Rabin–Millerův test. Jeho efektivita je založena na efektivitě mocnění modulo n. Při volbě malých a lze toto mocnění provádět efektivněji, takže test proběhne rychleji. Účinnost takového zrychleného postupu by bylo dobré podepřít teoretickým výsledkem, že bází není více než čtvrtina i v nějakém počátečním úseku intervalu $0, 1, \ldots, p-1$. O takovém výsledku nevím, přesto se zmíněné urychlení používá.

Přistupme nyní k odhadům počtu bází složených čísel. Budeme potřebovat několik přípravných tvrzení.

I.2.1 Lemma. *Ať $A_{i,j}$ jsou po dvou disjunktní množiny, kde $1 \leq i \leq k$ a $0 \leq j \leq m$. Ať $A_{i,j}$ má $\max(2^{j-1}, 1)$ prvků. Pak existuje právě $1 + \frac{2^{mk}-1}{2^k-1} = \frac{2^{mk}+2^k-2}{2^k-1}$ posloupností $(a_1, \ldots, a_k)$, že lze nalézt j, pro které $a_i \in A_{i,j}$ pro všechna i, $1 \leq i \leq k$.*

Důkaz. Pro $j = 0$ máme jednu možnost, pro $j > 0$ máme $2^{k(j-1)}$ možností. Celkem tedy $1 + \sum_{0 \leq j \leq m-1} (2^k)^j = 1 + \frac{2^{km}-1}{2^k-1}$. □

I.2.2 Lemma. *Ať $G_1, \ldots, G_k$ jsou cyklické 2-grupy a ať $G = G_1 \times \cdots \times G_k$. Počet všech $(a_1, \ldots, a_k) \in G$, u nichž se řády $a_i \in G_i$ shodují, je $\leq |G|/2^{k-1}$. K rovnosti dochází pouze když G je cyklická nebo elementární abelovská.*

Důkaz. Ať 2^m je nejnižší z řádů všech grup G_i a ať $H_i \leq G_i$ jsou řádu 2^m. Položme $H = H_1 \times \cdots \times H_k$ a předpokládejme, že tvrzení platí pro H. Ať je $H < G$. Žádný prvek $G \setminus H$ nemá požadovanou vlastnost, takže takových prvků je $|H|/2^{k-1} < |G|/2^{k-1}$.

Buď nyní $H = G$. Ať je $|H_1| = |H_2| = \cdots = |H_k| = 2^m$. Definujme $A_{i,j} \subseteq H_i$ tak, že v $A_{i,j}$ leží prvky řádu 2^j. Z lemmatu I.2.1 plyne, že těchto prvků je $1 + (2^{km} - 1)/(2^k - 1)$. Chceme tedy ukázat, že je

$$2^{k-1}\left(1 + \frac{2^{km}-1}{2^k-1}\right) \leq 2^{km},$$

přičemž rovnost nastává pouze pokud $k = 1$ nebo $m = 1$. Vztah můžeme zapsat jako

$$2^{k-1}((2^k - 1) + (2^{km} - 1)) = 2^{2k-1} + 2^{k(m+1)-1} - 2^k \leq 2^{k(m+1)} - 2^{km},$$

tedy jako

$$2^{km} + 2^{2k-1} \leq 2^{k(m+1)-1} + 2^k = 2^{km+(k-1)} + 2^k.$$

Případy $k = 1$ a $m = 1$ jsou zřejmé. V ostatních případech je

$$2^{km+(k-1)} \geq 2^{km} + 2^{km} > 2^{km} + 2^{2k-1}.$$

□

I.2.3 Lemma. *Uvažme $n = p_1^{e_1} \cdots p_k^{e_k} > 1$, kde $p_1 < \cdots < p_k$ jsou lichá prvočísla a $e_1, \ldots, e_k$ jsou kladná. Označme $\mu : \mathbb{Z}_n^* \cong \mathbb{Z}_{p_1^{e_1}}^* \times \cdots \times \mathbb{Z}_{p_k^{e_k}}^*$ kanonický isomorfismus, ve kterém $\mu(a) = (a_1, \ldots, a_n)$, $a \equiv a_i \pmod{p_i^{e_i}}$, $1 \leq i \leq k$, a ať $B \leq \mathbb{Z}_n^*$ a $C \leq \mathbb{Z}_n^*$ jsou po řadě maximální 2-podgrupa a maximální podgrupa lichého řádu. Konečně ať $B_i \leq \mathbb{Z}_{p_i^{e_i}}^*$ je také maximální 2-podgrupa. Každé $a \in \mathbb{Z}_n^*$ lze jednoznačně psát jako bc, kde $b \in B$ a $c \in C$. Uvažme podmínku, že*

některá lichá mocnina a je rovna 1 nebo některá mocnina a je rovna -1.

Prvek $a \in \mathbb{Z}_n^*$ tuto podmínku splňuje právě když $\mu(a) = (b_1, \ldots, b_k) \in B_1 \times \cdots \times B_k$ má tu vlastnost, že řády všech b_i, $1 \leq i \leq k$, se shodují.

Důkaz. Prvočísla p_i jsou lichá, a proto je každá z grup $\mathbb{Z}_{p_i^{e_i}}^*$ cyklická, $1 \leq i \leq k$. Požadavek, aby se nějaká lichá mocnina a rovnala 1, je totožný s požadavkem $a \in C$, tedy $b = 1$. Zbytek je jasný z toho, že $\mu(-1) = (-1, \ldots, -1)$. □

I.2.4 Lemma. *Ať $n = p_1^{e_1} \cdots p_k^{e_k} > 1$, kde $p_1 < \cdots < p_k$ jsou prvočísla a e_1, …, e_k jsou kladná. Potom počet $a \in \mathbb{Z}_n^*$ takových, že žádné p_i, $1 \leq i \leq k$, nedělí řád a v $\mathbb{Z}_n^*$, je $\leq (p_1 - 1) \cdots (p_k - 1)$.*

Důkaz. Ať μ je stejný isomorfismus jako v lemmatu I.2.3. Pokud $a \in \mathbb{Z}_n^*$ splňuje požadovanou podmínku, pak a_i musí padnout do $(p_i - 1)$-prvkové podgrupy $\mathbb{Z}_{p_i^{e_i}}^*$. □

I.2.5 Lemma. *Ať $n = p_1^{e_1} \cdots p_k^{e_k} > 1$, kde $p_1 < \cdots < p_k$ jsou lichá prvočísla a e_1, …, e_k jsou kladná. Je-li $a \in \mathbb{Z}_n$ takové, že n je silné pseudoprvočíslo v bázi a, tak je $a \in \mathbb{Z}_n^*$ a jeho řád v $\mathbb{Z}_n^*$ není dělitelný žádným p_i, $1 \leq i \leq k$.*

Důkaz. Vhodná mocnina a je rovna 1, a proto musí být každá báze a invertibilní. Je-li $a \in \mathbb{Z}_n^*$ řádu dělitelného p_i, tak $a^{n-1} \not\equiv 1$, neboť p_i nedělí $n - 1$. □

Přistupme nyní k odhadu počtu bází složeného čísla. Ať tedy $n = p_1^{e_1} \cdots p_k^{e_k}$ je složené číslo, kde $p_1 < \cdots < p_k$ jsou prvočísla a e_1, …,e_k jsou kladná. Je-li $a \in \mathbb{Z}_n^*$ báze, tak jistě splňuje podmínku lemmatu I.2.3 (všimněte si však, že tato podmínka poukazuje na výrazně silnější vlastnosti než jsou vlastnosti báze).

Z lemmat I.2.2, I.2.3 a I.2.5 tedy vyplývá, že bází je $\leq \varphi(n)/2^{k-1} < n/2^{k-1}$. Pro $k \geq 3$ tak dostáváme hledaný odhad $< n/4$.

Je-li $k = 1$ a $n = p^e$, p liché prvočíslo, tak z lemmat I.2.4 a I.2.5 plyne, že bází je $\leq (p-1)$. Protože je vždy $4(p-1) < p^2 \leq p^e$, dostáváme i zde okamžitě odhad $< n/4$.

Zbývá případ $n = p_1^{e_1} p_2^{e_2}$. Je-li ovšem $p_i - 1 < p_i^{e_i}/4$ pro některé i, $1 \leq i \leq 2$, je jistě $(p_1-1)(p_2-1) < (p_1^{e_1} p_2^{e_2})/4$, takže pro $e_1 > 1$ nebo $e_2 > 1$ lze opět použít lemmata I.2.4 a I.2.5 Ať tedy $e_1 = 1 = e_2$.

Mějme $n - 1 = 2^t q$, $p_1 - 1 = 2^{t_1} q_1$ a $p_2 - 1 = 2^{t_2} q_2$, kde q, q_1 a q_2 jsou lichá. Ať G_i jsou cyklické grupy řádu t_i, $1 \leq i \leq 2$, a ať H_i jsou jejich podgrupy řádu $t_3 = \min(t_1, t_2)$. Počet prvků $(b_1, b_2) \in H_1 \times H_2$ takových, že b_1 a b_2 mají stejný řád, je odhadnut lemmatem I.2.2 hodnotou $2^{2t_3 - 1}$. Pokud je $H_1 \times H_2$ vlastní podgrupa $G_1 \times G_2$, tak je počet bází $\leq 2^{2t_3+1}/4 = 2|H_1 \times H_2|/4 \leq |G_1 \times G_2|/4$. S ohledem na lemma I.2.3 to znamená, že nejvýše čtvrtina invertibilních prvků $\mathbb{Z}_n^*$ je pak bází. Tím pádem je bází méně než $n/4$, takže zbývá vyřešit případ $t_1 = t_2$.

Pro jeho řešení se jeví být klíčovým vztah

$$n - 1 = p_1 p_2 - 1 = (p_1 - 1)(p_2 - 1) + (p_1 - 1) + (p_2 - 1).$$

Z něj vidíme, že nemůže současně platit, aby $p_1 - 1$ dělilo $n - 1$ a $p_2 - 1$ dělilo $n - 1$. Kdyby tomu totiž tak bylo, tak by $p_1 - 1$ dělilo $p_2 - 1$ a naopak, takže bychom dostali $p_1 = p_2$. Z uvedeného vztahu dále vidíme, že 2^{t_1} dělí $p_1 p_2 - 1$ (neboť předpokládáme $t_1 = t_2$), takže je $t_1 \leq t$. Proto alespoň jedno q_i, $1 \leq i \leq 2$, nedělí q. Všechny báze leží v podgrupě $\mathbb{Z}_n^*$, která obsahuje prvky exponentu $2^t q$, a která je, jak jsme právě nahlédli, indexu alespoň 3. Podle lemmat I.2.2 a I.2.3 pouze polovina prvků této podgrupy může být bází, takže v $\mathbb{Z}_n^*$ jich může být nejvýše šestina. Odhad $< n/4$ tedy platí i v tomto případě. Můžeme uzavřít:

I.2.6 Tvrzení. *Buď n liché složené číslo. Pak počet bází $a \in \mathbb{Z}_n$, vůči kterým je n silné pseudoprvočíslo, je vždy menší než $n/4$.*

I.3 Pocklington–Lehmerův test

Hovořívá se též o Pocklingtonově–Lehmerově $N - 1$ testu, kde N je číslo, jež se po aplikaci Rabin–Millerova testu jeví být prvočíslem s vysokým stupněm pravděpodobnosti. Pocklington-Lehmerův test je typickým představitelem testů prvočíselnosti.

Místo N zde pišme n. Představme si, že $n - 1 = fu$ a že pro každý dělitel d čísla n platí $d \equiv 1 \pmod{f}$. Pak nemůže existovat vlastní dělitel d čísla n, který by splňoval $d \leq f$. Pokud je $f \geq \sqrt{n}$ a n je složené, tak takový vlastní dělitel existuje, takže k určení prvočíselnosti n stačí ověřit, že $d \equiv 1 \pmod{f}$ pro každý dělitel d čísla n. K tomu stačí ověřit $d \equiv 1 \pmod{p_i^{r_i}}$, kde $f = p_1^{r_1} \cdots p_k^{r_k}$ je prvočíselný rozklad f. Ovšem ověřovat $d \equiv 1 \pmod{p_i^{r_i}}$ pro všechny vlastní dělitele d čísla n by bylo absurdní, neboť nevíme, zda takoví dělitelé existují. Nicméně vztah $d \equiv 1 \pmod{p_i^{r_i}}$ lze obdržet za splnění určitých podmínek z následujícího tvrzení.

I.3.1 Tvrzení. *Ať je n kladné a ať p dělí $n-1$. Ať $a < n$ splňuje $a^{n-1} \equiv 1 \pmod{n}$ a ať $a^{(n-1)/p} - 1$ je nesoudělné s n (takže speciálně $a^{(n-1)/p} \not\equiv 1 \pmod{n}$). Ať m je největší takové, že p^m dělí $n-1$. Potom pro každý dělitel d čísla n platí $d \equiv 1 \pmod{p^m}$.*

Důkaz. Vztah $d \equiv 1 \pmod{p^m}$ zjevně stačí dokázat pro prvočíselné dělitele n. Ať je $d = q$ prvočíslo. Protože a je v $\mathbb{Z}_n$ invertibilní, je s n nesoudělné, a proto q nedělí a. Tudíž $a^{q-1} \equiv 1 \pmod{q}$. Uvažme nejmenší $e > 0$ takové, že $a^e \equiv 1 \pmod{q}$. Pak e dělí $q-1$, e nedělí $(n-1)/p$ (neboť $a^{(n-1)/p} \not\equiv 1 \pmod{q}$) a e dělí $n-1$. Tudíž maximální mocnina p, která dělí $n-1$ i e, je shodná, a rovna m. Proto p^m dělí $q-1$. □

Test tedy spočívá ve faktorizaci $n-1$, $n-1 = fu$, a v nalezení vhodných $a = a_p$ pro každé prvočíslo dělící f, $f \geq \sqrt{n}$. Při hledání a_p se opět využívá rychlost mocnění modulo n. Protože se neprochází všechny mocniny a, je otázka jak moc lze informací získaných o jednom a využít v dalších hledáních. Žádné poznatky v tomto směru mi nejsou známy. Zdá se mi zřejmé, že nesoudělnost $a^{(n-1)/p} - 1$ s n se ověřuje pouze poté, kdy se zjistí, že není $a^{(n-1)/p} \equiv 1 \pmod{n}$. Pokud se nalezne soudělnost, tak je nalezen vlastní dělitel n.

Číslo $f \geq \sqrt{n}$ lze zmenšit při použití následujícího tvrzení. Odpovídající postup se zdá být na místě, pokud má $n-1$ hodně prvočíselných dělitelů, ze kterých známe všechny dostatečně malé (což se například může stát, pokud při zkusmém dělení objevíme dostatek malých prvočíselných dělitelů).

I.3.2 Tvrzení. *Předpokládejme, že $n-1 = fu$, kde $f > 1$ a $u > 1$ jsou nesoudělné, a že existuje b, jež je menší než všechny prvočíselné dělitele u a splňuje $bf \geq \sqrt{n}$. Ať pro každé p dělící f lze najít $a < n$, jež splňuje $a^{n-1} \equiv 1 \pmod{n}$, přičemž $a^{(n-1)/p} - 1$ je nesoudělné s n. Pokud navíc existuje takové $a_u < n$, že $a_u^{n-1} \equiv 1 \pmod{n}$ a že $a_u^f - 1$ je s n nesoudělné, tak je n prvočíslo.*

Důkaz. Ať je $d = q$ nějaký prvočíselný dělitel n, $q \leq \sqrt{n}$. Z tvrzení I.3.1 plyne $q \equiv 1 \pmod{f}$. Uvažme nejmenší $e > 0$ takové, že $a^e \equiv 1 \pmod{q}$, kde $a = a_u$. Víme, že a je v $\mathbb{Z}_n$ invertibilní, takže q nedělí a a e dělí $q-1$. Současně e dělí $n-1$ a nedělí $f = (n-1)/u$. Tudíž e a u jsou soudělné, a jejich největší společný dělitel t je větší než b. Protože f a u jsou nesoudělné, tak $q \equiv 1 \pmod{e}$ a $q \equiv 1 \pmod{f}$ společně dávají $q \equiv 1 \pmod{ft}$, odkud $q > ft \geq fb \geq \sqrt{n}$. □

I.4 Pollardova metoda faktorizace

Představme si nejprve, že na m-prvkové množině M náhodně volíme zobrazení $f\colon M \to M$, přičemž sledujeme chování iterovaných obrazů nějakého pevně zvoleného bodu x. Nejmenší $i \geq 0$ takové, že existuje nekonečně mnoho $i' \geq 0$, jež splňují $f^i(x) = f^{i'}(x)$, nazvěme *preperiodou* f a nejmenší j takové, že $f^{i+j}(x) = f^i(x)$, nazveme *periodou* f (chápáno relativně k x). Pokud dráhu bodu x kreslíme z levého dolního rohu, pak tato dráha svým tvarem připomíná řecké písmeno ρ. Z tohoto důvodu se popisované metodě říká Pollardova ρ-metoda.

Lze dokázat, že průměrná délka periody je asymptoticky rovna $\sqrt{(\pi m)/8}$, a podobně i průměrná délka periody se asymptoticky blíží jistému násobku $\sqrt{m}$.

Zatím jsme ještě neřekli, jaká je vazba uvedených úvah na faktorizaci čísla n. Předpokládejme, že toto číslo není prvočíslo (to lze například zjistit testy složenosti a prvočíselnosti), a označme $p = m$ nějaký jeho hypotetický prvočíselný dělitel.

Předpokládejme, že $g : \mathbb{Z}_n \to \mathbb{Z}_n$ je nějaké zobrazení, u kterého lze předpokládat takový nedostatek jednoduše popsatelných zákonitostí, že z hlediska periodicity se chová podobně jako náhodná funkce. Vlastně nás takové chování nezajímá, přísně vzato, u zobrazení g, ale u zobrazení $f{:}\mathbb{Z}_p{\to}\mathbb{Z}_p$, které je určeno tak, aby pro $a \in \mathbb{Z}_p$ a $b \in \mathbb{Z}_n$, $b \equiv a \pmod p$, bylo $f(a) \equiv g(b) \pmod p$. V této definici je zahrnut předpoklad $g(b_1) \equiv g(b_2) \pmod p$ pro $b_1 \equiv b_2 \pmod p$; kdykoliv je splněn, lze f definovat. Jeho splnění je zaručeno, je-li $g = g(x)$ dáno nějakým polynomem ze $\mathbb{Z}_n[x]$. Pokud vím, tak v praxi se vystačí s polynomy $x^2 + c$, kde $c \in \{-1, 1\}$.

V dalším textu x neznamená polynomiální proměnnou, ale onen náhodně zvolený prvek $x \in \mathbb{Z}_p$, jehož iterace pomocí f sledujeme. Ovšem tak jako nevolíme f, ale g, tak nevolíme $x \in \mathbb{Z}_p$, ale $y \in \mathbb{Z}_n$, přičemž x splňuje, pro hypotetické prvočíslo p, vztah $x \equiv y \pmod p$.

Vtip celé metody spočívá v tom, že $\sqrt{p}$ je výrazně menší než p. Pokud se chování f dostatečně blíží průměrnému chování náhodně zvoleného zobrazení $\mathbb{Z}_p{\to}\mathbb{Z}_p$, tak budou existovat $i_1 < i_2$ splňující $f^{i_1}(x) = f^{i_2}(x)$, kde i_2 nepřesáhne hodnotu nějakého malého násobku $\sqrt{p}$. Prvočíslo p sice neznáme, avšak vztah $f^{i_1}(x) = f^{i_2}(x)$ se promítne do $\mathbb{Z}_n$ tak, že n a $g^{i_2}(y) - g^{i_1}(y)$ mají největšího společného dělitele různého od 1. Pokud tedy místo testu na rovnost $f^{i_1}(x) = f^{i_2}(x)$ budeme testovat největší společný dělitel n a $g^{i_2}(y) - g^{i_1}(y)$, máme naději, že při relativně malém počtu iterací zobrazení g odhalíme nějakého netriviálního dělitele n.

Tím je podstata Pollardova ρ-testu vyložena, zbývá se zmínit o jednom úskalí a vyložit některé důležité algoritmické aspekty. Úskalí spočívá v tom, že g a f mohou mít stejnou periodu, tedy že může být $g^{i_2}(y) \equiv g^{i_1}(y) \pmod n$. Taková situace poukazuje na fakt, že g nebo f nevykazují chování dostatečně blízké chování průměrnému, takže v takovém případě se volí funkce g znovu a jinak. Algoritmickými aspekty zde je pak míněn především způsob, kterým dochází k testu $f^{i_1}(x) = f^{i_2}(x)$ (budeme ho uvažovat v této formě, byť víme, že je reprezentován testem na největšího společného dělitele).

Bylo by jistě neefektivní při každé iteraci f provádět srovnání se všemi předchozími hodnotami. Původní postup, navržený Pollardem a Floydem, spočíval v tom, že se průběžně počítá jak $g^k(y)$, tak $g^{2k}(y) = (g^2)^k(y)$. Ve chvíli, kdy je $k \geq i$ (kde i je preperioda) a současně j dělí k (kde j je perioda), tak máme $f^k(x) = f^{2k}(x)$. Stačí tedy pouze testovat tuto rovnost — první zachycené k bude rovné $j\lceil i/j \rceil$, zjištěné k bude tedy v průměru v polovině periody (měřeno od okamžiku vstupu do periodické části — jde ovšem o heuristický odhad, který předpokládá nezávislost j a i/j).

Pollard–Floydův postup vykazuje určitou neefektivitu v tom, že testování rovnosti předchází trojí výpočet funkce g (počítáme $g(g^{i-1}(y))$, $g(g^{2i-2}(y))$ a

$g(g^{2i-1}(y))$. Brent navrhl jiný postup, který takový výpočet nepožaduje. Uvažme nejmenší e takové, že $f^{2^e-1}(x)$ padne do periodické části a že převyšuje délku periody (je tedy $2^e - 1 \geq i$ a $2^e \geq j$). Potom $f^{2^e-1}(x) = f^{2^e-1+j}(x)$, přičemž $2^e \leq 2^e - 1 + j < 2^{e+1}$. Hodnotu e sice neznáme, avšak vidíme, že můžeme postupovat následovně:

Pro $s = 0, 1, \ldots$ postupně zaznamenáme hodnotu $g^{2^s-1}(y)$ a srovnáváme ji s hodnotami $g^k(y)$, kde $2^s \leq k < 2^{s+1}$. Po jejich vyčerpání nahradíme $g^{2^s-1}(y)$ hodnotou $g^{2^{s+1}-1}(y)$ a pokračujeme obdobně dále. Hledanou rovnost odhalíme v prvém cyklu, ve kterém $e = s$ splňuje výše uvedené podmínky. Hodnota e je zjevně rovna $\lceil \lg \max(i+1, j) \rceil$. V průměru tedy lze očekávat (je-li můj naivní odhad správný), že $f^{2^e-1}(x)$ se bude nacházet zhruba v polovině periodické části. V Brentově postupu je třeba počítat $g^k(y)$ až do $k = 2^e - 1 + j$, což je v průměru (jak se zdá) o periodu více, nežli v případě Pollard-Floydova postupu. V praxi jsou oba postupy údajně co do účinnosti srovnatelné. Nicméně, jak ukážeme v tvrzení I.4.1, Brentův postup lze modifikovat tak, že nesrovnáváme $g^{2^s-1}(y)$ se všemi $g^k(y)$, $2^s \leq k < 2^{s+1}$, ale pouze s těmi, které splňují $3 \cdot 2^{s-1} = (3 \cdot 2^s)/2 \leq 2^{s+1}$. Tím se efektivita Brentova postupu zvýší natolik, že jasně převyšuje efektivitu Pollard–Floydova přístupu.

K implementaci Pollardovy ρ-metody budiž ještě poznamenáno, že výpočet největších společných dělitelů se doporučuje dávkovat — místo toho, aby se po řadě počítali největší společní dělitelé čísla n a čísel $g_1, \ldots, g_r$, tak se spočítá součin $g_1 \cdots g_r$ modulo n, a pak jeho největší společný dělitel s n. Pokud vyjde hodnota 1, tak je patrné, že všichni největší společní dělitelé jsou rovni 1. To je v algoritmu ten nejčastější případ — pokud je nalezen netriviální společný dělitel, algoritmus dosáhl svého cíle a není problém dodatečně zjistit, kde má netriviální společný dělitel svůj původ. Jako vhodná velikost dávky r se doporučuje hodnota $r = 20$.

Konečně ještě zmiňme, že při volbě c v polynomu $g(x) = x^2 + c$ je třeba se vystříhat hodnoty $c = -2$. Je-li $x = z + 1/z$, je $x^2 - 2 = z^2 + 1/z^2$, což vede na chování výrazně odlišné od chování průměrného zobrazení.

Pollardova ρ-metoda se údajně stále používá jako alternativa vůči zkusmému dělení — malé prvočíselné dělitele totiž odhalí rychleji než mocnější metody vzniklé později.

I.4.1 Tvrzení. *Ať je $f: M \to M$ zobrazení, $x \in M$, ať i a j po řadě označují (konečnou) preperiodu a periodu vzhledem k x. Pak existují $k \geq 0$ a $s \geq 0$, jež splňují $f^k(x) = f^{2^s-1}(x)$ a $3 \cdot 2^{s-1} \leq k < 2^{s+1}$. Nejmenší takové k je rovno 3, pokud $i = 0$ a $j = 1$, je rovno $2^r + j - 1$, pokud $i = 0$ a $j > 1$, a v ostatních případech je rovno $2^k + j\lceil (2^t + 1)/j \rceil - 1$. Přirozená čísla r a t jsou přitom určena vztahy $2^{r-1} < \max(i+1, j) \leq 2^r$ a $2^t \leq i < 2^{t+1}$.*

Důkaz. Hledáme nejmenší $k \geq 1$ takové, že pro nějaké $e \geq 0$ platí $3 \cdot 2^{e-1} \leq k < 2^{e+1}$ a $f^{2^e-1}(x) = f^k(x)$. Nerovnost $2^e + 2^{e-1} = 3 \cdot 2^{e-1} \leq k$ značí $2^e < k < 2^{e+1}$, z čehož plyne, že k určuje e jednoznačně a že musí být $e \geq 1$.

Případ $i = 0$ a $j = 1$ odpovídá situaci, kdy k je konstantní. Pak pro $e = 1$ dostáváme podmínku $3 \leq k < 4$, takže $k = 3$ je nejmenší možné řešení. V dalším již případ $i = 0$, $j = 1$ připouštět nebudeme.

Vztah $f^{2^e-1}(x) = f^k(x)$ lze zapsat jako dvojici podmínek, a to jednak $2^e - 1 \geq i$ a jednak podmínku, že $k - 2^e + 1$ je kladný násobek j. Z $k - 2^e + 1 \geq j$ a $2^{e+1} - 1 \geq k$ máme $2^e \geq j$, což spolu s $2^e \geq i + 1$ dává $e \geq r$. Volba $e = r$ je tedy nejmenší možná.

V případě $i = 0$ volíme pak $k = 2^r - 1 + j$ rovněž jako nejmenší možné. Vztahy $2^{r-1} < j \leq 2^r$ přitom vedou na $2^r + 2^{r-1} - 1 < k \leq 2^r + 2^r - 1$, čili na $3 \cdot 2^{r-1} \leq k < 2^{r+1}$, jak bylo požadováno.

Ať je $i > 0$. Je-li $2^t + 1 \leq j$, je $\lceil (2^t+1)/j \rceil = 1$ a $j \leq \max(i+1, j) \leq 2^r$. Je-li $2^t + 1 > j$, je $i + 1 \geq 2^t + 1 > j$, takže r je určeno nerovnicemi $2^{r-1} < i + 1 \leq 2^r$, jež jsou ekvivalentní nerovnostem $2^{r-1} \leq i < 2^r$, odkud $r - 1 = t$.

V případě $2^t + 1 \leq j$ máme $k = 2^r - 1 + j < 2^{r+1}$ a ze současné platnosti $2^t < j$ a $2^t < i + 1$ (tedy za platnosti $2^t < \max(i+1, j)$) plyne $t \geq r - 1$, takže $j > 2^t \geq 2^{r-1}$ a $2^r + 2^{r-1} \leq 2^r + j - 1 = k$. Vidíme, že zvolené k vyhovuje požadavkům tvrzení a že je nejmenší možné.

Buď nyní $2^t + 1 > j$. Pak $t = r - 1$ a platí $j\lceil (2^t + 1)/j \rceil = j(\lfloor 2^t/j \rfloor + 1) \leq 2^t + j \leq 2^{t+1} = 2^r$, takže $k = 2^r + j\lceil (2^t + 1)/j \rceil - 1 < 2^{r+1}$. Vidíme, že $e = r$ je nejmenší možná volba a že zbývá ověřit, zda $\lceil (2^t + 1)/j \rceil$ je skutečně nejmenší s takové, že $3 \cdot 2^{r-1} \leq (2^r - 1) + js$, tedy že jde o nejmenší s, jež splňuje $2^{r-1} + 1 \leq js$. Takovým nejmenším s ovšem je $\lceil (2^{r-1} + 1)/j \rceil$, což se shoduje s naší volbou, neboť $t = r - 1$. $\square$

I.5 O $p - 1$ metodě

Tato metoda faktorizace vede k cíli, pokud faktorizované číslo n má prvočíselné dělitele p určitých zvláštních vlastností. V základním tvaru $p - 1$ testu jde o jedinou vlastnost, totiž o to, aby $p - 1$ mělo pouze malé prvočíselné dělitele, kde „malé“, znamená menší než nějakou mez B, do které jsme schopni nebo ochotni prvočísla tabelovat a postupně procházet. Modifikace testu tuto vlastnost buď oslabují, nebo vycházejí od jiných hodnot než $p - 1$ (například $p + 1$, $p^2 + 1$, $p^2 + p + 1$ nebo $p^2 - p + 1$).

Číslo m se nazývá B-*hladké*, pokud všechny prvočíselné dělitele m jsou $\leq B$. Číslo m nazveme B-*mocné* (anglicky B-powersmooth), pokud všechny prvočíselné mocniny, které dělí m, jsou $\leq B$.

Zvolme nějaké číslo a (například $a = 2$). Je-li p prvočíselný dělitel n (přičemž předpokládáme $a < p$), tak je $a^{p-1} \equiv 1 \pmod{p}$. Tím pádem je $a^k \equiv 1 \pmod{p}$ pro každý násobek k čísla $p - 1$. V $p - 1$ testu se za k postupně volí čísla $p_1^{s_1}$, $p_2^{s_2}$, …, kde $p_1 = 2$, p_2, … jsou prvočísla $\leq B$ a s_i jsou nejvyšší exponenty, pro které platí $p_i^{s_i} \leq B$. Nevypočítávají se přímo hodnoty a^k (jež jsou velmi brzy zcela mimo dosah), ale tyto hodnoty modulo n. Pro každou z nich se pak zjišťuje nejvyšší společný dělitel a^k s n. Z předpokladu B-mocnosti $p-1$ vyplývá, že $p-1$ jistě dělí poslední z postupně uvažovaných hodnot k (jež je rovna nejmenšímu společnému násobku všech čísel $\leq B$), takže pokud má n takový dělitel p, že $p-1$ je B-mocné, tak pro všechna k od jisté hodnoty h je největší společný dělitel čísel $a^k - 1$ a n různý od 1. Je-li různý od n, dostáváme hledanou faktorizaci čísla n. Případ, kdy je roven n, je poměrně řídký a lze mu čelit volbou jiného počátečního a. Jestliže totiž n dělí a^{k-1}, kde $k = k' p_h^{s_h}$, jestliže n má rozklad $q_1^{t_1} \cdots q_r^{t_r}$ (takže ono hypotetické $p \mid n$ je rovno některému z q_j) a jestliže $b < n$ volíme tak, aby bylo $b \equiv a^{k'} \pmod{n}$, tak z předpokladu, že n dělí a^{k-1} a že b je nesoudělné s n, plyne $b^{p_h^{s_h}} \equiv 1 \pmod{q_j^{t_j}}$ pro každé j, $1 \le j \le k$. V takové situaci lze ještě zjistit nejmenší nenulový exponent $s \le s_h$, pro který je $b^{p_h^s}$ soudělné s n. Jestliže jejich největší společný dělitel je různý od n, tak jsme n rozštěpili. Jestliže $b^{p_h^s}$ dělí n, tak je multiplikativní řád b modulo $q_j^{t_j}$ roven p_h^s pro každé j, $1 \le j \le r$. To se samo o sobě jeví jako nepříliš časté, navíc jde o situaci, ve které řád a modulo $q_j^{t_j}$ nedělí $k' = p_1^{s_1} \cdots p_{s_h-1}^{s_{h-1}}$ pro žádné j, $1 \le j \le r$.

Základní myšlenka $p-1$ testu vychází, jak jsme viděli, z předpokladu, že multiplikativní řády a modulo nějaký prvočíselný dělitel čísla n lze sestavit z malých prvočísel. Případy, ve kterých $p-1$ test neposkytuje uspokojivou rychlost není obtížné najít — stačí například nalézt prvočísla p a q taková, že $2p+1$ a $2q+1$ jsou opět prvočísla, a položit $n = (2p+1)(2q+1)$. Jsou-li p a q přibližně stejně velká, tak se $p-1$ test nechová lépe nežli test zkusmým dělením.

Při implementaci $p-1$ testu se doporučuje (podobně jako u Pollardova ρ-testu) dávkovat násobením hodnoty $a^k - 1$ dříve než provedeme test na největšího společného dělitele s n (je-li dávkovaný součin s n nesoudělný, je to stejné, jako když je každý činitel nesoudělný).

Běžně se mez B údajně pohybuje okolo 10^6. Nyní uvedeme vylepšení $p-1$ testu, které pracuje jednak s mezí $B = B_1$, jednak s mezí $B_2 > B_1$, jež bývá okolo 10^8.

Vylepšený test má dvě fáze, z nichž prvá se shoduje s již popsaným postupem a pracuje s mezí B_1. Druhá fáze se vztahuje k mezi B_2. Z hlediska datových struktur je důležité, že zatímco prvočísla do p_1 tabelujeme explicitně, prvočísla mezi B_1 a B_2 máme zaznamenány ve formě tabulky rozdílů mezi každým prvočíslem p, $B_1 \le p < B_2$, a prvočíslem následujícím. Tyto rozdíly jsou vzhledem k velikosti prvočísla malé.

Vylepšený $p-1$ test hledá takové prvočíselné dělitele p čísla n, že $p-1$ je buď B_1-mocné (to je prvá fáze testu) nebo, že $p-1 = fq$, kde f je B_1-mocné a q je prvočíslo mezi B_1 a B_2 (údajně existuje i varianta, kdy se pouze předpokládá, že f je B_1-hladké).

V druhé fázi test probíhá v zásadě stejně jako v prvé. Na konci prvé fáze jsme zjistili (pokud nevedla k rozštěpení n), že $b-1$ je nesoudělné s n, kde $b < n$ a $b \equiv a^k \pmod{n}$, pro k nejmenší společný násobek čísel $1, \ldots, B_1$. Má-li n prvočíselného dělitele předpokládaného tvaru, je $a^{kq} \equiv 1 \pmod{n}$ pro nějaké prvočíslo q, $B_1 < q \le B_2$. Potřebujeme znát hodnoty $a^{pq} \equiv b^q \pmod{n}$, abychom mohli počítat největší společný dělitel čísel $a^{kq} - 1$ a n. Ovšem k výpočtu $b^q \mod n$ lze využít předchozího kroku a vynásobit již známou hodnotu hodnotou b^d, kde $q = q_1 + d$ a q_1 je prvočíslo předcházející q. Vtip modifikovaného testu spočívá v tom, že hodnoty $b^d \pmod{n}$ se spočítají předem, což je výpočtově výhodné, neboť maximální možné d je relativně malé. V průběhu testu se tedy pouze násobí předem spočítanými hodnotami, a tím je dáno, že hodnotu B_2 lze volit mnohem větší než hodnotu B_1. Platí se za to ovšem tím, že pravděpodobně n nerozštěpíme (byť to není zcela vyloučené) pokud každé $p-1$, jež je B_2-mocné, je dělitelné qq', kde q a q' jsou prvočísla větší než B_1.

Na závěr ještě něco málo k úvodní poznámce, že místo $p-1$ lze uvažovat i $p+1$, p^2+p+1 nebo p^2-p+1. Jde o to, že a lze uvažovat nejen jako prvek $\mathbb{Z}_p = \mathbb{F}_p$, ale i jako prvek $\mathbb{F}_{p^2}$, $\mathbb{F}_{p^3}$, $\mathbb{F}_{p^4}$ atd. V $\mathbb{F}_{p^2}$ je $a^{(p^2-1)} = a^{(p-1)(p+1)} = 1$, čili $a^{p+1} = 1$ v $\mathbb{F}^*_{p^2}/\mathbb{F}^*_p$. Potřebujeme tedy modelovat chování grupy $\mathbb{F}^*_{p^2}/\mathbb{F}^*_p$ pomocí výpočtů modulo n tak, abychom zkoumáním největších společných dělitelů opět dosáhli rozštěpení n. Podrobnosti budou uvedeny později.

II O kvadratických a číselných sítech

II.1 Úvodní myšlenky

Koncem šedesátých let a během sedmdesátých let byla za nejúčinnější metodu považována metoda CFRAC (continued fraction factoring algorithm). Tato metoda má podobně jako pozdější (a účinnější) metody kvadratického a číselného síta (quadratic and number field sieve) dvě fáze: v prvé se hledá velké množství kongruencí modulo N a druhá fáze má charakter řešení velkého počtu lineárních rovnic nad dvouprvkovým tělesem. Základní metodou řešení lineárních rovnic je Gaussova eliminace, a proto se té druhé fázi též říká gaussovská.

Hlavním motivem všech tří metod je hledání (a nalezení) takových čísel x a y, že je $x^2 \equiv y^2 \pmod N$, kde N je číslo, jež chceme rozštěpit. Pak N dělí $x^2 - y^2 = (x-y)(x+y)$, takže v případě, kdy N nedělí ani $x-y$ ani $x+y$ (což platí například když $1 \le x < y < N/2$), tak největší společný dělitel N s $x-y$ nebo s $x+y$ dává vlastního dělitele N.

Hledat taková x a y metodou pokusů a omylů by bylo bezcílným blouděním, místo toho se ve všech třech metodách hledají x_1, x_2, … taková, že pro daný soubor „malých“ prvočísel p_1, …, p_m sestavíme kongruence tvaru

$$x_k^2 \equiv (-1)^{e_{0k}} p_1^{e_{1k}} \cdots p_m^{e_{mk}} \pmod N.$$

Máme-li jich dostatečný počet, řekněme n, zkoumáme netriviální řešení homogenní soustavy rovnic s neznámými ε_1, …, ε_n tvaru

$$\varepsilon_1 e_{i1} + \varepsilon_2 e_{i2} + \cdots + \varepsilon_n e_{in} \equiv 0 \pmod 2.$$

Běží nám tedy o nalezení libovolného 0,1–vektoru $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)$, jenž splňuje $E\varepsilon^T = 0$, kde E je 0,1–matice o $m+1$ řádcích a n sloupcích, jejíž hodnota na pozici (i,k) je rovna $e_{ik} \pmod 2$, přičemž výpočet $E\varepsilon^T$ probíhá nad dvouprvkovým tělesem $\{0,1\}$.

Položíme-li $w_i = \sum_{k=1}^n \varepsilon_k e_{ik}$ (čili $(w_0, \dots, w_m)^T$ je rovno $(e_{ij})\varepsilon^T$ počítáno v $\mathbb{Z}$), budou čísla w_i sudá. Položme $v_i = w_i/2$, $0 \le i \le m$, a ať

$$x = x_1^{\varepsilon_1} \cdots x_n^{\varepsilon_n} \qquad \text{a} \qquad y = (-1)^{v_0} p_1^{v_1} \cdots p_m^{v_m}.$$

Potom $x^2 = x_1^{2\varepsilon_1} \cdots x_n^{2\varepsilon_n}$ a

$$\begin{aligned} y^2 &= (-1)^{w_0} p_1^{w_1} \cdots p_m^{w_m} = (-1)^{\Sigma \varepsilon_k e_{0k}} p_1^{\Sigma \varepsilon_k e_{1k}} \cdots p_m^{\Sigma \varepsilon_k e_{mk}} = \\ &= \prod_{k=1}^n \left((-1)^{e_{0k}} p_1^{e_{1k}} \cdots p_m^{e_{mk}}\right)^{\varepsilon_k} \equiv \prod_{k=1}^n (x_k^2)^{\varepsilon_k} = x^2. \end{aligned}$$

Řečeno jinými slovy: z vyjádření x_1^2, x_2^2, …, x_n^2 modulo N pomocí -1 a prvočísel p_i, $1 \le i \le m$, chceme vybrat podmnožinu takovou, že součet koeficientů u -1 a p_i je vždy sudý. Hodnoty x_j^2 z této podmnožiny vynásobíme a „přirozeně" odmocníme. V čem je skryté, že většinou nedostaneme modulo N tu samou hodnotu jako prostý součin čísel x_j? Důvod je v tom, že rovnice sestavujeme tak, aby matice E neměla nulové sloupce. Každý sloupec odpovídá vyjádření nějakého x_k^2 a nenulovost sloupce tedy značí, že některá z hodnot e_{ik}, $0 \le i \le m$, je lichá, což znamená, že modulo N je vazba mezi x_k a x_k^2 daná složitěji nežli skrze „přirozené" odmocňování půlením keoficientů. Speciálně vidíme, že čísla x_1, …, x_n je třeba volit větší než $\sqrt{N}/2$.

Všechny tři metody se liší způsobem hledání kongruencí $x^2 \equiv (-1)^{e_0} p_1^{e_1} \cdots p_m^{e_m}$. Je-li matice E již sestavena, je potřeba použít některou z technik práce s velmi řídkými maticemi.

Zde načrtneme zhruba způsob hledání kongruencí v případě metody CFRAC, ostatními metodami se budeme zabývat podrobněji v dalších oddílech.

Metody se liší tím, jak se získají kongruence tvaru $x^2 \equiv t \pmod{N}$, či — jinak vyjádřeno — pro jakou metodu faktorizace t tyto kongruence sestavujeme. Naším cílem totiž není pouze nalézt kongruenci $x^2 \equiv t$, ale toto t rozložit na $(-1)^{e_0} p_1^{e_1} \cdots p_m^{e_m}$, kde p_1, …, p_m jsou z nějaké předem dané množiny prvočísel (říká se jí *faktorizační báze*).

Zatímco v sítových metodách se kongruence volí tak, aby bylo proséváním možno rozhodnout, zda t je součinem prvočísel z faktorizační báze, v metodě CFRAC se prostě hledají kongruence $x^2 \equiv t$ takové, aby vyšlo t malé (neboť pro malá t lze doufat v snadné nalezení faktorizace, a tím ověřit, zda se t rozkládá nad naší faktorizační bází).

V metodě CFRAC se hledají metodou řetězových zlomků dobré racionální aproximace $P/Q \approx \sqrt{kN}$ pro měnící se hodnotu k. Z $P^2/Q^2 \approx kN$ vyplývá, že $t = P^2 - kNQ^2$ je relativně malé, přičemž $P^2 \equiv t \pmod{N}$.

Ptejme se, zda je důvod, aby všechny případy vztahů $a^2 \equiv t \pmod{N}$, kde t je malé, byly nalezeny tímto způsobem. Ze vztahu $a^2 \equiv t \pmod{N}$ vyplývá, že existují k a d takové, že $a^2 \equiv t + kd^2N$, kde k je nedělitelné čtverci (tak říkáme číslu, které nedělí žádné p^2, p prvočíslo — všimněte si, že i k z předchozího odstavce je rozumné volit jako číslo nedělitelné čtverci). Protože t je malé, je malé i $(a/d)^2 - kN = t/d^2$, takže $a/d \approx \sqrt{kN}$ je v jistém smyslu dobrá racionální aproximace čísla $\sqrt{kN}$. To je důvod, proč lze očekávat, že opačným postupem (tj. hledáním P a Q jak uvedeno výše) skutečně obdržíme všechny kongruence $x^2 \equiv t \pmod{N}$, kde t je malé.

Existují algoritmy, které racionální aproximaci P/Q rychle spočítají (a to právě metodou řetězových zlomků). Zde je uvádět nebudeme (alespoň ne v této verzi textu).

II.2 Myšlenka kvadratického síta

Kongruence tvaru

$$y^2 \equiv (-1)^{e_0} p_1^{e_1} \cdots p_m^{e_m},$$

kde $p_1, \ldots, p_m$ patří do vhodné faktorizační báze, budeme hledat pomocí kvadratických polynomů $Q(x) = Ax^2 + 2Bx + C$. Pro jednoduchost zápisu pišme $p_0 = -1$. Volbu koeficientů A, B a C provádíme tak, aby $Q(x)$ bylo jako celé číslo přímo rovno $p_0^{e_0} \cdots p_m^{e_m}$. Z hlediska použití v budovaném systému lineárních rovnic je potřeba, aby některý z koeficientů $e_0, \ldots, e_m$ byl lichý. V takovém případě nemůže v celých číslech platit $y^2 = Q(x)$, takže vztah x a y musí být založen na netriviálním použití modulární aritmetiky. V případech, kterým se budeme věnovat podrobněji, je $A\alpha^2 \equiv 1 \pmod{N}$ pro vhodné α, čili A je modulo N invertibilní čtverec. Dále je $B^2 - AC = hN$, pro nějaké $h \geq 1$. (Metody výběru polynomu $Q(x)$, které uvedeme, mají vždy $h = 1$, v obecnější části však připustíme i jiné hodnoty.) Dostáváme

$$AQ(x) = (Ax + B)^2 - (B^2 - AC) \equiv (Ax + B)^2 \pmod{N},$$

což dává $Q(x) \equiv ((Ax + B)\alpha)^2 \pmod{N}$, takže lze klást $y = (A\alpha + B)\alpha$.

Nejpřirozenější se jeví volba

$$Q(x) = (\lfloor\sqrt{N}\rfloor + x)^2 - N,$$

ve které $A = 1$, $B = \lfloor\sqrt{N}\rfloor$ a $C = \lfloor\sqrt{N}\rfloor^2 - N$. Základní metodou kvadratického síta se míní právě práce s tímto jedním polynomem. Jenom malá část hodnot $Q(x)$ se ovšem v dané faktorizační bázi rozkladá, a k jejich nalezení je třeba uvažovat velmi mnoho hodnot x. Přitom faktorizační báze nemůže být příliš velká, protože pak by byla i příliš velká potřebná soustava lineárníh rovnic. Vhodným se proto ukázalo uvažovat místo jednoho polynomu Q takových polynomů více. Mluví se pak o *kvadratickém sítu více polynomů* (Multiple Polynomial Quadratic Sieve, čili MPQS).

O volbě koeficientů A, B a C při použití této metody pojednáme později; nyní se soustředíme na to nejdůležitější v obou metodách, totiž na použití síta. Tomu ovšem nutně musí předcházet úvahy o dělitelnosti $Q(x)$ nějakým prvočíslem p z faktorizační báze, či — obecněji — úvahy o dělitelnosti číslem p^k, $k \geq 1$.

Podstatným je přitom případ, kdy p nedělí A. (Případům, kdy p dělí A, se lze buď zcela vyhnout, anebo je ošetřit analogickým způsobem.) Předpokládejme tedy, že p nedělí A. Potom $p^k \mid Q(x)$ právě když $p^k \mid AQ(x)$, a to nastane právě když $(Ax + B)^2 \equiv hN \pmod{p^k}$.

Tento vztah je možný pouze když hN je čtverec modulo p — proto se do faktorizační báze vkládají pouze taková prvočísla. Podobně se při práci s p^k předpokládá, že k nepřesáhlo nejvyšší možnou hodnotu, při které je hN čtverec modulo p^k.

Je-li $a^2 \equiv hN \pmod{p^k}$, kde p a hN jsou nesoudělná, je a modulo p^k rovna dvěma možným hodnotám v případě p lichého a v případě $p^k = 4$. Označme je a_{p^k} a b_{p^k}. V případě $p^k = 2$ je jediné možné řešení $a_2 = 1$, zatímco v případě $p = 2$ a $k \geq 3$ máme čtyři možná řešení a_{2^k}, b_{2^k}, c_{2^k} a d_{2^k}. Tato řešení ovšem existují pouze tehdy, pokud hN je čtverec modulo 2^k, tedy pokud $hN \equiv 1 \pmod{8}$. V případě $p^k = 4$ je odpovídající podmínka tvaru $hN \equiv 1 \pmod{4}$. Z recipročních vztahů pro Kroneckerův symbol $\left(\frac{a}{b}\right)$ (jenž zobecňuje Legenderův symbol) plyne rychlý algoritmus, pomocí kterého lze zjistit v obecném případě, zda hN je čtverec modulo p^k. Pokud tomu tak je, lze použít některý ze známých rychlých algoritmů pro nalezení nějakého řešení rovnice $a^2 \equiv hN \pmod{p^k}$. Ostatní řešení vyplývají ze znalosti prvků řádu 2 v grupě $\mathbb{Z}^*_{p^k}$.

Řešení rovnice $(Ax + B)^2 \equiv hN \pmod{p^k}$ získáme tedy zjevně tak, že vyjdeme od hodnot a, jež splňují $a^2 \equiv hN \pmod{p^k}$, a položíme $x = (a - B)A^{-1}$, kde A^{-1} se chápe modulo p^k. Z hodnot a_{p^k} a b_{p^k} (a případně též c_{p^k} a d_{p^k}) dostaneme vztahem $x = (a - B)A^{-1}$ hodnoty $a^{(Q)}_{p^k}$ a $b^{(Q)}_{p^k}$ (a případně též $c^{(Q)}_{p^k}$ a $d^{(Q)}_{p^k}$). Veškerá řešení v celých číslech kongruence $(Ax + B)^2 \equiv hN \pmod{p^k}$ jsou od nich odvozena přičtením násobku čísla p^k.

Znamená to, že máme jednu až čtyři hodnoty x, pro které je $p^k \mid Q(x)$, a víme, že všechny další takové sčítance odpovídají hodnotám $x + rp^k$, kde x je některé z výchozích $a^{(Q)}_{p^k}$, $b^{(Q)}_{p^k}$, …, a r je celé. Toto pozorování dovoluje konečně přistoupit k popisu použití principu síta.

Zvolíme *prosívací interval*, ze kterého vybíráme hodnoty x. Naším cílem je zjistit, pro která x z tohoto intervalu se $Q(x)$ plně rozkládá nad danou faktorizační bází. Tu volíme tak, aby obsahovala prvočísla až do určité meze $p_{\max}$, ovšem pouze ta, jež vyhovují výše uvedeným kritériím na čtverce. Rovněž exponenty k volíme s ohledem na tato kritéria do jisté hodnoty tak, aby bylo $p^k \le p_{\max}$. Ke každému x z prosívacího intervalu spočítáme $\ln|Q(x)|$, ovšem velmi hrubě — stačí zaokrouhlená celočíselná hodnota. Podobně pro každé p z faktorizační báze předem připravíme odhad hodnoty $\ln p$.

V operační paměti se udržuje pro každé x hodnota, jež vzniká úpravami počáteční hodnoty rovné výše uvedenému odhadu $\ln|Q(x)|$. Postupně se procházejí prvočísla p z faktorizační báze a pro každé p všechny jeho relevantní mocniny p^k. Zpracování určitého p^k znamená odečtení odhadu $\ln p$ ve všech případech, kdy p^k dělí $Q(x)$. Je-li k nejvyšší takový exponent, byl z odhadu $\ln|Q(x)|$ odečten odhad $\ln p$ celkem k–krát, tedy byl odečten odhad $k \ln p = \ln p^k$. Je-li $|Q(x)| = p_1^{k_1} \cdots p_m^{k_m}$, je $\ln|Q(x)| = \sum k_i \ln p_i$, takže $Q(x)$ se nad danou faktorizační bází rozkládá, pokud po skončení průběhu popsaného algoritmu bude uvažovanému x odpovídat hodnota blízká nule. Přesněji vyjádřit, co tato blízkost nule znamená, je ovšem dosti obtížné, neboť záleží na způsobu vyvážení snahy o nalezení všech případů, kdy se $Q(x)$ nad danou faktorizační bází rozkládá, a snahy o co největší účinnost, která se projevuje mírou zaokrouhlení $\ln p$. Pokud se vystačí s $\lfloor \ln p \rfloor$, což se zdá být standardní postup, tak u mocnin malých prvočísel mohou kumulované chyby vést k mylnému zařazení x z hlediska rozkladu $Q(x)$ nad danou faktorizační bází. Vyložíme, proč taková situace nezpůsobuje potíže, a také jak lze v celém postupu využít případy, kdy výsledná hodnota odpovídající určitému x je natolik malá, aby bylo zřejmé, že se $Q(x)$ nad faktorizační bází asi plně nerozkládá, že však jeho částečný rozklad $Q(x) = fs$, kde f lze nad danou bází faktorizovat, poskytuje s relativně malé.

Předně je třeba opětovně zdůraznit, že hodnot $Q(x)$, jež se nad faktorizační bází rozkládají, je relativně velmi málo. Proto lze v případech, kdy u x máme malou výslednou hodnotu, uplatnit postup dělením hodnotami p^k (či pouze p), a nespoléhat na logaritmické aproximace. Přitom samozřejmě lze využít kongruencí typu $x \equiv a_{p^k}^{(Q)} \pmod{p^k}$, a neprovádět dělení v případě, že tato kongruence není splněna.

Práce s mocninami prvočísel se zdá být méně účinná než práce s většími prvočísly faktorizační báze. Proto se například v prvé popsané implementaci MPQS mocniny p^k s $k \geq 2$ vůbec při prosívání neuvažovaly. Jinou možností je neprovádět prosívání malých prvočísel (řekněme do 100), zvolit větší toleranci pro zařazení $Q(x)$ do rozkládání dělením a najít hodnoty $Q(x)$, jež se nad faktorizační bází opravdu rozkládají, teprve v této fázi.

Volba větší tolerance umožňuje i jistou nadstavbu dosavadního postupu, pomocí které lze počet nalezených kongruencí ještě zvětšit. Pokud totiž zaznamenáváme i rozklady typu $Q(x) = fs$, kde f se rokládá nad faktorizační bází, a s je relativně malé, dochází (díky narozeninovému paradoxu) k situacím, ve kterých $Q(x_1) = f_1 s_1$, $Q(x_2) = f_2 s_2$ a $s_1 = s_2$. Označíme-li společnou hodnotu s, máme $y_1^2 \equiv f_1 s \pmod{N}$, $y_2^2 \equiv f_2 s \pmod{N}$, a tedy $(y_1 y_2 s^{-1})^2 \equiv f_1 f_2 \pmod{N}$. Hodnota s^{-1} je vztažena k $\mathbb{Z}_N^*$, což je na místě — kdyby s nebylo invertibilní, nalezli bychom hodnotu netriviálně soudělnou s N.

II.3 Volba koeficientů

Při práci s více polynomy je důležité je umět volit efektivně tak, aby jejich volba nepředstavovala větší výpočetní zátěž. Té se samozřejmě nelze plně vyhnout — například při výpočtu $a_{p^k}^{(Q)} = (-B + a_{p^k})A^{-1}$ musíme modulo p^k invertovat koeficient A polynomu $Q(x) = Ax^2 + 2Bx + C$.

Volba koeficientů A, B a C se řídí snahou o minimalizaci hodnot $Q(x)$. Proto je přirozené prosívací interval, řekněme délky $2M$, volit tak, aby v jeho středu leželo minimum paraboly $x \mapsto Q(x)$, tedy hodnota $-B/A$ (vždy předpokládáme $A > 0$). Volíme tedy prosívací interval $[-B/A - M, -B/A + M]$ tak, že pro každé x z tohoto intervalu je $Q(-B/A) \leq Q(x) \leq Q(-B/A + M)$. Přitom $Q(-B/A) = C - B^2/A$ a $Q(-B/A + M) = C - B^2/A + M^2 A$.

Položme $N = B^2 - AC$. Potom $Q(-B/A) = -N/A$ a $Q(-B/A + M) = -N/A + M^2 A$. Budeme hledat takové A, aby maximum z absolutních hodnot $Q(-B/A)$ a $Q(-B/A + M)$ bylo co nejmenší. Je-li $Q(-B/A + M) \leq 0$, bude toto maximum rovno N/A. Přitom $-N/A + M^2 A \leq 0$ právě když $A \leq \sqrt{N}/M$ a pak $N/A \geq M\sqrt{N}$. Je-li $Q(-B/A + M) \leq 0$, je tedy hledané maximum rovno alespoň $M\sqrt{N}$.

Ukážeme, že při volbě $Q(-B/A + M) > 0$, tedy $A > \sqrt{N}/M$, lze dosáhnout menší hodnoty maxima $Q(x)$, $-B/A - M \leq x \leq -B/A + M$. Funkce $A \mapsto N/A$ je klesající a funkce $A \mapsto -N/A + M^2 A$ je pro $A > \sqrt{N}/M$ rostoucí. Nejmenší možné hodnoty maxima proto dosáhneme v průseku obou funkcí, tedy pro $A = \sqrt{2N}/M$. Pak $Q(-B/A + M) = -Q(-B/A) = M\sqrt{N/2}$. Nahlédli jsme, že celočíselný koeficient A je vhodné volit tak, aby platilo $A \approx \sqrt{2N}/M$.

Při generování polynomů Q lze tedy postupovat tak, že se volí A blízké $\sqrt{2N}/M$. jež je nesoudělné s N, a pro které je N čtverec modulo A. Z poslední podmínky plyne existence B, které splňuje $B^2 \equiv N \pmod{A}$. Dopočítáním dostaneme $C = (B^2 - N)/A$. K určení B použijeme některý z algoritmů pro rychlé odmocňování. Tyto algoritmy ovšem předpokládají znalost prvočíselného rozkladu koeficientu A. Nejjednodušší postup je volit A vždy prvočíslo.

Jiný postup je založen na volbě A jako součinu několika (řekněme deseti) středně velkých prvočísel, jež nedělí N a modulo které je N čtverec. Výhodou tohoto postupu je větší variabilita při volbě koeficientu B. Při stejném A lze totiž pak B volit 2^t způsoby, kde t je počet prvočísel v rozkladu A. Použitím více polynomů Q se stejnou volbou A se ušetří čas potřebný pro invertování A modulo p^k ve výpočtech $a_{p^k}^{(Q)}$.