South Bohemia Mathematical Letters
Volume 25, (2017), No. 1, 34-38.

POLLARDŮV ρ – ALGORITMUS PRO FAKTORIZACI PŘIROZENÝCH ČÍSEL

JAROSLAV HORA

ABSTRAKT. S rozvojem moderních šifrovacích metod jde ruku v ruce i zvýšený zájem o moderní algoritmy pro hledání rozkladu složeného čísla v součin prvočísel. Faktorizační metoda, navržená J. M. Pollardem v r. 1975, byla svého času nejlepším známým algoritmem pro rozklad složených přirozených čísel. Jde o metodu využívající pravděpodobnost a lze ji vysvětlit i studentům středních škol.

ÚVOD

Je dobře známo, že v posledních desetiletích velmi vzrostl zájem o problematiku rozkladu (faktorizace) přirozených čísel a o další otázky z oblasti teorie čísel, a to zejména v souvislosti s moderním šifrováním zpráv (RSA šifrovací metoda). Běžná metoda pokusného dělení (trial division) vyžaduje při hledání případných vlastních dělitelů přirozeného čísla N zkoušet (prvo)číselné dělitele od 2 až do $\left[\sqrt{N}\right]$. Nejsou –li žádní, je N prvočíslem, jinak jsme nalezli vlastní dělitel složeného čísla N.

Metoda pokusného dělení je tedy zároveň testem prvočíselnosti i technikou pro nalezení vlastního dělitele přirozeného čísla N. Pro účely běžné školní výuky, kdy volíme „malá“ přirozená čísla N, je dobře využitelná: známe „seznam“ několika malých prvočísel a požadovaných pokusných dělení nebývá mnoho. Je – li ale N „velké“ přirozené číslo, vyjeví se nevýhody metody – my lidé (ale ani počítače) nemáme v paměti dostatečně velký seznam prvočísel, a hlavně je zřejmé, že testovacích dělení bude muset proběhnout velice mnoho. Ukazuje se, že metoda opakovaného dělení je pomalá a pro velká N ji nelze nechat v úplnosti proběhnout.

Seznamme se tedy s jinou faktorizační metodou, kterou navrhl John Pollard již v roce 1975, tedy před nástupem počítačové techniky. Jde o metodu Monte Carlo, tj. postup využívající pravděpodobnosti. Idea metody je dostupná středoškolským studentům.

1. NAROZENINOVÝ PARADOX

Příklad 1.: Sešla se skupina n osob, $n \geq 2$. Vypočtěte pravděpodobnost, že alespoň dva z nich budou mít narozeniny stejný den v roce. (Pro jednoduchost předpokládejme, že rok má 365 dní, tj. nebereme v úvahu přestupné roky).

Řešení: Určeme pravděpodobnost komplementárního jevu A, že žádné dvě osoby z n zúčastněných nebudou mít narozeniny týž den v roce.

Spočtěme nejprve, kolik je všech možných případů pro data narození těchto n osob. Pro den narození první osoby je 365 možností, pro dvě osoby je tedy 365^2 možností, pro n osob 365^n možností.

Vypočtěme dále, kolik z těchto případů je příznivých. První osoba se může narodit kterýkoli den v roce. Je -li ale již známo její datum narození a nemá -li druhá osoba mít narozeniny týž den, „zbývá“ pro její datum narození již jen 365 – 1 = 364 možností. Nemá -li datum narození

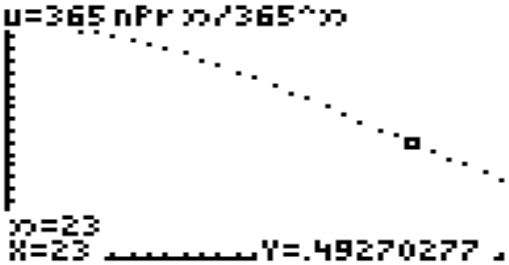
Key words and phrases. Faktorizace přirozených čísel, metoda opakovaného dělení, narozeninový paradox, Pollardův ρ – algoritmus.

žádných dvou osob z n zúčastněných připadnout na stejný den v roce, je 365.364. .... .(365 – n + 1) příznivých případů. Proto

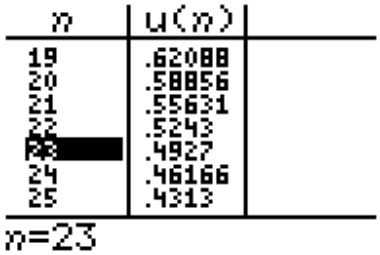
$$P(A) = \frac{365.(365-1).(365-2)....(365-n+1)}{365^n} \qquad (1)$$

je pravděpodobnost jevu, že žádné dvě osoby z *n* zúčastněných nebudou mít narozeniny týž den v roce.

Výpočet numerické hodnoty P(A) pro jednotlivé hodnoty *n* je vhodné přenechat nějakému pomocníku. V devadesátých letech minulého století jsme s oblibou používali grafické kalkulátory třídy TI 92 či program Derive, nyní je k dispozici řada dalších programů.



OBR. 1



OBR. 2

Pro *n* = 23 je P(A) = 0,49270, tj. pravděpodobnost jevu, že mezi 23 osobami nemají žádné dvě narozeniny týž den v roce, je již menší než 50 %. Je tedy „pravděpodobnější“, že nastane jev komplementární, tj. že aspoň dva lidé z této skupiny slaví narozeniny týž den v roce. Dále je možné si nechat několik hodnot vypsat do tabulky. Obr. 3 zachycuje displej kalkulátoru, na němž je uvedeno několik hodnot P(A) z (1) pro *n* = 19, 20, ..., 25.

Proč se mluví o „narozeninovém paradoxu“, když výpočet je korektní a nic paradoxního na něm není? Kdybychom o předpověď výsledku požádali nematematika, Hilbertova OMZU, obyčejného muže z ulice, dočkali bychom se asi o hodně vyššího odhadu čísla *n*. Celou věc si můžeme představit ještě jedním způsobem. Kdybychom házeli nikoli běžnou hrací kostkou, ale pravidelným mnohostěnem o 365 stěnách, pak jsme určili pravděpodobnost, že dojde k opakování již hozeného čísla. K opakování dojde „brzy“ a právě tato skutečnost se s výhodou využije v Pollardově ρ – algoritmu. K jeho popisu teď přistoupíme.

2. VSAĎ – VYHRAJEŠ SNAD!

Podle vyprávění mých rodičů byl tento slogan prvorepublikovým reklamním heslem jisté loterie. (Heslo bylo vcelku korektní, v dnešní době je třeba vzbudit zájem publika enormními výhrami pro náhodného jednotlivce a dosáhnout dojmu, že taková výhra čeká na každého). Ale i když je Pollardův ρ – algoritmus metodou Monte Carlo, tedy pravděpodobnostní, je vymyšlen korektně a vyhrajeme „skoro vždy“.

Máme faktorizovat složené přirozené číslo *N*, které není mocninou prvočísla. (Poslední podmínka se zřejmě snadno otestuje a rovněž existují testy na složenost přirozeného čísla). Zkonstruujeme jistou posloupnost $\{x_n\}$ přirozených čísel takto. Buď x_0 libovolné přirozené číslo. Další členy posloupnosti počítejme z kongruence $x_{i+1} \equiv x_i^2 + 1 \pmod N$.

Jak se uvádí v [2]: „The choice of this iteration function is black magic, but linear polynomials do not work, and higher degree polynomials are more costly to evaluate, and one cannot prove more about them than about $x^2 + 1$."

Nyní již můžeme přistoupit k formulaci Pollardovy ρ – metody. Nechť N je přirozené číslo, které chceme faktorizovat. Zvolme přirozené číslo x_0 a vypočtěme jistý počet členů posloupnosti nezáporných celých čísel $\{x_i\}$, definovaných následovně:

$x_{i+1} \equiv x_i^2 + 1 \pmod N$.

Předpokládejme, že prvočíslo p je nejmenším prvočíselným dělitelem čísla N. Dále předpokládejme, že $\{x_i\}$ je posloupností pseudonáhodných čísel modulo p. Dále, nechť pro jisté dva členy x_n, x_m, $n > m$ této posloupnosti platí, že $x_n \equiv x_m \pmod p$. (Příklad 1 nám dává naději, že na výskyt takovýchto členů posloupnosti $\{x_i\}$ nebudeme muset čekat dlouho). Platí tedy $p|(x_n - x_m)$, $p \mid N$, tedy největší společný dělitel $D(x_n - x_m, N) > 1$. Pokud zároveň $x_n \not\equiv x_m \pmod N$, je $D(x_n - x_m, N) < N$ a nalezli jsme netriviální vlastní dělitel čísla N. Navíc je výhodné, že tento dělitel lze nalézt Eukleidovým algoritmem, což je velice efektivní metoda.

Krajně neuspokojivé by ale bylo, kdybychom vskutku museli počítat $D(x_n - x_m, N)$ pro všechny dvojice $n, m \in N$, $n > m$. Objem výpočtů by narostl tak, že by to zcela znehodnotilo praktické využití navrhované metody. Naštěstí lze velkou část těchto propočtů ušetřit. Postačí totiž testovat jen dvojice x_{2i}, x_i , $i \in \mathrm{N}$, jak posléze ukážeme. Nyní zapišme nástin algoritmu pro Pollardovu ρ – metodu a pro malý kalkulátor:

Je dáno přirozené číslo N, které chceme rozložit.

1. Zvolme přirozené číslo x_0.
2. Vypočtěme $x_{i+1} \equiv x_i^2 + 1 \pmod N$, $i = 0, 1, \ldots$
3. Pro jistý počet $i \in N$ vypočtěme $D(x_{2i} - x_i, N)$.
4. Opakujme to do té doby, než $D(x_{2i} - x_i, N)$ je netriviálním dělitelem čísla N– **úspěch.**
   Pokud proces běží přes daný časový limit – **neúspěch**.

Příklad 2: Rozložme číslo $N = 527$ pomocí Pollardovy ρ – metody.

Řešení: Volme kupř. $x_0 = 3$ a vypočtěme dalších deset členů posloupnosti $\{x_i\}$. Máme $x_1 = 10$, $x_2 = 101$, $x_3 = 189$, $x_4 = 413$, $x_5 = 349$, $x_6 = 65$, $x_7 = 10$, $x_8 = 101$, $x_9 = 189$, $x_{10} = 413$. Dále

$D(x_2 - x_1, \mathrm{N}) = D(101 - 10, 527) = 1$,

$D(x_4 - x_2, \mathrm{N}) = D(413 - 101, 527) = 1$,

$D(x_6 - x_3, \mathrm{N}) = D(65 - 189, 527) = 31$.

Nalezli jsme netriviální dělitel čísla N = 527, je 527 **= 31 . 17**.

Teď také vidíme, odkud se vzalo písmeno ρ v názvu metody. První člen $x_0 = 3$ tvoří předperiodu, pak dochází k zacyklení, neboť $x_1 = x_7 = 10$. Kdybychom si získanou posloupnost zakreslili na papír ve tvaru uzlového grafu, dostali bychom zřejmě obrázek, který připomíná řecké písmeno ρ. (Jindy může být předperioda delší a podobnost grafického znázornění členů posloupnosti s písmenem ρ výraznější).

Pro úspěch metody je rozhodující snížení výpočetní náročnosti, kdy nemusíme testovat počítat $D(x_n - x_m, N)$ pro všechny dvojice $n, m \in \mathrm{N}$, $n > m$, ale jen pro dvojice tvaru x_{2i}, x_i , $i \in \mathrm{N}$. To souvisí s tzv. Floydovým trikem pro hledání cyklu, který se v posloupnosti $\{x_i\}$ mod p objeví.

Věta: Nechť p je prvočíslo dělící číslo N a nechť x_0 je dané přirozené číslo. Nechť v posloupnosti $\{x_i\}$, kde $x_{i+1} \equiv x_i^2 \pmod p$, $i = 0, 1, \ldots$, existují taková $m, n \in \mathrm{N}$, $m < n$, že $x_n \equiv x_m \pmod p$. Potom pro jisté $t \in \mathrm{N}$ platí $x_{2t} \equiv x_t \pmod p$.

Návod k důkazu: Postupujme podle tohoto schématu:

1. Pišme $n = m + d$, $d \geq 1$. Ukažme, že $x_{m+1} \equiv x_{m+d+1} \pmod p$ a dále indukcí, že $x_{m+r} \equiv x_{m+d+r} \pmod p$ pro všechna $r \in \mathrm{N}$.

2. Mezi čísly m, $m + 1$, ..., $m + d - 1$ je právě jedno násobkem čísla d. Předpokládejme, že k je ten index z množiny $\{0, 1, ..., d - 1\}$, pro který $d \mid m + k$. Potom $m + k = d \,.\, e$ pro jisté $e \in \mathrm{N}$, $x_{ed} \equiv x_{m+k} \equiv x_{m+k+d} \equiv x_{ed+d} \pmod p$.

3. Dokažme, že obdobně platí $x_{ed} \equiv x_{ed+2d} \pmod p$, $x_{ed} \equiv x_{ed+3d} \pmod p$, ..., $x_{ed} \equiv x_{ed+ed} \equiv x_{2}$ed $\pmod p$. Položíme -li nyní $t = ed$, máme $x_{2t} \equiv x_t \pmod p$, což bylo dokázati.

Jak jsme naznačili, existují případy, kdy Pollardova ρ – metoda selže.

Příklad 3: Pokusme se rozložit $N = 1241$ s tím, že $x_0 = 6$.

Řešení: Napišme několik prvních členů posloupnosti $\{x_i\}$: $x_0 = 6$, $x_1 = 37$, $x_2 = 129$, $x_3 = 509$, $x_4 = 954$, $x_5 = 464$, $x_6 = 604$, $x_7 = 1204$, $x_8 = 129$, $x_9 = 509$, $x_{10} = 954$, $x_{11} = 464$, $x_{12} = 604$, $x_{13} = 1204$, $x_{14} = 129$, $x_{15} = 509$, $x_{16} = 954$ atd.

Dále $\mathrm{D}(x_2 - x_1, 1241) = \mathrm{D}(129 - 37, 1241) = 1$, $\mathrm{D}(x_4 - x_2, 1241) = \mathrm{D}(954 - 129, 1241) = 1$, $\mathrm{D}(x_6 - x_3, 1241) = \mathrm{D}(604 - 509, 1241) = 1$, $\mathrm{D}(x_8 - x_4, 1241) = \mathrm{D}(129 - 954, 1241) = 1$, $\mathrm{D}(x_{10} - x_5, 1241) = \mathrm{D}(954 - 464, 1241) = 1$. Poté ale $\mathrm{D}(x_{12} - x_6, 1241) = \mathrm{D}(145 - 145, 1241) = 1241$. Následuje $\mathrm{D}(x_{14} - x_7, 1241) = \mathrm{D}(129 - 1204, 1241) = 1$, $\mathrm{D}(x_{16} - x_8, 1241) = \mathrm{D}(954 - 129, 1241) = 1$, $\mathrm{D}(x_{18} - x_9, 1241) = \mathrm{D}(604 - 509, 1241) = 1$, $\mathrm{D}(x_{20} - x_{10}, 1241) = \mathrm{D}(129 - 954, 1241) = 1$, $\mathrm{D}(x_{22} - x_{11}, 1241) = \mathrm{D}(954 - 464, 1241) = 1$. Poté opět $\mathrm{D}(x_{24} - x_{12}, 1241) = \mathrm{D}(604 - 604, 1241) = 1241$ a začíná být patrné, že existují (vcelku vzácné) případy, kdy Pollardova ρ – metoda neposkytne výsledek.

Kdybychom ale volili $x_0 = 7$, dostali bychom $x_1 = 50$, $x_2 = 19$, $x_3 = 362$, $x_4 = 740$, $x_5 = 320$, $x_6 = 639$, $x_7 = 33$, $x_8 = 1090$, $x_9 = 464$, $x_{10} = 604$, $x_{11} = 1204$, $x_{12} = 129$, $x_{13} = 509$, $x_{14} = 954$.

Poté máme $\mathrm{D}(x_2 - x_1, 1241) = \mathrm{D}(19 - 50, 1241) = 1$, $\mathrm{D}(x_4 - x_2, 1241) = \mathrm{D}(740 - 19, 1241) = 1$, $\mathrm{D}(x_6 - x_3, 1241) = \mathrm{D}(639 - 362, 1241) = 1$, $\mathrm{D}(x_8 - x_4, 1241) = \mathrm{D}(1090 - 740, 1241) = 1$, $\mathrm{D}(x_{10} - x_5, 1241) = \mathrm{D}(604 - 320, 1241) = 1$, $\mathrm{D}(x_{12} - x_6, 1241) = \mathrm{D}(129 - 639, 1241) = 17$.

Snadno se zjistí, že $1241 = 17.73$ je hledaný rozklad. Změna čísla x_0 tedy pomohla.

Pollardova ρ – metoda by mohla být v případě malých N dostupná i pro žáky ZŠ, protože vyžaduje provádění operací s přirozenými čísly, které jsou žákům známy. Studentům SŠ se zájmem o výpočetní techniku a programování by mohla poskytnout pokročilejší podněty.

Závěrem poznamenejme, že začátkem osmdesátých let minulého století byla Pollardova ρ – metoda nejlepším dostupným algoritmem pro faktorizaci přirozených čísel. V roce 1981 se např. Brentovi a Pollardovi zdařilo pomocí této metody rozložit osmé Fermatovo číslo $\mathrm{F}_8 = 2^{2^8} + 1 = p_{16}.p_{62}$, tedy v součin dvou prvočísel, majících 16, resp. 62 cifer. Je známo (viz [1]) i porovnání časové náročnosti metody opakovaného dělení a Pollardovy ρ – metody: je –li n (zhruba) počet cifer čísla N, je časová náročnost metody opakovaného dělení úměrná $2^{n/2}$, kdežto u Pollardova postupu $2^{n/4}$ a to je pro velká n významný rozdíl.

ZÁVĚR

Dnes jsou k dispozici lepší faktorizační metody – jenže jsou složité a asi je nebudeme moci vysvětlit našim žákům.

LITERATURA

[1] von zur Gathen, J, Gerhard, J.: *Modern Computer Algebra*, Cambridge University Press, 1999.

[2] Childs, L. N.: *A Concrete Introduction to Higher Algebra*, Springer, New York, third ed., 2009.

KATEDRA MATEMATIKY, FYZIKY A TECHNICKÉ VÝCHOVY FPE ZČU PLZEŇ, ČESKÁ REPUBLIKA
E-mail address: horajar@kmt.zcu.cz