Centrum voor Wiskunde en Informatica
Centre for Mathematics and Computer Science

H.J.J. te Riele

Factoriseren en Primaliteitstesten, een Inleiding

Afdeling Numerieke Wiskunde Notitie NM-N8804 Oktober

The Centre for Mathematics and Computer Science is a research institute of the Stichting Mathematisch Centrum, which was founded on February 11, 1946, as a nonprofit institution aiming at the promotion of mathematics, computer science, and their applications. It is sponsored by the Dutch Government through the Netherlands Organization for the Advancement of Pure Research (Z.W.O.).

Factoriseren en Primaliteitstesten, een Inleiding

Herman J.J. te Riele
Centre for Mathematics and Computer Science
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

Een inleidend overzicht wordt gegeven van algoritmen voor het vinden van de ontbinding in priemfactoren van een gegeven natuurlijk getal. Hierbij wordt onderscheid gemaakt tussen factorisatie- en primaliteitsalgoritmen. Veel van de besproken algoritmen zijn gebaseerd op het gebruik van kettingbreuken, en op het rekenen in eindige groepen.

1980 Mathematics Subject Classification (1985 Revision): Primary 11A51; Secondary 11Y05, 11Y11.
Key Words & Phrases: Factorization, Primality, Continued fraction.

1. INLEIDING

De zogenaamde Fundamentele Stelling van de Rekenkunde zegt dat ieder natuurlijk getal N op precies één manier kan worden geschreven in de vorm

$$N = \prod_{i=1}^{s} p_i^{e_i}, \tag{1}$$

waarbij $p_1, p_2, \cdots, p_s$ priemgetallen zijn met $p_1 < p_2 < \cdots < p_s$ en waarbij $e_1, e_2, \cdots, e_s$ natuurlijke getallen zijn (de exponenten). Als $s = e_1 = 1$, dan is N een priemgetal. Het is een klassiek probleem om, gegeven N, de representatie (1) te vinden (ook wel genoemd de 'ontbinding' van N in priemfactoren). Een voor de hand liggende aanpak is om te proberen of N misschien deelbaar is door, achtereenvolgens, de priemgetallen 2, 3, 5, 7, enz. Als een priemdeler gevonden is wordt deze 'eruit gedeeld' en de procedure wordt voortgezet met het quotient. Deze algoritme kan worden beëindigd als er een getal M overblijft dat geen priemdelers $\leqslant M^{1/2}$ bezit: in dat geval moet M priem zijn, want ieder samengesteld getal heeft tenminste één priemdeler $\leqslant$ de wortel uit dat getal.

VOORBEELD. $N = 2444$. Dit getal heeft twee priemdelers 2. Het quotient 611 is niet deelbaar door 3, 5, 7, 11, maar deling door 13 geeft het quotient 47 met rest 0, dus $611 = 13 \times 47$. Het quotient 47 heeft geen priemdelers $\leqslant 47^{1/2}$ ($= 6.88..$), want die zijn al eerder geprobeerd, dus 47 is priem. Conclusie: de representatie (1) van 2144 is: $2144 = 2^2 \times 13 \times 47$. □

Helaas groeit de hoeveelheid werk, nodig om (1) te vinden, snel met N, in het bijzonder als N priem is. Neem bijvoorbeeld het getal $N = 2^{31} - 1 = 2147483647$. De wortel uit N is 46340.9.. en er zijn 4792 priemgetallen onder 46340. Het kost ons dus, in het ergste geval (d.w.z. als N priem is), 4792 'trials' om de representatie (1) van N te vinden, nog afgezien van de vraag hoe we aan de 4792 'trial' — priemgetallen moeten komen (N is inderdaad priem. In sectie 5 zullen we uitleggen hoe men, gebruikmakend van de speciale vorm die N heeft, met veel minder werk kan laten zien dat N priem is).

Gelukkig zijn er in de loop der eeuwen veel efficientere methoden om (1) te bepalen gevonden. De

Notitie NM-N8804
Centrum voor Wiskunde en Informatica
Postbus 4079, 1009 AB Amsterdam

ontdekking, ruim tien jaar geleden, van bepaalde cryptografische codes waarvan de moeite om ze te 'breken' bepaald wordt door de moeite die men moet doen om van bepaalde grote getallen N de representatie (1) te vinden, heeft een extra stimulans gegeven aan dit onderzoek.

2. De kleine stelling van Fermat

In 1640 schreef Pierre de Fermat een brief aan zijn vriend Bernard Frenicle de Bessy waarin hij het volgende resultaat meedeelde: als N een priemgetal is en als a een geheel getal is zo dat $ggd(a,N)=1$, dan geldt

$$a^{N-1} \equiv 1 \pmod{N}; \tag{2}$$

m.a.w., a^{N-1} geeft, na deling door N, de rest 1. Voorbeeld: kies $N=11$, $a=2$, dan is $2^{11-1}=1024$ en $1024=93\times 11+1$, dus inderdaad is de rest van 2^{11-1} bij deling door 11 gelijk aan 1. Deze zgn. *Kleine Stelling van Fermat* kan worden gebruikt om te beslissen of een getal samengesteld is, zonder dat we ook maar één priemdeler van dit getal hoeven te kennen: als a zo is dat $ggd(a,N)=1$ en als $a^{N-1} \pmod{N}$ niet gelijk is aan 1, dan kan N niet priem zijn. Voor grotere waarden van N lijkt het een flinke klus om $a^{N-1} \pmod{N}$ uit te rekenen, maar dit kan heel efficient gebeuren door gebruik te maken van de binaire representatie van de exponent $N-1$. We illustreren dit aan de hand van een voorbeeld: kies $N=899$ en $a=2$. De binaire representatie van 898 is 1110000010; we starten nu met $2^1 \pmod{N}$, rekenen vervolgens $2^{11(\text{binair})} \pmod{N}$ uit, daarna $2^{111(\text{binair})} \pmod{N}$ en bouwen zo verder de gewenste exponent op door steeds een nieuw binair cijfer aan de tot dan toe verkregen exponent toe te voegen. Voor de berekening betekent dat: voor ieder binair cijfer in de exponent een kwadratering (mod N) uitvoeren, en als dat cijfer een 1 was: ook nog vermenigvuldigen (mod N) met 2. Zo krijgen we het volgende rekenschema:

	exponent e		$2^e \pmod{899}$	
	1		2	
	11	$2 * 2$	$\equiv 4$	$4 * 2 = 8$
	111	$8 * 8$	$\equiv 64$	$64 * 2 = 128$
	1110	$128 * 128$	$\equiv 202$	
	11100	$202 * 202$	$\equiv 349$	
	111000	$349 * 349$	$\equiv 436$	
	1110000	$436 * 436$	$\equiv 407$	
	11100000	$407 * 407$	$\equiv 233$	
	111000001	$233 * 233$	$\equiv 349$	$349 * 2 = 698$
(898 =)	1110000010	$698 * 698$	$\equiv 845$	

We vinden dus dat $2^{898} \equiv 845 \pmod{899}$, dus het getal 899 kan niet priem zijn. Het aantal stappen dat we moeten nemen om de berekening uit te voeren is van de orde van grootte van ${}^2\log N$ (in dit voorbeeld: 9), en dit is veel goedkoper dan bv. $N-1$ keer a, modulo N, met zichzelf te vermenigvuldigen. Het voordeel van de 'binaire' berekening is ook dat we slechts hoeven te rekenen met getallen die hoogstens zo groot worden als het kwadraat van N.

Opmerking. In de Kleine Stelling van Fermat is sprake van de ggd (grootste gemeenschappelijke deler) van twee getallen. Deze deler kan efficient worden uitgerekend met behulp van de zgn. *Algoritme van Euclides.* Deze algoritme is gebaseerd op de relatie: $ggd(a,b)=ggd(a \bmod b, b)$. Aangezien $0 \leqslant a \bmod b < b$ kunnen we deze relatie, na verwisseling van de twee argumenten, herhaald toepassen, tot we $a \pmod{b}=0$ vinden. Het aantal uit te voeren stappen is van de orde van grootte van $\log(\max(a,b))$.

VOORBEELD. $ggd(8991,3293)=ggd(3293,3405)$ (want $8991(\text{mod } 3293)=2405)=ggd(2405,888)=ggd(888,629)=ggd(629,259)=ggd(259,111)=ggd(111,37)=ggd(37,0)=37$. □

Met de Kleine Stelling van Fermat kunnen we dus snel zekerheid krijgen over de vraag of het getal N samengesteld is. Helaas kunnen we niet, als we vinden dat $a^{N-1}\equiv 1(\text{mod } N)$, concluderen dat N priem is: voor zeer veel getallen N zou dat een juiste conclusie zijn, maar er zijn uitzonderingen! Deze getallen worden a-pseudopriemgetallen genoemd. Dat zijn dus *samengestelde* getallen N met een a zo dat $ggd(a,N)=1$ en $a^{N-1}\equiv 1(\text{mod } N)$. In zo'n situatie zouden we kunnen proberen een andere a te vinden waarvoor $a^{N-1}\not\equiv 1(\text{mod } N)$, maar er zijn helaas ook samengestelde getallen, zogenaamde Carmichael-getallen, waarvoor zulke a's niet bestaan, m.a.w. dat zijn samengestelde getallen N met de eigenschap dat voor *alle* a met $ggd(a,N)=1$ geldt dat $a^{N-1}\equiv 1(\text{mod } N)$. Het kleinste 2-pseudopriemgetal is $341=11*31$ en het kleinste Carmichael-getal is $561=3*11*17$. Onder de grens $25*10^9$ zijn er 1091987405 priemgetallen en 21853 oneven 2-pseudopriemgetallen. Het aantal Carmichael-getallen onder deze grens is 2163. Als we voor een willekeurig oneven getal $n<25*10^9$ vinden dat (2) geldt met $a=2$, en als we dat getal priem zouden verklaren, dan is de kans dat we het bij het verkeerde eind hebben dus erg klein. Deze methode is verder verfijnd en heeft geleid tot zgn. probabilistische priemtests die getallen priem verklaren met een uiterst kleine kans op een vergissing (maar zonder wiskundige zekerheid!). In sectie 5 komen we hier nog op terug.

3. KETTINGBREUKEN

Bij sommige factorisatiemethoden spelen *kettingbreuken* een belangrijke rol. Stel $z=z_0$ is een of ander reëel getal, dan definiëren we $q_0:=[z_0]$ (d.w.z. het grootste gehele getal dat in z is begrepen), $z_{i+1}:=1/(z_i-q_i)$, $q_{i+1}:=[z_{i+1}]$ $(i=0,1,2,\cdots)$. Dan kunnen we schrijven:

$$z_0 = q_0 + \cfrac{1}{q_1 + \cfrac{1}{\ddots \cfrac{1}{q_m + \cfrac{1}{z_{m+1}}}}}.$$

Een dergelijke uitdrukking heet de *enkelvoudige kettingbreuk* van z, en wordt kortweg aangeduid met $[q_0,q_1,\cdots,q_m,z_{m+1}]$. Het getal $C_m:=[q_1,q_2,\cdots,q_m]$ is een goede, rationale, benadering van z en wordt de m-de convergent van z genoemd. Dit getal $C_m:=A_m/B_m$ kan als volgt uit $q_0,q_1,\cdots,q_m$ worden berekend: $A_{-2}:=B_{-1}:=0$, $A_{-1}:=B_{-2}:=1$, en

$$A_{i+1} := q_{i+1}A_i+A_{i-1}, \quad B_{i+1} := q_{i+1}B_i+B_{i-1} \quad (i=-1,0,1,\cdots). \tag{3}$$

Een belangrijke klasse van kettingbreuken vormen die van de wortel uit een getal N dat zelf geen kwadraat is. We starten nu met $z_0=(P_0+N^{1/2})/Q_0$ waarbij $P_0=0$ en $Q_0=1$ (en dus $q_0=[N^{1/2}]$). Als we nu definiëren:

$$P_{i+1} := q_iQ_i-P_i \quad \text{en} \quad Q_{i+1} := (N-P_{i+1}^2)/Q_i \quad (i=0,1,2,\cdots), \tag{4}$$

dan geldt $z_{i+1}=(P_{i+1}+N^{1/2})/Q_{i+1}$ en alle P_j en Q_j zijn gehele getallen. Ook geldt

$$0<P_j<N^{\frac{1}{2}}, \quad 0<Q_j<2N^{\frac{1}{2}} \quad \text{voor alle} \quad j>0. \tag{5}$$

Hieruit volgt dat de kettingbreukontwikkeling van $N^{1/2}$ altijd *periodiek* is. Immers, volgens (5) is het totaal aantal mogelijke verschillende fracties z_i ten hoogste gelijk aan $[N^{1/2}][2N^{1/2}]<2N$. Dus na minder dan $2N$ stappen zal er een fractie z_i optreden die al eerder is voorgekomen.

Verder geldt voor teller en noemer A_n en B_n van de convergent C_n de belangrijke relatie:

$$A_j^2-NB_j^2 = (-1)^{j+1}Q_{j+1}. \tag{6}$$

Voor de feitelijke berekening van de kettingbreukontwikkeling van $N^{1/2}$ worden vaak de volgende formules gebruikt:

$$\begin{aligned} P_{i+1} &= [N^{\frac{1}{2}}]-R_i, \\ Q_{i+1} &= Q_{i-1}+q_i(P_i-P_{i+1}), \\ q_{i+1} &= [(P_{i+1}+[N^{\frac{1}{2}}])/Q_{i+1}], \\ R_{i+1} &= P_{i+1}+[N^{\frac{1}{2}}] \pmod{Q_{i+1}}, \end{aligned} \qquad i = 0,1,\cdots \tag{7}$$

waarbij $R_0=0$ en $Q_{-1}=(N-P_0^2)/Q_0$.

VOORBEELD. Voor $N=143$ vinden we de volgende waarden voor P_i, Q_i, q_i, R_i, A_i en B_i ($i=0,1,2,3$):

i	P_i	Q_i	q_i	R_i	A_i	B_i
−2					0	1
−1		143			1	0
0	0	1	11	0	11	1
1	11	22	1	0	12	1
2	11	1	22	0	275	23
3	11	22	1	0	287	24

Voor $i=1$ en $i=3$ hebben (P_i,Q_i) dezelfde waarden, waaruit de periodiciteit van de kettingbreukontwikkeling blijkt.

Bij het uitvoeren van de algoritmen die we hier beschrijven zal het vaak voorkomen dat de gehele getallen waarmee we willen rekenen groter zijn dan de woordlengte van de gebruikte computer. Dat betekent dat we zgn. multi-precisie routines nodig hebben voor het uitvoeren van de elementaire rekenkundige operaties optellen, aftrekken, vermenigvuldigen en delen. Een goede beschrijving van de vele hiervoor bekende technieken kan men vinden in de boeken van Knuth: The Art of Computer Programming (zie literatuurlijst).

4. FACTORISATIEMETHODEN

4.1. Geheugenzuinige methoden

In de inleiding is al de trial-and-error factorisatiemethode genoemd. Hiervoor moet men de beschikking hebben over de priemen tot een zekere grens. Indien men een computer met weinig geheugen heeft kan deze methode ook worden uitgevoerd met behulp van een rij van oneven getallen die (bijna) alle priemgetallen bevat, en die op eenvoudige wijze gegenereerd kan worden. Neem bv. de rij van oneven getallen met termen die achtereenvolgens, mod 30, gelijk zijn aan 1, 7, 11, 13, 17, 19, 23 en 29. Als we de priemen 2, 3 en 5 apart behandelen, kunnen we de rij starten bij 7, en de volgende termen met behulp van de incrementen 4, 2, 4, 2, 4, 6, 2 en 6 opbouwen. Na het laatste increment 6 gebruikt te hebben beginnen we weer vooraan met het increment 4, 2, enz. Dit soort algoritmen wordt wel een wiel genoemd, met acht spaken (de acht increments). Er zijn verfijningen met zelfs 5760 spaken bedacht, maar het aantal bewerkingen dat men moet doen blijft van de orde van grootte van $N^{1/2}$.

Een andere geheugenzuinige algoritme waarvan het aantal bewerkingen van de orde van grootte van $N^{1/4}$ lijkt te zijn (dit is niet bewezen, maar wel experimenteel bevestigd) is afkomstig van D. Shanks. Deze algoritme is gebaseerd op de kettingbreukontwikkeling van $N^{1/2}$ en is door Shanks op een HP41-C pocketcalculator geprogrammeerd. De algoritme luidt als volgt:

Stap 1. Zoek in de kettingbreukontwikkeling van $z_0 = N^{1/2}$ naar een Q_m met m even en $Q_m = t^2$ (een echt kwadraat). Definieer $\tilde{P}_0 := P_{m+1}$ en $\tilde{Q}_0 := t$.

Stap 2. Bereken de kettingbreukontwikkeling van $z_0 = (\tilde{P}_0 + N^{1/2})/\tilde{Q}_0$ totdat $\tilde{P}_n = \tilde{P}_{n+1}$. Dan is $\tilde{Q}_n$ of $\tilde{Q}_n/2$ een factor van N.

Het zou te ver voeren om hier uit te leggen waarom deze algoritme werkt.

VOORBEELD. $N = 2771$. In stap 1 vinden we voor P_i, Q_i, q_i en R_i:

i	P_i	Q_i	q_i	R_i
-1		2771		
0	0	1	52	0
1	52	67	1	37
2	15	38	1	29
3	23	59	1	16
4	36	25	3	13
5	39			

Dus $Q_4 = t^2 = 25$; verder is $P_5 = 39$, dus we kiezen $\tilde{P}_0 = 39$ en $\tilde{Q}_0 = 5$ en we vinden in stap 2 voor $\tilde{P}_i$, $\tilde{Q}_i$, $\tilde{q}_i$ en $\tilde{R}_i$ ($\tilde{Q}_{-1} = (2771 - 39^2)/5 = 250$):

i	$\tilde{P}_i$	$\tilde{Q}_i$	$\tilde{q}_i$	$\tilde{R}_i$
-1		250		
0	39	5	18	1
1	51	34	3	1
2	51			

Dus $\tilde{P}_2 = \tilde{P}_1$ en $\tilde{Q}_1/2 = 17$ is een factor van N met $N = 17 \times 163$. □

Vanwege de relaties (5) kunnen met behulp van deze algoritme op een HP41-C getallen to 10^{20} nog in redelijke tijd ontbonden worden.

Een andere geheugenzuinige algoritme, waarmee goede resultaten zijn geboekt (zoals de ontbinding van het achtste Fermat-getal $2^{256}+1$) is een algoritme van Pollard die, als regel, de kleinste priemfactor p van N in $O(p^{1/2})$ operaties kan vinden. Een eenvoudige versie van deze algoritme ziet er als volgt uit: zij $f(x) := x^2 + 1$ en definiëer een rij gehele getallen $\{x_i\}$, $i = 0,1,2, \cdots$, door het voorschrift $x_{i+1} := f(x_i) \pmod N$, met x_0 nog vrij te kiezen. Op grond van een waarschijnlijkheidstechnische redenering mag men verwachten dat $x_{2j} \equiv x_j \pmod p$ voor een of andere $j = O(p^{1/2})$. Dus als $G_j = ggd(x_{2j} - x_j, N)$, dan zal G_j vroeg of laat, maar gemiddeld na $O(p^{1/2})$ stappen, een factor van N opleveren. In de praktijk rekent men niet steeds G_j uit, maar men bouwt het product $Q_j \equiv \Pi_{i=1}^{j}(x_{2i} - x_i) \pmod N$ op. Het enige dat men hoeft te onthouden is het triple $T_i = (x_i, x_{2i}, Q_i)$, $(i = 1,2,3, \cdots)$. T_{i+1} wordt uit T_i berekend door middel van de relaties:

$$x_{i+1} \equiv x_i^2 + 1 \pmod N \quad \text{en} \quad x_{2i+2} \equiv x_{2i+1}^2 + 1 \equiv (x_{2i}^2 + 1)^2 + 1 \pmod N.$$

Het geheugenbeslag wordt hier dus tot een minimum beperkt. Van tijd tot tijd, bijvoorbeeld na iedere 100 stappen, wordt $ggd(Q_i, N)$ uitgerekend.

VOORBEELD. Voor $N=2771$, $x_0=6$ vinden we

i	$T_i = (x_i, x_{2i}, Q_i)$
1	(37, 1370, 1333)
2	(1370, 2263, 1610)
3	(934, 808, 2194)
4	(2263, 1523, 246)
5	(362, 2416, 962)
6	(808, 893, 1411)

en $ggd(Q_6, 2771)=17$. □

Er zijn ook geheugenzuinige algoritmen die met name geschikt zijn om priemfactoren p van een getal N te vinden waarbij $p-1$ opgebouwd is uit kleine priemfactoren; analoog voor priemfactoren p waarbij $p+1$ uit kleine priemfactoren bestaat.

4.2. General purpose methoden

De krachtigste bekende factorisatiemethoden, die geschikt zijn voor willekeurige getallen (en die dus niet afhankelijk zijn van de grootte van de te vinden priemdeler p, of van de structuur van $p-1$), zijn in feite gebaseerd op de simpele waarneming dat, als we twee getallen x en y kennen waarvoor geldt:

$$x^2 \equiv y^2 \pmod{N}, \tag{8}$$

N een deler van $(x-y)(x+y)$ moet zijn; en als we een beetje geluk hebben, zal N twee delers d en N/d hebben, die niet beide in $x-y$ zitten, noch beide in $x+y$. En dan kunnen we één van die twee delers (en daarna natuurlijk ook de tweede) vinden door $ggd(x-y, N)$ uit te rekenen.

Het probleem is om de getallen x en y te vinden. Het idee hierbij is dat als we een aantal waarden van u kennen waarvoor geldt dat $u^2 \equiv a \pmod{N}$, voor kleine waarden van $|a|$, m.a.w. als we een aantal kleine kwadraatresten a van N kennen, dan blijkt het soms mogelijk om deze congruenties te combineren tot een van de vorm (8). Er zijn verschillende methoden bedacht om deze kleine kwadraatresten te vinden en we zullen er hier een tweetal bespreken: de kettingbreukmethode en de kwadratische zeef-methode. We zullen ons hier slechts tot de principes van de methodes beperken.

De eerste maakt gebruik van de kettingbreukontwikkeling van $N^{1/2}$. Definiëer de zgn. factorbasis $FB=(p_1,p_2,\cdots,p_r)$ als een verzameling van priemgetallen met de eigenschap dat de vergelijking $x^2 \equiv N \pmod{p}$ oplosbaar is voor alle $p \in FB$. Reken nu in de kettingbreukontwikkeling van $N^{1/2}$, met behulp van (7), de waarden uit van $Q_1,Q_2,Q_3,\cdots$, tesamen met de overeenkomstige waarden van $A_0,A_1,A_2,\cdots \pmod{N}$. Selecteer die Q_i's die alleen zijn opgebouwd uit priemfactoren uit FB (door trial-and-error m.b.v. deze priemen) en probeer uit deze Q's, gebruikmakend van (6), een relatie van de vorm (8) te vinden. We zullen een voorbeeld geven om de lezer enig 'gevoel' te geven voor wat er precies gebeurt.

VOORBEELD. $N=45313$; het is gemakkelijk in te zien dat de vgl. $x^2 \equiv N \pmod{p}$ voor $p=2$, 5 en 7 wel, maar voor $p=5$ géén oplossingen heeft. Voor de factorbasis kiezen we daarom: $FB=\{2,3,7\}$. Als we m.b.v. (7) de kettingbreuk van $N^{1/2}$ uitrekenen vinden we $Q_2=56=2^3\times 7$, $Q_8=Q_{10}=288=2^5\times 3^2$, en voor de bijbehorende A_i vinden we $A_1=213$, $A_7=16206$ en $A_9=19483$. Volgens (6) geldt er dan: $A_7^2 \equiv Q_8 \pmod{N}$ en $A_9^2 \equiv Q_{10} \pmod{N}$. Als we deze twee congruenties met elkaar vermenigvuldigen, vinden we: $(A_7\times A_9)^2 \equiv 288^2 \pmod{N}$ en $ggd(A_7\times A_9-288, N)=113$, dus $N=45313=113\times 401$. □

De tweede methode gaat uit van het kwadratisch polynoom $Q(x):=(x+M)^2-N$, $x=0, \pm 1, \pm 2, \cdots$ met $M=[N^{1/2}]$ en dezelfde factorbasis *FB* zoals boven gedefiniëerd. Er geldt natuurlijk

$$(x+M)^2 \equiv Q(x) \pmod{N} \tag{9}$$

en als x klein is t.o.v. M dan is $Q(x) \approx 2xM$ en dus is $Q(x)$ ook klein t.o.v. N. De kwadraatresten die we hier krijgen zijn weliswaar wat groter dan bij de kettingbreukmethode hierboven beschreven, maar er is een belangrijk verschil: het factoriseren van deze kwadraatresten kan gebeuren zonder 'trial-and-error', maar met een zgn. zeef. Het feit namelijk dat $Q(x)$ een kwadratisch polynoom in x is, maakt het mogelijk om 'in een klap' voor een gegeven waarde van een priem p, een rekenkundige rij van x-waarden aan te geven (met verschil p) waarvoor $Q(x)$ deelbaar is door p. Als we dit op een handige manier bijhouden, dan blijkt het mogelijk om veel sneller dan bij de kettingbreukmethode de benodigde kwadraatresten te vinden die alleen zijn samengesteld uit priemfactoren uit de factorbasis.

VOORBEELD. $N=7169$. Het kwadratisch polynoom is hier $Q(x)=(x+84)^2-N$. De vgl. $Q(x) \equiv 0 \pmod{p}$ is voor $p=2$, 5 en 7 oplosbaar, maar niet voor $p=3$. (oplossingen: voor $p=2$: $x \equiv 1 \pmod 2$, voor $p=5$: $x \equiv 3$ of 4 (mod 5), voor $p=7$: $x \equiv 1$ of 6 (mod 7)). Voor $x=3$ vinden we nu: $Q(3)=400$, zodat $87^2 \equiv 400=20^2 \pmod N$. We hebben hier geluk omdat we direct een congruentie van de vorm (8) vinden en niet verschillende congruenties hoeven te combineren om (8) te vinden. We vinden nu: $ggd(87-20, 7169)=67$, dus $7169=67\times 107$. □

In het algemeen is het niet zo eenvoudig om uit een aantal gegeven congruenties van de vorm $X_i^2 \equiv \pm Y_i \pmod N$, waarbij Y_i slechts opgebouwd is uit priemfactoren uit een gegeven factorbasis, een congruentie van de vorm (8) te vinden. Als er voldoende congruenties voorhanden zijn (d.w.z. 2 meer dan het aantal priemen in de factorbasis) dan is het altijd mogelijk om een subset van deze congruenties te vinden waarvan het product van de vorm (8) is. Als we de exponenten van de ontbinding van Y_i in priemfactoren uit de factorbasis in een aparte vector zetten, dan komt het probleem neer op het vinden van een subset van de exponent-vectoren die, mod 2 opgeteld, de nulvector opleveren. Deze subset kan worden gevonden met behulp van Gauss-eliminatie.

Het grootste tot nu toe m.b.v. de kwadratische zeef ontbonden getal bestaat uit 96 cijfers, en dit getal is het product van 2 priemfactoren, ieder van 48 cijfers.

4.3. Groepentheoretische methoden

Tot nu toe hebben we geen aandacht besteed aan een belangrijke klasse van factorisatiemethoden, nl. die welke gebaseerd zijn op berekeningen in eindige groepen. We zullen hier in het kort twee voorbeelden van dit soort methoden bespreken, nl. Pollard's $p-1$ methode en Lenstra's elliptische kromme methode (ECM).

Bij Pollard's $p-1$ methode wordt gerekend in de groep van integers modulo een priem p: deze vormen een eindige cyclische abelse groep onder vermenigvuldiging. De Kleine Stelling van Fermat kan in deze context geformuleerd worden als de Stelling van Lagrange voor eindige groepen die zegt dat ieder element van een multiplicatieve groep de identiteit oplevert, als we dit element tot de macht van de orde van de groep brengen. Gegeven een priemfactor p van N, als we $b \equiv a^{p-1} \pmod N$ zouden kunnen uitrekenen, dan zouden we vinden $ggd(b-1, N)=p$.

Helaas kennen we p niet, en dus ook niet $p-1$, de orde van de groep. Echter, als het getal $p-1$ slechts uit kleine priemfactoren zou zijn opgebouwd, dan kan de volgende algoritme toch succes opleveren:

Stap 1. Kies een willekeurig getal a;
Stap 2. Laat M het product zijn van 'alle' kleine priemen tot 'hoge' machten;
Stap 3. Bereken $b \equiv a^M \pmod N$;
Stap 4. Bereken $ggd(b-1, N)$.

Als $p-1$ een deler is van M dan zullen we in stap 4 een deler van N vinden. Dat we in stap 3 mod N

en niet mod p rekenen is voor het toepassen van Fermat's Kleine Stelling niet relevant.

Alvorens de elliptische kromme methode van H.W. Lenstra Jr. te beschrijven geven we eerst enkele resultaten uit de theorie van de elliptische krommen.

Zij F een lichaam van karakteristiek ongelijk aan 2 of 3. Dan bestaat er voor ieder paar van integers (A, B) een elliptische kromme over het lichaam F bestaande uit de punten (x,y) uit $F \star F$ die voldoen aan de vergelijking

$$E: y^2 = x^3 + Ax + B.$$

Deze punten vormen een eindige abelse groep (de zgn. Mordell-Weil groep), additief geschreven, met de volgende groepswetten:

1) de identitieit van de groep is het punt op oneindig;
2) de inverse van het punt (x,y) is het punt $(x, -y)$;
3) als $P_1=(x_1,y_1)$, $P_2=(x_2,y_2)$, en $P_3=(x_3,y_3)$ drie punten zijn die op dezelfde rechte lijn $Y=mX+b$ liggen, dan is $P_1+P_2+P_3=0$ (de identiteit van de groep).

Als het lichaam F de verzameling van integers modulo een priem p is, dan is de groep van punten eindig, en hij heeft een orde die begrensd wordt door $p+1-2.p^{1/2}$ en $p+1+2.p^{1/2}$. Als we punten op een kromme beschouwen modulo een of ander samengesteld getal N, dan krijgen we een groep die de directe som is van de groepen voor de factoren van N. Het is dan meestal gemakkelijker om de homogene versie van de elliptische kromme te beschouwen:

$$E: y^2z = x^3 + Axz^2 + Bz^3.$$

In dit geval hebben punten op de kromme drie coordinaten (x,y,z) en de identiteit van de groep is het punt (0,1,0).

De elliptische kromme methode ziet er nu als volgt uit:

Stap 1. Kies een willekeurige kromme E door een willekeurig paar van gehele getallen (A,B) (mod N) te kiezen;
Stap 2. Kies een willekeurig punt P op de kromme (mod N);
Stap 3. Zij M het product van alle 'kleine' priemen tot 'hoge' machten;
Stap 4. Berekenen $M.P=Q=(x,y,z)$ in de Mordell-Weil groep;
Stap 5. Bereken $ggd(N,z)$.

Als we proberen om N te factorisern waarbij p een priemfactor van N is, en als de orde van de kromme (mod p) een deler is van M, dan zal het punt Q de identiteit van de groep zijn (mod p) (en niet de identiteit van de groep (mod N)). Maar dan zal de z-coordinaat een factor p bevatten, en deze zal in stap 5 worden gevonden.

Als we Pollard's $p-1$ methode met de ECM-methode vergelijken, dan merken we op dat in beide gevallen de orde van de groep 'ongeveer' p is, en dat de methoden zullen slagen als de orde van de groep alleen maar deelbaar is door kleine priemfactoren (men zegt ook wel dat de orde in dat geval 'glad' is). Echter, bij de $p-1$ methode slagen we als de orde van een enkele specifieke groep glad is, terwijl we bij de elliptische kromme methode uit vele krommen kunnen kiezen. Als er dan ergens 'in de buurt van p' een 'glad' getal ligt, dan hebben we een grotere kans om een groep van deze orde te vinden. Karakteristiek voor de ECM-methode is dan ook dat de hierboven beschreven stappen 1 tot en met 5 voor vele verschillende krommen worden uitgevoerd. Aangezien de berekeningen op verschillende krommen onafhankelijk van elkaar zijn, leent ECM zich uitstekend voor implementatie op gedistribueerde computer systemen, of op parallelle (multiprocessor) systemen.

Met behulp van ECM zijn priemfactoren van $\leqslant$ 37 cijfers gevonden, en samen met de kwadratische zeef behoort ECM tot de krachtigste bekende factorisatie-methoden.

5. PRIMALITEITSTESTEN

5.1. Monte Carlo-achtige tests

In het slot van sectie 2 hebben we al een Monte Carlo-achtige priemtest genoemd, namelijk de test of (2) geldt voor $a=2$. Hier zullen we een verbeterde test bespreken. Daarvoor definiëren we een *sterke a-pseudopriem* als een oneven samengesteld getal N met de eigenschap dat, als we schrijven $N-1=2^s t$ met t oneven, óf geldt $a^t \equiv 1 \pmod N$ óf $a^{2^r t} \equiv -1 \pmod N$ voor een of andere r met $0 \leqslant r < s$. Uit deze definitie volgt dat een sterke a-pseudopriem een 'gewone' a-pseudopriem (in sectie 2 gedefiniëerd) is, maar niet omgekeerd. Kort geleden is ontdekt dat het getal $S=151*751*28351=3215031751$ het enige getal kleiner dan $25*10^9$ is dat een sterke a-pseudopriem is voor $a=2,3,5$ en 7. Dat levert ons een eenvoudige primaliteitstest op voor alle oneven getallen kleiner dan $25*10^9$: verifiëer of N een sterke a-pseudopriem is voor $a=2,3,5$ en 7. Zo ja, dan is N priem als $N \neq S$, en anders is N samengesteld.

Voor iedere gegeven a bestaan er oneindig veel sterke a-pseudopriemen, maar geen enkel oneven samengesteld getal N is een sterke a-pseudopriem voor meer dan 1/4 van de getallen $a<N$. Als we dus nagaan of N een sterke pseudopriem is voor k willekeurig gekozen getallen a dan zijn er twee mogelijkheden: als N niet een sterke pseudopriem is voor één van die a's, dan is N samengesteld; in het andere geval kunnen we N priem verklaren, waarbij de kans dat we het bij het verkeerde eind hebben hoogstens $(1/4)^k$ is.

Kunnen we iets meer concluderen als we zouden weten dat N een a-pseudopriem is voor *alle* getallen a onder een bepaalde grens? Het antwoord wordt gegeven door het volgende resultaat: als de zgn. Gegeneraliseerde Riemann Hypothese juist is, dan moet er een $b<70(\log N)^2$ bestaan waarvoor het oneven samengestelde getal N géén sterke b-pseudopriem is. In de Gegeneraliseerde Riemann Hypothese (GRH) wordt aangenomen dat alle complexe nulpunten van bepaalde functies reëel deel 1/2 hebben. Deze Hypothese is tot op heden bevestigd, noch weerlegd, maar voor het speciale geval van de 'gewone' Riemann Hypothese bestaat een aanzienlijke numerieke evidentie voor de juistheid ervan. Als N dus een sterke b-pseudopriem is voor alle $b<70(\log N)^2$, dan is N priem, mits GRH waar is.

5.2. De Lucas-test voor Mersenne-getallen

In deze sectie zullen we de zgn. Lucas-test bescrijven; dat is een primaliteitstest voor getallen van een speciale vorm, de zgn. Mersenne-getallen $M_n := 2^n - 1$. De Lucas-test luidt als volgt: stel $n>2$, $T_0 := 4$ en $(k>0)$ $T_k := T_{k-1}^2 - 2 \pmod{M_n}$. Dan is M_n priem dan en slechts dan als M_n een deler is van T_{n-2}.

Het aantal voor deze test benodigde stappen is $n-2=O(\log M_n)$. Voor het voorbeeld M_{31} genoemd in de inleiding moeten we dus 29 stappen uitvoeren om primaliteit te bewijzen in plaats van de 4792 delingen door de priemgetallen onder $M_n^{1/2}$. Wel zijn de stappen uit de Lucas-test duurder dan de delingen door de kleine priemgetallen.

VOORBEELD. $N=2^{13}-1=8191$. Voor T_k, $k=1,2,\cdots,11$ vinden we achtereenvolgens: 14, 194, 4870, 3953, 5970, 1857, 36, 1294, 3470, 128, 0 $(=T_{13-2})$, waaruit volgt dat N priem is. □

De Lucas-test is gebruikt om de grootste bekende priemgetallen op te sporen: het record staat momenteel op M_{216091}, een getal van 65050 cijfers.

5.3. Een test gebaseerd op de ontbinding van $N-1$

De Kleine Stelling van Fermat kan, zoals in sectie 2 is aangegeven, niet gebruikt worden om wiskundige zekerheid te verkrijgen over het priem zijn van een getal. Die zekerheid kan wel verkregen worden, als strengere voorwaarden worden opgelegd. Dit is het geval bij de

STELLING VAN E. LUCAS. Stel dat $N-1=\Pi_{j=1}^{s} q_j^{e_j}$ de representatie (1) van $N-1$ is. Als er een

geheel getal a kan worden gevonden zo dat $a^{(N-1)/q_j} \not\equiv 1 \pmod N$ voor alle priemdelers q_j van $N-1$, en $a^{N-1} \equiv 1 \pmod N$, dan is N priem. □

VOORBEELD. $N=487$, $N-1=2.3^5$; met de in sectie 2 beschreven binaire algoritme vinden we dat $2^{(N-1)/2}=2^{243} \equiv 1 \pmod{487}$, dus met $a=2$ kunnen we de stelling niet toepassen. Kiezen we echter $a=3$, dan vinden we dat $3^{(N-1)/2}=3^{243} \equiv -1 \pmod{487}$ en dat $3^{(N-1)/3}=3^{162} \equiv 232 \pmod{487}$. Er volgt dat $3^{N-1} \equiv 1 \pmod{487}$, dus N is priem.

Het idee achter de stelling van Lucas is het volgende. De orde van de groep van primitieve restklassen (mod N) is gelijk aan $\phi(N)$; dit is $=N-1$, als N priem is en $<N-1$ als N samengesteld is. Bovendien is, als N priem is, de bewuste groep een cyclische groep, en deze bevat een generator van de orde $N-1$. Als N samengesteld is, dan heeft ieder element van de groep een orde hoogstens gelijk aan $\phi(N)<N-1$. In de stelling van Lucas wordt, in feite, een element uit de groep gezocht van de orde $N-1$. Als we zo'n element vinden (ook wel primitieve wortel van N genoemd), dan moet dus N priem zijn..

Er is echter één mogelijk bezwaar: we moeten de volledige ontbinding in priemfactoren van $N-1$ kennen! Dit kan soms onoverkomelijke problemen geven als N groot is, en $N-1$ twee of meerdere grote priemfactoren bevat. Dit heeft aanleiding gegeven tot varianten van de stelling van Lucas waarbij $N-1$ niet volledig ontbonden hoeft te worden.

We geven hier, als voorbeeld, de volgende stelling van *Proth, Pocklington en Lehmer:*

Stel dat $N-1$ gedeeltelijk in priemfactoren is ontbonden als $N-1=FR$ waarbij alle priemfactoren van F bekend zijn, en $ggd(F,R)=1$. Stel verder dat er een a bestaat zo dat $a^{N-1} \equiv 1 \pmod N$ en $ggd(a^{(N-1)/q}-1,N)=1$ voor iedere priemdeler q van F. Dan is iedere priemfactor p van N gelijk aan $1 \pmod F$.

GEVOLG. Als $F>N^{1/2}$ dan is N priem.

Er zijn enkele verfijningen in deze stelling aangebracht, waardoor, onder additionele voorwaarden, kan worden geconcludeerd dat N priem is als $F>N^{1/3}$.

Er zijn ook varianten gevonden die gebaseerd zijn op de ontbinding in priemfactoren van $N+1$, of N^2+1, of $N^2 \pm N+1$, maar het bezwaar blijft steeds dat we vroeg of laat tegen een voorbeeld aanlopen, waarbij we de gevraagde ontbinding niet kunnen vinden, en daardoor de stelling niet kunnen toepassen.

5.4. De Adleman-Ponerance-Rumely-Cohen-Lenstra test

Pas enkele jaren geleden zijn er efficiente primaliteitstests gevonden waarbij het bovengenoemde probleem niet kan optreden. We zullen hier een vereenvoudigde beschrijving van de onderliggende ideeën geven.

Om te bewijzen dat N priem is gaat men als volgt te werk. Stel p_i is het i-de priemgetal ($p_1=2$, $p_2=3$, $\cdots$) en zij m een of ander natuurlijk getal. Beschouw nu alle priemgetallen van de vorm $\Pi_{j=1}^{m} p_j^{e_j}+1$, waarbij $e_j=0$ of 1 en noem deze priemgetallen $q_1, q_2, \cdots, q_k$. Zij nu $m(N)$ de kleinste waarde van m waarvoor het product van deze k priemgetallen groter is dan $N^{1/2}$. Er zijn tabellen van deze waarden van m gemaakt; bijvoorbeeld: als $2.5*10^{43}<N<5*10^{89}$ dan is $m(N)=6$. Dus ook voor grote waarden van N blijft $m(N)$ nog betrekkelijk klein. De priemen $p_1, \cdots, p_m$ heten de 'begin'-priemen en de hieruit opgebouwde priemen $q_1, \cdots, q_k$ de 'Euclidische' priemen. We nemen aan dat N niet deelbaar is door een van de begin- of Euclidische priemen. Zij r nu een deler van N met $r<N^{1/2}$. Als we nu $r \pmod{q_j}$ zouden kennen voor $j=1, \cdots, k$, dan zouden we r zelf kunnen vinden m.b.v. de zogenaamde Chinese Reststelling. (Deze zegt dat er precies één zo'n r bestaat met $0 \leqslant r < \Pi q_j$ en dit laatste product is groter dan $N^{1/2}$). Zij g_j een primitieve wortel van q_j en zij $I_j=I_j(r)$ het kleinste natuurlijke getal waarvoor geldt: $g_j^{I_j} \equiv r \pmod{q_j}$. Als we nu I_j zouden kennen

dan zouden we ook $r(\mathrm{mod}\, q_j)$ kunnen uitrekenen (m.b.v. g_j). Als we ook $I_j(\mathrm{mod}\, p)$, voor alle priemen p die een deler zijn van $q_j - 1$, zouden kennen, dan zouden we, door opnieuw gebruik te maken van de Chinese Reststelling, I_j kunnen bepalen. Samenvattend: als we $I_j(\mathrm{mod}\, p)$ zouden kennen voor alle priemen p die q_j delen, dan zouden we ook r kunnen vinden.

De moeilijkheid bij deze methode schuilt in het vinden van de $I_j(\mathrm{mod}\, p)$. Het zou te ver voeren om hier uit de doeken te doen hoe dat kan gebeuren. We volstaan met te vermelden dat uiteindelijk een aantal 'educated guesses' voor de $I_j(\mathrm{mod}\, p)$, voor alle p die $q_j - 1$ delen, en voor $1 \leqslant j \leqslant k$, worden gedaan en dat steeds wordt gecontroleerd of de bijbehorende r een deler van N is.

De hier beschreven ideeën zijn afkomstig van Adleman, Pomerance en Rumely. Cohen en H.W. Lenstra Jr. hebben deze ideeën verder uitgewerkt tot een algemeen bruikbare priemtest waarmee van willekeurige getallen tot ca. 200 cijfers nog binnen redelijke tijd het priem zijn kan worden bewezen.

LITERATUUR

Er bestaat een aanzienlijke hoeveelheid literatuur op het gebied van primaliteitstesten en factorisatie. We zullen hier enkele overzichtsartikelen en belangrijke boeken noemen, waarmee de lezer zijn eigen weg verder kan bepalen.

D.A. BUELL (1987) Factoring: Algorithms, Computations, and Computers. *J. Supercomp. 1,* 191-216.

J.D. DIXON (1984). Factorization and primality tests. *Amer. Math. Monthly 91,* 333-352.

D. KNUTH. *The Art of Computer Programming, Vol. 1: Fundamental Algorithms; Vol. 2: Seminumerical Algorithms,* Addison-Wesley, 1975 resp. 1969.

C. POMERANCE (1981).. Recent Developments in Primality Testing. *Math. Intell. 3,* 97-105.

C. POMERANCE (1982). The search for prime numbers. *Scientific Amer.,* 122-130.

H. RIESEL (1985). *Prime Numbers and Computer Methods for Factorization,* Birkhäuser.

H.C. WILLIAMS (1984). Factoring on a Computer. *Math. Intell. 6,* 29-36.

H.C. WILLIAMS (1982). The influence of computers in the development of number theory. *Comp. and Math. with Applics. 8,* 75-93.