



Dass das Problem, die Primzahlen von den zusammengesetzten zu unterscheiden und letztere in ihre Primfaktoren zu zerlegen zu den wichtigsten und nützlichsten der ganzen Arithmetik gehört und den Fleiss und die Weisheit der Geometer der Antike und der Neuzeit beschäftigt hat, ist so bekannt, dass es überflüssig ist, viel darüber zu sagen.

CARL FRIEDRICH GAUSS

Disquisitiones Arithmeticae (1801)

Art. 329

Die in

J. Pollard:
A Monte Carlo Method for Factorization,
BIT, vol.15 (1975), 331-335.

vorgeschlagene Methode markiert, zusammen mit der
"Kettenbruchmethode" von

M. Morrison, J. Brillhart:
A Method of Factoring and the Factorization of F_7 ,
Mathematics of Computation, vol. 29 (1975), 183-205.

den Beginn der Neuzeit des Faktorisierens ganzer Zahlen.

- ▶ Mit einer Variante dieser Methode haben J.M. POLLARD und R.P. BRENT 1980 die 8. Fermat-Zahl

$$F_8 = 2^{2^8} + 1 \approx 1.11579 \cdot 10^{77}$$

in zwei 16- bzw. 62-stellige Primfaktoren zerlegt.

Kombinatorische Überlegungen

- ▶ A endliche Menge, $f : A \rightarrow A$ Transformation; für jedes $a \in A$ wird die f -Bahn von a , i.e., die Folge der f -Iterierten von a :

$$a = \underbrace{a_0 \xrightarrow{f} a_1 \xrightarrow{f} a_2 \rightarrow \dots \rightarrow a_{\lambda-1} \xrightarrow{f} a_\lambda}_{\text{Vorperiode}} \underbrace{\xrightarrow{f} a_{\lambda+1} \xrightarrow{f} \dots \xrightarrow{f} a_{\lambda+\mu-1} \xrightarrow{f} a_{\lambda+\mu}}_{\text{Periode}} = a_\lambda$$

schliesslich-periodisch.

λ = Länge der Vorperiode

μ = Länge der Periode

$\rho = \lambda + \mu$ die ρ -Länge von a bezüglich f , d.h. die f -Bahn von a enthält genau ρ verschiedene Elemente

λ, μ sind die minimalen Indices $i \geq 0, j > 0$ mit $a_i = a_{i+j}$, wobei

$$a_n = f^n(a) = \underbrace{f(f(\dots f(a)\dots))}_{n\text{-mal}} \quad (n \geq 0)$$

Probleme:

- ▶ wie kann man möglichst effizient ein Element der f -Periode von a bestimmen?
- ▶ wie kann man die ρ -Länge von a unter f abschätzen?

Naiver Ansatz:

- ▶ für $j = 1, 2, 3, \dots$ vergleiche a_j mit a_1, a_2, \dots, a_{j-1}
bis eine Wiederholung eines Elements auftritt.
- ▶ Dieses Verfahren liefert λ und μ (und somit ρ) exakt, hat aber das Problem eines potentiell ungeheuren Speicherbedarfs.

Floyd's Trick ("cycle detection")

- ▶ Für $i < j$ gilt:

$$a_i = a_j \iff (\lambda \leq i \text{ und } \mu | j - i)$$

also insbesondere für $t > 0$:

$$a_t = a_{2t} \iff (\lambda \leq t \text{ und } \mu | t) \quad (*)$$

- ▶ Idee: bestimme durch Iteration der Abbildung

$$(x, y) \mapsto (f(x), f(f(y)))$$

die Folge

$$(a_0, a_0), (a_1, a_2), (a_2, a_4), (a_3, a_6), \dots$$

und teste nach jedem Schritt auf Gleichheit der beiden Komponenten, d.h. bestimme den ersten Zeitpunkt, zu dem eine "Kollision" zwischen den Folgen $(a_n)_{n \geq 0}$ und $(a_{2n})_{n \geq 0}$ eintritt.

- ▶ Behauptung:

$$t = \begin{cases} \mu \cdot \lceil \frac{\lambda}{\mu} \rceil & \text{falls } \lambda \neq 0 \\ \mu & \text{falls } \lambda = 0 \end{cases}$$

und es ist

$$t \leq \rho \leq 2t.$$

(Begründung!)

- ▶ Sobald man $a_t = a_{2t}$ kennt, kann man μ problemlos bestimmen (einmal durch den Zyklus laufen).
- ▶ Den Wert λ erhält man aus der Tatsache

$$f^\lambda(a_0) = a_\lambda = a_{t+\lambda} = f^\lambda(a_t)$$

d.h. aus der Iteration $(x, y) \mapsto (f(x), f(y))$ mit Startwert $(x, y) = (a_0, a_t)$, indem man die erste Kollision feststellt.

- ▶ Wichtig: das Verfahren arbeitet mit “konstantem Speicher”.

Erwartungswert der ρ -Länge

- ▶ A sei Menge von N Elementen, gezogen werden Elemente $a_1, a_2, a_3, \dots \in A$ (“mit Zurücklegen”) unter Gleichverteilung.
- ▶ Frage: wie oft muss man im Mittel ziehen bis man ein schon früher gezogenes Element nochmals zieht (d.h. bis eine “Kollision” auftritt) ?
- ▶ Hinweis: diese Frage spielt in vielen algorithmischen Überlegungen eine wichtige Rolle, z.B.
 - ▶ Hashing
 - ▶ Deterministische Generierung von “Zufallszahlen”
 - ▶ Probabilistische Algorithmen
 - ▶ Kryptosysteme

► Elementare Wahrscheinlichkeiten:

$$\begin{aligned}
 Q(N, k) &= \text{Wkeit } [a_1, a_2, \dots, a_k \text{ paarweise verschieden}] \\
 &= 1 \cdot \left(1 - \frac{1}{N}\right) \cdot \left(1 - \frac{2}{N}\right) \dots \left(1 - \frac{k-1}{N}\right) \\
 &= \frac{\binom{N}{k} \cdot k!}{N^k} = \frac{N!}{(N-k)! N^k}
 \end{aligned}$$

$$\begin{aligned}
 P(N, k) &= \text{Wkeit } [a_1, \dots, a_{k-1} \text{ pw verschieden und} \\
 &\quad a_k \in \{a_1, \dots, a_{k-1}\}] \\
 &= \frac{k-1}{N} Q(N, k-1) = Q(N, k-1) - Q(N, k)
 \end{aligned}$$

► Gesucht ist also

$$\begin{aligned}
 \sum_{k \geq 1} k P(N, k) &= \sum_{k \geq 1} k (Q(N, k-1) - Q(N, k)) \\
 &= \underbrace{Q(N, 0)}_{=1} + \sum_{k \geq 1} Q(N, k)
 \end{aligned}$$

- Man kann zeigen, dass dies

$$\sum_{k \geq 0} Q(N, k) = \sum_{k \geq 1} k P(N, k) \sim \sqrt{\frac{\pi N}{2}} + \frac{2}{3}$$

ist. Als Hinweis: mit $1 - x \leq e^{-x}$ hat man

$$Q(N, k) = \prod_{1 \leq \ell < k} \left(1 - \frac{\ell}{n}\right) = e^{-\frac{1}{N} \sum_{1 \leq \ell < k} \ell} = e^{-\frac{1}{N} \binom{k}{2}} \leq e^{-\frac{1}{N} (k-1)^2}$$

Andererseits gilt (Unter- und Obersummen des Integrals)

$$\sum_{k \geq 1} e^{-k^2/N} \leq \int_0^{\infty} e^{-x^2/2N} dx \leq \sum_{k \geq 0} e^{-k^2/N}$$

Der Wert des Integrals ist aber bekannt:

$$\int_0^{\infty} e^{-x^2/2N} dx = \sqrt{\frac{\pi N}{2}}$$

- ▶ Folgerung: betrachtet man eine zufällig gewählte Funktion $f : A \rightarrow A$ und ein zufällig gewähltes Element $a \in A$, so wird die mittlere ρ -Länge (Erwartungswert) der ρ -Länge von a unter f $\mathcal{O}(\sqrt{N})$ sein.

wer's genau wissen will:

→ R. SEDGEWICK, PH. FLAJOLET,
An Introduction to the Analysis of Algorithms,
Addison-Wesley, 1996

Abschnitte 4.8 und 8.7, 8.8, Stichwort: Ramanujans Q -Funktion.

Heuristik zur ρ -Länge:

Wann ist $Q(N, k) = \text{W.keit} [a_1, a_2, \dots, a_k \text{ pw. verschieden}] \approx 1/2$?

$$\begin{aligned}
 Q(N, k) &= \left(1 - \frac{1}{N}\right)\left(1 - \frac{2}{N}\right)\dots\left(1 - \frac{k-1}{N}\right) \sim \frac{1}{2} \\
 &\quad \downarrow \updownarrow \\
 \sum_{1 \leq j < k} \ln\left(1 - \frac{j}{N}\right) &\sim \ln(1/2) \\
 &\quad \downarrow \quad \text{(falls } k \ll N) \\
 \frac{k(k-1)}{2N} &= \sum_{1 \leq j < k} \frac{j}{N} \sim \ln 2 \\
 &\quad \downarrow \\
 k &\sim \underbrace{\sqrt{2 \ln 2}}_{1.2} \cdot \sqrt{N}
 \end{aligned}$$

“Geburtstags-Paradoxon”

- ▶ Für $N = 365$ ist $\sqrt{2 \ln 2} \cdot \sqrt{N}$ etwa $22.9 \dots$, d.h. hat man eine Gesellschaft von ≥ 23 Personen, deren Geburtstage über das Jahr gleichverteilt sind, so ist die Wahrscheinlichkeit $> 50\%$, dass zwei Personen am gleichen Tag Geburtstag haben.
- ▶ Bei ≥ 35 Personen ist die Wahrscheinlichkeit bereits $> 80\%$.

POLLARDS Idee

- ▶ Daten
 - ▶ N die zu faktorisierende Zahl, (keine Primzahl)
 - ▶ p ein (echter) Primteiler von N (unbekannt!)
 - ▶ $f(X) \in \mathbb{Z}[X]$ ein “geeignetes” Polynom,
 - ▶ $a \in \mathbb{Z}_N = \{0, 1, \dots, N-1\}$ ein zufällig gewählter Startwert.
- ▶ Man iteriert die Abbildung

$$f_N : \mathbb{Z}_N \rightarrow \mathbb{Z}_N : x \mapsto f(x) \bmod N$$

und erzeugt so eine Folge (Bahn von a unter f) in \mathbb{Z}_N :

$$a = a_0 \xrightarrow{f_N} a_1 \xrightarrow{f_N} a_2 \xrightarrow{f_N} a_3 \xrightarrow{f_N} \dots$$

In $\mathbb{Z}_p = \{0, 1, 2, \dots, p-1\}$ geschieht folgendes

- ▶ Die Abbildung

$$f_p : \mathbb{Z}_p \rightarrow \mathbb{Z}_p : x \mapsto f(x) \bmod p$$

erzeugt das modulo- p -Bild von $(a_k)_{k \geq 0}$:

- ▶ Dies ist die Folge $(b_k)_{k \geq 0}$ in \mathbb{Z}_n mit

$$b_0 \xrightarrow{f_p} b_1 \xrightarrow{f_p} b_2 \xrightarrow{f_p} b_3 \xrightarrow{f_p} \dots$$

und

$$b_k \equiv a_k \bmod p \quad (k \geq 0)$$

- ▶ Beide Folgen sind schliesslich-periodisch.
- ▶ Ziel: Kollision der Folge $(b_k)_{k \geq 0}$ entdecken, bevor die erste Kollision in $(a_k)_{k \geq 0}$ eintritt.

- ▶ Genauer: wenn es gelingt, Indices $i < j$ zu finden mit

$$b_i = b_j \quad \text{und} \quad a_i \neq a_j$$

so ist p gemeinsamer Teiler von $a_j - a_i$ und von N , d.h.

$$p \mid \text{ggT}((a_j - a_i) \bmod N, N)$$

und dieser ggT ist nichttrivialer Teiler von N .

- ▶ Beachte: die Eigenschaft " $b_i = b_j$ ", d.h. " $a_i \equiv a_j \pmod{p}$ ", kann nicht direkt überprüft werden, da p nicht bekannt ist - aber man kann die Situation gezielt herbeiführen:

"cycle detection"!

► Zur Realisierung:

- Bereits einfache quadratische Polynome wie $f(X) = X^2 + c$ ($c \neq 0, -2$) erweisen sich (experimentell) als “geeignet”, d.h. ihre “Reduktion mod p ” f_p erscheint genügend “zufällig”, so dass Floyds “cycle detection”-Trick in $\mathcal{O}(\sqrt{p})$ ein n mit $b_n = b_{2n}$, d.h. $a_n \equiv a_{2n} \pmod{p}$, findet.
- Verfahren: mit f wie oben berechne (modulo N)

$$(a_1, a_2) \mapsto (a_2, a_4) \mapsto (a_3, a_6) \mapsto \dots$$

und teste jeweils, ob $\text{ggT}(a_n - a_{2n} \pmod{N}, N)$ einen nichttrivialen Teiler von N liefert.

► Variationen:

- Je nach zusätzlicher Information über mögliche Teiler von N kann es sinnvoll sein, mit anderen Polynomen zu rechnen. (Nebenbei: lineare Polynome tun's nicht)
- Man muss nicht nach jedem f -Iterationsschritt den ggT berechnen, man kann auch die Differenzen $a_{2n} - a_n \bmod N$ für mehrere n "auf sammeln" (Produkt $\bmod N$ bilden) und dann erst den ggT berechnen

► Komplexität:

- Die Laufzeit von POLLARDS Methode ist $\mathcal{O}(\sqrt{p})$, wenn p der kleinste Primfaktor von N ist, also $\mathcal{O}(N^{1/4})$. Das ist experimentell belegt, aber der Beweis verwendet unbewiesene Annahmen über die Zufälligkeitseigenschaften der f .

Beispiel: $N = 2183$ $f = X^2 + 1$ $a = 2$

$$a_0 = 2$$

$$a_1 = 5$$

$$a_2 = 26$$

$$a_3 = 677$$

$$a_4 = -100$$

$$a_5 = -914$$

$$a_6 = -692$$

$$\text{ggT}(a_2 - a_1, N) = 1$$

$$\text{ggT}(a_4 - a_2, N) = 1$$

$$\text{ggT}(a_6 - a_3, N) = 37$$

$$b_0 = 2$$

$$b_1 = 5$$

$$b_2 = -11$$

$$b_3 = 11$$

$$b_4 = 11$$

$$b_5 = 11$$

$$b_6 = 11$$

In der Tat: $2183 = 37 \cdot 59$

Beispiel: $N = 91643$ $f = X^2 - 1$ $a = 3$

$$a_0 = 3$$

$$b_0 = 3$$

$$a_1 = 8$$

$$b_1 = 8$$

$$a_2 = 63$$

$$\text{ggT}(a_2 - a_1, N) = 1$$

$$b_2 = 63$$

$$a_3 = 3968$$

$$b_3 = 13$$

$$a_4 = 74070$$

$$\text{ggT}(a_4 - a_2, N) = 1$$

$$b_4 = 55$$

$$a_5 = 65061$$

$$b_5 = 86$$

$$a_6 = 35193$$

$$\text{ggT}(a_6 - a_3, N) = 1$$

$$b_6 = 50$$

$$a_7 = 83746$$

$$b_7 = 13$$

$$a_8 = 45368$$

$$\text{ggT}(a_8 - a_4, N) = 113$$

$$b_8 = 55$$

In der Tat: $91643 = 113 \cdot 811$