Kapitel 6
Das RSA-Verfahren

§1: New directions in cryptography

Bei allen bisher betrachteten Codes verläuft die Entschlüsselung entweder genauso oder zumindest sehr ähnlich wie die Verschlüsselung; insbesondere kann jeder, der eine Nachricht verschlüsseln kann, jede andere entsprechend verschlüsselte Nachricht auch entschlüsseln. Man bezeichnet diese Verfahren daher als *symmetrisch.*

Der Nachteil eines symmetrischen Verfahrens besteht darin, daß in einem Netzwerk jeder Teilnehmer mit jedem anderen einen Schlüssel vereinbaren muß. In militärischen Netzen war dies traditionellerweise so geregelt, daß das gesamte Netz denselben Schlüssel benutzte, der in einem Codebuch für jeden Tag im voraus festgelegt war; in kommerziellen Netzen wie beispielsweise einem Mobilfunknetz ist dies natürlich unmöglich.

1976 publizierten MARTIN HELLMAN, damals Assistenzprofessor in Stanford, und sein Forschungsassistent WHITFIELD DIFFIE eine Arbeit mit dem Titel *New directions in cryptography* (IEEE Trans. Inform. Theory **22**, 644–654), in der sie vorschlugen, den Vorgang der *Ver*schlüsselung und den der *Ent*schlüsselung völlig voneinander zu trennen: Es sei schließlich nicht notwendig, daß der Sender einer verschlüsselten Nachricht auch in der Lage sei, diese zu *ent*schlüsseln.

Der Vorteil eines solchen Verfahrens wäre, daß jeder potentielle Empfänger nur einen einzigen Schlüssel bräuchte und dennoch sicher sein könnte, daß nur er selbst seine Post entschlüsseln kann. Der Schlüssel müßte nicht einmal geheimgehalten werden, da es ja nicht schadet, wenn jedermann Nachrichten *ver*schlüsseln kann. In einem Netzwerk mit n Teilnehmern bräuchte man also nur n Schlüssel, um es jedem Teilnehmer zu erlauben, mit jeden anderen so zu kommunizieren, und diese Schlüssel könnten sogar in einem öffentlichen Verzeichnis stehen. Bei einem symmetrischen Kryptosystem wäre der gleiche Zweck nur erreichbar mit $\frac{1}{2}n(n-1)$ Schlüsseln, die zudem noch durch ein sicheres Verfahren wie etwa ein persönliches Treffen oder durch vertrauenswürdige Boten ausgetauscht werden müßten.

DIFFIE und HELLMAN machten nur sehr vage Andeutungen, wie so ein System mit öffentlichen Schritten aussehen könnte. Es ist zunächst einmal klar, daß ein solches System keinerlei Sicherheit gegen einen BAYESschen Gegner bieten kann, denn die Verschlüsselungsfunktion ist eine bijektive Abbildung zwischen endlichen Mengen, und jeder, der die Funktion kennt, kann zumindest im Prinzip auch ihre Umkehrfunktion berechnen.

Wer im Gegensatz zum BAYESschen Gegner nur über begrenzte Ressourcen verfügt, kann diese Berechnung allerdings möglicherweise nicht mit realistischem Aufwand durchführen, und nur darauf beruht die Sicherheit eines Kryptosystems mit öffentlichen Schlüsseln. DIFFIE und HELLMAN bezeichnen eine Funktion, deren Umkehrfunktion nicht mit vertretbarem Aufwand berechnet werden kann, als *Einwegfunktion* und schlagen als Verschlüsselungsfunktion eine solche Einwegfunktion vor.

Damit hat man aber noch kein praktikables Kryptosystem, denn bei einer echten Einwegfunktion ist es auch für den legitimen Enpfänger nicht möglich, seinen Posteingang zu entschlüsseln. DIFFIE und HELLMAN schlagen deshalb eine Einwegfunktion mit *Falltür* vor, wobei der legitime Empfänger zusätzlich zu seinem öffentlichen Schlüssel noch über einen geheimen Schlüssel verfügt, mit dem er (und nur er) diese Falltür öffnen kann.

Natürlich hängt alles davon ab, ob es solche Einwegfunktionen mit Falltür wirklich gibt. DIFFIE und HELLMAN gaben keine an, und es gab unter den Experten einige Skepsis bezüglich der Möglichkeit, solche Funktionen zu finden.

Tatsächlich aber gab es damals bereits Systeme, die auf solchen Funktionen beruhten, auch wenn sie nicht in der offenen Literatur dokumentiert waren: Die britische *Communications-Electronics Security Group* (CESG) hatte bereits Ende der sechziger Jahre damit begonnen, nach entsprechenden Verfahren zu suchen, um die Probleme des Militärs mit dem Schlüsselmanagement zu lösen, aufbauend auf (impraktikablen) Ansätzen von AT&T zur Sprachverschlüsselung während des zweiten Weltkriegs. Die Briten sprachen nicht von Kryptographie mit öffentlichen Schlüsseln, sondern von *nichtgeheimer Verschlüsselung,* aber das Prinzip war das gleiche.

Erste Ideen dazu sind in einer auf Januar 1970 datierten Arbeit von JAMES H. ELLIS zu finden, ein praktikables System in einer auf den 20. November 1973 datierten Arbeit von CLIFF C. COCKS. Wie im Milieu üblich, gelangte nichts über diese Arbeiten an die Öffentlichkeit; erst 1997 veröffentlichten die *Goverment Communications Headquarters* (GCHQ), zu denen CESG gehört, einige Arbeiten aus der damaligen Zeit unter `http://www.cesg.gov.uk/about/nsecret/`.

Im akademischen Bereich gab es ein Jahr nach Erscheinen der Arbeit von DIFFIE und HELLMAN das erste Kryptosystem mit öffentlichen Schlüsseln: Drei Professoren am Massachussetts Institute of Technology fanden nach rund vierzig erfolglosen Ansätzen 1977 schließlich jenes System, das heute nach ihren Anfangsbuchstaben mit RSA bezeichnet wird: RON RIVEST, ADI SHAMIR und LEN ADLEMAN.

Das System wurde 1983 von der eigens dafür gegründeten Firma RSA Computer Security Inc. patentiert und mit großem kommerziellem Erfolg vermarktet. Das Patent lief zwar im September 2000 aus, die Firma ist aber weiterhin erfolgreich im Kryptobereich tätig; sie hatte beispielsweise auch einen Kandidaten für AES entwickelt, der es immerhin bis in die Endrunde schaffte.

RSA ist übrigens identisch mit dem von COCKS vorgeschlagenen System, so daß einige Historiker auch Zweifel an den Behauptungen der GCHQ haben. Die Beschreibung durch RIVEST, SHAMIR und ADLEMAN erschien 1978 unter dem Titel *A method for obtaining digital signatures and public-key cryptosystems* in Comm. ACM **21**, 120–126.

§2: Algebraische Vorbereitungen

Das RSA-Verfahren hängt ab von Multiplikationen modulo einer großen Zahl N, die Produkt zweier Primzahlen p und q ist; wir wollen uns daher zunächst etwas mit dem Rechnen modulo einer solchen Zahl N beschäftigen.

Wir wissen bereits, daß die ganzen Zahlen modulo einer Primzahl p einen Körper bilden, den Körper $\mathbb{F}_p$ mit p Elementen. Eine weitere für RSA wichtige Eigenschaft des Rechnens modulo p ist der folgende Satz:

Kleiner Satz von Fermat: Für jedes $a \in \mathbb{Z}$ ist
$$a^p \equiv a \bmod p\,;$$
ist a nicht durch p teilbar, ist auch
$$a^{p-1} \equiv 1 \bmod p\,.$$

Beweis: Für zwei Zahlen $x, y \in \mathbb{Z}$ ist
$$(x+y)^p = \binom{p}{0}x^p + \binom{p}{1}x^{p-1}y + \cdots + \binom{p}{p-1}xy^{p-1} + \binom{p}{p}y^p$$
mit
$$\binom{p}{j} = \frac{p!}{j!(p-j)!}\,.$$
Für $j = 0$ und $j = p$ ist $\binom{p}{0} = \binom{p}{p} = 1$, ansonsten sind j und $p - j$ beide kleiner als p. Im Nenner steht daher dann kein Faktor p, der das p aus dem Zähler wegkürzen könnte, so daß alle $\binom{p}{j}$ mit $1 \le j \le p-1$ durch p teilbar sind, und
$$(x+y)^p \equiv x^p + y^p \bmod p\,.$$
Induktiv folgt, daß eine entsprechende Gleichung auch für Summen mit mehr als zwei Summanden gilt, insbesondere gilt also für beliebig viele Summanden eins, daß
$$(1 + \cdots + 1)^p \equiv 1^p + \cdots + 1^p = 1 + \cdots + 1 \bmod p\,.$$
Damit ist $a^p \equiv a \bmod p$ für jede natürliches Zahl a. Für $a = 0$ gilt die Beziehung auch, und da für jedes positive a
$$0 = \bigl(a + (-a)\bigr)^p = a^p + (-a)^p = a + (-a)^p \bmod p$$

ist, gilt sie auch für negative a.

Falls a nicht durch p teilbar ist, gibt es, da $\mathbb{F}_p$ ein Körper ist, ein b mit $ba \equiv 1 \bmod p$; mit diesem b ist

$$a^{p-1} \equiv (ba)a^{p-1} = ba^p \equiv ba \equiv 1 \bmod p\,.$$

■

Korollar: Falls die natürliche Zahl e keinen Teiler mit $p-1$ gemeinsam hat, gibt es eine natürliche Zahl d, so daß

$$\left(a^e\right)^d \equiv a \bmod p$$

für alle $a \in \mathbb{Z}$.

Beweis: Nach dem erweiterten EUKLIDischen Algorithmus läßt sich der größte gemeinsame Teiler eins von e und $p-1$ als ganzzahlige Linearkombination dieser beiden Zahlen schreiben, es gibt also ganze Zahlen d und r, so daß

$$d \cdot e + r \cdot (p-1) = 1$$

ist. Hier könnte d eventuell noch negativ sein, aber indem wir nötigenfalls noch ein Vielfaches der Gleichung

$$(p-1) \cdot e - e \cdot (p-1) = 0$$

dazuaddieren, können wir annehmen, daß d positiv ist (und r dann natürlich negativ). Für nicht durch p teilbares a ist dann

$$\left(a^e\right)^d = a^{de} = a^{1-r(p-1)} = a \cdot \left(a^{p-1}\right)^{-r} \equiv a \cdot 1^{-r} = a \bmod p\,,$$

wie behauptet.

Für durch p teilbares a sind beide Seiten der zu beweisenden Kongruenz durch p teilbar, so daß sie auch in diesem Fall richtig ist.

■

Für RSA interessiert, wie bereits erwähnt, vor allem das Rechnen modulo des Produkts zweier Primzahlen; hier gilt entsprechend

Satz: Ist $N = pq$ Produkt zweier verschiedener Primzahlen p und q, so gilt
a) Für jede ganze Zahl a ist $a^{1+(p-1)(q-1)} \equiv a \bmod N$.
b) Falls die natürliche Zahl e keinen Teiler mit $(p-1)(q-1)$ gemeinsam hat, gibt es eine natürliche Zahl d, so daß

$$\left(a^e\right)^d \equiv a \bmod p$$

für alle $a \in \mathbb{Z}$.
c) Die Berechnung von $(p-1)(q-1)$ aus N ist äquivalent zur Faktorisierung von N.

Beweis: a) Nach dem kleinen Satz von FERMAT ist für jedes nicht durch p teilbare a und jedes r

$$a^{1+r(p-1)} \equiv a \bmod p\,.$$

Für Vielfache von p sind beide Seiten durch p teilbar, so daß die Gleichung tatsächlich für alle ganzen Zahlen a gilt.

Entsprechendes gilt natürlich auch für die Primzahl q, und damit ist für jede ganze Zahl s

$$a^{1+s(p-1)(q-1)} \equiv a \bmod p \qquad \text{und} \qquad a^{1+(p-1)(q-1)} \equiv a \bmod q\,.$$

Diese beiden Kongruenzen besagen, daß die Differenz zwischen linker und rechter Seite sowohl durch p als auch durch q teilbar ist, also auch durch deren Produkt N, d.h.

$$a^{1+s(p-1)(q-1)} \equiv a \bmod N\,.$$

b) Geht wegen *a)* genau wie beim Primzahlfall im obigen Korollar.

c) Falls die Faktorisierung $N = pq$ bekannt ist, läßt sich natürlich leicht $(p-1)(q-1)$ berechnen. Ist umgekehrt

$$(p-1)(q-1) = pq - p - q + 1 = N + 1 - (p+q)$$

bekannt, so kennt man außer dem Produkt N auch die Summe M der beiden Primfaktoren, und diese sind die leicht berechenbaren Lösungen der quadratischen Gleichung $x^2 - Mx + N = 0$.

■

§3: Das RSA-Verfahren zur Verschlüsselung und für elektronische Unterschriften

Zur praktischen Durchführung des RSA-Verfahrens wählt man sich zwei verschiedene Primzahlen p, q, die unbedingt geheim gehalten werden müssen, und eine natürliche Zahl e, die keinen gemeinsamen Teiler mit $(p-1)(q-1)$ hat. Dann veröffentlicht man die beiden Zahlen $N = pq$ und e als öffentlichen Schlüssel.

Sodann berechnet man zu $(p-1)(q-1)$ gemäß obigem Satz nach dem EUKLIDischen Algorithmus eine Zahl d, so daß

$$\left(a^e\right)^d \equiv a \bmod p$$

für alle a; diese Zahl ist der geheime Schlüssel.

Jeder, der den öffentlichen Schlüssel (N, e) kennt, kann Nachrichten verschlüsseln: Er bricht die Nachricht auf in Blöcke, die durch ganze Zahlen zwischen null und $N-1$ dargestellt werden können, berechnet für jeden so dargestellten Block den Chiffretext $b \equiv a^e \bmod N$, der als Zahl zwischen null und $N-1$ interpretiert und an den Inhaber des geheimen Schlüssels geschickt wird. (Wir schreiben in solchen Fällen in Zukunft kurz $b = a^e \bmod N$, wobei „mod" in diesem Zusammenhang als die Berechnung des Divisionsrests bei Division durch N zu interpretieren ist.)

Der Empfänger berechnet $b^d \bmod N$; da

$$b^d \equiv a^{ed} \equiv a \bmod N$$

ist, entschlüsselt dies die Nachricht.

Im Gegensatz zu symmetrischen Kryptoverfahren endet die Nützlichkeit des RSA-Verfahrens nicht mit der bloßen Möglichkeit einer Verschlüsselung; das Verfahren kann beispielsweise auch benutzt werden, um in Zugangskontrollsystemen, vor Geldautomaten oder bei einer Bestellung im Internet die Identität einer Person Zu beweisen: Nur der Inhaber des geheimen Schlüssels d kann zu einem gegebenen a eine Zahl b berechnen, für die $b^e \equiv a \bmod N$ ist.

Falls also der jeweilige Gegenüber eine Zufallszahl a erzeugt und als Antwort das zugehörige b verlangt, kann er anhand eines öffentlichen Schlüsselverzeichnisses die Richtigkeit von b überprüfen und sich so von der Identität seines Partners überzeugen. Im Gegensatz zu Kreditkarteninformation oder Paßwörtern ist dieses Verfahren auch immun gegen Abhören: Falls jedesmal ein neues zufälliges a erzeugt wird, nützt ein einmal abgehörtes b nichts.

Trotzdem ist das Verfahren in dieser Form nicht als Ersatz zur Übertragung von Kreditkarteninformation oder ähnlichem geeignet, da der Gegenüber anhand des öffentlichen Schlüssels jederzeit zu einer willkürlich gewählten Zahl b die Zahl $a = b^e \bmod N$ erzeugen kann um dann zu behaupten, er habe b als Antwort darauf empfangen. Man müßte also beispielsweise noch zusätzlich verlangen, daß die Zahl a eine spezielle Form hat, etwa daß die vordere Hälte der Ziffernfolge identisch mit der hinteren Hälfte ist.

Praktische Bedeutung hat vor allem eine weitere Variante: die elektronische Unterschrift. Her geht es darum, daß der Empfänger erstens davon überzeugt wird, daß eine Nachricht tatsächlich vom behaupteten Absender stammt, und daß er dies zweitens auch einem Dritten gegenüber beweisen kann. (In Deutschland sind solche elektronischen Unterschriften, sofern gewisse formale Voraussetzungen erfüllt sind, rechtsverbindlich.)

Um einen Nachrichtenblock a mit $0 \leq a < N$ zu unterschreiben, berechnet der Inhaber des öffentlichen Schlüsseln (N, e) mit seinem geheimen Schlüssel d die Zahl

$$b = a^d \bmod N$$

und sendet das Paar (a, b) an den Empfänger. Dieser überprüft, ob

$$b^e \equiv a \bmod N\,;$$

falls ja, akzeptiert er dies als unterschriebene Nachricht a. Da er ohne Kenntnis des geheimen Schlüssels d nicht in der Lage ist, den Block (a, b) zu erzeugen, kann er auch gegenüber einem Dritten beweisen, daß der Absender die Nachricht a unterschrieben hat.

Für kurze Nachrichten ist dieses Verfahren in der vorgestellten Form praktikabel; in vielen Fällen kann man sogar auf die Übermittlung von a verzichten, da $b^e \bmod N$ für ein falsch berechnetes b mit an Sicherheit grenzender Wahrscheinlichkeit keine sinnvolle Nachricht ergibt.

Falls die Übermittelte Nachricht geheimgehalten werden soll, müssen a und b natürlich noch vor der Übertragung mit dem öffentlichen Schlüssel des Empfängers oder nach irgendeinem anderen Kryptoverfahren verschlüsselt werden.

Bei langen Nachrichten ist die Verdoppelung der Nachrichtenlänge nicht mehr akzeptabel, und selbst, wenn man auf die Übertragung von a verzichten kann, ist das Unterschreiben jedes einzelnen Blocks sehr aufwendig. Deshalb unterschreibt man meist nicht die Nachricht selbst, sondern einen daraus extrahierten Hashwert. Dieser Wert muß natürlich erstens von der gesamten Nachricht abhängen, und zweitens muß es für den Empfänger (praktisch) unmöglich sein, zwei Nachrichten zu erzeugen, die zum gleichen Hashwert führen. Die Theorie (und Praxis) dieser sogenannten kollisionsfreien Hashfunktionen sei für den Augenblick zurückgestellt zugunsten der Sicherheit von RSA.

§4: Faktorisierungsverfahren

Der offensichtliche Angriff auf RSA ist die Faktorisierung der öffentlich bekannten Zahl N; sobald man diese in ihre beiden Primfaktoren p und q zerlegt hat, ist das Verfahren gebrochen. Wir wollen daher in diesem Paragraphen sehen, welche Möglichkeiten es gibt, N in seine Primfaktoren zu zerlegen.

a) Mögliche Ansätze zur Faktorisierung

Grundsätzlich gibt es zwei Klassen von Verfahren, mit denen man einen Teiler einer natürlichen Zahl N finden kann: Einmal Verfahren, deren erwartete Laufzeit von der Länge des Faktors abhängt, zum anderen solche, deren Laufzeit nur von N abhängt.

Bei der Anwendung von RSA wird man, um Verfahren der ersten Kategorie auszubremsen, p und q ungefähr gleich groß wählen, so daß diese Verfahren im schlechtestmöglichen Fall arbeiten müssen.

Die einfachste Art der Faktorisierung ist das Abdividieren von Primzahlen; zumindest für kleine Primfaktoren (mindestens bis etwa 2^{16}) ist dies auch die schnellste und effizienteste Methode, da die anderen Verfahren Schwierigkeiten haben, Produkte kleiner Primzahlen zu trennen.

Für etwas größere Faktoren bis zu etwa acht Dezimalstellen ist die POLLARDsche Monte-Carlo-Methode oder ρ-Methode sehr gut geeignet: Man erzeugt mit einem quadratischen Generator (z.B. $x_{i+1} = x_i^2 + c \bmod N$) Zufallszahlen und berechnet deren ggT mit der zu faktorisierenden Zahl. Da der EUKLIDische Algorithmus im Vergleich zur Erzeugung der Zufallszahlen relativ teuer ist, empfiehlt es sich, die erzeugten Zufallszahlen zunächst modulo N miteinander zu multiplizieren und dann erst in etwa jedem hundertsten Schritt den ggT von N mit diesem Produkt zu berechnen. (Dies setzt voraus, daß alle sehr kleinen Faktoren bereits abdividiert sind; sonst ist die Gefahr zu groß, daß im Produkt von hundert Zufallszahlen mehr als ein Primfaktor steckt.) Bei Faktoren mit mehr als acht Dezimalstellen wird die Methode schnell langsamer, so daß man dann zu alternativen Verfahren übergehen sollte.

Die nächste Klasse von Verfahren beruht auf gruppentheoretischen Überlegungen, im wesentlichen dem kleinen Satz von FERMAT im Falle zyklischer Gruppen und Verallgemeinerungen auf weitere Gruppen wie die multiplikative Gruppe eines Körpers $\mathbb{F}_{p^2}$ oder einer elliptischen Kurve. Diese Verfahren sind sehr effizient, wenn die Gruppenordnung nur relativ kleine Primteiler hat. Aus diesem Grund wurde früher häufig empfohlen, daß für die Primteiler p eines RSA-Modulus N sowohl $p-1$ als auch $p+1$ jeweils mindestens einen „großen“ Primfaktor haben sollen. Da diese Verfahren jedoch ihre Stärke bei Faktoren mit einer Länge von bis etwa 30 oder 35 Dezimalstellen haben und ein N mit 70 Dezimalstellen, wie wir bald sehen werden, heute auch aus anderen Gründen völlig unsicher ist, gilt diese Empfehlung heute nicht mehr, und wir müssen uns nicht genauer mit diesen Verfahren befassen.

Umso interessanter ist dagegen ein Verfahren, dessen Stärke bei nahe beieinander liegenden Primfaktoren liegt. Es wurde von PIERRE DE FERMAT vorgeschlagen und beruht auf der Formel

$$x^2 - y^2 = (x+y)(x-y)\,.$$

Ist $N = pq$ Produkt zweier ungerader Primzahlen, so ist

$$N = (x+y)(x-y) \qquad \text{mit} \qquad x = \frac{p+q}{2} \qquad \text{und} \qquad y = \frac{p-q}{2}\,;$$

zusammen mit obiger Formel folgt, daß dann

$$N + y^2 = x^2$$

ist.

FERMAT berechnet nun für $y = 0, 1, 2, \ldots$ die Zahlen $N + y^2$; falls er auf ein Quadrat x^2 stößt, berechnet er

$$\mathrm{ggT}(N, x+y) \qquad \text{und} \qquad \mathrm{ggT}(N, x-y)\,.$$

Wenn er Pech hat, sind dies die beiden Zahlen eins und N, wenn er Glück hat, sind es p und q. Für zufällig gewählte Paare (x, y) kommt beides jeweils mit fünfzigprozentiger Wahrscheinlichkeit vor.

Falls p und q nahe beieinander liegen, führt schon ein kleines y zur korrekten Faktorisierung; bei der Wahl der Primzahlen muß also darauf geachtet werden, daß sie zwar die gleiche *Größenordnung* haben, aber nicht zu weit beieinander liegen. Liegt etwa p in der Größenordnung von $2q$, hat die Differenz die gleiche Größenordnung wie p, und FERMATS Verfahren bräuchte etwa p Rechenschritte, wird also vom Aufwand her vergleichbar mit der Faktoriserung durch Abdividieren. Somit kann man sich auch gegen diese Attacke recht gut schützen.

Wirklich gefährlich sind eine Klasse von Siebverfahren, die auf FERMATS Methode aufbauen. Diese Verfahren sind die schnellsten derzeit bekannten zur Faktorisierung von RSA-Moduln; die Wahl einer sicheren Ziffernlänge hängt also davon ab, welche Zahlen diese Verfahren faktorisieren können.

b) Das quadratische Sieb

Das quadratische Sieb ist der Grundalgorithmus der ganzen Klasse; es ist logisch einfacher, allerdings vor allem für große Zahlen auch deutlich langsamer als die Variationen, mit denen wir uns im nächsten Abschnitt beschäftigen werden.

Bei allen diesen Verfahren geht es darum, Zahlenpaare (x, y) zu finden, für die

$$x^2 \equiv y^2 \bmod N$$

ist. Für diese erwarten wir, daß in etwa der Hälfte aller Fälle

$$\mathrm{ggT}(x+y, N) \qquad \text{und} \qquad \mathrm{ggT}(x-y, N)$$

nichttriviale Teiler von N sind.

Bei quadratischen Sieb betrachten wir dazu das Polynom

$$f(x) = \left(x + \left[\sqrt{N}\right]\right)^2 - N\,.$$

Offensichtlich ist für jedes x

$$f(x) \equiv \left(x + \left[\sqrt{N}\right]\right)^2 \bmod N\,,$$

wobei links und rechts verschiedene Zahlen stehen. Insbesondere steht links im allgemeinen keine Quadratzahl.

Falls wir allerdings Werte $x_1, x_2, \ldots, x_r$ finden können, für die das Produkt der $f(x_i)$ eine Quadratzahl ist, dann ist

$$\prod_{i=1}^{r} f(x_i) \equiv \prod_{i=1}^{r} \left(x + \left[\sqrt{N}\right]\right)^2 \bmod N$$

eine Relation der gesuchten Art.

Um die x_i zu finden, betrachten wir eine Menge $\mathcal{B}$ von Primzahlen, die sogenannte Faktorbasis. Typischerweise enthält $\mathcal{B}$ für die Faktorisierung einer etwa hundertstelligen Zahl etwa 100–120 Tausend Primzahlen, deren größte somit, wie die folgende Tabelle zeigt, im einstelligen Millionenbereich liegt.

n	n-te Primzahl	n	n-te Primzahl
100 000	1 299 709	600 000	8 960 453
200 000	2 750 159	700 000	10 570 841
300 000	4 256 233	800 000	12 195 257
400 000	5 800 079	900 000	13 834 103
500 000	7 368 787	1 000 000	15 485 863

Beim quadratischen Sieb interessieren nur x-Werte, für die $f(x)$ als Produkt von Primzahlen aus $\mathcal{B}$ (und eventuell auch Potenzen davon) darstellbar ist. Ist

$$f(x_i) = \prod_{p\in\mathcal{B}} p^{e_{ip}} ,$$

so ist

$$\prod_{i=1}^{r} f(x_i)^{\varepsilon_i} = \prod_{p\in\mathcal{B}} p^{\sum_{i=1}^{r} \varepsilon_i e_{ip}}$$

genau dann ein Quadrat, wenn

$$\sum_{i=1}^{r} \varepsilon_i e_{ip}$$

für alle $p \in \mathcal{B}$ gerade ist. Dies hängt natürlich nur ab von den ε_i mod 2 und den e_{ip} mod 2; wir können ε_i und e_{ip} daher als Elemente des Körpers mit zwei Elementen auffassen und bekommen dann über $\mathbb{F}_2$ das Gleichungssystem

$$\sum_{i=1}^{r} \varepsilon_i e_{ip} = 0 \quad \text{für alle } p \in \mathcal{B} .$$

Betrachten wir die ε_i als Variablen, ist dies ein homogenes lineares Gleichungssystem in r Variablen mit soviel Gleichungen, wie es Primzahlen in der Faktorbasis gibt. Dieses Gleichungssystem hat nichttriviale Lösungen, falls die Anzahl der Variablen die der Gleichungen übersteigt, falls es also mehr Zahlen x_i gibt, für die $f(x_i)$ über der Faktorbasis faktorisiert werden kann, als Primzahlen in der Faktorbasis.

Für jede nichttriviale Lösung ist

$$\prod_{i=1}^{r} f(x_i)^{\varepsilon_i} = \prod_{i=1}^{r} \left(x + \left[\sqrt{N}\right]\right)^{2\varepsilon_i} \bmod N$$

eine Relation der Form $x^2 \equiv y^2 \bmod N$, die mit einer Wahrscheinlichkeit von etwa ein halb zu einer Faktorisierung von N führt. Falls wir zehn linear unabhängige Lösungen des Gleichungssystems betrachten, führt also mit einer Wahrscheinlichkeit von etwa 99,9% mindestens eine davon zu einer Faktorisierung.

Da ε_i nur die Werte 0 und 1 annimmt, stehen in obigem Produkt natürlich keine echten Potenzen: Man multipliziert einfach nur die Faktoren miteinander, für die $\varepsilon_i = 1$ ist. Außerdem interessieren nicht die links- und rechtsstehenden Quadrate, sondern deren Quadratwurzeln; tatsächlich also berechnet

$$x = \prod_{p\in\mathcal{B}} p^{\frac{1}{2}\sum_{i=1}^{r} \varepsilon_i e_{ip}} \qquad \text{und} \qquad y = \prod_{i=1}^{r} \left(x + \left[\sqrt{N}\right]\right)^{\varepsilon_i} .$$

wobei beide Produkte nur modulo N berechnet werden müssen.

Zum besseren Verständnis des Verfahrens wollen wir versuchen, damit die Zahl 15 zu faktorisieren. Dies ist zwar eine sehr untypische Anwendung, da das quadratische Sieb üblicherweise erst für mindestens etwa vierzigstellige Zahlen angewandt wird, aber zumindestens das Prinzip sollte auch damit klarwerden.

Als Faktorbasis verwenden wir die Menge

$$\mathcal{B} = \{2, 3, 7, 11\} ;$$

die Primzahl fünf fehlt, da $3 \cdot 5 = 15$ ist und daher bei einer Faktorbasis, die sowohl drei als auch fünf enthält, die Gefahr zu groß ist, daß die linke wie auch die rechte Seite der Kongruenz durch fünfzehn teilbar ist. Bei realistischen Anwendungen muß man auf solche Überlegungen keine Rücksicht nehmen, denn dann sind die Elemente der Faktorbasis höchstens siebenstellig und somit erheblich kleiner als die gesuchten Faktoren.

Wir berechnen $f(x)$ für $x = 1, 2, \ldots,$ bis wir einige Funktionswerte haben, die über der Faktorbasis faktorisiert werden können. Die faktorisierbaren Werte sind in folgender Tabelle zusammengestellt:

x	$x + \left[\sqrt{N}\right]$	$f(x)$	*Faktorisierung*
1	4	1	
3	6	21	$3 \cdot 7$
5	8	49	7^2
6	9	66	$2 \cdot 3 \cdot 11$
10	13	154	$2 \cdot 7 \cdot 11$
54	57	3234	$2 \cdot 3 \cdot 7^2 \cdot 11$

Die erste und die dritte Zeile sind selbst schon Relationen der gesuchten Art, nämlich

$$4^2 \equiv 1 \bmod 15 \qquad \text{und} \qquad 8^2 \equiv 7^2 \bmod 15\,.$$

Die zweite Relation ist nutzlos, denn $8-7=1$ und $8+7=15$. Die erste dagegen führt zur Faktorisierung, denn

$$\mathrm{ggT}(4+1,15)=5 \qquad \text{und} \qquad \mathrm{ggT}(4-1,15)=3\,.$$

Da dies aber ein Zufall ist, der bei großen Werten von N so gut wie nie vorkommt, wollen wir das ignorieren und mit den Relationen zu $x=3,6,10$ und 51 arbeiten:

$$\begin{array}{rll} 6^2 & \equiv 3\cdot 7 & \bmod 15\\ 9^2 & \equiv 2\cdot 3\cdot 11 & \bmod 15\\ 13^2 & \equiv 2\cdot 7\cdot 11 & \bmod 15\\ 54^2 & \equiv 2\cdot 3\cdot 7^2\cdot 11 & \bmod 15 \end{array}$$

Multipliziert man die ersten drei dieser Relationen miteinander, folgt

$$(6\cdot 9\cdot 13)^2 \equiv (2\cdot 3\cdot 7\cdot 11)^2 \bmod 15$$

oder $702^2 \equiv 462^2 \bmod 15$. Da

$$\mathrm{ggT}(702-462,15)=\mathrm{ggT}(240,15)=15$$

ist, bringt das leider nichts.

Wir erhalten auch dann rechts ein Quadrat, wenn wir das Produkt der ersten, dritten und vierten Relation bilden; dies führt auf

$$(6\cdot 13\cdot 57)^2 \equiv (2\cdot 3\cdot 7^2\cdot 11)^2 \bmod 15$$

oder $4446^2 \equiv 3234^2 \bmod 15$. Hier ist

$$\mathrm{ggT}(4446-3234,15)=\mathrm{ggT}(1212,15)=3\,,$$

womit wir die Zahl 15 faktorisiert haben – wenn auch nicht unbedingt auf die einfachstmögliche Weise.

Bei realistischen Beispielen sind die Funktionswerte $f(x)$ deutlich größer als die Primzahlen aus der Faktorbasis; außerdem liegen die

vollständig faktorisierbaren Zahlen viel dünner als hier: Bei der Faktorisierung einer hundertstelligen Zahl etwa muß man davon ausgehen, daß nur etwa jeder 10^9-te Funktionswert über der Faktorbasis zerfällt.

Daher ist es wichtig, ein Verfahren zu finden, mit dem diese wenigen Funktionswerte schnell und einfach bestimmt werden können. Das ist zum Glück möglich:

Der Funktionswert $f(x)$ ist genau dann durch p teilbar, wenn

$$f(x) \equiv 0 \bmod p$$

ist. Da für $x,y,a,b\in\mathbb{Z}$ und mit $x\equiv y \bmod p$ und $a\equiv b \bmod p$ gilt

$$a+x\equiv b+y \bmod p \qquad \text{und} \qquad a\cdot x\equiv b\cdot y \bmod p$$

und für $n\in\mathbb{N}$ auch

$$x^n \equiv y^n \bmod p\,,$$

ist für jedes Polynom f mit ganzzahligen Koeffizienten

$$f(x)\equiv f(y) \bmod p\,.$$

Ist also insbesondere $f(x)\equiv 0 \bmod p$, so ist auch

$$f(x+kp)\equiv 0 \bmod p \quad \text{für alle } k\in\mathbb{Z}\,.$$

Es genügt daher, im Bereich $0\le x<p-1$ nach Werten zu suchen, für die $f(x)$ durch p teilbar ist.

Dazu kann man f auch als Polynom über dem Körper mit p Elementen betrachten und nach Nullstellen in diesem Körper suchen. Für Polynome großen Grades und große Werte von p kann dies recht aufwendig sein; hier, bei einem quadratischen Polynom, müssen wir natürlich einfach eine quadratische Gleichung lösen: In $\mathbb{F}_p$ wie in jedem anderen Körper auch gilt

$$f(x)=\left(x-\left[\sqrt{N}\right]\right)^2-N=0 \iff \left(x-\left[\sqrt{N}\right]\right)^2=N\,,$$

und diese Gleichung ist genau dann lösbar, wenn es ein Element $w\in\mathbb{F}_p$ gibt mit Quadrat N, wenn also in $\mathbb{Z}$

$$w^2\equiv N \bmod p$$