Diskretna matematika

(Matematičke osnove kriptografije javnog ključa)

Bilješke s predavanja

Andrej Dujella

Sadržaj

1 Elementarna teorija brojeva **2**
- 1.1 Djeljivost 2
- 1.2 Kongruencije 14
- 1.3 Kvadratni ostatci 31
- 1.4 Diofantske jednadžbe 38

2 Algebarske strukture **48**
- 2.1 Polugrupe i grupe 48
- 2.2 Prsteni i polja 61
- 2.3 Konačna polja 69

3 Kriptografija **74**
- 3.1 Kratki uvod u kriptografiju 74
- 3.2 Data Encryption Standard i Advanced Encryption Standard 76
- 3.3 RSA kriptosustav 86
- 3.4 Ostali kriptosustavi s javnim ključem 95
  - 3.4.1 Rabinov kriptosustav 95
  - 3.4.2 Kriptosustavi zasnovani na problemu diskretnog logaritma 97
  - 3.4.3 Merkle-Hellmanov kriptosustav . . . 102
  - 3.4.4 McElieceov kriptosustav 106
  - 3.4.5 NTRU kriptosustav 107

Poglavlje 1

Elementarna teorija brojeva

1.1 Djeljivost

Teorija brojeva je grana matematike koja se ponajprije bavi proučavanjem svojstava skupa prirodnih brojeva

$$\mathbb{N} = \{1, 2, 3, 4, \dots\}.$$

Jedno od osnovnih svojstava skupa $\mathbb{N}$ je da su na njemu definirane operacije zbrajanja i množenja koje zadovoljavaju zakone komutativnosti, asocijativnosti i distributivnosti:

$$m + n = n + m, \quad m \cdot n = n \cdot m, \quad \text{(komutativnost)},$$

$$(k + m) + n = k + (m + n), \quad (k \cdot m) \cdot n = k \cdot (m \cdot n), \quad \text{(asocijativnost)},$$

$$k \cdot (m + n) = k \cdot m + k \cdot n, \quad \text{(distributivnost)},$$

za sve prirodne brojeve k, m, n. Pored toga, na skupu $\mathbb{N}$ imamo uređaj takav da za svaka dva različita elementa m, n iz $\mathbb{N}$ vrijedi ili $m < n$ ili $n < m$. Nadalje, svaki neprazan podskup od $\mathbb{N}$ ima najmanji element, te vrijedi princip matematičke indukcije: *Ako je $S \subseteq \mathbb{N}$ za koji vrijedi da je $1 \in S$, te da $k \in S \Rightarrow k + 1 \in S$, onda je $S = \mathbb{N}$.* Ova svojstva ćemo u daljnjem često koristiti.

Osim svojstava skupa $\mathbb{N}$, proučavat ćemo i svojstva skupa cijelih brojeva $0, \pm 1, \pm 2, \pm 3, \dots$ kojeg ćemo označavati sa $\mathbb{Z}$, te skupa racionalnih brojeva, tj. brojeva oblika $\frac{p}{q}$ za $p \in \mathbb{Z}$, $q \in \mathbb{N}$, kojeg ćemo označavati s $\mathbb{Q}$.

Pojam djeljivosti je jedan od najjednostavnijih, ali ujedno i najvažnijih pojmova u teoriji brojeva. Stoga ćemo s njim započeti naša razmatranja.

Definicija 1.1. *Neka su $a \neq 0$ i b cijeli brojevi. Kažemo da je b* djeljiv *s a, odnosno da a* dijeli *b, ako postoji cijeli broj x takav da je $b = ax$. To zapisujemo sa $a|b$. Ako b nije djeljiv sa a, onda pišemo $a \nmid b$.*

Ako $a|b$, onda još kažemo da je a djelitelj *od b, a da je b* višekratnik *od a. Oznaka $a^k \| b$ će nam značiti da $a^k | b$, ali $a^{k+1} \nmid b$.*

Uočimo da je relacija "biti djeljiv" relacija parcijalnog uređaja na skupu $\mathbb{N}$. To znači da vrijedi:

- $m|m$, (refleksivnost),

- $k|m$ i $m|n \Rightarrow k|n$, (tranzitivnost),

- $m|n$ i $n|m \Rightarrow m = n$, (antisimetričnost),

za prirodne brojeve k, m, n. No, to nije relacija parcijalnog uređaja na skupu $\mathbb{Z}$ jer $m|n$ i $n|m$ povlači da je $a = \pm b$, pa ne vrijedi antisimetričnost.

Teorem 1.1 (Teorem o dijeljenju s ostatkom)**.** *Za proizvoljan prirodan broj a i cijeli broj b postoje jedinstveni cijeli brojevi q i r takvi da je $b = qa + r$, $0 \leq r < a$.*

Dokaz: Promotrimo skup $\{b - am : m \in \mathbb{Z}\}$. Najmanji nenegativni član ovog skupa označimo sa r. Tada je po definiciji $0 \leq r < a$ i postoji $q \in \mathbb{Z}$ takav da je $b - qa = r$, tj. $b = qa + r$.

Da bi dokazali jedinstvenost od q i r, pretpostavimo da postoji još jedan par q_1, r_1 koji zadovoljava iste uvjete. Pokažimo najprije da je $r_1 = r$. Pretpostavimo da je npr. $r < r_1$. Tada je $0 < r_1 - r < a$, dok je s druge strane $r_1 - r = a(q - q_1) \geq a$. Prema tome je $r_1 = r$, pa je stoga i $q_1 = q$. □

Definicija 1.2. *Neka su b i c cijeli brojevi. Cijeli broj a zovemo* zajednički djelitelj *od b i c ako $a|b$ i $a|c$. Ako je barem jedan od brojeva b i c različit od nule, onda postoji samo konačno mnogo zajedničkih djelitelja od b i c. Najveći među njima zove se* najveći zajednički djelitelj *od b i c i označava se s* $\mathrm{nzd}(b, c)$ *(ili* $\gcd(b, c)$ *od "greatest common divizor", ili samo (b, c)). Slično se definira najveći zajednički djelitelj brojeva $b_1, b_2, \ldots, b_n$ koji nisu svi jednaki nuli, te se označava s* $\mathrm{nzd}(b_1, b_2, \ldots, b_n)$.

Uočimo da je $\mathrm{nzd}(b, c) \geq 1$.

Teorem 1.2.

$$\mathrm{nzd}(b, c) = \min(\{bx + cy : x, y \in \mathbb{Z}\} \cap \mathbb{N})$$

Dokaz: Neka je $g = \mathrm{nzd}(b, c)$, te neka je l najmanji pozitivni član skupa $S = \{bx + cy : x, y \in \mathbb{Z}\}$. To znači da postoje cijeli brojevi x_0 i y_0 takvi da je $l = bx_0 + cy_0$.

Pokažimo da $l|b$ i $l|c$. Pretpostavimo da npr. $l \nmid b$. Tada po Teoremu 1.1 postoje cijeli brojevi q i r takvi da je $b = lq + r$ i $0 < r < l$. Sada je

$$r = b - lq = b - q(bx_0 + cy_0) = b(1 - qx_0) + c(-qy_0) \in S,$$

što je u suprotnosti s minimalnošću od l. Dakle, $l|b$, a na isti način se pokazuje da $l|c$. To znači da je $l \leq g$.

Budući da je $g = \text{nzd}(b, c)$, to postoje $\beta, \gamma \in \mathbb{Z}$ takvi da je $b = g\beta$, $c = g\gamma$, pa je $l = bx_0 + cy_0 = g(\beta x_0 + \gamma y_0)$. Odavde slijedi da je $g \leq l$, pa smo dokazali da je $g = l$. □

Ako se cijeli broj d može prikazati u obliku $d = bx + cy$, onda je $\text{nzd}(b, c)$ djelitelj od d. Posebno, ako je $bx + cy = 1$, onda je $\text{nzd}(b, c) = 1$.

Ako je d zajednički djeljitelj od b i c, onda $d | \text{nzd}(b, c)$. Zaista, d dijeli b i c, pa onda dijeli i $bx + cy$, te tvrdnja slijedi iz Teorema 1.2.

Definicija 1.3. *Reći ćemo da su cijeli brojevi a i b i* relativno prosti *ako je* $\text{nzd}(a, b) = 1$. *Za cijele brojeve $a_1, a_2, \ldots, a_n$ reći ćemo da su* relativno prosti *ako je* $\text{nzd}(a_1, a_2, \ldots, a_n) = 1$, *a da su* u parovima relativno prosti *ako je* $\text{nzd}(a_i, a_j) = 1$ *za sve* $1 \leq i, j \leq n$, $i \neq j$.

Na primjer, brojevi 6, 10 i 15 su relativno prosti, ali nisu u parovima relativno prosti.

Propozicija 1.3. *Ako je* $\text{nzd}(a, m) = \text{nzd}(b, m) = 1$, *onda je* $\text{nzd}(ab, m) = 1$.

Dokaz: Po Teoremu 1.2 postoje $x_0, y_0, x_1, y_1 \in \mathbb{Z}$ takvi da je $1 = ax_0 + my_0 = bx_1 + my_1$. Odavde je $ax_0bx_1 = (1 - my_0)(1 - my_1) = 1 - my_2$, gdje je $y_2 = y_0 + y_1 - my_0y_1$. Sada iz $abx_0x_1 + my_2 = 1$ zaključujemo da je $\text{nzd}(ab, m) = 1$. □

Propozicija 1.4.

$$\text{nzd}(a, b) = \text{nzd}(a, b + ax)$$

Dokaz: Označimo $\text{nzd}(a, b) = d$, $\text{nzd}(a, b + ax) = g$. Po Teoremu 1.2 postoje $x_0, y_0 \in \mathbb{Z}$ takvi da je $d = ax_0 + by_0$, odnosno

$$d = a(x_0 - xy_0) + (b + ax)y_0.$$

Odavde slijedi da $g|d$. Pokažimo sada da $d|g$. Budući $d|a$ i $d|b$ imamo da $d|(b + ax)$. Dakle, d je zajednički djelitelj od a i $b + ax$, pa po Teoremu 1.2 imamo da $d|g$.

Pošto su brojevi d i g pozitivni po definiciji, iz $d|g$ i $g|d$ slijedi da je $d = g$. □

Teorem 1.5 (Euklidov algoritam)**.** *Neka su b i $c > 0$ cijeli brojevi. Pretpostavimo da je uzastopnom primjenom Teorema 1.1 dobiven niz jednakosti*

$$\begin{aligned}
b &= cq_1 + r_1, \quad 0 < r_1 < c, \\
c &= r_1q_2 + r_2, \quad 0 < r_2 < r_1, \\
r_1 &= r_2q_3 + r_3, \quad 0 < r_3 < r_2, \\
&\ldots \\
r_{j-2} &= r_{j-1}q_j + r_j, \quad 0 < r_j < r_{j-1}, \\
r_{j-1} &= r_jq_{j+1}.
\end{aligned}$$

Tada je $\mathrm{nzd}(b,c)$ *jednak* r_j*, posljednjem ostatku različitom od nule. Vrijednosti od* x_0 *i* y_0 *u izrazu* $\mathrm{nzd}(b,c) = bx_0 + cy_0$ *mogu se dobiti izražavanjem svakog ostatka* r_i *kao linearne kombinacije od* b *i* c*.*

Dokaz: Po Propoziciji 1.4 imamo

$$\begin{aligned}\mathrm{nzd}(b,c) &= \mathrm{nzd}(b-cq_1,c) = \mathrm{nzd}(r_1,c) = \mathrm{nzd}(r_1,c-r_1q_2) = \mathrm{nzd}(r_1,r_2)\\ &= \mathrm{nzd}(r_1-r_2q_3,r_2) = \mathrm{nzd}(r_3,r_2).\end{aligned}$$

Nastavljajući ovaj proces, dobivamo: $\mathrm{nzd}(b,c) = \mathrm{nzd}(r_{j-1},r_j) = \mathrm{nzd}(r_j,0) = r_j$.

Indukcijom ćemo dokazati da je svaki r_i linearna kombinacija od b i c. To je točno za r_1 i r_2, pa pretpostavimo da vrijedi za r_{i-1} i r_{i-2}. Budući da je r_i linearna kombinacija od r_{i-1} i r_{i-2}, po pretpostavci indukcije dobivamo da je i linearna kombinacija od b i c. □

Euklidov algoritam je jedan od najstarijih, ali ujedno i jedan od najvažnijih algoritama u teoriji brojeva. Uz pretpostavku da je $b > c \geq 0$, te uz dogovor da nam "$b \bmod c$" označava ostatak pri dijeljenju broj b s c, možemo ga sažeto zapisati ovako:

Euklidov algoritam:

while $(c > 0,$
$\quad (b,c) = (c, b \bmod c)\ )$
return b

Primjer 1.1. *Odredimo* $d = \mathrm{nzd}(252,198)$ *i prikažimo* d *kao linearnu kombinaciju brojeva* 252 *i* 198*.*

Rješenje:

$$\begin{aligned}252 &= 198\cdot 1 + 54\\ 198 &= 54\cdot 3 + 36\\ 54 &= 36\cdot 1 + 18\\ 36 &= 18\cdot 2\end{aligned}$$

Dakle, $(252,198) = 18$. Nadalje, imamo:

$$\begin{aligned}18 &= 54 - 36\cdot 1 = 54 - (198 - 54\cdot 3)\cdot 1 = 4\cdot 54 - 1\cdot 198\\ &= 4\cdot(252 - 198\cdot 1) - 1\cdot 198 = 4\cdot 252 - 5\cdot 198.\end{aligned}$$

♢

Rješenja jednadžbe $bx+cy = \mathrm{nzd}(b,c)$ mogu se efikasno dobiti na slijedeći način: ako je

$$\begin{aligned}r_{-1} &= b, \quad r_0 = c; \quad r_i = r_{i-2} - q_i r_{i-1};\\ x_{-1} &= 1, \quad x_0 = 0; \quad x_i = x_{i-2} - q_i x_{i-1};\\ y_{-1} &= 0, \quad y_0 = 1; \quad y_i = y_{i-2} - q_i y_{i-1},\end{aligned}$$

onda je

$$bx_i + cy_i = r_i, \quad \text{za } i = -1, 0, 1, \ldots, j+1.$$

Ova formula je točna za $i = -1$ i $i = 0$, pa tvrdnja trivijalno slijedi indukcijom, jer obje strane formule zadovoljavaju istu rekuzivnu relaciju. Posebno, vrijedi:

$$bx_j + cy_j = \text{nzd}(b, c).$$

Primjer 1.2. *Odredimo $g = \text{nzd}(3587, 1819)$ i nađimo cijele brojeve x, y takve da je $3587x + 1819y = g$.*

Rješenje:

$$\begin{aligned} 3587 &= 1819 \cdot 1 + 1768 \\ 1819 &= 1768 \cdot 1 + 51 \\ 1768 &= 51 \cdot 34 + 34 \\ 51 &= 34 \cdot 1 + 17 \\ 34 &= 17 \cdot 2 \end{aligned}$$

i	-1	0	1	2	3	4
q_i			1	1	34	1
x_i	1	0	1	-1	35	-36
y_i	0	1	-1	2	-69	71

Dakle, $g = 17$, te $3587 \cdot (-36) + 1819 \cdot 71 = 17$. $\diamondsuit$

Zadatak 1.1. *Odredite $g = \text{nzd}(423, 198)$ i nađite cijele brojeve x, y takve da je $423x + 198y = g$.*

Zadatak 1.2. *Odredite cijele brojeve x, y takve da je*
a) $71x + 50y = 1$, *b)* $93x + 81y = 3$.

Verzija Euklidovog algoritma koja rauna ne samo $\text{nzd}(b, c)$, već i cijele brojeve x i y takve da je $bx + cy = \text{nzd}(b, c)$ naziva se prošireni Euklidov algoritam. Sa $\lfloor x \rfloor$ ćemo označavati najveći cijeli broj koji je $\leq x$. Tada je $\left\lfloor \frac{b}{c} \right\rfloor$ kvocijent pri dijeljenju b s c.

Prošireni Euklidov algoritam:

$(x, y, g, u, v, w) = (1, 0, b, 0, 1, c)$;
while$(w > 0,$
  $q = \lfloor g/w \rfloor$;
  $(x, y, g, u, v, w) = (u, v, w, x - qu, y - qv, g - qw)$)
return (x, y, g)

Propozicija 1.6. *Za broj koraka j u Euklidovom algoritmu vrijedi $j < 2 \log_2 c$.*

Dokaz: Pogledajmo i-ti korak. Imamo $r_i \leq \frac{r_{i-1}}{2}$ ili $\frac{r_{i-1}}{2} < r_i < r_{i-1}$. U ovom drugom slučaju imamo $q_{i+1} = 1$ i $r_{i+1} = r_{i-1} - r_i < \frac{r_{i-1}}{2}$. Dakle, u svakom slučaju je $r_{i+1} < \frac{r_{i-1}}{2}$. Odavde je

$$1 \leq r_j < \frac{r_{j-2}}{2} < \frac{r_{j-4}}{4} < \cdots < \frac{r_0}{2^{j/2}}$$

ako je j paran, a

$$2 \leq r_{j-1} < \frac{r_{j-3}}{2} < \cdots < \frac{r_0}{2^{(j-1)/2}}$$

ako je j neparan.

Dakle, u svakom slučaju je $c = r_0 > 2^{j/2}$, pa je $j < 2 \log_2 c$. □

Malo preciznijom analizom, koja uključuje Fibonaccijeve brojeve, može se dokazati da je za brojeve $1 \leq b, c \leq N$ broj koraka u Euklidovu algoritmu za računanje $\mathrm{nzd}(b, c)$ manji ili jednak

$$\left\lceil \frac{\ln(\sqrt{5}N)}{\ln((1+\sqrt{5})/2)} \right\rceil - 2 \approx 2.078 \ln N + 1.672 .$$

Poznato je da je prosječan broj koraka u Euklidovu algoritmu za brojeve b i c iz skupa $\{1, \ldots, N\}$ približno jednak

$$\frac{12 \ln 2}{\pi^2} \ln N + 0.14 \approx 0.843 \ln N + 0.14 .$$

Pojednostavljeno rečeno, broj koraka je $O(\ln N)$. Kako svaki korak Euklidova algoritma zahtijeva jedno dijeljenje brojeva $\leq N$, dobivamo da je složenost Euklidovog algoritma $O(\ln^3 N)$. Ovu ocjenu možemo poboljšati ako uočimo da u svakom koraku radimo sa sve manjim brojevima. Tako da je broj operacija

$$\begin{aligned} O(\ln b \cdot \ln q_1 \, + \, & \ln c \cdot \ln q_2 + \ln r_1 \cdot \ln q_3 + \cdots + \ln r_{n-2} \cdot \ln q_n) \\ & = O(\ln N \cdot (\ln q_1 + \ln q_2 + \cdots + \ln q_n)) \\ & = O(\ln N \cdot \ln(q_1 q_2 \cdots q_n)) = O(\ln^2 N) \end{aligned}$$

(posljednju jednakost dobijemo množeći sve lijeve i sve desne strane u jednakostima u Euklidovu algoritmu).

Zadatak 1.3. *Dokažite da, uz oznake iz Teorema 1.5, za* $i = 0, 1, \ldots, j+1$ *vrijedi* $x_{i-1}y_i - x_i y_{i-1} = (-1)^i$, *te* $\mathrm{nzd}(x_i, y_i) = 1$.

Propozicija 1.7. *Uz oznake iz Teorema 1.5, vrijedi:* $|x_j| \leq \frac{c}{2g}$, $|y_j| \leq \frac{b}{2g}$, *gdje je* $g = \mathrm{nzd}(b, c)$.

Dokaz: Pokažimo indukcijom da je $(-1)^i x_i \leq 0$, $(-1)^i y_i \geq 0$ za $i = -1, 0, 1, \ldots, j+1$. Za $i = -1, 0$ tvrdnja vrijedi po definiciji, a ako pretpostavimo da vrijedi za $i-2$, $i-1$, onda iz $x_i = x_{i-2} - q_i x_{i-1}$ slijedi $(-1)^i x_i = (-1)^{i-2} x_{i-2} + (-1)^{i-1} q_i x_{i-1} \leq 0$. Za y_i je dokaz sasvim analogan.

Prema tome, $|x_i| = |x_{i-2}| + q_i |x_{i-1}|$, $|y_i| = |y_{i-2}| + q_i |y_{i-1}|$. Nadalje, budući da je $r_{j+1} = 0$, imamo $\frac{b}{g} x_{j+1} = -\frac{c}{g} y_{j+1}$, pa iz $(\frac{b}{g}, \frac{c}{g}) = 1$ i $(x_{j+1}, y_{j+1}) = 1$ slijedi $|x_{j+1}| = \frac{c}{g}$, $|y_{j+1}| = \frac{b}{g}$. Uvrstimo li ovo u $|x_{j+1}| = |x_{j-1}| + q_{j+1}|x_j|$, $|y_{j+1}| = |y_{j-1}| + q_{j+1}|y_j|$ i uvažimo da je $q_{j+1} \geq 2$ (zbog $r_j < r_{j-1}$), dobivamo traženi rezultat. □

Prikazat ćemo još jedan algoritam za računanje najvećeg zajedničkog djelitelja, tzv. "binarni gcd algoritam". Kod njega se umjesto dijeljenja koriste samo operacije oduzimanja i pomaka (dijeljenja sa 2). Kao rezultat dobivamo algoritam koji ima veći broj koraka, ali su ti koraci jednostavniji. U samom algoritmu susrećemo dvije ideje. Prva je da iako je faktorizacija brojeva općenito težak problem, izdvajanje potencija broja 2 je vrlo jednostavno. Druga ideja je zamjena dijeljenja oduzimanjem, a povezana je s činjenicom da u originalnom Euklidovom algoritmu vrlo često umjesto dijeljenja zapravo imamo oduzimanje, jer je pripadni kvocijent jednak 1. Može se pokazati da vjerojatnost da je Euklidov kvocijent jednak q iznosi

$$P(q) = \log_2 \Big(1 + \frac{1}{(q+1)^2 - 1}\Big).$$

Tako je $P(1) \approx 0.415$, $P(2) \approx 0.170$, $P(3) \approx 0.093$, ... Dakle, u 41.5% slučajeva kvocijent je jednak 1. Ovi rezultati su u uskoj vezi s ranije navedenim prosječnim brojem koraka u Euklidovom algoritmu.

Označimo s $v_2(k)$ najveću potenciju broja 2 koja dijeli k.

Binarni gcd algoritam:

$\beta = \min\{v_2(a), v_2(b)\}$
$a = a/2^{v_2(a)}$; $b = b/2^{v_2(b)}$
while$(a \neq b,$
    $(a, b) = (\min\{a, b\}, |b - a|/2^{v_2(b-a)})$)
return $2^\beta a$

Definicija 1.4. *Prirodan broj $p > 1$ se zove* prost *ako p nema niti jednog djelitelja d takvog da je $1 < d < p$. Ako prirodan broj $a > 1$ nije prost, onda kažemo da je* složen.

Teorem 1.8. *Svaki prirodan broj $n > 1$ može se prikazati kao produkt prostih brojeva (s jednim ili više faktora).*

Dokaz: Dokazat ćemo teorem matematičkom indukcijom. Broj 2 je prost. Pretpostavimo da je $n > 2$, te da tvrdnja teorema vrijedi za sve m, $2 \leq m < n$. Želimo dokazati da se i n može prikazati kao produkt prostih faktora. Ako je n prost, nemamo što dokazivati. U protivnom je $n = n_1 n_2$, gdje je $1 < n_1 < n$ i $1 < n_2 < n$. Po pretpostavci indukcije, n_1 i n_2 su produkti prostih brojeva, pa stoga i n ima to svojstvo. □

Iz Teorema 1.8 slijedi da svaki prirodan broj n možemo prikazati u obliku

$$n = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_r^{\alpha_r},$$

gdje su $p_1, \ldots, p_r$ različiti prosti brojevi, a $\alpha_1, \ldots, \alpha_r$ prirodni brojevi. Ovakav prikaz broja n zvat ćemo *kanonski rastav* broja n na proste faktore.

Propozicija 1.9. *Ako je p prost broj i $p|ab$, onda $p|a$ ili $p|b$. Općenitije, ako $p|a_1 a_2 \cdots a_n$, onda p dijeli barem jedan faktor a_i.*

Dokaz: Ako $p \nmid a$, onda je $\text{nzd}(p, a) = 1$, pa postoje cijeli brojevi x i y takvi da je $ax + py = 1$. Sada je $abx + pby = b$, pa p dijeli b.

Općenitiju tvrdnju dokazujemo indukcijom. Pretpostavimo da tvrdnja vrijedi za produkte s manje od n faktora. Sada ako $p|a_1(a_2 \cdots a_n)$, onda $p|a_1$ ili $p|a_2 a_3 \cdots a_n$. Ako $p|a_2 a_3 \cdots a_n$, onda po induktivnoj pretpostavci $p|a_i$ za neki $i = 2, \ldots, n$. □

Teorem 1.10 (Osnovni teorem aritmetike)**.** *Faktorizacija svakog prirodnog broja $n > 1$ na proste faktore je jedinstvena do na poredak prostih faktora.*

Dokaz: Pretpostavimo da n ima dvije različite faktorizacije. Dijeleći s prostim brojevima koji su zajednički objema reprezentacijama, dobit ćemo jednakost oblika

$$p_1 p_2 \cdots p_r = q_1 q_2 \cdots q_s,$$

gdje su p_i, q_j prosti brojevi, ne nužno različiti, ali takvi da se niti jedan prost broj s lijeve strane ne pojavljuje na desnoj strani, tj. $p_i \neq q_j$ za sve i,j. Međutim, to je nemoguće jer iz $p_1|q_1 q_2 \cdots q_s$, po Propoziciji 1.9, slijedi pa p_1 dijeli barem jedan q_j. No, to znači da je $p_1 = q_j$, kontradikcija. □

Napomena 1.1. Primijetimo da analogon Teorema 1.10 ne vrijedi za cijele brojeve u (nekim) kvadratnim poljima. Kao primjer nejednoznačne faktorizacije na proste faktore u prstenu $\mathbb{Z}[\sqrt{-6}] = \{a + b\sqrt{-6} : a, b \in \mathbb{Z}\}$ navedimo ove dvije faktorizacije broja 10:

$$10 = 2 \cdot 5 = (2 + \sqrt{-6})(2 - \sqrt{-6}).$$

U primjenama Teorema 1.10 često ćemo prirodan broj a pisati u obliku $a = \prod_p p^{\alpha(p)}$, gdje je $\alpha(p) \geq 0$ i podrazumijevamo da je $\alpha(p) = 0$ za skoro sve proste brojeve p. Ako je $a = 1$, onda je $\alpha(p) = 0$ za sve p.

Ako je $a = \prod_p p^{\alpha(p)}$, $b = \prod_p p^{\beta(p)}$, $c = \prod_p p^{\gamma(p)}$ i $ab = c$, onda je po Teoremu 1.10, $\alpha(p) + \beta(p) = \gamma(p)$ za sve p. Dakle, ako $a|c$, onda je $\alpha(p) \leq \gamma(p)$. Obratno, ako je $\alpha(p) \leq \gamma(p)$, onda možemo definirati prirodan broj $b = \prod_p p^{\beta(p)}$ sa $\beta(p) = \gamma(p) - \alpha(p)$. Tada je $ab = c$, pa $a|c$. Prema tome, dobili smo da vrijedi

$$a|c \iff \alpha(p) \leq \gamma(p), \quad \forall p. \tag{1.1}$$

Kao posljedicu formule (1.1) dobivamo formulu

$$\mathrm{nzd}(a,b) = \prod_p p^{\min(\alpha(p),\beta(p))}. \tag{1.2}$$

Definicija 1.5. *Neka su $a_1, a_2, \ldots, a_n$ cijeli brojevi različiti od nule. Najmanji prirodan broj c za koji vrijedi da $a_i|c$ za sve $i = 1, 2, \ldots, n$ zove se* najmanji zajednički višekratnik brojeva $a_1, a_2, \ldots, a_n$ *i označava s* $\mathrm{nzv}(a_1, a_2, \ldots, a_n)$ *(ili* $\mathrm{lcm}(a_1, a_2, \ldots, a_n)$ *od "least common multiple", ili samo s* $[a_1, a_2, \ldots, a_n]$*).*

Iz (1.1) slijedi da je

$$\mathrm{nzv}(a,b) = \prod_p p^{\max(\alpha(p),\beta(p))}. \tag{1.3}$$

Propozicija 1.11.

$$\mathrm{nzd}(a,b) \cdot \mathrm{nzv}(a,b) = |ab|$$

Dokaz: Po Teoremu 1.10 i formulama (1.2) i (1.3), dovoljno je provjeriti da za sve realne brojeve x, y vrijedi:

$$\min(x,y) + \max(x,y) = x + y.$$

Zaista, ako je $x \leq y$, onda je $\min(x,y) + \max(x,y) = x + y$, a ako je $x > y$, onda je $\min(x,y) + \max(x,y) = y + x = x + y$. □

Reći ćemo da je prirodan broj a *(potpun) kvadrat* ako se može zapisati u obliku n^2, $n \in \mathbb{N}$. Iz Teorema 1.10 vidimo da je a potpun kvadrat ako i samo ako su svi eksponenti $\alpha(p)$ parni. Kažemo da je a *kvadratno slobodan* ako je 1 najveći kvadrat koji dijeli a. Stoga je a kvadratno slobodan ako i samo ako su svi eksponenti $\alpha(p)$ jednaki 0 ili 1. Konačno, ako je p prost, onda je $p^k \| a$ ekvivalentno s $k = \alpha(p)$.

Primjer 1.3. *Neka su a i b prirodni brojevi takvi da je $(a,b) = 1$, te da je ab potpun kvadrat. Dokazati da su tada a i b potpuni kvadrati.*

Rješenje: Neka je $a = \prod_p p^{\alpha(p)}$, $b = \prod_p p^{\beta(p)}$. Budući da je ab potpun kvadrat, broj $\alpha(p) + \beta(p)$ je paran za sve p. S druge strane, $(a,b) = 1$ povlači da je za sve p barem jedan od brojeva $\alpha(p)$, $\beta(p)$ jednak 0. No, to znači da su brojevi $\alpha(p)$ i $\beta(p)$ parni za sve p, pa su a i b potpuni kvadrati. ♢

Primjer 1.4. *Dokazati da svaki složen broj n ima prosti faktor $p \leq \sqrt{n}$.*

Rješenje: Neka je p najmanji djelitelj od n koji je veći od 1. Tada je p očito prost i postoji $m \in \mathbb{N}$ takav da je $n = p \cdot m$. Budući da je $m \geq p$, dobivamo da je $p \leq \sqrt{n}$. ♢

Primjer 1.4 možemo iskoristiti za generiranje tablice prostih brojeva tzv. *Eratostenovim sitom*. Recimo, na primjer, da želimo napraviti tablicu prostih brojeva ≤ 200. Napišemo sve prirodne brojeve od 2 do 200. Prekrižimo sve prave višekratnike broja 2, pa broja 3, pa broja 5. U svakom koraku, prvi neprekriženi broj je prost, te u idućem koraku križamo njegove prave višekratnike (prvi novoprekriženi broj će biti njegov kvadrat, jer su svi manji višekratnici već ranije prekriženi). U našem slučaju, nakon križanja višekratnika od 7, 11 i 13, tablica je gotova (jer je $17 > \sqrt{200}$).

Teorem 1.12 (Euklid)**.** *Skup svih prostih brojeva je beskonačan.*

Dokaz: Pretpostavimo da su $p_1, p_2, \ldots, p_k$ svi prosti brojevi. Promotrimo broj

$$n = 1 + p_1 p_2 \cdots p_k.$$

Uočimo da n nije djeljiv ni sa p_1, ni sa p_2, ..., ni sa p_k. Dakle, svaki prosti faktor p od n je različit od $p_1, \ldots, p_k$. Budući da je n ili prost ili ima prosti faktor, dobili smo prost broj različit od $p_1, \ldots, p_k$, što je kontradikcija. □

Za $x \in \mathbb{R}$, sa $\pi(x)$ ćemo označavati broj prostih brojeva koji su $\leq x$. Osnovni rezultat o distribuciji prostih brojeva je teorem o prostim brojevima (engl. Prime Number Theorem - PNT) koji kaže da je

$$\pi(x) \sim \frac{x}{\ln x}\,.$$

Ovu činjenicu je prvi naslutio Gauss, a dokazali su je neovisno Hadamard i de la Vallée Poussin 1896. godine.

Još bolja aproksimacija za funkciju $\pi(x)$ je logaritamsko-integralna funkcija

$$\mathrm{li}(x) = \int_2^x \frac{1}{\ln t}\, dt.$$

Po l'Hôpitalovu pravilu neposredno dobivamo da je

$$\lim_{x\to\infty} \frac{\mathrm{li}(x)}{x/\ln(x)} = 1.$$

Stoga je teorem o prostim brojevima ekvivalentan sa $\pi(x) \sim \mathrm{li}(x)$.

Primjer 1.5. *Dokazati da prostih brojeva oblika $4k + 3$ ima beskonačno mnogo.*

Rješenje: Pri dijeljenju sa 4 neparni prosti broj može dati ostatak 1 ili 3. Produkt brojeva oblika $4k+1$ i sam ima taj oblik. Zaista,

$$(4s+1)(4t+1) = 4(4st+s+t)+1.$$

Neka su sada $p_1, p_2, \ldots, p_n$ svi prosti brojevi oblika $4k+3$. Promotrimo broj

$$4p_1p_2\cdots p_n - 1.$$

Ako bi svi njegovi prosti faktori bili oblika $4k+1$, onda bi i on sam imao taj oblik. Prema tome, on ima barem jedan prosti faktor p oblika $4k+3$. Očito je $p \neq p_i$, za $i = 1, 2, \ldots, n$, pa smo dobili kontradikciju. ♢

Primjer 1.6. *Dokazati da za svaki prirodan broj n postoji n uzastopnih složenih brojeva.*

Rješenje: To su npr. brojevi

$$(n+1)!+2,\ (n+1)!+3,\ \ldots,\ (n+1)!+n,\ (n+1)!+n+1,$$

jer je $(n+1)!+j$ djeljivo sa j za $j = 2, 3, \ldots, n+1$. ♢

Propozicija 1.13. *Potencija s kojom prosti broj p ulazi u rastav broja $n!$ na proste faktore jednaka je*

$$\left\lfloor \frac{n}{p} \right\rfloor + \left\lfloor \frac{n}{p^2} \right\rfloor + \left\lfloor \frac{n}{p^3} \right\rfloor + \cdots .$$

Dokaz: U produktu $n! = 1 \cdot 2 \cdot 3 \cdots n$ ima točno $\left\lfloor \frac{n}{p} \right\rfloor$ faktora koji su višekratnici od p. Među njima je točno $\left\lfloor \frac{n}{p^2} \right\rfloor$ onih koji su također višekratnici od p^2, itd. Svaki faktor u produktu $n!$ koji je višekratnik od p^m, u gornjoj sumi je brojem točno m puta: kao višekratnik od $p, p^2, \ldots, p^m$. □

Primjer 1.7. *Odrediti s koliko nula završava broj* 562! .

Rješenje: Trebamo naći najveću potenciju broja 10 koja dijeli 562!. Budući da je $10 = 2 \cdot 5$ i $2 < 5$, dovoljno je naći najveću potenciju prostog broja 5 koja dijeli 562!. Prema Propoziciji 1.13, taj broj je jednak

$$\left\lfloor \frac{562}{5} \right\rfloor + \left\lfloor \frac{562}{25} \right\rfloor + \left\lfloor \frac{562}{125} \right\rfloor = 112 + 22 + 4 = 138.$$

♢

Zadatak 1.4. *S koliko nula završava broj* 2008! *?*

Primjer 1.8. *Dokazati da ne postoji polinom $f(x)$ s cjelobrojnim koeficijentima, stupnja ≥ 1, takav da je $f(n)$ prost za sve $n \in \mathbb{N}$.*

Rješenje: Neka je $f(1) = p$. Tada je p prost broj. Budući da je $f(1 + kp) - f(1)$ djeljivo sa $(1 + kp) - 1 = kp$ (jer $x - y$ dijeli $x^m - y^m$), slijedi da $p|f(1 + kp)$, za svaki $k \in \mathbb{N}$. Međutim, $f(1 + kp)$ je prost, pa mora biti $f(1 + kp) = p$, $\forall k \in \mathbb{N}$. Budući da polinom $f(x) - p$ ima beskonačno mnogo nultočaka, on mora biti nulpolinom, pa je $f(x) = p$, što je u suprotnosti s pretpostavkom da je $\operatorname{st} f \geq 1$. $\diamondsuit$

Puno teži problem je odrediti polinome $f(x)$ takve da je $f(n)$ prost za beskonačno mnogo prirodnih brojeva n. Zna se da to vrijedi za linearne polinome $f(x) = ax + b$ ako je $(a, b) = 1$ (Dirichletov teorem o prostim brojevima u aritmetičkom nizu). No, već za polinom $f(x) = x^2 + 1$, to je otvoreno pitanje. Hipoteza je da tvrdnja vrijedi za sve polinome koji su ireducibilni i za koje ne postoji prirodan broj $d > 1$ takav da $d|f(n)$, $\forall n \in \mathbb{N}$.

Primjer 1.9. *Neka je broj $2^k + 1$ prost. Dokazati da je tada $k = 0$ ili $k = 2^n$ za neki $n \geq 0$.*

Rješenje: Pretpostavimo da k ima neki neparan prosti faktor p. Tada iz $k = p \cdot m$ slijedi da je broj

$$2^k + 1 = (2^m)^p + 1^p = (2^m + 1)(2^{m(p-1)} - 2^{m(p-2)} + \cdots + 1)$$

djeljiv s $2^m + 1$, pa nije prost. $\diamondsuit$

Brojevi $f_n = 2^{2^n} + 1$ nazivaju se *Fermatovi brojevi*. Fermat je smatrao da su svi oni prosti. Zaista, $f_0 = 3$, $f_1 = 5$, $f_2 = 17$, $f_3 = 257$ i $f_4 = 65537$ su prosti. Međutim, $f_5 = 2^{32} + 1$ je složen. Pokažimo to!

$$\begin{aligned} 2^{32} + 1 &= 2^4 \cdot 2^{28} + 1 = (641 - 5^4) \cdot 2^{28} + 1 = 641 \cdot 2^{28} - (5 \cdot 2^7)^4 + 1 \\ &= 641 \cdot 2^{28} - (641 - 1)^4 + 1 \\ &= 641 \cdot (2^{28} - 641^3 + 4 \cdot 641^2 - 6 \cdot 641 + 4) \end{aligned}$$

Prema tome, $641|f_5$.

Hipoteza je da je samo konačno mnogo Fermatovih brojeva prosto.

Zadatak 1.5. *Dokažite da za $m \neq n$ vrijedi $(f_m, f_n) = 1$. Pokažite da ova činjenica povlači da prostih brojeva ima beskonačno mnogo.*

Primjer 1.10. *Neka je broj $2^n - 1$ prost. Dokazati da je tada i broj n prost.*

Rješenje: Pretpostavimo da je broj n složen, tj. $n = ab$, $a > 1$, $b > 1$. Tada je broj $2^n - 1 \overset{=}{} (2^a)^b - 1^b$ djeljiv s $2^a - 1$, pa nije prost. $\diamondsuit$

Brojevi $M_p = 2^p - 1$, gdje je p prost, zovu se *Mersennovi brojevi*. Neki Mersennovi brojevi su prosti, kao npr. $M_7 = 127$, a neki su složeni, kao npr. $M_{11} = 2047 = 23 \cdot 89$. Hipoteza je da Mersennovih brojeva koji su prosti ima beskonačno mnogo. Najveći poznati prosti Mersennov broj je $M_{43112609}$. To je ujedno i najveći danas poznati prosti broj (ima 12978189 znamenaka; otkrili su ga 23.8.2008. Smith, Woltman i Kurowski u okviru GIMPS projekta).

1.2 Kongruencije

Teoriju kongruencija uveo je u svom djelu *Disquisitiones Arithmeticae* iz 1801. godine Carl Friedrich Gauss (1777-1855), jedan od najvećih matematičara svih vremena. On je također uveo i oznaku za kongruenciju koju i danas rabimo.

Definicija 1.6. *Ako cijeli broj $m \neq 0$ dijeli razliku $a-b$, onda kažemo da je a* kongruentan *b modulo m i pišemo $a \equiv b \pmod{m}$. U protivnom, kažemo da a nije kongruentan b modulo m i pišemo $a \not\equiv b \pmod{m}$.*

Budući da je $a - b$ djeljivo s m ako i samo ako je djeljivo s $-m$, bez smanjenja općenitosti možemo se usredotočiti na pozitivne module i kod nas će ubuduće modul m biti prirodan broj. Kongruencije imaju mnoga svojstva zajednička s jednakostima.

Propozicija 1.14. *Relacija "biti kongruentan modulo m"je relacija ekvivalencije na skupu $\mathbb{Z}$.*

Dokaz: Treba provjeriti refleksivnost, simetričnost i tranzitivnost.

(1) Iz $m|0$ slijedi $a \equiv a \pmod{m}$.

(2) Ako je $a \equiv b \pmod{m}$, onda postoji $k \in \mathbb{Z}$ takav $a - b = mk$. Sada je $b - a = m \cdot (-k)$, pa je $b \equiv a \pmod{m}$.

(3) Iz $a \equiv b \pmod{m}$ i $b \equiv c \pmod{m}$ slijedi da postoje $k, l \in \mathbb{Z}$ takvi da je $a - b = mk$ i $b - c = ml$. Zbrajanjem dobivamo $a - c = m(k + l)$, što povlači $a \equiv c \pmod{m}$. □

Još neka od jednostavnih svojstava kongruencija dana su u sljedećoj propoziciji.

Propozicija 1.15. *Neka su a, b, c, d cijeli brojevi.*

(1) Ako je $a \equiv b \pmod{m}$ i $c \equiv d \pmod{m}$, onda je $a + c \equiv b + d \pmod{m}$, $a - c \equiv b - d \pmod{m}$, $ac \equiv bd \pmod{m}$.

(2) Ako je $a \equiv b \pmod{m}$ i $d|m$, onda je $a \equiv b \pmod{d}$.

(3) Ako je $a \equiv b \pmod{m}$, onda je $ac \equiv bc \pmod{mc}$ za svaki $c \neq 0$.

Dokaz: (1) Neka je $a - b = mk$ i $c - d = ml$. Tada je $(a + c) - (b + d) = m(k + l)$ i $(a - c) - (b - d) = m(k - l)$, pa je $a + c \equiv b + d \pmod{m}$ i $a - c \equiv b - d \pmod{m}$. Zbog $ac - bd = a(c - d) + d(a - b) = m(al + dk)$ slijedi da je $ac \equiv bd \pmod{m}$.

(2) Neka je $m = de$. Tada iz $a - b = mk$ slijedi $a - b = d \cdot (ek)$, pa je $a \equiv b \pmod{d}$.

(3) Iz $a - b = mk$ slijedi $ac - bc = (mc) \cdot k$, pa je $ac \equiv bc \pmod{mc}$. □

Propozicija 1.16. *Neka je f polinom s cjelobrojnim koeficijentima. Ako je $a \equiv b \pmod{m}$, onda je $f(a) \equiv f(b) \pmod{m}$.*

Dokaz: Neka je $f(x) = c_n x^n + c_{n-1}x^{n-1} + \cdots + c_0$, gdje su $c_i \in \mathbb{Z}$. Budući da je $a \equiv b \pmod m$, uzastopnom primjenom Propozicije 1.15.1) dobivamo: $a^2 \equiv b^2 \pmod m$, $a^3 \equiv b^3 \pmod m$, $\ldots$, $a^n \equiv b^n \pmod m$. Tada je $c_i a^i \equiv c_i b^i \pmod m$ i konačno:

$$c_n a^n + c_{n-1}a^{n-1} + \cdots + c_0 \equiv c_n b^n + c_{n-1}b^{n-1} + \cdots + c_0 \pmod m.$$

□

Teorem 1.17. *Vrijedi:* $ax \equiv ay \pmod m$ *ako i samo ako* $x \equiv y \pmod{\frac{m}{\mathrm{nzd}(a,m)}}$. *Specijalno, ako je* $ax \equiv ay \pmod m$ *i* $\mathrm{nzd}(a,m) = 1$, *onda je* $x \equiv y \pmod m$.

Dokaz: Ako je $ax \equiv ay \pmod m$, onda postoji $z \in \mathbb{Z}$ takav da je $ay - ax = mz$. Sada imamo: $\frac{a}{\mathrm{nzd}(a,m)}(y-x) = \frac{m}{\mathrm{nzd}(a,m)}z$, tj. $\frac{m}{\mathrm{nzd}(a,m)}$ dijeli $\frac{a}{\mathrm{nzd}(a,m)}(y-x)$. No, brojevi $\frac{a}{\mathrm{nzd}(a,m)}$ i $\frac{m}{\mathrm{nzd}(a,m)}$ su relativno prosti, pa zaključujemo da $\frac{m}{\mathrm{nzd}(a,m)}$ dijeli $y-x$, tj. da je $x \equiv y \pmod{\frac{m}{\mathrm{nzd}(a,m)}}$.

Obrnuto, ako je $x \equiv y \pmod{\frac{m}{\mathrm{nzd}(a,m)}}$, onda po Propoziciji 1.15.3) dobivamo $ax \equiv ay \pmod{\frac{am}{\mathrm{nzd}(a,m)}}$. No, $\mathrm{nzd}(a,m)$ je djelitelj od a, pa po Propoziciji 1.15.2) dobivamo $ax \equiv ay \pmod m$. □

Definicija 1.7. *Skup* $\{x_1, \cdots, x_m\}$ *se zove* potpuni sustav ostataka *modulo* m *ako za svaki* $y \in \mathbb{Z}$ *postoji točno jedan* x_j *takav da je* $y \equiv x_j \pmod m$. *Drugim riječima, potpuni sustav ostataka dobivamo tako da iz svake klase ekvivalencije modulo* m *uzmemo po jedan član.*

Očito je da postoji beskonačno mnogo potpunih sustava ostataka modulo m. Jedan od njih je tzv. sustav najmanjih nenegativnih ostataka:

$$\{0, 1, \ldots, m-1\}.$$

Pored njega, često se koristi i sustav apsolutno najmanjih ostataka. Ako je m neparan broj, apsolutno najmanji ostatci su

$$-\frac{m-1}{2}, -\frac{m-3}{2}, \ldots, -1, 0, 1, \ldots, \frac{m-3}{2}, \frac{m-1}{2},$$

a ako je m paran, onda su to

$$-\frac{m-2}{2}, -\frac{m-4}{2}, \ldots, -1, 0, 1, \ldots, \frac{m-2}{2}, \frac{m}{2}.$$

Teorem 1.18. *Neka je* $\{x_1, \ldots, x_m\}$ *potpuni sustav ostataka modulo* m, *te neka je* $\mathrm{nzd}(a,m) = 1$. *Tada je* $\{ax_1, \ldots, ax_m\}$ *također potpuni sustav ostataka modulo* m.

Dokaz: Dovoljno je dokazati da je $ax_i \not\equiv ax_j \pmod{m}$ za $i \neq j$. Pretpostavimo da je $ax_i \equiv ax_j \pmod{m}$. Tada Teorem 1.17 povlači da je $x_i \equiv x_j \pmod{m}$, tj. $i = j$. □

Neka je $f(x)$ polinom s cjelobrojnim koeficijentima. Rješenje kongruencije $f(x) \equiv 0 \pmod{m}$ je svaki cijeli broj x koji je zadovoljava. Ako je x_1 neko rješenje ove kongruencije, a $x_2 \equiv x_1 \pmod{m}$, onda je, po Propoziciji 1.16, x_2 također rješenje. Dva rješenja x i x' smatramo ekvivalentnim ako je $x \equiv x' \pmod{m}$. Broj rješenja kongruencije je broj neekvivalentnih rješenja.

Teorem 1.19. *Neka su a i m prirodni, te b cijeli broj. Kongruencija $ax \equiv b \pmod{m}$ ima rješenja ako i samo ako $d = \mathrm{nzd}(a, m)$ dijeli b. Ako je ovaj uvjet zadovoljen, onda gornja kongruencija ima točno d rješenja modulo m.*

Dokaz: Ako kongruencija $ax \equiv b \pmod{m}$ ima rješenja, onda postoji $y \in \mathbb{Z}$ tako da je $ax - my = b$. Odavde je očito da $\mathrm{nzd}(a, m) | b$. Pretpostavimo sada da $d = \mathrm{nzd}(a, m)$ dijeli b. Stavimo $a' = \frac{a}{d}$, $b' = \frac{b}{d}$, $m' = \frac{m}{d}$. Sada trebamo riješiti kongruenciju $a'x \equiv b' \pmod{m'}$. No, ona ima točno jedno rješenje modulo m'. Zaista, budući da je $\mathrm{nzd}(a', m') = 1$, po Teoremu 1.18, kad x prolazi potpunim sustavom ostataka modulo m' i $a'x$ prolazi tim istim sustavom, tj. svaki ostatak modulo m' (pa tako i b') se dobiva točno za jedan x iz potpunog sustava ostataka modulo m'.

Jasno je da ako je x' neko rješenje od $a'x' \equiv b' \pmod{m'}$, onda su sva rješenja od $ax \equiv b \pmod{m}$ u cijelim brojevima dana sa $x = x' + nm'$, za $n \in \mathbb{Z}$, a sva međusobno neekvivalentna rješenja sa $x = x' + nm'$, gdje je $n = 0, 1, \ldots, d-1$. Dakle, ako d dijeli b, onda kongruencija $ax \equiv b \pmod{m}$ ima točno d rješenja modulo m. □

Iz Teorema 1.19 slijedi da ako je p prost broj i a nije djeljiv s p, onda kongruencija $ax \equiv 1 \pmod{p}$ uvijek ima rješenje i to rješenje je jedinstveno. To znači da a ima "multiplikativni inverz modulo p". Ovo pak povlači da skup ostataka $\{0, 1, \ldots, p-1\}$ pri dijeljenju sa p, uz zbrajanje i množenje (mod p), čini polje. To polje se obično označava sa $\mathbb{Z}_p$ ili $\mathbb{F}_p$.

Postavlja se pitanje kako riješiti kongruenciju $a'x \equiv b' \pmod{m'}$, gdje je $\mathrm{nzd}(a', m') = 1$. Budući da je $\mathrm{nzd}(a', m') = 1$, to postoje brojevi $u, v \in \mathbb{Z}$ takvi da je $a'u + m'v = 1$ i u, v se mogu naći pomoću Euklidovog algoritma. Sada je $a'u \equiv 1 \pmod{m'}$, pa je $x \equiv ub' \pmod{m'}$.

Primjer 1.11. *Riješimo kongruenciju* $555x \equiv 15 \pmod{5005}$.

Rješenje: Budući da je $(555, 5005) = 5$ i $5|15$, treba riješiti kongruenciju

$$111x \equiv 3 \pmod{1001}.$$

Primijenimo Euklidov algoritam:

$$\begin{aligned} 1001 &= 111 \cdot 9 + 2 \\ 111 &= 2 \cdot 55 + 1 \\ 2 &= 1 \cdot 2 \end{aligned}$$

i	-1	0	1	2
q_i			9	55
y_i	0	1	-9	496

Dakle, rješenje kongruncije $111u \equiv 1 \pmod{1001}$ je $u \equiv 496 \pmod{1001}$. Stoga je rješenje od $111x \equiv 3 \pmod{1001}$, $x \equiv 1488 \equiv 487 \pmod{1001}$. Konačno, rješenje polazne kongruencije je

$$x \equiv 487, 1488, 2489, 3490, 4491 \pmod{5005}.$$

♢

Zadatak 1.6. *Riješite kongruencije*
a) $589x \equiv 209 \pmod{817}$, *b)* $49x \equiv 5000 \pmod{999}$.

Kineski teorem o ostacima (engl. Chinese Remainder Theorem - CRT) govori o rješenju sustava linearnih kongruencija. Ime mu se vezuje uz kineskog matematičara iz prvog stoljeća Sun Tzua. Smatra se da je teorem već tada korišten u kineskoj vojsci za prebrojavanje vojnika. Pretpostavimo da treba prebrojiti grupu od približno 1000 vojnika. Vojnici se rasporede npr. u 3, 4, 5 i 7 kolona, te se zabilježi koliko je vojnika ostalo kao "višak" u zadnjem redu. Tako dobivamo sustav od četiri kongruencije s modulima 3, 4, 5 i 7, a taj sustav, prema sljedećem teoremu, ima jedinstveno rješenje između 800 i 1200.

Teorem 1.20 (Kineski teorem o ostatcima)**.** *Neka su $m_1, m_2, \ldots, m_r$ u parovima relativno prosti prirodni brojevi, te neka su $a_1, a_2, \ldots, a_r$ cijeli brojevi. Tada sustav kongruencija*

$$x \equiv a_1 \pmod{m_1}, \quad x \equiv a_2 \pmod{m_2}, \quad \ldots, \quad x \equiv a_r \pmod{m_r} \tag{1.4}$$

ima rješenja. Ako je x_0 jedno rješenje, onda su sva rješenja od (1.4) dana sa $x \equiv x_0 \pmod{m_1 m_2 \cdots m_r}$.

Dokaz: Neka je $m = m_1 m_2 \cdots m_r$, te neka je $n_j = \frac{m}{m_j}$ za $j = 1, \ldots, r$. Tada je $\mathrm{nzd}(m_j, n_j) = 1$, pa postoji cijeli broj x_j takav da je $n_j x_j \equiv a_j \pmod{m_j}$. Promotrimo broj

$$x_0 = n_1 x_1 + \cdots + n_r x_r.$$

Za njega vrijedi: $x_0 \equiv 0 + \cdots + 0 + n_j x_j + 0 + \cdots + 0 \equiv a_j \pmod{m_j}$. Prema tome, x_0 je rješenje od (1.4).

Ako su sada x, y dva rješenja od (1.4), onda je $x \equiv y \pmod{m_j}$ za $j = 1, \ldots, r$, pa jer su m_j u parovima relativno prosti, dobivamo da je $x \equiv y \pmod{m}$. □

Složenost algoritma opisanog u dokazu Kineskog teorema o ostacima je $O(\ln^2 m)$.

Primjer 1.12. *Riješimo sustav:*

$$x \equiv 2 \pmod 5, \quad x \equiv 3 \pmod 7, \quad x \equiv 4 \pmod{11}.$$

Rješenje: Uz oznake iz Teorema 1.20 imamo da je $x_0 = 77x_1 + 55x_2 + 35x_3$, gdje x_1, x_2, x_3 zadovoljavaju

$$77x_1 \equiv 2 \pmod 5, \quad 55x_2 \equiv 3 \pmod 7, \quad 35x_3 \equiv 4 \pmod{11},$$

odnosno

$$2x_1 \equiv 2 \pmod 5, \quad 6x_2 \equiv 3 \pmod 7, \quad 2x_3 \equiv 4 \pmod{11}.$$

Stoga možemo uzeti $x_1 = 1$, $x_2 = 4$, $x_3 = 2$, što daje $x_0 = 367$. Prema tome, sva rješenja našeg sustava dana su sa $x \equiv 367 \pmod{385}$. ♢

Zadatak 1.7. *Riješite sustav kongruencija*

$$x \equiv 5 \pmod 7, \quad x \equiv 7 \pmod{11}, \quad x \equiv 3 \pmod{13}.$$

Primjer 1.13. *Riješimo sustav kongruencija*

$$x \equiv 3 \pmod{10}, \quad x \equiv 8 \pmod{15}, \quad x \equiv 5 \pmod{84}.$$

Rješenje: Uočimo da brojevi 10, 15 i 84 nisu u parovima relativno prosti, pa ne možemo Kineski teorem o ostatcima primjeniti direktno, a može se dogoditi da takav sustav uopće nema rješenja. Sada postupamo ovako. Naš sustav je ekvivalentan sa

$$x \equiv 3 \pmod 2, \quad x \equiv 3 \pmod 5, \quad x \equiv 8 \pmod 3, \quad x \equiv 8 \pmod 5,$$

$$x \equiv 5 \pmod 4, \quad x \equiv 5 \pmod 3, \quad x \equiv 5 \pmod 7.$$

Dakle, moduli su nam potencije prostih brojeva i sada usporedimo kongruencije koje odgovaraju istom prostom broju:

$$x \equiv 3 \pmod 2, \quad x \equiv 5 \pmod 4 \iff x \equiv 1 \pmod 4,$$

$$x \equiv 8 \pmod 3, \quad x \equiv 5 \pmod 3 \iff x \equiv 2 \pmod 3,$$

$$x \equiv 3 \pmod 5, \quad x \equiv 8 \pmod 5 \iff x \equiv 3 \pmod 5,$$

$$x \equiv 5 \pmod 7.$$

Prema tome, naš sustav je ekvivalentan sa sustavom

$$x \equiv 1 \pmod 4, \quad x \equiv 2 \pmod 3, \quad x \equiv 3 \pmod 5, \quad x \equiv 5 \pmod 7$$

na kojeg možemo doslovno primijeniti Kineski teorem o ostatcima. Imamo: $m = 4 \cdot 3 \cdot 5 \cdot 7 = 420$, $n_1 = 105$, $n_2 = 140$, $n_3 = 84$, $n_4 = 60$,

$$\begin{aligned} 105x_1 &\equiv 1 \pmod 4 \iff x_1 \equiv 1 \pmod 4 \implies x_1 = 1, \\ 140x_2 &\equiv 2 \pmod 3 \iff 2x_1 \equiv 2 \pmod 3 \implies x_2 = 1, \\ 84x_3 &\equiv 3 \pmod 5 \iff 4x_3 \equiv 3 \pmod 5 \implies x_3 = 2, \\ 60x_4 &\equiv 5 \pmod 7 \iff 4x_4 \equiv 5 \pmod 7 \implies x_4 = 3. \end{aligned}$$

Dakle, rješenje je

$$x \equiv 105 \cdot 1 + 140 \cdot 1 + 84 \cdot 2 + 60 \cdot 3 = 593 \equiv 173 \pmod{420}.$$

♢

Kineski teorem o ostatcima ima brojne primjene. Jedan od razloga jest to da on omogućava da se računanje po jednom velikom modulu zamjeni s nekoliko neovisnih računanja po puno manjim modulima, što je jako dobra osnova za "paralelizaciju" računanja.

U primjenama su često m_i-ovi fiksni, dok a_i-ovi variraju. U takvoj situaciji dobro je onaj dio algoritma koji ne ovisi o a_i-ovima izračunati unaprijed. Sljedeći algoritam koristi tu ideju, a također vodi računa o racionalnom korištenju brojeva n_i koji mogu biti jako veliki.

Garnerov algoritam za CRT:

for $(1 \le i \le k-1$,
  $\mu_i = \prod_{j=1}^{i} m_j$;
  $c_i = \mu_i^{-1} \bmod m_{i+1}$ $)$
$m = \mu_{k-1} m_k$

$x = a_1$
for $(1 \le i \le k-1$,
  $y = ((a_{i+1} - x)c_i) \bmod m_{i+1}$;
  $x = x + y\mu_i$ $)$
$x = x \mod m$

Ovaj algoritam "rješava" module jedan po jedan. Tako da nakon i-tog koraka u petlji, x zadovoljava $x \equiv a_j \pmod{m_j}$ za $j = 1, 2, \ldots, i+1$.

Ukoliko treba rješiti neki jednokratni problem, onda naravno nema koristi od prethodnog računanja s m_i-ovima. U takvoj se situaciji preporuča induktivna uporaba originalnog algoritma za sustav od dvije kongruencije. Naime, ako želimo riješiti sustav

$$x \equiv a_1 \pmod{m_1}, \quad x \equiv a_2 \pmod{m_2},$$

onda jednom primjenom Euklidovog algoritma dobivamo oba željena inverza iz $um_1 + vm_2 = 1$. Tada je $x = um_1a_2 + vm_2a_1 \bmod m_1m_2$ rješenje sustava.

Induktivni algoritam za CRT:

$m = m_1$; $x = a_1$
for $(2 \leq i \leq k$,
    nađi u, v takve da je $um + vm_i = 1$;
    $x = uma_i + vm_i x$;
    $m = mm_i$;
    $x = x \bmod m$ $)$

Definicija 1.8. Reducirani sustav ostataka *modulo* m *je skup cijelih brojeva* r_i *sa svojstvom da je* $\mathrm{nzd}(r_i, m) = 1$, $r_i \not\equiv r_j \pmod{m}$ *za* $i \neq j$, *te da za svaki cijeli broj* x *takav da je* $\mathrm{nzd}(x, m) = 1$ *postoji* r_i *takav da je* $x \equiv r_i \pmod{m}$. *Jedan reducirani sustav ostataka modulo* m *je skup svih brojeva* $a \in \{1, 2, \ldots, m\}$ *takvih da je* $\mathrm{nzd}(a, m) = 1$. *Jasno je da svi reducirani sustavi ostataka modulo* m *imaju isti broj elemenata. Taj broj označavamo s* $\varphi(m)$, *a funkciju* φ *zovemo* Eulerova funkcija. *Drugim riječima,* $\varphi(m)$ *je broj brojeva u nizu* $1, 2, \ldots, m$ *koji su relativno prosti sa* m.

Teorem 1.21. *Neka je* $\{r_1, \ldots, r_{\varphi(m)}\}$ *reducirani sustav ostataka modulo* m, *te neka je* $\mathrm{nzd}(a, m) = 1$. *Tada je* $\{ar_1, \ldots, ar_{\varphi(m)}\}$ *također reducirani sustav ostataka modulo* m.

Dokaz: Direktno iz Teorema 1.3 i 1.18. □

Teorem 1.22 (Eulerov teorem)**.** *Ako je* $\mathrm{nzd}(a, m) = 1$, *onda je* $a^{\varphi(m)} \equiv 1 \pmod{m}$.

Dokaz: Neka je $\{r_1, r_2, \ldots, r_{\varphi(m)}\}$ reducirani sustav ostataka modulo m. Budući da je, po Teoremu 1.21, $\{ar_1, ar_2, \ldots, ar_{\varphi(m)}\}$ također reducirani sustav ostataka modulo m, zaključujemo da je

$$\prod_{j=1}^{\varphi(m)} (ar_j) \equiv \prod_{i=1}^{\varphi(m)} r_i \pmod{m},$$

odnosno,

$$a^{\varphi(m)} \prod_{j=1}^{\varphi(m)} r_j \equiv \prod_{i=1}^{\varphi(m)} r_i \pmod{m}.$$

Kako je $\mathrm{nzd}(r_i, m) = 1$, primjenom Teorema 1.17, dobivamo $a^{\varphi(m)} \equiv 1 \pmod{m}$. □

Teorem 1.23 (Mali Fermatov teorem)**.** *Neka je* p *prost broj. Ako* $p \nmid a$, *onda je* $a^{p-1} \equiv 1 \pmod{p}$. *Za svaki cijeli broj* a *vrijedi* $a^p \equiv a \pmod{p}$.

Dokaz: Očito je $\varphi(p) = p - 1$, pa tvrdnja teorema slijedi iz Teorema 1.22. □

Primjer 1.14. *Odredimo zadnje dvije znamenke u decimalnom zapisu broja* 3^{400}.

Rješenje: Budući da je $\varphi(25) = 20$, imamo $3^{20} \equiv 1 \pmod{25}$, pa je $3^{400} \equiv 1 \pmod{25}$. Također je $3^2 \equiv 1 \pmod{4}$, pa je $3^{400} \equiv 1 \pmod{4}$. Dakle, $3^{400} \equiv 1 \pmod{100}$, pa su zadnje dvije znamenke 01. ♢

Zadatak 1.8. *Odredite zadnje dvije znamenke broja* 2^{1000}.

Definicija 1.9. *Kažemo da je složen broj* n pseudoprost *u bazi* b *(kraće:* n *je psp(*b*)) ako je*

$$b^n \equiv b \pmod{n}.$$

Primjer 1.15. Broj $341 = 11 \cdot 31$ je psp(2), jer je $2^{340} \equiv 32^{68} \equiv (-1)^{68} \equiv 1 \pmod{11}$ i $2^{340} \equiv 32^{68} \equiv 1^{68} \equiv 1 \pmod{31}$, pa je $2^{340} \equiv 1 \pmod{341}$.

Slično, $91 = 7 \cdot 13$ je psp(3), jer je $3^{90} \equiv 27^{30} \equiv (-1)^{30} \equiv 1 \pmod{7}$ i $3^{90} \equiv 27^{30} \equiv 1^{30} \equiv 1 \pmod{13}$, pa je $3^{90} \equiv 1 \pmod{91}$. ♢

Poznato je da za svaki prirodan broj $b \geq 2$ postoji beskonačno mnogo pseudoprostih brojeva u bazi b. Postojanje pseudoprostih brojeva nam pokazuje da testiranje samo s jednom bazom nije dovoljno da bismo zaključili da je broj prost. Zato možemo pokušati kombinirati više baza. Tako je npr. 341 psp(2), a nije psp(3), dok je 91 psp(3), a nije psp(2). No, broj $561 = 3 \cdot 11 \cdot 17$ je pseudoprost u svakoj bazi. Takvi brojevi se nazivaju *Carmichaelovi brojevi*. Ukoliko je poznata faktorizacija od n, onda je lako ustanoviti je li on Carmichaelov broj. Naime, *Korseltov kriterij* kaže da je n Carmichaelov ako i samo ako je n složen, kvadratno slobodan i za svaki prosti faktor p od n vrijedi da $p - 1$ dijeli $n - 1$. Odavde neposredno slijedi da n mora biti produkt od barem tri različita prosta broja. Zaista, kako je n kvadratno slobodan, on mora biti produkt različitih prostih brojeva. Ostaje za vidjeti zašto n ne može biti produkt dvaju prostih brojeva. Pretpostavimo da je $n = pq$, $p < q$. Tada je $n - 1 = pq - 1 \equiv p - 1 \not\equiv 0 \pmod{(q-1)}$, što je u suprotnosti s Korseltovim kriterijem.

Poznato je da postoji beskonačno mnogo Carmichaelovih brojeva. Označimo s $C(x)$ broj Carmichaelovih brojeva koji su $\leq x$. Alford, Granville i Pomerance su 1994. godine dokazali da je $C(x) > x^{2/7}$. Erdős je postavio slutnju da za svaki $\varepsilon > 0$ postoji $x_0(\varepsilon)$ takav da je $C(x) > x^{1-\varepsilon}$ za $x \geq x_0(\varepsilon)$.

Postojanje Carmichaelovih brojeva pokazuje važan nedostatak testiranja prostosti na osnovu Malog Fermatova teorema. Sada ćemo pokazati kako se malim modificiranjem testa taj nedostatak može ukloniti.

Neka je n neparan prirodan broj, $\text{nzd}(b, n) = 1$, te $b^{n-1} \equiv 1 \pmod{n}$. Budući da je $n - 1$ paran, možemo pokušati "vaditi drugi korijen" iz ove kongruencije, tj. računati $b^{(n-1)/2}$, $b^{(n-1)/4}, \ldots$. Pretpostavimo da u i-tom koraku prvi put dobijemo na desnoj strani nešto različito od 1, recimo

$b^{(n-1)/2^i} \equiv a \pmod{n}$. Tada, ako je n prost, onda mora biti $a = -1$ jer je $b^{(n-1)/2^{i-1}} \equiv 1 \pmod{n}$, a jedina rješenja kongruencije $x^2 \equiv 1 \pmod{n}$, ako je n prost, su $x \equiv \pm 1 \pmod{n}$. Dakle, kombinirajući Mali Fermatov teorem sa svojstvom kongruencije $x^2 \equiv 1 \pmod{p}$ dobivamo jači zahtjev od onog iz definicije pseudoprostih brojeva.

Definicija 1.10. *Neka je n neparan složen broj, te neka je $n - 1 = 2^s \cdot t$, gdje je t neparan. Ako za cijeli broj b vrijedi*

$$b^t \equiv 1 \pmod{n} \text{ ili postoji } r < s \text{ takav da je } b^{2^r \cdot t} \equiv -1 \pmod{n}, \quad (1.5)$$

onda kažemo da je n jak pseudoprost broj u bazi b *(ili da je n spsp(b)).*

Ako uvjet (1.5) nije ispunjen za neki b, $0 < b < n$, tada je broj n složen. U tom slučaju broj b nazivamo *svjedok složenosti od n.*

Svaki spsp(b) je ujedno i psp(b). Obrat ne vrijedi. Npr. $n = 341$ je psp(2), ali nije spsp(2). Zaista, $340 = 2^2 \cdot 85$, dok je $2^{85} \equiv 32 \pmod{341}$ i $2^{170} \equiv 1 \pmod{341}$. Kao primjer jakog pseudoprostog broja navedimo npr. da je 91 spsp(10) jer je $10^{45} \equiv -1 \pmod{91}$. Pojam jakog pseudoprostog broja uveo je Selfridge 1974. godine. No, pravu snagu testu zasnovanom na njemu daje sljedeći teorem koji su neovisno dokazali Monier i Rabin 1980. godine, i koji donosimo bez dokaza.

Teorem 1.24. *Neka je n neparan složen broj. Tada je n jak pseudoprost broj u bazi b za najviše $(n-1)/4$ baza b, $0 < b < n$.*

Teorem 1.24 nam pokazuje da u slučaju jakih pseudoprostih brojeva ne postoji analogon Carmichaelovih brojeva. Dakle, nemoguće je da složen broj bude jak pseudoprost broj u svakoj bazi.

Miller-Rabinov test prostosti. Neka je n neparan broj za kojeg želimo ustanoviti je li prost ili složen. Neka je $n - 1 = 2^s t$, gdje je t neparan. Na slučajan način izaberemo b, $0 < b < n$. Izračunamo $b^t \bmod n$. Ako dobijemo ± 1, zaključujemo da je n prošao test, te biramo sljedeći b. U protivnom, uzastopno kvadriramo b^t modulo n sve dok ne dobijemo rezultat -1. Ako dobijemo -1, onda je n prošao test. Ako nikad ne dobijemo -1, tj. ako dobijemo da je $b^{2^{r+1}t} \equiv 1 \pmod{n}$, ali $b^{2^r t} \not\equiv -1 \pmod{n}$, onda sigurno znamo da je n složen. Ako n prođe test za k b-ova, onda je vjerojatnost da je n složen $\leq \frac{1}{4^k}$.

Npr. za $k = 20$ je vjerojatnost da je n složen manja od 10^{-12}. Tako dobiveni "vjerojatno prosti brojevi" se nazivaju još i "industrijski prosti brojevi" (koliko smo za njih spremni "platiti", tj. koliko veliki k uzmemo, toliku "kvalitetu" možemo i očekivati). Poznato je da ne postoji niti jedan broj manji od 10^{12} koji je istovremeno spsp(b) za $b = 2, 3, 5, 7$ i 11. Napomenimo još da se ocjena iz Teorema 1.24 može značajno poboljšati za velike brojeve

n. Tako je vjerojatnost da je 500-bitni broj, koji prođe samo jedan test, složen manja od $1/4^{28}$.

Složenost jednog Miller-Rabinova testa je $O(\ln^3 n)$. Naime, $b^t \bmod n$ se može izračunati u $O(\ln^3 n)$ bitnih operacija, a potom za računanje $b^{2t}, b^{4t}, \ldots, b^{2^{s-1}t}$ uzastopnim kvadriranjem trebamo također $O(\ln^3 n)$ bitnih operacija.

Uz pretpostavku da vrijedi tzv. proširena Riemannova slutnja (ERH), Miller-Rabinov test postaje polinomijalni deterministički algoritam za dokazivanje prostosti. Naime, može se pokazati da ako je n složen broj, onda uz pretpostavku da vrijedi ERH postoji barem jedna baza $b < 2\ln^2 n$ za koju ne vrijedi (1.5). Dakle, uz pretpostavku da vrijedi ERH, složenost ovog algoritma je $O(\ln^5 n)$. To je i bio originalni Millerov test iz 1976. godine.

Definicija 1.11. *Funkciju* $\vartheta : \mathbb{N} \to \mathbb{C}$ *za koju vrijedi*

1) $\vartheta(1) = 1$,

2) $\vartheta(mn) = \vartheta(m)\vartheta(n)$ *za sve* m, n *takve da je* $\mathrm{nzd}(m, n) = 1$,

zovemo multiplikativna funkcija.

Teorem 1.25. *Eulerova funkcija* φ *je multiplikativna. Nadalje, za svaki prirodan broj* $n > 1$ *vrijedi* $\varphi(n) = n\prod_{p|n}(1 - \frac{1}{p})$.

Dokaz: Neka su m, n relativno prosti prirodni brojevi, te neka a i b prolaze skupom svih reduciranih ostataka modulo m, odnosno modulo n. Naš je cilj pokazati da tada $an + bm$ prolazi skupom svih reduciranih ostataka modulo mn. Ako to pokažemo, dobit ćemo da je $\varphi(m)\varphi(n) = \varphi(mn)$.

Budući da je $\mathrm{nzd}(a, m) = 1$ i $\mathrm{nzd}(b, n) = 1$, broj $an + bm$ je relativno prost s m i s n, pa stoga i s mn. Nadalje, svaka dva broja gornjeg oblika su međusobno nekongruentni modulo mn. Zaista, iz $an + bm \equiv a'n + b'm \pmod{mn}$ slijedi $(a - a')n \equiv (b' - b)m \pmod{mn}$. Odavde $m|a - a'$, $n|b' - b$, pa je $a = a'$, $b = b'$. Stoga nam još preostaje pokazati da ako je $\mathrm{nzd}(c, mn) = 1$, onda je $c \equiv an + bm \pmod{mn}$ za neke a, b. Budući je $\mathrm{nzd}(m, n) = 1$, postoje cijeli brojevi x, y takvi da je $mx + ny = 1$. Očito je $\mathrm{nzd}(cy, m) = 1$, $\mathrm{nzd}(cx, n) = 1$, pa brojevi a i b definirani sa $cy \equiv a \pmod{m}$, $cx \equiv b \pmod{n}$ imaju tražena svojstva.

Neka je sada $n = p_1^{\alpha_1} \cdots p_k^{\alpha_k}$. Jedini brojevi u nizu $1, 2, \ldots, p_i^{\alpha_i}$ koji nisu relativno prosti s $p_i^{\alpha_i}$ su brojevi $p_i, 2p_i, \ldots, p_i^{\alpha_i - 1} \cdot p_i$. Stoga je $\varphi(p_i^{\alpha_i}) = p_i^{\alpha_i} - p_i^{\alpha_i - 1}$. Zbog multiplikativnosti od φ, imamo

$$\begin{aligned}\varphi(n) &= \varphi(\prod_{i=1}^{k} p_i^{\alpha_i}) = \prod_{i=1}^{k} \varphi(p_i^{\alpha_i}) = \prod_{i=1}^{k} p_i^{\alpha_i}\left(1 - \frac{1}{p_i}\right) \\ &= n\prod_{i=1}^{k}\left(1 - \frac{1}{p_i}\right) = n\prod_{p|n}\left(1 - \frac{1}{p}\right).\end{aligned}$$

□

Zadatak 1.9. *Za koje prirodne brojeve n je broj $\varphi(n)$ neparan?*

Primjer 1.16. *Odredimo sve prirodne brojeve n za koje vrijedi $\varphi(n) = 12$.*

Rješenje: Ako je $n = p_1^{\alpha_1} \cdots p_r^{\alpha_r}$, onda je

$$\varphi(n) = p_1^{\alpha_1 - 1}(p_1 - 1) \cdots p_r^{\alpha_r - 1}(p_r - 1).$$

Iz $(p_i - 1)|12$ slijedi $p_i \in \{2, 3, 5, 7, 13\}$. Ako je $p_i = 2$, onda je $\alpha_i \leq 3$; ako je $p_i = 3$, onda je $\alpha_i \leq 2$; a ako je $p_i \neq 2, 3$, onda je $\alpha_i = 1$. Imamo četiri mogućnosti (s k označavamo broj oblika $2^\alpha 3^\beta$):

1) $n = 13 \cdot k \implies \varphi(n) = 12 \cdot \varphi(k) = 12$
$\varphi(k) = 1 \implies k = 1$ ili $k = 2 \implies n = 13$ ili $n = 26$;

2) $n = 7 \cdot k \implies \varphi(n) = 6 \cdot \varphi(k) = 12$
$\varphi(k) = 2 \implies k = 3$, $k = 4$ ili $k = 6 \implies n = 21$, $n = 28$ ili $n = 42$;

3) $n = 5 \cdot k \implies \varphi(n) = 4 \cdot \varphi(k) = 12$
$\varphi(k) = 3$, što nema rješenja;

4) $n = 2^\alpha 3^\beta$
Lako se provjeri da nema rješenja za $\alpha = 0$ ni za $\beta = 0$. Ako je $\alpha, \beta > 0$, onda je $\varphi(n) = 2^{\alpha-1}3^{\beta-1} \cdot 2 = 12 = 2^2 \cdot 3$, što povlači da je $\alpha = 2$, $\beta = 2$, tj. $n = 36$.

Rješenja su: $n = 13, 21, 26, 28, 36, 42$. ♢

Često uz multiplikativnu funkciju f vežemo funkciju $g(n) = \sum_{d|n} f(d)$. Pokažimo da je g također multiplikativna. Neka je $\text{nzd}(m, n) = 1$. Tada je

$$\begin{aligned} g(mn) &= \sum_{d|m} \sum_{d'|n} f(dd') = \sum_{d|m} \sum_{d'|n} f(d) f(d') = \left(\sum_{d|m} f(d) \right) \left(\sum_{d'|n} f(d') \right) \\ &= g(m)g(n). \end{aligned}$$

Definicija 1.12. *Neka je n prirodan broj. S $\tau(n)$ ćemo označavati broj pozitivnih djelitelja broja n, a sa $\sigma(n)$ sumu svih pozitivnih djelitelja broja n.*

Jasno je da vrijedi $\tau(n) = \sum_{d|n} 1$, $\sigma(n) = \sum_{d|n} d$. Stoga su funkcije τ i σ multiplikativne. Budući da je $\tau(p^j) = j + 1$, $\sigma(p^j) = 1 + p + p^2 + \cdots p^j = \frac{p^{j+1}-1}{p-1}$, dobivamo sljedeće formule za τ i σ:

$$\begin{aligned} \tau(p_1^{\alpha_1} \cdots p_k^{\alpha_k}) &= \prod_{i=1}^{k} (\alpha_i + 1), \\ \sigma(p_1^{\alpha_1} \cdots p_k^{\alpha_k}) &= \prod_{i=1}^{k} \frac{p_i^{\alpha_i+1} - 1}{p_i - 1}. \end{aligned}$$

Teorem 1.26.

$$\sum_{d|n} \varphi(d) = n$$

Dokaz: Funkcija $g(n) = \sum_{d|n} \varphi(d)$ je multiplikativna, pa je dovoljno provjeriti da je $g(p^\alpha) = p^\alpha$, jer će onda biti

$$g(n) = g(p_1^{\alpha_1} \cdots p_k^{\alpha_k}) = g(p_1^{\alpha_1}) \cdots g(p_k^{\alpha_k}) = p_1^{\alpha_1} \cdots p_k^{\alpha_k} = n.$$

Imamo:

$$\begin{aligned} g(p^\alpha) &= \varphi(1) + \varphi(p) + \varphi(p^2) + \cdots + \varphi(p^\alpha) \\ &= 1 + (p-1) + (p^2 - p) + \cdots + (p^\alpha - p^{\alpha-1}) = p^\alpha. \end{aligned}$$

□

Teorem 1.27 (Wilson)**.** *Ako je p prost broj, onda je* $(p-1)! \equiv -1 \pmod{p}$.

Dokaz: Za $p = 2$ i $p = 3$ kongruencija je očito zadovoljena. Stoga smijemo pretpostaviti da je $p \geq 5$. Grupirajmo članove skupa $\{2, 3, \ldots, p-2\}$ u parove (i, j) sa svojstvom $i \cdot j \equiv 1 \pmod{p}$. Očito je $i \neq j$ jer bi inače broj $(i-1)(i+1)$ bio djeljiv sa p, a to je nemoguće zbog $0 < i - 1 < i + 1 < p$. Tako dobivamo $\frac{p-3}{2}$ parova i ako pomnožimo odgovarajućh $\frac{p-3}{2}$ kongruencija, dobit ćemo

$$2 \cdot 3 \cdots (p-2) \equiv 1 \pmod{p},$$

pa je

$$(p-1)! \equiv 1 \cdot 1 \cdot (p-1) \equiv -1 \pmod{p}.$$

□

Očito je da vrijedi i obrat Wilsonovog teorema. Zaista, neka vrijedi

$$(p-1)! \equiv -1 \pmod{p}$$

i pretpostavimo da p nije prost. Tada p ima djelitelj d, $1 < d < p$, i d dijeli $(p-1)!$. No, tada d mora dijeliti i -1, što je kontradikcija.

Teorem 1.28. *Neka je p prost broj. Tada kongruencija $x^2 \equiv -1 \pmod{p}$ ima rješenja ako i samo ako je $p = 2$ ili $p \equiv 1 \pmod{4}$.*

Dokaz: Ako je $p = 2$, onda je $x = 1$ jedno rješenje.

Ako je $p \equiv 1 \pmod{4}$, onda iz Wilsonovog teorema imamo:

$$[1 \cdot 2 \cdots \frac{p-1}{2}] \cdot [(p-1)(p-2) \cdots (p - \frac{p-1}{2})] \equiv [(\frac{p-1}{2})!]^2 \equiv -1 \pmod{p},$$

pa je $x = (\frac{p-1}{2})!$ jedno rješenje.

Neka je $p \equiv 3 \pmod{4}$. Pretpostavimo da postoji $x \in \mathbb{Z}$ takav da je $x^2 \equiv -1 \pmod{p}$. Tada je $x^{p-1} \equiv (-1)^{\frac{p-1}{2}} \equiv -1 \pmod{p}$, što je u suprotnosti s Malim Fermatovim teoremom. □

Primjer 1.17. *Dokažimo da postoji beskonačno mnogo prostih brojeva oblika* $4k+1$.

Rješenje: Neka su $p_1, p_2, \ldots, p_n$ svi prosti brojevi oblika $4k+1$. Promotrimo broj

$$m = 4p_1^2 p_2^2 \cdots p_n^2 + 1.$$

Neka je p neki prosti faktor od m. Tada kongruencija $x^2 \equiv -1 \pmod{p}$ ima rješenje $x = 2p_1 p_2 \cdots p_n$, pa p mora biti oblika $4k+1$. Očito je $p \neq p_i$, $i = 1, 2, \ldots, n$, pa smo dobili kontradikciju. □

Teorem 1.29 (Lagrange)**.** *Neka je $f(x)$ polinom s cjelobrojnim koeficijentima stupnja n. Pretpostavimo da je p prost broj, te da vodeći koeficijent od f nije djeljiv s p. Tada kongruencija $f(x) \equiv 0 \pmod{p}$ ima najviše n rješenja modulo p.*

Dokaz: Za $n = 1$ tvrdnja teorema vrijedi po Teoremu 1.19. Pretpostavimo da tvrdnja vrijedi za sve polinome stupnja $n-1$, te neka je f polinom stupnja n. Za svaki $a \in \mathbb{Z}$ imamo $f(x) - f(a) = (x-a)g(x)$, gdje je g polinom stupnja $n-1$ s cjelobrojnim koeficijentima i s istim vodećim koeficijentom kao f. Zato ako kongruencija $f(x) \equiv 0 \pmod{p}$ ima rješenje $x = a$, onda sva rješenja ove kongruencije zadovoljavaju $(x-a)g(x) \equiv 0 \pmod{p}$. No, po induktivnoj prepostavci kongruencija $g(x) \equiv 0 \pmod{p}$ ima najviše $n-1$ rješenja, pa kongruencija $f(x) \equiv 0 \pmod{p}$ ima najviše n rješenja. □

Definicija 1.13. *Neka su a i n relativno prosti prirodni brojevi. Najmanji prirodni broj d sa svojstvom da je $a^d \equiv 1 \pmod{n}$ zove se* red *od a modulo n. Još se kaže da a* pripada eksponentu *d modulo n.*

Propozicija 1.30. *Neka je d red od a modulo n. Tada za prirodan broj k vrijedi $a^k \equiv 1 \pmod{n}$ ako i samo ako $d|k$. Posebno, $d|\varphi(n)$.*

Dokaz: Ako $d|k$, recimo $k = d \cdot l$, onda je $a^k \equiv (a^d)^l \equiv 1 \pmod{n}$.

Obratno, neka je $a^k \equiv 1 \pmod{n}$. Podijelimo k sa d, pa dobivamo $k = q \cdot d + r$, gdje je $0 \leq r < d$. Sada je

$$1 \equiv a^k \equiv a^{qd+r} \equiv (a^d)^q \cdot a^r \equiv a^r \pmod{n},$$

pa zbog minimalnosti od d slijedi da je $r = 0$, tj. $d|k$. □

Primjer 1.18. *Dokažimo da svaki prosti djelitelj Fermatovog broja $2^{2^n}+1$, za $n > 1$, ima oblik $p = k \cdot 2^{n+1} + 1$.*

Rješenje: Iz $2^{2^n} + 1 \equiv 0 \pmod{p}$ slijedi $2^{2^n} \equiv -1 \pmod{p}$ i $2^{2^{n+1}} \equiv 1 \pmod{p}$, pa slijedi da 2 pripada eksponentu 2^{n+1} modulo p. Budući da je $\varphi(p) = p-1$, slijedi da $2^{n+1}|p-1$, tj. postoji $k \in \mathbb{N}$ takav da je $p = k \cdot 2^{n+1} + 1$. ♢

Definicija 1.14. *Ako je red od a modulo n jednak $\varphi(n)$, onda se a zove* primitivni korijen *modulo n.*

Ako postoji primitivni korijen modulo n, onda je grupa reduciranih ostataka modulo n ciklička. Slijedeći teorem pokazuje da je grupa $(\mathbb{Z}_p^*, \cdot_p)$ ciklička.

Teorem 1.31. *Ako je p prost broj, onda postoji točno $\varphi(p-1)$ primitivnih korijena modulo p.*

Dokaz: Svaki od brojeva $1, 2, \ldots, p-1$ pripada modulo p nekom eksponentu d, koji je djelitelj od $\varphi(p) = p-1$. Označimo sa $\psi(d)$ broj brojeva u nizu $1, 2, \ldots, p-1$ koji pripadaju eksponentu d. Tada je

$$\sum_{d|p-1} \psi(d) = p-1.$$

Dovoljno je dokazati da ako je $\psi(d) \neq 0$, onda je $\psi(d) = \varphi(d)$. Zaista, po Teoremu 1.26 je

$$\sum_{d|p-1} \varphi(d) = p-1,$$

pa ako bi bilo $\psi(d) = 0 < \varphi(d)$ za neki d, onda bi suma $\sum_{d|p-1} \psi(d)$ bila manja od $p-1$. Stoga je $\psi(d) \neq 0$ za svaki d, pa ako pokažemo da to povlači da je $\psi(d) = \varphi(d)$, onda ćemo dobiti da vrijedi $\psi(p-1) = \varphi(p-1)$, što se i tvrdilo u teoremu.

Dokažimo sada tvrdnju da $\psi(d) \neq 0$ povlači $\psi(d) = \varphi(d)$. Neka je $\psi(d) \neq 0$, te neka je a broj koji pripada eksponentu d modulo p. Promotrimo kongruenciju

$$x^d \equiv 1 \pmod{p}.$$

Ona ima rješenja $a, a^2, \ldots, a^d$ i po Lagrangeovom teoremu to su sva rješenja. Pokažimo da brojevi a^m, za $1 \leq m \leq d$ i $(m,d) = 1$, predstavljaju sve brojeve koji pripadaju eksponentu d modulo p. Zaista, svaki od njih ima red d, jer ako je $a^{md'} \equiv 1 \pmod{p}$, onda $d | md'$, pa $d | d'$. Ako je b bilo koji broj koji pripada eksponentu d modulo p, onda je $b \equiv a^m$ za neki m, $1 \leq m \leq d$. Budući da je

$$b^{\frac{d}{(m,d)}} \equiv (a^d)^{\frac{m}{(m,d)}} \equiv 1 \pmod{p},$$

to je $(m,d) = 1$. Dakle, dobili smo da je $\psi(d) = \varphi(d)$. □

Teorem 1.32. *Neka je p neparan prost broj, te neka je g primitivni korijen modulo p. Tada postoji $x \in \mathbb{Z}$ takav da je $g' = g + px$ primitivni korijen modulo p^j za sve $j \in \mathbb{N}$.*

Dokaz: Imamo $g^{p-1} = 1 + py$, za neki $y \in \mathbb{Z}$. Po binomnom teoremu je

$$g'^{p-1} = 1 + py + (p-1)pxg^{p-2} + \binom{p-1}{2} p^2 x^2 g^{p-3} + \cdots + p^{p-1} x^{p-1},$$

tj. $g'^{p-1} = 1 + pz$, gdje je $z \equiv y + (p-1)g^{p-2}x \pmod{p}$. Koeficijent uz x nije djeljiv sa p, pa možemo odabrati x tako da bude $(z,p) = 1$. Tvrdimo da tada g' ima traženo svojstvo. Dokažimo to.

Pretpostavimo da g' pripada eksponentu d modulo p^j. Tada d dijeli $\varphi(p^j) = p^{j-1}(p-1)$. No, g' je primitivni korijen modulo p, pa $p-1$ dijeli d. Dakle, $d = p^k(p-1)$ za neki $k < j$. Nadalje, imamo

$$(1+pz)^p = 1 + p^2 z_1, \quad (1+pz)^{p^2} = (1+p^2 z_1)^p = 1 + p^3 z_2, \quad \ldots,$$

$$(1+pz)^{p^k} = 1 + p^{k+1} z_k,$$

gdje je $(z_i, p) = 1$ za $i = 1, \ldots, k$. Budući da je $g'^d \equiv 1 \pmod{p^j}$, odavde zaključujemo da je $j = k+1$, što povlači da je $d = \varphi(p^j)$. □

Teorem 1.33. *Za prirodan broj n postoji primitivni korijen modulo n ako i samo ako je $n = 2$, 4, p^j ili $2p^j$, gdje je p neparan prost broj.*

Dokaz: Jasno je da je 1 primitivni korijen modulo 2, te da je 3 primitivni korijen modulo 4. Neka je g primitivni korijen modulo p^j. Odaberimo među brojevima g i $g + p^j$ onaj koji je neparan. Tada je on primitivni korijen modulo $2p^j$ jer je $\varphi(2p^j) = \varphi(p^j)$.

Ostaje još dokazati nužnost. Neka je najprije $n = 2^j$ za $j \geq 3$. Tada za neparan broj a vrijedi $a^2 \equiv 1 \pmod{8}$. Budući da $8 | a^2 - 1$ i $2 | a^2 + 1$ imamo $a^4 \equiv 1 \pmod{16}$. Ponavljajući ovaj argument dobivamo: $a^{2^{j-2}} \equiv 1 \pmod{2^j}$ za $j \geq 3$. Budući da je $\varphi(2^j) = 2^{j-1}$, dokazali smo da ne postoji primitivni korijen modulo 2^j za $j \geq 3$.

Konačno, neka je $n = n_1 n_2$, gdje je $(n_1, n_2) = 1$, $n_1 > 2$, $n_2 > 2$. Brojevi $\varphi(n_1)$ i $\varphi(n_2)$ su parni, pa imamo

$$a^{\frac{1}{2}\varphi(n)} \equiv \left(a^{\varphi(n_1)}\right)^{\frac{1}{2}\varphi(n_2)} \equiv 1 \pmod{n_1},$$

$$a^{\frac{1}{2}\varphi(n)} \equiv \left(a^{\varphi(n_2)}\right)^{\frac{1}{2}\varphi(n_1)} \equiv 1 \pmod{n_2}.$$

Stoga je $a^{\frac{1}{2}\varphi(n)} \equiv 1 \pmod{n}$, što znači da ne postoji primitivni korijen modulo n. □

Rcimo nešto o tome kako se nalazi primitivni korijen. Možemo krenuti redom i testirati je li $g = 2$, $g = 3$, ... primitivni korijen. Pritom ne treba testirati brojeve oblika g_0^k, $k \geq 2$, jer ako g_0 nije primitivni korijen, onda to ne može biti ni g_0^k. Testiranje je li g primitivni korijen se zasniva na sljedećoj očitoj činjenici: g je primitivni korijen ako i samo ako za svaki prosti faktor q od $p-1$ vrijedi $g^{(p-1)/q} \not\equiv 1 \pmod{p}$.

Može se postaviti pitanje kolika je vjerojatnost da već $g = 2$ bude primitivan korijen. S tim u vezi spomenimo poznatu Artinovu slutnju koja kaže da za prirodan broj a, koji nije potencija nekog prirodnog broja, vrijedi $\nu_a(N) \sim A \cdot \pi(N)$, gdje je $\pi(N)$ broj prostih brojeva $\leq N$, $\nu_a(N)$ broj

prostih brojeva $\leq N$ za koje je a primitivni korijen, dok je A Artinova konstanta

$$\prod_{p \text{ prost}} \left(1 - \frac{1}{p(p-1)}\right) \approx 0.3739558.$$

Poznato je da tzv. generalizirana Riemannova slutnja (GRH) povlači Artinovu slutnju, a također povlači i ocjenu $O(\ln^6 p)$ za najmanji primitivni korijen modulo p.

Primjer 1.19. *Nađimo najmanji primitivni korijen*
a) modulo 5, *b) modulo* 11, *c) modulo* 23.

Rješenje: a) $2^2 \not\equiv 1 \pmod 5 \implies$ 2 je primitivni korijen modulo 5.
b) $2^2 \not\equiv 1 \pmod{11}$, $2^5 \not\equiv 1 \pmod{11} \implies$ 2 je primitivni korijen modulo 11.
c) $2^{11} = 32 \cdot 64 \equiv 9 \cdot (-5) \equiv 1 \pmod{23}$, $3^{11} = 27^3 \cdot 9 \equiv 64 \cdot 9 \equiv (-5) \cdot 9 \equiv 1 \pmod{23}$, $4^{11} = (2^{11})^2 \equiv 1 \pmod{23}$, $5^{11} = (25)^5 \cdot 5 \equiv 32 \cdot 5 \equiv 9 \cdot 5 \equiv -1 \not\equiv 1 \pmod{23} \implies$ 5 je primitivni korijen modulo 23. ♢

Zadatak 1.10. *Nađite najmanji primitivni korijen*
a) modulo 13, *b) modulo* 17, *c) modulo* 41.

Definicija 1.15. *Neka je g primitivni korijen modulo n. Lako se vidi da tada brojevi g^l, $l = 0, 1, \ldots, \varphi(n) - 1$ tvore reducirani sustav ostataka modulo n. Stoga za svaki cijeli broj a takav da je* $\operatorname{nzd}(a, n) = 1$ *postoji jedinstveni l takav da je $g^l \equiv a \pmod n$. Eksponent l se zove* indeks *od a u odnosu na g i označava se sa* $\operatorname{ind}_g a$ *ili* $\operatorname{ind} a$.

Teorem 1.34.
1) $\operatorname{ind} a + \operatorname{ind} b \equiv \operatorname{ind}(ab) \pmod{\varphi(n)}$
2) $\operatorname{ind} 1 = 0$, $\operatorname{ind} g = 1$
3) $\operatorname{ind}(a^m) \equiv m \operatorname{ind} a \pmod{\varphi(n)}$ *za* $m \in \mathbb{N}$
4) $\operatorname{ind}(-1) = \frac{1}{2}\varphi(n)$ *za* $n \geq 3$

Dokaz: Svojstva 1) – 3) slijede direktno iz definicije, a svojstvo 4) slijedi iz $g^{2 \operatorname{ind}(-1)} \equiv (-1)^2 \equiv 1 \pmod n$ i $2 \operatorname{ind}(-1) < 2\varphi(n)$. □

Uočimo da su svojstva indeksa 1) – 3) potpuno analogna svojstvima logaritamske funkcije.

Propozicija 1.35. *Ako je $(n, p-1) = 1$, onda kongruencija $x^n \equiv a \pmod p$ ima jedinstveno rješenje.*

Dokaz: Iz $x^n \equiv a \pmod p$, po Teoremu 1.34, dobivamo

$$n \operatorname{ind} x \equiv \operatorname{ind} a \pmod{p-1},$$

pa jer je $(n, p-1) = 1$, ova kongruencija ima jedinstveno rješenje. □

Primjer 1.20. *Riješimo kongruenciju* $x^5 \equiv 2 \pmod 7$.

Rješenje: Imamo: $3^2 \equiv 2 \pmod 7$, $3^3 \equiv 6 \pmod 7$, $3^6 \equiv 1 \pmod 7$. Stoga je 3 primitivni korijen modulo 7 i $\mathrm{ind}_3\, 2 = 2$. Dakle, dobivamo kongruenciju

$$5\,\mathrm{ind}_3\, x \equiv 2 \pmod 6,$$

čije je rješenje $\mathrm{ind}_3\, x = 4$, pa je $x \equiv 3^4 \equiv 4 \pmod 7$. ♢

Primjer 1.21. *Riješimo kongruenciju* $5x^4 \equiv 3 \pmod{11}$.

Rješenje: Iz Primjera 1.19 znamo da je 2 primitivni korijen modulo 11. Nadalje je $2^4 \equiv 5 \pmod{11}$, $2^8 \equiv 3 \pmod{11}$, pa dobivamo

$$\mathrm{ind}_2\, 5 + 4\mathrm{ind}_2\, x \equiv \mathrm{ind}_2\, 3 \pmod{10}, \quad 4\mathrm{ind}_2\, x \equiv 8 - 4 \equiv 4 \pmod{10}.$$

Prema tome, trebamo riješiti kongruenciju $2\mathrm{ind}_2\, x \equiv 2 \pmod 5$. Odavde je $\mathrm{ind}_2\, x \equiv 1$ ili $6 \pmod{10}$, pa su rješenje $x \equiv 2 \pmod{11}$ i $x \equiv 2^6 \equiv 9 \pmod{11}$. ♢

Zadatak 1.11. *Riješite kongruencije*
a) $2x^8 \equiv 5 \pmod{13}$, *b)* $x^6 \equiv 5 \pmod{17}$, *c)* $x^{12} \equiv 37 \pmod{41}$.

Primjer 1.22. *Riješimo kongruenciju* $3^x \equiv 2 \pmod{23}$.

Rješenje: Iz Primjera 1.19 znamo da je 5 primitivni korijen modulo 23. Nadalje je $5^2 \equiv 2 \pmod{23}$, $5^5 \equiv 2^2 \cdot 5 \equiv -3 \pmod{23}$, $5^{11} \equiv -1 \pmod{23}$, što povlači da je $5^{16} \equiv 3 \pmod{23}$. Imamo:

$$x\mathrm{ind}_5\, 3 \equiv \mathrm{ind}_5\, 2 \pmod{22}, \quad 16x \equiv 2 \pmod{22}.$$

Sada je $(16, 22) = 2$, po dobivamo $8x \equiv 1 \pmod{11}$, odakle je $x \equiv 7 \pmod{11}$. Dakle, rješenja su $x \equiv 7, 18 \pmod{22}$. ♢

Zadatak 1.12. *Riješite kongruenciju* $7^x \equiv 6 \pmod{17}$.

1.3 Kvadratni ostatci

Definicija 1.16. *Neka je* $\text{nzd}(a, m) = 1$. *Ako kongruencija* $x^2 \equiv a \pmod{m}$ *ima rješenja, onda kažemo da je* a kvadratni ostatak *modulo* m. *U protivnom kažemo da je* a kvadratni neostatak *modulo* m.

Primjer 1.23. *Kvadratni ostatci modulo* 5 *su* 1 *i* 4, *a neostatci su* 2 *i* 3.

Teorem 1.36. *Neka je* p *neparan prost broj. Reducirani sustav ostataka modulo* p *sastoji se od* $\frac{p-1}{2}$ *kvadratnih ostataka i* $\frac{p-1}{2}$ *kvadratnih neostataka.*

Dokaz: Svaki kvadratni ostatak modulo p kongruentan je kvadratu nekog od brojeva

$$-\frac{p-1}{2}, \ldots, -1, 1, \ldots, \frac{p-1}{2},$$

tj. kongruentan je nekom od brojeva $1^2, 2^2, \ldots, (\frac{p-1}{2})^2$. Preostaje pokazati da je ovih $\frac{p-1}{2}$ brojeva međusobno nekongruentno modulo p. Pa pretpostavimo da je $k^2 \equiv l^2 \pmod{p}$, gdje je $1 \leq k < l \leq \frac{p-1}{2}$. Tada je $(l-k)(l+k) \equiv 0 \pmod{p}$, pa je $l - k \equiv 0 \pmod{p}$ ili $l + k \equiv 0 \pmod{p}$, što je u suprotnosti s pretpostavkama na k i l, jer je $0 < l - k < p$ i $0 < l + k < p$. □

Tvrdnja Teorema 1.36 također slijedi i iz činjenice da je grupa da postoji primitivni korijen g modulo p. Sada je jasno da su $g^0, g^2, g^4, \ldots, g^{p-3}$ kvadratni ostatci, a $g^1, g^3, g^5, \ldots, g^{p-2}$ kvadratni neostatci.

Zadatak 1.13. *Odredite sve kvadratne ostatke modulo* 7 *i modulo* 17.

Definicija 1.17. *Neka je* p *neparan prost broj. Po definiciji,* Legendreov simbol $(\frac{a}{p})$ *je jednak* 1 *ako je* a *kvadratni ostatak modulo* p, -1 *ako je* a *kvadratni neostatak modulo* p, *a* 0 *ako* $p|a$.

Dakle, broj rješenja kongruencije $x^2 \equiv a \pmod{p}$ je jednak $1 + (\frac{a}{p})$.

Teorem 1.37 (Eulerov kriterij)**.**

$$\left(\frac{a}{p}\right) \equiv a^{\frac{p-1}{2}} \pmod{p}$$

Dokaz: Ako je $(\frac{a}{p}) = 0$, onda $p|a$, pa je tvrdnja očito zadovoljena.

Ako je $(\frac{a}{p}) = 1$, onda postoji $x_0 \in \mathbb{Z}$ takav da je $x_0^2 \equiv a \pmod{p}$. Sada je iz Malog Fermatovog teorema $a^{\frac{p-1}{2}} \equiv x_0^{p-1} \equiv 1 \equiv (\frac{a}{p}) \pmod{p}$.

Neka je $(\frac{a}{p}) = -1$. Za svaki $i \in \{1, \ldots, p-1\}$ odaberimo $j \in \{1, \ldots, p-1\}$ tako da vrijedi $i \cdot j \equiv a \pmod{p}$ (to je moguće po Teoremu 1.18). Uočimo da je $i \neq j$, budući da kongruencija $x^2 \equiv a \pmod{p}$ nema rješenja. Dakle, skup $\{1, \ldots, p-1\}$ se raspada na $\frac{p-1}{2}$ parova (i, j) za koje vrijedi $i \cdot j \equiv a \pmod{p}$. Množenjem ovih $\frac{p-1}{2}$ kongruencija, te koristeći Wilsonov teorem, dobivamo

$$a^{\frac{p-1}{2}} \equiv (p-1)! \equiv -1 \pmod{p}.$$

□

Propozicija 1.38.
1) Ako je $a \equiv b \pmod{p}$*, onda je* $(\frac{a}{p}) = (\frac{b}{p})$.
2) $(\frac{a}{p})(\frac{b}{p}) = (\frac{ab}{p})$
3) Ako je $(a,p) = 1$*, onda je* $(\frac{a^2}{p}) = 1$.
4) $(\frac{1}{p}) = 1, \quad (\frac{-1}{p}) = (-1)^{\frac{p-1}{2}}$.

Dokaz: 1) Ako je $a \equiv b \pmod{p}$, onda kongruencija $x^2 \equiv a \pmod{p}$ ima rješenja ako i samo ako rješenja ima kongruencija $x^2 \equiv b \pmod{p}$.
2) Iz

$$\left(\frac{a}{p}\right)\left(\frac{b}{p}\right) \equiv a^{\frac{p-1}{2}} b^{\frac{p-1}{2}} \equiv (ab)^{\frac{p-1}{2}} \equiv \left(\frac{ab}{p}\right) \pmod{p}$$

slijedi $(\frac{a}{p})(\frac{b}{p}) = (\frac{ab}{p})$.
3) Kongruencija $x^2 \equiv a^2 \pmod{p}$ očito ima rješenje $x = a$.
4) Prva tvrdnja je specijalni slučaj od 3), dok druga slijedi uvrštavanjem $a = -1$, u Eulerov kriterij. □

Postavlja se pitanje kako izračunati Legendreov simbol. Jedna mogućnost je pomoću Eulerova kriterija. Koristeći efikasne metode za modularno potenciranje, Eulerov kriterij nam omogućava da Legendreov simbol izračunamo uz $O(\ln^3 p)$ bitnih operacija. No, postoji i efikasniji algoritam, čija je složenost $O(\ln^2 p)$, a koji je vrlo sličan Euklidovu algoritmu. Taj algoritam je zasnovan na *Gaussovu kvadratnom zakonu reciprociteta*, koji glasi

$$\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) = (-1)^{(p-1)(q-1)/4}$$

za različite proste brojeve p i q. Dakle, ovaj zakon nam omogućava da $(\frac{p}{q})$ zamijenimo s $(\frac{q}{p})$, što je posebno korisno ukoliko je $p < q$. Međutim, za primjenu ovog zakona oba parametra moraju biti prosti brojevi, što dolazi od zahtjeva u definiciji Legendreova simbola da jedan od parametara (donji) bude prost. To nas vodi do potrebe uvođenja poopćenja Legendreova simbola kod kojeg parametri neće morati biti prosti.

Definicija 1.18. *Neka je* m *neparan prirodan broj i* $m = \prod_{i=1}^{k} p_i^{\alpha_i}$ *njegov rastav na proste faktore, te neka je* a *proizvoljan cijeli broj.* Jacobijev simbol $(\frac{a}{m})$ *se definira sa*

$$\left(\frac{a}{m}\right) = \prod_{i=1}^{k} \left(\frac{a}{p_i}\right)^{\alpha_i},$$

gdje $(\frac{a}{p_i})$ *predstavlja Legendreov simbol.*

Jasno je da ako je m prost, onda se Jacobijev i Legendreov simbol podudaraju. Ako je $\mathrm{nzd}(a,m) > 1$, onda je $(\frac{a}{m}) = 0$. Ako je a kvadratni ostatak modulo m, onda je a kvadratni ostatak modulo p_i za svaki i. Zato je $(\frac{a}{p_i}) = 1$

za svaki i, pa je i $(\frac{a}{m}) = 1$. Međutim, $(\frac{a}{m}) = 1$ ne povlači da je a kvadratni ostatak modulo m. Da bi a bio kvadratni ostatak modulo m nužno je i dovoljno da svi $(\frac{a}{p_i})$ budu jedanki 1. Spomenimo da postoji i općenitiji pojam, tzv. Kroneckerov simbol $(\frac{a}{b})$, koji se definira za proizvoljne cijele brojeve a i b (b može biti paran i negativan).

Navodimo osnovna svojstva Jacobijevog simbola koja se koriste u njegovom računanju:

1) $a \equiv b \pmod{m} \quad \Longrightarrow \quad \left(\frac{a}{m}\right) = \left(\frac{b}{m}\right)$.

2) $\left(\frac{ab}{m}\right) = \left(\frac{a}{m}\right)\left(\frac{b}{m}\right), \quad \left(\frac{a}{mn}\right) = \left(\frac{a}{m}\right)\left(\frac{a}{n}\right)$

3) $\left(\frac{-1}{m}\right) = (-1)^{(m-1)/2} = \begin{cases} 1, & \text{ako je } m \equiv 1 \pmod{4} \\ -1, & \text{ako je } m \equiv 3 \pmod{4} \end{cases}$

4) $\left(\frac{2}{m}\right) = (-1)^{(m^2-1)/8} = \begin{cases} 1, & \text{ako je } m \equiv 1,7 \pmod{8} \\ -1, & \text{ako je } m \equiv 3,5 \pmod{8} \end{cases}$

5) $\left(\frac{m}{n}\right)\left(\frac{n}{m}\right) = (-1)^{(m-1)(n-1)/4}$ ako su m i n relativno prosti

Algoritam za računanje Jacobijevog simbola $(\frac{a}{m})$

$a = a \bmod m$
$t = 1$
while $(a \neq 0)$ {
    while (a paran) {
        $a = a/2$;
        if $(m \equiv 3,5 \pmod{8})$ then $t = -t$ }
    $(a, m) = (m, a)$
    if $(a \equiv m \equiv 3 \pmod{4})$ then $t = -t$
    $a = a \bmod m$ }
if $(m = 1)$ then return t
    else return 0

Vidimo da je ovaj algoritam vrlo sličan Euklidovom algoritmu. Jedina bitna razlika je u posebnom tretiranju faktora 2, kojeg moramo izlučiti prije nego što zamjenimo gornji i donji parametar.

Primjer 1.24. *Izračunajmo* $\left(\frac{105}{317}\right)$.

Rješenje: Imamo: $\left(\frac{105}{317}\right) = \left(\frac{317}{105}\right) = \left(\frac{2}{105}\right) = 1$. ♢

Primjer 1.25. *Izračunajmo* $\left(\frac{-23}{83}\right)$.

Rješenje: Imamo:

$$\left(\frac{-23}{83}\right) = -\left(\frac{23}{83}\right) = \left(\frac{83}{23}\right) = \left(\frac{14}{23}\right) = \left(\frac{2}{23}\right)\left(\frac{7}{23}\right) = \left(\frac{7}{23}\right) = -\left(\frac{23}{7}\right)$$
$$= -\left(\frac{2}{7}\right) = -1.$$

♢

Zadatak 1.14. *Izračunati:* $\left(\frac{51}{71}\right)$, $\left(\frac{7}{227}\right)$.

Primjer 1.26. *a) Odredimo sve proste brojeve p takve da je -2 kvadratni ostatak modulo p.*

b) Dokažimo da postoji beskonačno mnogo prostih brojeva oblika $8k+3$.

Rješenje: a) Trebamo naći sve proste brojeva za koje vrijedi $\left(\frac{-2}{p}\right) = \left(\frac{-1}{p}\right)\left(\frac{2}{p}\right) = 1$. Imamo dvije mogućnosti:

1) $\left(\frac{-1}{p}\right) = 1 \quad \& \quad \left(\frac{2}{p}\right) = 1$. Prvi uvjet je ekvivalentan s $p \equiv 1 \pmod 4$, a drugi s $p \equiv 1, 7 \pmod 8$, što zajedno daje $p \equiv 1 \pmod 8$.

2) $\left(\frac{-1}{p}\right) = -1 \quad \& \quad \left(\frac{2}{p}\right) = -1$. Prvi uvjet je ekvivalentan s $p \equiv 3 \pmod 4$, a drugi s $p \equiv 3, 5 \pmod 8$, što zajedno daje $p \equiv 3 \pmod 8$.

Dakle, $p \equiv 1$ ili $3 \pmod 8$.

b) Neka su $p_1, p_2, \ldots, p_n$ svi prosti brojevi oblika $8k+3$. Promotrimo broj

$$m = p_1^2 p_2^2 \cdots p_n^2 + 2.$$

Prema a), svi prosti faktori od m su oblika $8k+1$ ili $8k+3$. No, $m \equiv 3 \pmod 8$, pa ne mogu svi faktori biti oblika $8k+1$. Dakle, postoji prosti faktor p oblika $8k+3$. Kako je očito $p \neq p_i$, $i = 1, 2, \ldots, n$, dobili smo kontradikciju. ♢

Zadatak 1.15. *Odredite sve proste brojeve p takve da je $\left(\frac{-3}{p}\right) = 1$. Dokažite da postoji beskonačno mnogo prostih brojeva oblika $6k+1$.*

Primjer 1.27. *a) Neka je $p \equiv 3 \pmod 4$ prost broj takav da je $q = 2p+1$ također prost. Dokažimo da je tada $2^p \equiv 1 \pmod q$.*

b) Pokažimo da Mersennov broj $M_{251} = 2^{251} - 1$ nije prost.

Rješenje: a) Kako je $\varphi(q) = q - 1 = 2p$, imamo da je $2^{2p} - 1 = (2^p - 1)(2^p + 1) \equiv 0 \pmod q$. Dakle, $2^p \equiv 1 \pmod q$ ili $2^p \equiv -1 \pmod q$. Po pretpostavci je $p = 4k+3$, $q = 8k+7$. Ako bi bilo $2^p \equiv 1 \pmod q$, to bi značilo da je $2^{4k+3} \equiv -1 \pmod q$, odnosno

$$x^2 \equiv -2 \pmod q,$$

za $x = 2^{2k+2}$, a to je nemoguće prema Primjeru 1.26.

b) Brojevi 251 i $2 \cdot 251 + 1 = 503$ su prosti i $251 \equiv 3 \pmod 4$, pa iz a) slijedi da $503 | M_{251}$, što znači da M_{251} nije prost. ♢

Eulerov kriterij može poslužiti kao osnova za test prostosti. Neparan složen broj n je *Eulerov pseudoprost broj u bazi* b (n je epsp(b)) ako zadovoljava Eulerov kriterij:

$$\left(\frac{b}{n}\right) \equiv b^{(n-1)/2} \pmod n$$

(ovdje $(\frac{b}{n})$ označava Jacobijev simbol). Može se pokazati da je svaki spsp(b) ujedno i epsp(b), pa je stoga tzv. Solovay-Strassenov test, koji je zasnovan na Eulerovim pseudoprostim brojevima, manje efikasan od Miller-Rabinova testa.

Pretpostavimo sada da je $(\frac{a}{p}) = 1$. To znači da postoji cijeli broj x takav da je

$$x^2 \equiv a \pmod p\,. \tag{1.6}$$

Postavlja se pitanje kako naći taj broj x, tj. kako efikasno izračunati kvadratni korijen od a modulo p. Ako je p vrlo mali, to možemo napraviti tako da ispitamo redom sve moguće ostatke modulo p. No, za imalo veće p-ove, to je vrlo neefikasan algoritam.

Odgovor na postavljeno pitanje je vrlo lak za brojeve specijalnog oblika. Zapravo, mogli bi reći da taj oblik i nije jako specijalan, budući da pola prostih brojeva ima takav oblik.

Propozicija 1.39. *Ako je* $p \equiv 3 \pmod 4$, *onda je* $x = a^{(p+1)/4}$ *rješenje kongruencije (1.6).*

Dokaz: Budući da je a kvadratni ostatak modulo p, iz Eulerovog kriterija imamo $a^{(p-1)/2} \equiv 1 \pmod p$, pa je

$$x^2 \equiv (a^{(p+1)/2}) \equiv a \cdot a^{(p-1)/2} \equiv a \pmod p.$$

□

Prethodnu propoziciju je moguće modificirati i na preostale proste brojeve, uz poznavanje barem jednog kvadratnog neostatka modulo p. Ako je $p \equiv 5 \pmod 8$, onda je broj 2 kvadratni neostatak modulo p. Upravo ta činjenica se koristi u sljedećoj propoziciji.

Propozicija 1.40. *Ako je* $p \equiv 5 \pmod 8$, *onda je jedan brojeva* $a^{(p+3)/8}$ *i* $2^{(p-1)/4} a^{(p+3)/8}$ *rješenje kongruencije (1.6).*

Dokaz: Ako je $p = 8k+5$, onda je $a^{4k+2} \equiv 1 \pmod p$. Odavde je $a^{2k+1} \equiv \pm 1 \pmod p$, pa je $a^{2k+2} \equiv \pm a \pmod p$. Ako u posljednjoj kongruenciji imamo predznak $+$, onda je $x = a^{k+1} = a^{(p+3)/8}$ rješenje kongruencije (1.6).

Ukoliko imamo predznak $-$, onda iskoristimo činjenicu da je $(\frac{2}{p}) = -1$. To povlači da je $2^{4k+2} \equiv -1 \pmod{p}$, pa za $x = 2^{(p-1)/4} a^{(p+3)/8}$ vrijedi

$$x^2 \equiv 2^{4k+2} a^{2k+2} \equiv (-1)(-a) \equiv a \pmod{p}.$$

□

Preostao je slučaj $p \equiv 1 \pmod{8}$. Taj slučaj je i najteži, zato što ne možemo eksplicitno napisati jedan kvadratni neostatak modulo p (iako znamo da ih ima "puno", tj. $(p-1)/2$). Opisat ćemo Tonellijev algoritam za nalaženje kvadratnog korijena u tom slučaju. Pretpostavimo da nam je poznat jedan kvadratni neostatak d modulo p. Ovo je teoretski najproblematičniji dio algoritma. Naime, nije poznat niti jedan (bezuvjetni) deterministički polinomijalni algoritam na nalaženje kvadratnog neostatka. Pretpostavimo li da vrijedi tzv. proširena Riemannova slutnja (ERH), onda postoji kvadratni neostatak manji od $2\ln^2 p$, pa nam jednostavno pretraživanje daje polinomijalni algoritam. U praksi ovo nije problem, jer je vjerojatnost da je slučajno izabrani broj kvadratni neostatak jednaka $1/2$. Tako je vjerojatnost da od 20 slučajno izabranih brojeva niti jedan nije kvadratni neostatak manja od 10^{-6}.

Neka je $p = 2^s t + 1$, gdje je t neparan. Prema Eulerovom kriteriju imamo:

$$a^{2^{s-1}t} \equiv 1 \pmod{p}, \quad a^{2^{s-2}t} \equiv \pm 1 \pmod{p}, \quad d^{2^{s-1}t} \equiv -1 \pmod{p}.$$

Dakle, postoji $t_2 \geq 0$ takav da je

$$a^{2^{s-2}t} d^{t_2 2^{s-1}} \equiv 1 \pmod{p}, \quad a^{2^{s-3}t} d^{t_2 2^{s-2}} \equiv \pm 1 \pmod{p}.$$

Analogno zaključujemo da postoji $t_3 \geq 0$ takav da je

$$a^{2^{s-3}t} d^{t_3 2^{s-2}} \equiv 1 \pmod{p}, \quad a^{2^{s-4}t} d^{t_3 2^{s-3}} \equiv \pm 1 \pmod{p}.$$

Nastavljajući ovaj postupak, na kraju dobijemo $t_s \geq 0$ takav da je

$$a^t d^{2t_s} \equiv 1 \pmod{p},$$

pa je $x = a^{(t+1)/2} d^{t_s}$ rješenje kongruencije (1.6).

Spojivši gornje dvije propozicije i Tonellijev algoritam, dobivamo sljedeći algoritam za računanje rješenja kongruencije $x^2 \equiv a \pmod{p}$.

Kvadratni korijen modulo p

$a = a \bmod p$
if $(p \equiv 3, 7 \pmod 8)$ then $\{$
    $x = a^{(p+1)/4} \bmod p$;
    return x $\}$

if $(p \equiv 5 \pmod 8)$ then $\{$
    $x = a^{(p+3)/8} \bmod p$;
    $c = x^2 \bmod p$;
    if $(c \neq a \bmod p)$ then $x = x \cdot 2^{(p-1)/4} \bmod p$;
    return x $\}$

Nađi broj $d \in \{2, 3, \cdots, p-1\}$ takav da je $\left(\frac{d}{p}\right) = -1$
Prikaži $p - 1 = 2^s t$, t neparan
$A = a^t \bmod p$
$D = d^t \bmod p$
$m = 0$
for $(0 \leq i \leq s-1)$ $\{$
    if $((AD^m)^{2^{s-1-i}} \equiv -1 \pmod p)$ then $m = m + 2^i$ $\}$
$x = a^{(t+1)/2} D^{m/2} \bmod p$
return x

1.4 Diofantske jednadžbe

Teorem 1.41. *Neka su a, b, c cijeli brojevi i $d = \text{nzd}(a, b)$. Ako $d \nmid c$, onda jednadžba*

$$ax + by = c \tag{1.7}$$

nema cjelobrojnih rješenja. Ako $d|c$, onda jednadžba (1.7) ima beskonačno mnogo cjelobrojnih rješenja. Ako je (x_1, y_1) jedno rješenje, onda su sva rješenja dana sa $x = x_1 + \frac{b}{d} \cdot t$, $y = y_1 - \frac{a}{d} \cdot t$, gdje je $t \in \mathbb{Z}$.

Dokaz: Ako (1.7) ima rješenja, onda očito $d|c$. Pretpostavimo sada da $d|c$ i promotrimo kongruenciju

$$ax \equiv c \pmod{b}. \tag{1.8}$$

Po Teoremu 1.19 ova kongruencija ima rješenja i ako je x_1 neko rješenje, onda su sva rješenja od (1.8) dana sa $x \equiv x_1 + \frac{b}{d} \cdot k \pmod{b}$, gdje je $k = 0, 1, \ldots, d-1$. Stoga su sva rješenja od (1.7) dana sa $x = x_1 + \frac{b}{d} \cdot t$, $t \in \mathbb{Z}$. Uvrstimo li ovo u (1.7), dobivamo $by = c - ax_1 - \frac{ab}{d} \cdot t = by_1 - \frac{ab}{d} \cdot t$, pa je $y = y_1 - \frac{a}{d} \cdot t$. □

Definicija 1.19. *Uređenu trojku prirodnih brojeva (x, y, z) zovemo* Pitagorina trojka *ako su x, y katete, a z hipotenuza nekog pravokutnog trokuta, tj. ako vrijedi*

$$x^2 + y^2 = z^2. \tag{1.9}$$

Ako su x, y, z relativno prosti, onda kažemo da je (x, y, z) primitivna Pitagorina trojka. *(Takav trokut zovemo (primitivni) Pitagorin trokut.)*

Uočimo najprije da je u svakoj primitivnoj Pitagorinoj trojki točno jedan od brojeva x, y neparan. Zaista, ako bi x i y bili parni, onda trojka ne bi bila primitivna, a ako bi x i y bili neparni, onda bi iz $x^2 + y^2 \equiv 2 \pmod{4}$ i $z^2 \equiv 0 \pmod{4}$ dobili kontradikciju.

Teorem 1.42. *Sve primitivne Pitagorine trojke (x, y, z) u kojima je y paran, dane su formulama*

$$x = m^2 - n^2, \quad y = 2mn, \quad z = m^2 + n^2, \tag{1.10}$$

gdje je $m > n$ i m, n su relativno prosti prirodni brojevi različite parnosti.

Dokaz: Jednadžbu (1.9) možemo pisati u obliku $y^2 = (z + x)(z - x)$. Neka je $y = 2c$. Brojevi $z + x$ i $z - x$ su parni, pa postoje prirodni brojevi a i b takvi da je $z + x = 2a$, $z - x = 2b$. Sada je

$$c^2 = ab.$$

Iz $z = a + b$, $x = a - b$, zaključujemo da je $(a, b) = 1$, pa postoje $m, n \in \mathbb{N}$, $(m, n) = 1$, takvi da je $a = m^2$, $b = n^2$. Odavde je

$$x = m^2 - n^2, \quad z = m^2 + n^2, \quad y = 2mn.$$

Brojevi m i n moraju biti različite parnosti jer je broj $x = m^2 - n^2$ neparan.

Lako se provjeri da brojevi x, y, z definirani sa (1.10) zadovoljavaju (1.9). Zaista,

$$(m^2 - n^2)^2 + (2mn)^2 = m^4 + 2m^2n^2 + n^4 = (m^2 + n^2)^2.$$

Treba još provjeriti da su relativno prosti. Pretpostavimo da je $(x, z) = d > 1$. Tada je d neparan, $d|(m^2+n^2)+(m^2-n^2) = 2m^2$ i $d|(m^2+n^2)-(m^2-n^2) = 2n^2$. No, ovo je u kontradikciji s pretpostavkom da su m i n, pa stoga i m^2 i n^2, relativno prosti. □

Iz Teorema 1.42 slijedi da su sve Pitagorine trojke dane identitetom:

$$[d(m^2 - n^2)]^2 + (2dmn)^2 = [d(m^2 + n^2)]^2. \tag{1.11}$$

Kod određivanja Pitagorinih trokuta sa zadanom stranicom, od koristi nam mogu biti sljedeće činjenice:

- Neparan prirodan broj k se može prikazati kao zbroj kvadrata dva relativno prosta broja $k = m^2 + n^2$, $\text{nzd}(m, n) = 1$ ako i samo ako svi prosti faktori p broja k zadovoljavaju $p \equiv 1 \pmod 4$ (jedan smjer slijedi iz $(\frac{-1}{p}) = -1$ ako je $p \equiv 3 \pmod 4$). Broj različitih prikaza je 2^{r-1}, gdje je r broj različitih prostih faktora od k. Posebno, za prost broj p takav da je $p \equiv 1 \pmod 4$ prikaz je jedinstven.
- Prirodan broj k se može prikazati kao razlika dva kvadrata $k = m^2 - n^2$ ako i samo ako $k \not\equiv 2 \pmod 4$ (jedan smjer slijedi iz $m^2 \equiv 0$ ili $1 \pmod 4$).

Primjer 1.28. *Nađimo sve Pitagorine trokute u kojima je jedna stranica jednaka a)* 39, *b)* 2003.

Rješenje: a) Sve Pitagorine trojke su dane identitetom (1.11). U ovom slučaju imamo tri mogućnosti: $d = 1$, $d = 3$, $d = 13$ (lako se vidi da ne može biti $d = 39$, jer ne postoji Pitagorin trokut sa stranicom 1).

Ako je $d = 1$, onda je $m^2 + n^2 \neq 39$, pa mora biti $m^2 - n^2 = (m-n)(m+n) = 39$. Odavde je $m - n = 1$, $m + n = 39$ ili $m - n = 3$, $m + n = 13$, što povlači da je $m = 20$, $n = 19$ ili $m = 8$, $n = 5$. Tako dobivamo Pitagorine trojke $(39, 760, 761)$ i $(39, 80, 89)$.

Ako je $d = 3$, onda je $m^2 - n^2 = 13$ ili $m^2 + n^2 = 13$, što povlači da je $m = 7$, $n = 6$ ili $m = 3$, $n = 2$. Dobivene trojke su $(39, 252, 255)$ i $(15, 36, 39)$.

Ako je $d = 13$, onda je $m^2 - n^2 = 3$. Odavde je $m = 2$, $n = 1$, što daje trojku $(39, 52, 65)$.

b) Broj 2003 je prost. Stoga je $d = 1$. Tada je $m^2 + n^2 \neq 2003$ jer je $2003 \equiv 3 \pmod 4$, dok iz $m^2 - n^2 = 2003$ slijedi $m = 1002$, $n = 1001$, te je jedina trojka $(2003, 2006004, 2006005)$. ♢

Zadatak 1.16. *Nađite sve Pitagorine trokute kojima je jedna stranica jednaka a)* 34, *b)* 2001.

Teorem 1.43. *Jednadžba $x^4 + y^4 = z^2$ nema rješenja u prirodnim brojevima. Drugim riječima, ne postoji pravokutni trokut kojem su duljine kateta kvadrati prirodnih brojeva.*

Dokaz: Pretpostavimo da takav trokut postoji i izaberimo među svim takvim trokutima onaj s najmanjoj hipotenuzom. Tako dobivamo Pitagorinu trojku (x^2, y^2, z). Pokažimo da su x i y relativno prosti. U protivnom bi bilo $x = a \cdot d$, $y = b \cdot d$, $d > 1$. Tada bi iz $z^2 = d^4(a^4 + b^4)$ slijedilo da postoji $c \in \mathbb{N}$ takav da je $z = d^2 \cdot c$, te bi dobili Pitagorinu trojku (a^2, b^2, c) s hipotenuzom manjom od z, što je kontradikcija.

Dakle, (x^2, y^2, z) je primitivna Pitagorina trojka, pa po Teoremu 1.42 (ako odaberemo da je y paran) postoje relativno prosti prirodni brojevi različite parnosti m i n tako da vrijedi

$$x^2 = m^2 - n^2, \quad y^2 = 2mn, \quad z = m^2 + n^2.$$

Iz $x^2 + n^2 = m^2$ slijedi da je n paran, a m neparan. Stavimo: $n = 2k$, $y = 2t$, pa dobivamo

$$t^2 = mk.$$

Odavde slijedi da postoje prirodni brojevi r i s takvi da je $m = r^2$ i $k = s^2$. Budući da je (x, n, m) primitivna Pitagorina trojka, po Teoremu 1.42 postoje u, v takvi da je $(u, v) = 1$, $n = 2uv$, $m = u^2 + v^2$. Sada iz $n = 2s^2$ slijedi da je $s^2 = uv$, pa postoje $a, b \in \mathbb{N}$ takvi da je $u = a^2$, $v = b^2$. Prema tome, $a^4 + b^4 = r^2$, pa je (a^2, b^2, r) Pitagorina trojka za čiju hipotenuzu vrijedi: $r < r^2 = m < m^2 + n^2 = z$, što je u suprotnosti s minimalnošću od z. □

Napomena 1.2. *Iz Teorema 1.43 slijedi da jednadžba $x^4 + y^4 = z^4$ nema rješenja u prirodnim brojevima. Ovo je specijalni slučaj tzv. Velikog Fermatovog teorema koji kaže da jednadžba $x^n + y^n = z^n$ nema rješenja u prirodnim brojevima za $n \geq 3$. Ovaj teorem je dokazao 1995. godine Andrew Wiles.*

Diofantska jednadžba oblika

$$x^2 - dy^2 = 1, \tag{1.12}$$

gdje je d prirodan broj koji nije potpun kvadrat, naziva se *Pellova jednadžba*. Slučaj kad je d potpun kvadrat isključujemo jer je tada očito da jednadžba (1.12) ima samo trivijalna rješenja $x = \pm 1$, $y = 0$. Zaista, ako je $d = \delta^2$,

onda iz $(x-\delta y)(x+\delta y)=1$ slijedi $x-\delta y=x+\delta y=\pm 1$. Jednadžba je dobila ime po engleskom matematičaru Johnu Pellu, kojem je Euler, po svemu sudeći pogrešno, pripisao zasluge za njezino rješavanje.

Kao prvi korak u proučavanju Pellove jednadžbe, dokazat ćemo da ona ima beskonačno mnogo rješenja u prirodnim brojevima. Koristit ćemo Dirichletov teorem iz diofantskih aproksimacija koji kaže da za svaki iracionalan broj α postoji beskonačno mnogo racionalnih brojeva $\frac{p}{q}$ sa svojstvom

$$\left|\alpha-\frac{p}{q}\right|<\frac{1}{q^2}. \tag{1.13}$$

Ideja dokaza je za dani prirodan broj Q promotriti $Q+1$ brojeva

$$0,\ 1,\ \{\alpha\}=\alpha-\lfloor\alpha\rfloor,\ \{2\alpha\},\ \ldots,\ \{(Q-1)\alpha\}$$

iz segmenta $[0,1]$ i podjelu tog segmenta na Q disjunktnih podintervala širine $1/Q$, te po Dirichletovom principu zaključiti da postoji podinterval koji sadrži barem dva, od promatranih $Q+1$ brojeva.

Lema 1.44. *Neka je d prirodan broj koji nije potpun kvadrat. Tada postoji cijeli broj k, $|k|<1+2\sqrt{d}$, sa svojstvom da jednadžba*

$$x^2-dy^2=k \tag{1.14}$$

ima beskonačno mnogo rješenja u prirodnim brojevima.

Dokaz: Po Dirichletovom teoremu, postoji beskonačno mnogo parova prirodnih brojeva (x,y) sa svojstvom

$$\left|\sqrt{d}-\frac{x}{y}\right|<\frac{1}{y^2},\quad \text{tj.}\quad \left|x-y\sqrt{d}\right|<\frac{1}{y}.$$

Za svaki takav par (x,y) vrijedi

$$|x+y\sqrt{d}|=|x-y\sqrt{d}+2y\sqrt{d}|<\frac{1}{y}+2y\sqrt{d}\le(1+2\sqrt{d})y,$$

pa je

$$|x^2-dy^2|=|x-y\sqrt{d}|\cdot|x+y\sqrt{d}|<1+2\sqrt{d}.$$

Budući da parova (x,y) s navedenim svojstvom ima beskonačno, a cijelih brojeva koji su po modulu manji od $1+2\sqrt{d}$ samo konačno, to postoji neki cijeli broj k, takav da je $|k|<1+2\sqrt{d}$, za kojeg jednadžba (1.14) ima beskonačno mnogo rješenja. □

Teorem 1.45. *Pellova jednadžba $x^2-dy^2=1$ ima barem jedno rješenje u prirodnim brojevima x i y.*

Dokaz: Beskonačno mnogo rješenja jednadžbe (1.14) možemo podijeliti u k^2 klasa, stavljajući rješenja (x_1, y_1) i (x_2, y_2) u istu klasu ako i samo ako je $x_1 \equiv x_2 \pmod{k}$ i $y_1 \equiv y_2 \pmod{k}$. Tada neka od tih klasa sadrži barem dva (u stvari beskonačno) različita rješenja (x_1, y_1), (x_2, y_2) (x_1, x_2 su različiti prirodni brojevi). Stavimo

$$x = \frac{x_1x_2 - dy_1y_2}{k}, \quad y = \frac{x_1y_2 - x_2y_1}{k}$$

("podijelimo rješenja" $x_2 + y_2\sqrt{d}$ i $x_1 + y_1\sqrt{d}$). Tvrdimo da je $x, y \in \mathbb{Z}$, $y \neq 0$ i $x^2 - dy^2 = 1$. Imamo: $x_1x_2 - dy_1y_2 \equiv x_1^2 - dy_1^2 \equiv k \equiv 0 \pmod{k}$, $x_1y_2 - x_2y_1 \equiv x_1y_1 - x_1y_1 \equiv 0 \pmod{k}$, pa su $x, y \in \mathbb{Z}$. Pretpostavimo da je $y = 0$, tj. $x_1y_2 = x_2y_1$. Tada je

$$k = x_2^2 - dy_2^2 = x_2^2 - d \cdot \frac{x_2^2y_1^2}{x_1^2} = \frac{x_2^2}{x_1^2}(x_1^2 - dy_1^2) = \frac{x_2^2}{x_1^2} \cdot k,$$

tj. $x_1^2 = x_2^2$, što je u suprotnosti s pretpostavkom da su x_1 i x_2 različiti prirodni brojevi. Konačno,

$$\begin{aligned} x^2 - dy^2 &= \frac{1}{k^2}\left[(x_1x_2 - dy_1y_2)^2 - d(x_1y_2 - x_2y_1)^2\right] \\ &= \frac{1}{k^2}(x_1^2x_2^2 + d^2y_1^2y_2^2 - dx_1^2y_2^2 - dx_2^2y_1^2) \\ &= \frac{1}{k^2}(x_1^2 - dy_1^2)(x_2^2 - dy_2^2) = \frac{1}{k^2} \cdot k \cdot k = 1. \end{aligned}$$

□

Za najmanje rješenja (x, y) u prirodnim brojevima Pellove jednadžbe (1.12) kažemo da je njeno *fundamentalno rješenje*. Označavamo ga sa (x_1, y_1), a često također i sa $x_1 + y_1\sqrt{d}$.

Teorem 1.46. *Pellova jednadžba $x^2 - dy^2 = 1$ ima beskonačno mnogo rješenja. Ako je (x_1, y_1) fundamentalno rješenje, onda su sva rješenja (u prirodnim brojevima) ove jednadžbe dana formulom*

$$x_n + y_n\sqrt{d} = (x_1 + y_1\sqrt{d})^n, \quad n \in \mathbb{N}, \tag{1.15}$$

tj.

$$\begin{aligned} x_n &= x_1^n + \binom{n}{2}dx_1^{n-2}y_1^2 + \binom{n}{4}d^2x_1^{n-4}y_1^4 + \cdots, \\ y_n &= nx_1^{n-1}y_1 + \binom{n}{3}dx_1^{n-3}y_1^3 + \binom{n}{5}d^2x_1^{n-5}y_1^5 + \cdots. \end{aligned}$$

Dokaz: Iz (1.15) slijedi $x_n - y_n\sqrt{d} = (x_1 - y_1\sqrt{d})^n$, pa množenjem dobivamo

$$x_n^2 - dy_n^2 = (x_1 - dy_1^2)^n = 1,$$

što znači da su (x_n, y_n) zaista rješenja (i ima ih beskonačno mnogo).

Pretpostavimo sada da je (s,t) rješenje koje nije oblika (x_n, y_n), $n \in \mathbb{N}$. Budući da je $x_1 + y_1\sqrt{d} > 1$ i $s + t\sqrt{d} > 1$, to postoji $m \in \mathbb{N}$ takav da je

$$(x_1 + y_1\sqrt{d})^m < s + t\sqrt{d} < (x_1 + y_1\sqrt{d})^{m+1}. \tag{1.16}$$

Pomnožimo li (1.16) sa $(x_1 + y_1\sqrt{d})^{-m} = (x_1 - y_1\sqrt{d})^m$, dobivamo

$$1 < (s + t\sqrt{d})(x_1 - y_1\sqrt{d})^m < x_1 + y_1\sqrt{d}.$$

Definirajmo $a, b \in \mathbb{Z}$ s $a + b\sqrt{d} = (s + t\sqrt{d})(x_1 - y_1\sqrt{d})^m$. Imamo: $a^2 - db^2 = (s^2 - dt^2)(x_1^2 - dy_1^2)^m = 1$. Iz $a + b\sqrt{d} > 1$ slijedi $0 < a - b\sqrt{d} < 1$, pa je $a > 0$ i $b > 0$. Stoga je (a, b) rješenje u prirodnim brojevima jednadžbe $x^2 - dy^2 = 1$ i $a + b\sqrt{d} < x_1 + y_1\sqrt{d}$, što je kontradikcija. □

Teorem 1.47. *Neka je (x_n, y_n), $n \in \mathbb{N}$ niz svih rješenja Pellove jednadžbe $x^2 - dy^2 = 1$ u prirodnim brojevima, zapisan u rastućem redosljedu. Uzmimo da je $(x_0, y_0) = (1, 0)$ (to je "trivijalno rješenje"). Tada vrijedi:*

$$x_{n+2} = 2x_1x_{n+1} - x_n, \quad y_{n+2} = 2x_1y_{n+1} - y_n, \quad n \geq 0.$$

Dokaz: Po Teoremu 1.46 je $x_n + y_n\sqrt{d} = (x_1 + y_1\sqrt{d})^n$. Odavde je

$$\begin{aligned}(x_{n+1} + y_{n+1}\sqrt{d})(x_1 + y_1\sqrt{d}) &= x_{n+2} + y_{n+2}\sqrt{d},\\ (x_{n+1} + y_{n+1}\sqrt{d})(x_1 - y_1\sqrt{d}) &= x_n + y_n\sqrt{d}.\end{aligned}$$

Sada imamo:

$$\begin{aligned}x_{n+2} &= x_1x_{n+1} + dy_1y_{n+1},\\ x_n &= x_1x_{n+1} - dy_1y_{n+1},\end{aligned}$$

odakle zbrajanjem dobivamo $x_{n+2} = 2x_1x_{n+1} - x_n$. Analogno je

$$\begin{aligned}y_{n+2} &= x_1y_{n+1} + y_1x_{n+1},\\ y_n &= x_1y_{n+1} - y_1x_{n+1},\end{aligned}$$

pa ponovo zbrajanjem dobivamo $y_{n+2} = 2x_1y_{n+1} - y_n$. □

Teoremi 1.46 i 1.47 nam pokazuju kako možemo generirati sva rješenja Pellove jednadžbe $x^2 - dy^2 = 1$ ukoliko znamo njeno fundamentalno (najmanje) rješenje. No, ostaje pitanje kako naći to fundamentalno rješenje. Ponekad rješenje možemo naći uvrštavajući redom $y = 1, 2, 3, \ldots$ i provjeravajući je li $dy^2 + 1$ kvadrat. Međutim, već i za relativno male d-ove fundamentalno rješenje može biti vrlo veliko, npr. za $d = 94$, fundamentalno rješenje je $2143295 + 221064\sqrt{94}$. Stoga je potrebno naći efikasniji način za njegovo nalaženje.

Jedan relativno efikasan algoritam za nalaženje fundamentalog rješenja dobit ćemo iz veze Pellovih jednadžbi s diofantskim aproksimacijama, te preko njih s verižnim razlomcima. Naime, svako netrivijano rješenje jednadžbe $x^2 - dy^2 = 1$ inducira jako dobru racionalnu aproksimaciju iracionalnog broja $\sqrt{d}$. Zaista,

$$\left|\sqrt{d} - \frac{x}{y}\right| = \frac{1}{y|x + y\sqrt{d}|} < \frac{1}{2\sqrt{d}y^2}. \tag{1.17}$$

Poznato je da se sve jako dobre racionalne aproksimacije realnog broja mogu dobiti iz njegovog razvoja u verižni razlomak.

Neka je $\alpha \in \mathbb{R}$. Izraz oblika

$$\alpha = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \ddots}},$$

gdje je $a_0 \in \mathbb{Z}$, te $a_1, a_2, \ldots \in \mathbb{N}$, zove se razvoj broja α u *jednostavni verižni (ili neprekidni) razlomak*. Verižni razlomak kraće zapisujemo kao $[a_0; a_1, a_2, \ldots]$. Brojevi $a_0, a_1, a_2, \ldots$ se zovu *parcijalni kvocijenti*, a definiraju se na sljedeći način:

$$a_0 = \lfloor \alpha \rfloor, \quad \alpha = a_0 + \frac{1}{\alpha_1}, \quad a_1 = \lfloor \alpha_1 \rfloor, \quad \alpha_1 = a_1 + \frac{1}{\alpha_2}, \quad a_2 = \lfloor \alpha_2 \rfloor, \ldots.$$

Postupak se nastavlja sve dok je $a_k \neq \alpha_k$. Razvoj u jednostavni verižni razlomak broja α je konačan ako i samo ako je α racionalan broj. Ako je $\alpha = \frac{m}{n}$, brojevi $a_0, a_1, a_2, \ldots$ su upravo kvocijenti iz Euklidovog algoritma primjenjenog na brojeve m i n.

Primjer 1.29. *Razvijmo broj $\frac{41}{47}$ u jednostavni verižni razlomak.*

Rješenje:

$$\begin{aligned} 47 &= 41 \cdot 1 + 6 \\ 41 &= 6 \cdot 6 + 5 \\ 6 &= 5 \cdot 1 + 1 \\ 5 &= 1 \cdot 5 \end{aligned}$$

Odavde je $\frac{47}{41} = [1; 6, 1, 5]$, pa je $\frac{41}{47} = [0; 1, 6, 1, 5]$. ♢

Racionalne brojeve

$$\frac{p_k}{q_k} = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \ddots + \cfrac{1}{a_k}}} = [a_0; a_1, \ldots, a_k]$$

zovemo *konvergente verižnog razlomka*. Brojnici i nazivnici konvergenti zadovoljavaju sljedeće rekurzije:

$$p_{n+2} = a_{n+2}p_{n+1} + p_n, \quad p_0 = a_0, \quad p_1 = a_0a_1 + 1, \quad (p_{-1} = 1,\ p_{-2} = 0),$$
$$q_{n+2} = a_{n+2}q_{n+1} + q_n, \quad q_0 = 1, \quad q_1 = a_1, \quad (q_{-1} = 0,\ q_{-2} = 1).$$

Indukcijom se lako dokazuje sljedeća važna relacija koja povezuje konvergente sa susjednim indeksima:

$$q_n p_{n-1} - p_n q_{n-1} = (-1)^n. \tag{1.18}$$

Relacija (1.18) povlači da je $\frac{p_{2k}}{q_{2k}} \leq \alpha$ i $\alpha \leq \frac{p_{2k+1}}{q_{2k+1}}$ za svaki k. Nadalje, može se dokazati da ako je α pozitivan, onda vrijede sljedeće nejednakosti:

$$\frac{p_0}{q_0} < \frac{p_2}{q_2} < \frac{p_4}{q_4} < \cdots \leq \alpha \leq \cdots < \frac{p_5}{q_5} < \frac{p_3}{q_3} < \frac{p_1}{q_1}.$$

Ako je α iracionalan, onda je $\lim_{n\to\infty} \frac{p_n}{q_n} = \alpha$.

Postavlja se pitanje koliko dobro konvergente aproksimiraju α. Odgovor je dan u sljedećim nejednakostima:

$$\frac{1}{q_n(q_n + q_{n+1})} < \left|\alpha - \frac{p_n}{q_n}\right| < \frac{1}{q_n^2}\,. \tag{1.19}$$

Vrijedi i svojevrsni obrat ove činjenice (Legendreov teorem): ako je $\frac{p}{q}$ racionalan broj koji zadovoljava nejednakost $\left|\alpha - \frac{p}{q}\right| < \frac{1}{2q^2}$, onda je $\frac{p}{q}$ konvergenta od α.

Iz nejednakosti (1.17) i Legendreovog teorema zaključujemo da za svako rješenje Pellove jednadžbe $x^2 - dy^2 = 1$ vrijedi da je $\frac{x}{y}$ neka konvergenta u razvoju od $\sqrt{d}$. Broj $\sqrt{d}$ je kvadratna iracionalnost, pa mu je razvoj periodičan. Štoviše, broj $\sqrt{d} + \lfloor\sqrt{d}\rfloor$ je reduciran (veći je od 1, a konjugat $-\sqrt{d} + \lfloor\sqrt{d}\rfloor$ mu je iz $\langle -1, 0\rangle$), pa mu je razvoj čisto periodičan. Odavde slijedi da $\sqrt{d}$ ima razvoj oblika:

$$\sqrt{d} = [a_0; \overline{a_1, a_2, \ldots, a_{\ell-1}, 2a_0}],$$

gdje je $a_0 = \lfloor\sqrt{d}\rfloor$. Nadalje, može se pokazati da vrijedi "palindromno svojstvo" $a_1 = a_{\ell-1}$, $a_2 = a_{\ell-2}$, $\ldots$.

Sada ćemo navesti algoritam za razvoj kvadratnih iracionalnosti u verižni razlomak. Neka je α kvadratna iracionalnost. Prikažemo je u obliku $\alpha = \frac{s_0+\sqrt{d}}{t_0}$, gdje su $d, s_0, t_0 \in \mathbb{Z}$, $t_0 \neq 0$, $d \neq \square$ i $t_0|(d - s_0^2)$. Ako je $\alpha = \sqrt{d}$, onda je jednostavno $s_0 = 0$, $t_0 = 1$. Sada brojeve a_i (tzv. parcijalne kvocijente) računamo rekurzivno na slijedeći način:

$$a_i = \left\lfloor \frac{s_i + a_0}{t_i} \right\rfloor, \quad s_{i+1} = a_i t_i - s_i, \quad t_{i+1} = \frac{d - s_{i+1}^2}{t_i}. \tag{1.20}$$

Uočimo da, iako je α iracionalan broj, ovaj algoritam radi samo s cijelim brojevima. Pokazuje se da su nizovi (s_i) i (t_i) ograničeni. Preciznije, dobije se da za dovoljno velike indekse i vrijedi

$$0 < s_i < \sqrt{d}, \quad 0 < t_i < s_i + \sqrt{d} < 2\sqrt{d}.$$

Na taj način se upravo i zaključuje da razvoj mora biti periodičan (jer moraju postojati različiti indeksi j, k takvi da je $(s_j, t_j) = (s_k, t_k)$). Odavde direktno dobivamo ocjenu za duljinu perioda u razvoju od $\sqrt{d}$: $\ell(d) < \sqrt{d} \cdot 2\sqrt{d} = 2d$. Preciznijom analizom odnosa između s_i i t_i (posebno kongruencije $s_i^2 \equiv d \pmod{t_i}$), dobije se ocjena $\ell(d) = O(\sqrt{d} \ln d)$, a slutnja je (povezana s čuvenom Riemannovom slutnjom) da vrijedi $\ell(d) = O(\sqrt{d} \ln \ln d)$.

Izjednačavanjem racionalnih i iracionalnih dijelova u jednakosti

$$\sqrt{d} = \frac{\frac{s_{n+1}+\sqrt{d}}{t_{n+1}} p_n + p_{n-1}}{\frac{s_{n+1}+\sqrt{d}}{t_{n+1}} q_n + q_{n-1}},$$

dobiva se relacija

$$p_n^2 - dq_n^2 = (-1)^{n+1} t_{n+1}, \quad \text{za sve } n \geq -1. \tag{1.21}$$

Ona nam pokazuje da rješenja Pellove jednadžbe $x^2 - dy^2 = 1$ odgovaraju onim n-ovima za koje je $(-1)^{n+1} t_{n+1} = 1$. Nije teško za vidjeti da je $t_i = 1$ ako i samo ako $\ell | i$ (ℓ je duljina perioda). Zato vrijedi

Teorem 1.48. *Neka je ℓ duljina perioda u razvoju od $\sqrt{d}$.*

Ako je ℓ paran, onda jednadžba $x^2 - dy^2 = -1$ nema rješenja, a sva rješenja od $x^2 - dy^2 = 1$ su dana sa $(x, y) = (p_{n\ell-1}, q_{n\ell-1})$, $n \in \mathbb{N}$. Posebno, fundamentalno rješenje je $(p_{\ell-1}, q_{\ell-1})$.

Ako je ℓ neparan, onda su sva rješenja jednadžbe $x^2 - dy^2 = -1$ dana sa $(x, y) = (p_{(2n-1)\ell-1}, q_{(2n-1)\ell-1})$, a sva rješenja jednadžbe $x^2 - dy^2 = 1$ sa $(x, y) = (p_{2n\ell-1}, q_{2n\ell-1})$, $n \in \mathbb{N}$. Posebno, fundamentalno rješenje od $x^2 - dy^2 = 1$ je $(p_{2\ell-1}, q_{2\ell-1})$.

Primjer 1.30. *Razvijmo broj $\sqrt{15}$ u jednostavni verižni razlomak.*

Rješenje: Imamo:

$$s_0 = 0,\ t_0 = 1,\ a_0 = 3,$$

$$s_1 = a_0 t_0 - s_0 = 3,\ t_1 = \frac{15 - s_1^2}{t_0} = 6,\ a_1 = \left\lfloor \frac{s_1 + \lfloor \sqrt{d} \rfloor}{t_1} \right\rfloor = \left\lfloor \frac{3 + \lfloor \sqrt{15} \rfloor}{6} \right\rfloor = 1,$$

$$s_2 = 3,\ t_2 = 1,\ a_2 = \left\lfloor \frac{3 + \lfloor \sqrt{15} \rfloor}{1} \right\rfloor = 6,$$

$$s_3 = 3,\ t_3 = 6.$$

Dakle, $(s_1, t_1) = (s_3, t_3)$, pa je $\sqrt{15} = [3; \overline{1, 6}]$. $\diamondsuit$

Primjer 1.31. *Nađimo sva rješenja jednadžbi* $x^2-15y^2=-1$ *i* $x^2-15y^2=1$*, za koja vrijedi* $1<x<1000$*.*

Rješenje: Prema Primjeru 1.30 je $\sqrt{15}=[3,\overline{1,6}]$. Dakle, period $r=2$ je paran, pa jednadžba $x^2-15y^2=-1$ nema rješenja. Najmanje rješenje jednadžbe $x^2-15y^2=1$ je očito $(x_1,y_1)=(4,1)$ (također ga možemo dobiti kao (p_1,q_1)). Dalje imamo: $x_2=8\cdot 4-1=31$, $y_2=8$; $x_3=8\cdot 31-4=244$, $y_3=63$, dok je već $x_4>1000$. ♢

Primjer 1.32. *Nađimo najmanja rješenja jednadžbi* $x^2-29y^2=-1$ *i* $x^2-29y^2=1$ *u prirodnim brojevima (ako postoje).*

Rješenje: Razvijanjem u verižni razlomak dobiva se

$$\sqrt{29}=[5;\overline{2,1,1,2,10}].$$

To znači da je period $r=5$ neparan, pa je najmanje rješenje od $x^2-29y^2=-1$ dano sa (p_4,q_4), a najmanje rješenje od $x^2-29y^2=1$ sa (p_9,q_9).

n	-1	0	1	2	3	4	5	6	7	8	9
a_n		5	2	1	1	2	10	2	1	1	2
p_n	1	5	11	16	27	70	727	1524	2251	3775	9801
q_n	0	1	2	3	5	13	135	283	418	701	1820

Dakle, $(p_4,q_4)=(70,13)$, $(p_9,q_9)=(9801,1820)$.
Uočimo da je $(70+13\sqrt{29})^2=9801+1820\sqrt{29}$. ♢

Zadatak 1.17. *Nađite najmanja rješenja u prirodnim brojevima jednadžbi*

$$x^2-13y^2=\pm 1,\quad x^2-14y^2=\pm 1,\quad x^2-31y^2=\pm 1.$$

Poglavlje 2

Algebarske strukture

2.1 Polugrupe i grupe

Definicija 2.1. *Neka je X neprazan skup.* Binarna operacija *na X je proizvoljna funkcija $\alpha : X \times X \to X$. Dakle, binarna operacija pridružuje uređenom paru (x, y) element $\alpha(x, y)$. Često se umjesto $\alpha(x, y)$ koriste oznake $x \circ y$, $x + y$, $x \cdot y$ ili xy.*

Binarna operacija je asocijativna *ako vrijedi*

$$x \circ (y \circ z) = (x \circ y) \circ z, \quad \forall x, y, z \in X.$$

Skup X s asocijativnom binarnom operacijom zove se polugrupa.

Homomorfizam *polugrupa $(X, \circ)$ i $(Y, \cdot)$ je preslikavanje $f : X \to Y$ za koje vrijedi*

$$f(x \circ y) = f(x) \cdot f(y), \quad \forall x, y \in X.$$

Primjer 2.1.

1. *Neka je S proizvoljan neprazan skup. Skup svih preslikavanja $S \to S$, uz kompoziciju kao binarnu operaciju je polugrupa.*
2. *$(\mathbb{R}, +)$, $(\mathbb{R}, \cdot)$, $(\mathbb{R}^+, \cdot)$ su polugrupe.*
3. *$(M_n, +)$, $(M_n, \cdot)$, gdje je M_n skup svih kvadratnih matrica (s realnim koeficijentima) reda n, su polugrupe.*
4. *Preslikavanje $\det : (M_n, \cdot) \to (\mathbb{R}, \cdot)$ je homomorfizam polugrupa.*
5. *Preslikavanje $\log : (\mathbb{R}^+, \cdot) \to (\mathbb{R}, +)$ je homomorfizam polugrupa.*

Definicija 2.2. Neka je $(X, \cdot)$ polugrupa. *Neutralni element* ili *jedinica* je element $e \in X$ takav da je $x \cdot e = e \cdot x = x$ za svaki $x \in X$.

Polugrupa u kojoj postoji neutralni element zove se *monoid*.

Primjer 2.2. *Primjeri monoida:*

1. $(\mathbb{N}, +)$ *nije monoid, ali* $(\mathbb{N} \cup \{0\}, +)$ *jest;*

2. $(\mathbb{N}, \cdot)$ *je monoid;*

3. $(M_n, +)$, $(M_n, \cdot)$ *su monoidi.*

Propozicija 2.1. *Ako u polugrupi* $(X, \cdot)$ *postoji neutralni element, onda je on jedinstven.*

Dokaz: Neka su $e, e' \in X$ neutralni elementi. Tada je $e' \cdot e = e$ i $e' \cdot e = e'$, pa je $e = e'$. □

Definicija 2.3. Monoid X je *grupa* ako za svaki $x \in X$ postoji $x^{-1} \in X$ takav da je $x \cdot x^{-1} = x^{-1} \cdot x = e$.

Propozicija 2.2. *Neka je* $(X, \cdot)$ *monoid, te* $x \in X$. *Ako je* x *invertibilan, onda je njegov inverz jedinstven.*

Dokaz: Neka su x' i x'' inverzi od x. Tada je $x \cdot x' = x' \cdot x = e$ i $x \cdot x'' = x'' \cdot x = e$. Nadalje imamo:

$$(x'' \cdot x) \cdot x' = e \cdot x' = x', \quad x'' \cdot (x \cdot x') = x'' \cdot e = x'',$$

pa je zbog asocijativnosti $x' = x''$. □

Teorem 2.3. *Neka je* $(X, \cdot)$ *grupa. Tada za sve* $a, b \in X$ *jednadžbe* $ax = b$ *i* $ya = b$ *imaju jedinstveno rješenje.*

Dokaz: Neka je $ax = b$. Tada je $a^{-1}ax = a^{-1}b$, pa je $x = a^{-1}b$. Dakle, ako rješenje postoji, onda je to $a^{-1}b$. Provjerimo da je to zaista rješenje:

$$a(a^{-1}b) = (aa^{-1})b = eb = b.$$

Tvrdnja za drugu jednadžbu dokazuje se sasvim analogno. □

Primjer 2.3. *Dokažimo da je* $G = \{a + b\sqrt{2} : a, b \in \mathbb{Q},\ a^2 + b^2 \neq 0\}$ *grupa s obzirom na množenje.*

Rješenje: Ako su $a + b\sqrt{2}, c + d\sqrt{2} \in G$, onda je $(a + b\sqrt{2})(c + d\sqrt{2}) = (ac + 2bd) + (ad + bc)\sqrt{2} \in G$, pa smo pokazali zatvorenost. Asocijativnost slijedi iz asocijativnosti množenja u $\mathbb{R}$. Broj $1 \in G$ je neutralni element s obzirom na množenje. Inverzni element od $a + b\sqrt{2} \in G$ je

$$\frac{1}{a + b\sqrt{2}} = \frac{a - b\sqrt{2}}{a^2 - 2b^2} = \frac{a}{a^2 - 2b^2} + \frac{-b}{a^2 - 2b^2}\sqrt{2} \in G,$$

jer je $a^2 - 2b^2 \neq 0$ (ako a, b nisu obadva jednaki 0) što slijedi iz činjenice da je broj $\sqrt{2}$ iracionalan. ♢

Primjer 2.4. *Neka je n prirodan broj. Dokažimo da je $K_n = \{z \in \mathbb{C} : z^n = 1\}$ grupa s obzirom na množenje kompleksnih brojeva.*

Rješenje: Pokažimo zatvorenost: $z_1, z_2 \in K_n$ povlači $(z_1 \cdot z_2)^n = z_1^n \cdot z_2^n = 1$, pa je $z_1 \cdot z_2 \in K_n$. Asocijativnost se nasljeđuje iz $\mathbb{C}$. Neutralni element je $1 \in K_n$. Inverz od $z \in K_n$ je $\frac{1}{z}$ i $(\frac{1}{z})^n = \frac{1}{z^n} = 1$, pa je $\frac{1}{z} \in K_n$ (uočimo da je $z \neq 0$).

Ovaj primjer pokazuje da za svaki $n \in \mathbb{N}$ postoji (konačna) grupa s n elemenata. ♢

Primjer 2.5. *Neka je $m \in \mathbb{N}$, te neka je $\mathbb{Z}_m = \{0, 1, 2, \ldots, m-1\}$ (ostatci pri dijeljenju s m). Definiramo binarnu operaciju $+_m$ na $\mathbb{Z}_m$ na sljedeći način. Za $x, y \in \mathbb{Z}_m$, neka je $x + y = qm + r$, $q \in \mathbb{Z}$, $r \in \mathbb{Z}_m$. Tada je $x +_m y = r$. Dokažimo da je $(\mathbb{Z}_m, +_m)$ abelova grupa.*

Rješenje: Zatvorenost je ispunjena po definiciji. Dokažimo da vrijedi asocijativnost. Neka je $x +_m y = r$, $(x +_m y) +_m z = s$, $y +_m z = t$. To znači da postoje cijeli brojevi k, l, p takvi da je $x + y = km + r$, $r + z = lm + s$, $y + z = pm + t$. Tada je

$$\begin{aligned} x + t &= x - pm + pm + t = -pm + x + y + z = -pm + km + r + z \\ &= -pm + km + lm + s = (-p + k + l)m + s. \end{aligned}$$

Stoga je $x +_m (y +_m z) = x +_m t = s = (x +_m y) +_m z$.

Neutralni element je 0. Element 0 je sam sebi inverz, a za $k > 0$ je $m - k$ inverz od k. Komutativnost slijedi iz komutativnosti zbrajanja u $\mathbb{Z}$. ♢

Primjer 2.6. *Na skupu $\mathbb{Z}_m \setminus \{0\}$ definiramo operaciju $\cdot_m$ na sljedeći način. Ako je $x \cdot y = qm + r$, $q \in \mathbb{Z}$, $r \in \mathbb{Z}_m$, onda je $x \cdot_m y = r$. Dokažimo da je $(\mathbb{Z}_m \setminus \{0\}, \cdot_m)$ grupa ako i samo ako je m prost.*

Rješenje: Ako je m složen, recimo $m = a \cdot b$, onda su $a, b \in \mathbb{Z}_m \setminus \{0\}$, ali $a \cdot_m b = 0$, pa operacija $\cdot_m$ nije zatvorena.

Ako je m prost, operacija $\cdot_m$ je zatvorena i asocijativna (dokazuje se slično kao u prethodnom primjeru). Neutralni element je 1. Pokažimo da $a \in \mathbb{Z}_m \setminus \{0\}$ ima inverz. Vrijedi $\text{nzd}(a, m) = 1$, pa postoje $u, v \in \mathbb{Z}$ takvi da je $au + mv = 1$. Štoviše, za svaki $t \in \mathbb{Z}$ je $a(u + tm) + m(v - ta) = 1$. Odaberimo t tako da je $z = u + tm \in \mathbb{Z}_m$. Tada je $a \cdot_m z = 1$, tj. $z = a^{-1}$. ♢

Npr. u $(\mathbb{Z}_5 \setminus \{0\}, \cdot_5)$ imamo sljedeću "tablicu množenja":

$\cdot_5$	1	2	3	4
1	1	2	3	4
2	2	4	1	3
3	3	1	4	2
4	4	3	2	1

Slično kao u prethodnom primjeru, pokazuje se za svaki prirodan broj m skup $\mathbb{Z}_m^* = \{a \in \mathbb{Z}_m : \text{nzd}(a, m) = 1\}$ čini grupu s obzirom na operaciju $\cdot_m$. Za p prost je $\mathbb{Z}_p^* = \mathbb{Z}_p \setminus \{0\}$.

Neka su X i Y grupe. *Homomorfizam grupa* $f : X \to Y$ se definira na isti način kao i homomorfizam polugrupa.

Propozicija 2.4. *Neka su X i Y grupe s neutralnim elementima e_X i e_Y, te neka je $f : X \to Y$ homomorfizam. Tada je*

(1) $f(e_X) = e_Y$,

(2) $f(x^{-1}) = (f(x))^{-1}$, $\forall x \in X$.

Dokaz:

(1) Iz $f(e_X) = f(e_X \cdot e_X) = f(e_X) \cdot f(e_X)$, slijedi $(f(e_X))^{-1} f(e_X) = f(e_X)$, tj. $e_Y = f(e_X)$, jer je $(f(e_X))^{-1}$ inverz od $f(e_X)$ u Y.

(2) Koristeći (1), dobivamo:

$$e_Y = f(e_X) = f(x \cdot x^{-1}) = f(x) \cdot f(x^{-1}),$$
$$e_Y = f(x^{-1} \cdot x) = f(x^{-1}) \cdot f(x),$$

pa je po definiciji inverza $(f(x))^{-1} = f(x^{-1})$.

□

Primjer 2.7. *Neka je S neprazan skup i neka je $B(S)$ skup svih bijekcija (permutacija) $S \to S$. Tada $B(S)$ čini grupu s obzirom na operaciju komponiranja funkcija Zaista, kompozicija bijekcija je bijekcija, komponiranje je asocijativno, identiteta je bijekcija, a inverz bijekcije je bijekcija. Ako je $S = \{1, 2, \ldots, n\}$, onda se $B(S)$ označava sa S_n i naziva* simetrična grupa stupnja n. *Vrijedi: $k(S_n) = n!$.*

Permutaciju f obično zapisujemo ovako:

$$\begin{pmatrix} 1 & 2 & 3 & \cdots & n \\ f(1) & f(2) & f(3) & \cdots & f(n) \end{pmatrix}.$$

Na primjer,

$$S_3 = \left\{ \begin{pmatrix} 1\ 2\ 3 \\ 1\ 2\ 3 \end{pmatrix}, \begin{pmatrix} 1\ 2\ 3 \\ 1\ 3\ 2 \end{pmatrix}, \begin{pmatrix} 1\ 2\ 3 \\ 2\ 1\ 3 \end{pmatrix}, \begin{pmatrix} 1\ 2\ 3 \\ 2\ 3\ 1 \end{pmatrix}, \begin{pmatrix} 1\ 2\ 3 \\ 3\ 1\ 2 \end{pmatrix}, \begin{pmatrix} 1\ 2\ 3 \\ 3\ 2\ 1 \end{pmatrix} \right\},$$

te vrijedi

$$\begin{pmatrix} 1\ 2\ 3 \\ 3\ 1\ 2 \end{pmatrix} \circ \begin{pmatrix} 1\ 2\ 3 \\ 2\ 1\ 3 \end{pmatrix} = \begin{pmatrix} 1\ 2\ 3 \\ 1\ 3\ 2 \end{pmatrix}, \quad \begin{pmatrix} 1\ 2\ 3 \\ 2\ 1\ 3 \end{pmatrix} \circ \begin{pmatrix} 1\ 2\ 3 \\ 3\ 1\ 2 \end{pmatrix} = \begin{pmatrix} 1\ 2\ 3 \\ 3\ 2\ 1 \end{pmatrix}.$$

Svaka se permutacija može prikazati kao produkt (tj. kompozicija) disjunktnih ciklusa, gdje $(i_1\, i_2 \cdots i_r)$ *znači da je* $f(i_1) = i_2$, $f(i_2) = i_3$, $\ldots$ $f(i_{r-1}) = i_r$, $f(i_r) = i_1$. *Na primjer,*

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 3 & 5 & 2 & 8 & 1 & 6 & 7 & 4 \end{pmatrix} = (1\,3\,2\,5)\,(4\,8)\,(6)\,(7) = (1\,3\,2\,5)\,(4\,8).$$

Ciklus duljine 2 *naziva se* transpozicija. *Svaka permutacija se može napisati kao produkt transpozicija. Zaista, dovoljno je vidjeti da se svaki ciklus može napisati u tom obliku:*

$$(i_1\, i_2\, \cdots i_r) = (i_1\, i_2)(i_2\, i_3) \cdots (i_{r-1}\, i_r).$$

Iako rastav permutacije na produkt traspozicija nije jedinstven, broj transpozicija u rastavu je uvijek iste parnosti. Stoga ima smisla definirati parnost permutacije. Za permutaciju kažemo da je parna *ako se može dobiti kao produkt (tj. kompozicija) parnog broja transpozicija.*

Definicija 2.4. *Grupa* $(X, \cdot)$ *je* komutativna *ili* abelova *ako je* $x \cdot y = y \cdot x$ *za sve* $x, y \in X$.

Obično se za abelove grupe koristi aditivni zapis: $x + y$.

Napomena 2.1. Simetrična grupa S_n, za $n \geq 3$, nije abelova. Zaista, u Primjeru 2.7 smo vidjeli da je

$$(1\,3\,2)\,(1\,2) = (2\,3) \neq (1\,2)\,(1\,3\,2) = (1\,3).$$

Primjer 2.8. *Primjeri komutativnih grupa:*

1. $(\mathbb{Z}, +)$, $(\mathbb{Q}, +)$, $(\mathbb{R}, +)$, $(\mathbb{C}, +)$;
2. $(\mathbb{Q}^*, \cdot)$, $(\mathbb{R}^*, \cdot)$, $(\mathbb{C}^*, \cdot)$, *gdje je* $X^* = X \setminus \{0\}$.

Definicija 2.5. *Neka je* X *grupa, te* Y *podskup od* X *sa svojstvom da je* $y^{-1} \in Y$ *za svaki* $y \in Y$ *i* $y \cdot y' \in Y$ *za sve* $y, y' \in Y$. *Tada kažemo da je* Y podgrupa *od* X. *Oznaka:* $Y \leq X$.

Napomena 2.2. $Y \subseteq X$ je podgrupa od X ako i samo ako vrijedi $y' \cdot y^{-1} \in Y$ za sve $y, y' \in Y$.

Definicija 2.6. *Neka je* X *grupa, te* $S \subseteq X$ *neki podskup. Presjek svih podgrupa od* X *koje sadrže skup* S *je također podgrupa od* X *koja se naziva* podgrupa generirana skupom S. *To je najmanja podgrupa od* X *koja sadrži* S. *Označava se sa* $X(S)$.

Primjer 2.9.

1. *U grupi* $(\mathbb{Z}, +)$ *gledamo* $S = \{1\}$. *Neka je* $Y \subseteq \mathbb{Z}$ *podgrupa koja sadrži* S. *Tvrdimo da je* $Y = \mathbb{Z}$. *Budući da je* $1 \in Y$, *a* Y *je podgrupa, to* Y *sadrži i* $1+1 = 2$. *Matematičkom indukcijom dobivamo da je* $n \in Y$, *za svaki* $n \in \mathbb{N}$. *Ali* Y *je grupa, pa sadrži neutralni element* 0, *a također i* $-n$, *za svaki* $n \in \mathbb{N}$. *Prema tome,* $Y = \mathbb{Z}$. *Dakle,* $(\mathbb{Z}, +)$ *je grupa s jednim generatorom (elementom* 1*).*

2. *Neka je* $G = \{1, -1\}$ *uz operaciju množenja. Tada je* G *generirana s elementom* -1. *Uočimo da* G *predstavlja invertibilne elemente monoida* $(\mathbb{Z}, \cdot)$.

Definicija 2.7. *Grupa generirana s jednim elementom zove se* ciklička grupa.

Primjer 2.10.

1. *Grupa* $(\mathbb{Z}, +)$ *je ciklička; generator joj je element* 1 *(a također i element* -1*).*

2. *Grupa* $(\mathbb{Z}_m, +_m)$ *je ciklička; generator joj je svaki element* $a \in \mathbb{Z}_m$ *koji je relativno prost sa* m.

3. *Grupa* $(\mathbb{Z}_p \setminus \{0\}, \cdot_p)$, *za prost broj* p, *je ciklička; generator joj je svaki primitivni korijen modulo* p.

Definicija 2.8. Neka je X grupa, $Y \leq X$, te $a \in X$. *Red podgrupe* Y je njezin kardinalni broj $k(Y)$. *Red elementa* a je red podgrupe $X(a)$ generirane elementom a, tj. najmanji prirodan broj r (ako takav postoji) sa svojstvom da je a^r jednak neutralnom elementu u grupi X.

Primjer 2.11. *Odrediti red*

a) elementa 6 *u grupi* $(\mathbb{Z}_7, +_7)$;

b) elementa 6 *u grupi* $(\mathbb{Z}_9, +_9)$;

c) elementa 6 *u grupi* $(\mathbb{Z}_{10}, +_{10})$.

Rješenje:

a) Red je 7, jer iz $6x \equiv 0 \pmod 7$ slijedi $x \equiv 0 \pmod 7$.

b) Red je 3, jer iz $6x \equiv 0 \pmod 9$ slijedi $2x \equiv 0 \pmod 3$, tj. $x \equiv 0 \pmod 3$.

c) Red je 5, jer iz $6x \equiv 0 \pmod{10}$ slijedi $3x \equiv 0 \pmod 5$, tj. $x \equiv 0 \pmod 5$.

$\diamondsuit$

Primjer 2.12. *Grupa* $(\mathbb{Z}_{20}^*, \cdot_{20})$ *nije ciklička.*

Rješenje: Red ove grupe je $\varphi(20) = 8$. Zaista,

$$\mathbb{Z}_{20}^* = \{1, 3, 7, 9, 11, 13, 17, 19\}.$$

No, u toj grupi ne postoji niti jedan element reda 8. Npr. red od 3 je jednak 4, jer je $3^2 \equiv 9 \pmod{20}$, $3^3 \equiv 7 \pmod{20}$, $3^4 \equiv 1 \pmod{20}$ (podrupa generirana s 3 je $\{1, 3, 7, 9\}$, dok je red od 9 jednak 2, jer je $9^2 \equiv 1 \pmod{20}$ (podrupa generirana s 9 je $\{1, 9\}$. Analogno je dobije da $3, 7, 13, 17$ imaju red 4, dok $9, 11, 19$ imaju red 2 (naravno, neutralni element 1 ima red 1).

Drugim riječima, pokazali smo da ne postoji primitivni korijen modulo 20. ♢

Neka je $Y \leq X$ podgrupa. Definiramo relaciju $\sim$ na X ovako:

$$x \sim x' \text{ ako i samo ako postoji } y \in Y \text{ takav da je } x = x'y.$$

Očito je $\sim$ relacija ekvivalencije. Klase ekvivalencije označavamo sa $[x]$.

Propozicija 2.5. $[x] = xY = \{xy : y \in Y\}$

Dokaz: Uzmimo $x' \in [x]$. Tada postoji $y \in Y$ takav da je $x' = xy$, pa je $x' \in xY$. Tako smo dokazali da je $[x] \subseteq xY$.

Za svaki $y \in Y$ je $xy \sim x$, pa je $xy \in [x]$. Stoga je i $xY \subseteq [x]$. □

Uočimo da je $[e] = eY = Y$ i to je jednina među klasama koja je podgrupa od X. Skup $[x]$ zove se *lijeva klasa* grupe X po podgrupi Y. Vrijedi:

$$X = \bigcup_{x \in X} [x] = \bigcup_{x \in X} xY.$$

Propozicija 2.6. *Svaka lijeva klasa* xY *ima isti kardinalni broj.*

Dokaz: Dokazat ćemo da je preslikavanje $\phi : Y \to xY$ definirano sa $\phi(y) = xy$ bijekcija. Neka je $xy \in xY$. Tada je $\phi(y) = xy$, pa je ϕ surjekcija. Ako je $\phi(y) = \phi(y')$, onda je $xy = xy'$, što povlači $x^{-1}xy = x^{-1}xy'$, tj. $y = y'$. Dakle, ϕ je i injekcija. □

Definicija 2.9. Kvocijentni skup $X/\sim$ naziva se *lijevi kvocijentni skup* grupe X po podgrupi Y i označava se X/Y. Kardinalni broj skupa X/Y naziva se *indeks podgrupe* Y u grupi X i označava se sa $[X : Y]$.

Propozicija 2.7. $k(X) = [X : Y] \cdot k(Y)$

Dokaz: Slijedi neposredno iz Propozicije 2.6. □

Korolar 2.8 (Lagrangeov teorem)**.** *Ako je grupa* X *konačna, a* $Y \leq X$ *njezina podgrupa, onda red od* Y *dijeli red od* X*. Nadalje, red svakog elementa* $x \in X$ *dijeli red od* X*.*

Napomena 2.3. Eulerov teorem je specijalni slučaj Lagrangeovog teorema. Zaista, red grupe $(\mathbb{Z}_m^*, \cdot_m)$ je $\varphi(m)$, pa za svaki $a \in \mathbb{Z}_m^*$ vrijedi $a^{\varphi(m)} \equiv 1 \pmod{m}$.

Sasvim analogno se definiraju *desne klase* Yx. Jasno je da je

$$k(Yx) = k(Y) = k(xY), \quad \forall x \in X.$$

Međutim, ne mora vrijediti da je $Yx = xY$.

Definicija 2.10. Podgrupa $Y \leq X$ je *normalna* ako je $xY = Yx$ za svaki $x \in X$. Oznaka: $Y \lhd X$.

Napomena 2.4. Ako je grupa X abelova, onda je svaka njezina podgrupa normalna. U svakoj grupi X postoje barem dvije normalne podgrupe: $\{e\}$ i X. To su tzv. trivijalne normalne podgrupe.

Za grupu X kažemo da je *prosta* ako nema netrivijalnih normalnih podgrupa.

Propozicija 2.9. *Sljedeće tvrdnje su ekvivalentne:*

1) Podgrupa $Y \leq X$ je normalna.

2) $xY \subseteq Yx$, $\forall x \in X$.

3) $xYx^{-1} \subseteq Y$, $\forall x \in X$.

Dokaz: Očito 1) $\Rightarrow$ 2). Dokažimo da 2) $\Rightarrow$ 3). Iz $xY \subseteq Yx$ slijedi $xYx^{-1} \subseteq Yxx^{-1} = Y$. Preostaje dokazati da 3) $\Rightarrow$ 1). Iz $xYx^{-1} \subseteq Y$, množenjem sa x zdesna, dobivamo da za svaki $x \in X$ vrijedi $xY \subseteq Yx$. Uvrstimo u zadnju relaciju x^{-1} umjesto x, pa dobivamo da $x^{-1}Y \subseteq Yx^{-1}$. Odavde je $x^{-1}Yx \subseteq Y$, što povlači $Yx \subseteq xY$. Time smo dokazali da je $xY = Yx$. □

Primjer 2.13. *Neka je* $H = \left\{ \begin{bmatrix} 1 & 0 \\ b & 1 \end{bmatrix} : b \in \mathbb{R} \right\}$. *Dokazati da je H podgrupa od $GL(2, \mathbb{R})$ (grupe regularnih matrica reda 2 s realnim koeficijentima), ali nije normalna podrupa.*

Rješenje: Neka su $A, B \in H$, $A = \begin{bmatrix} 1 & 0 \\ b & 1 \end{bmatrix}$, $B = \begin{bmatrix} 1 & 0 \\ d & 1 \end{bmatrix}$. Tada je $AB^{-1} = \begin{bmatrix} 1 & 0 \\ b-d & 1 \end{bmatrix} \in H$, pa je H podgrupa od $GL(2, \mathbb{R})$.

Uzmimo sada element $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ iz $GL(2, \mathbb{R})$ i $B = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$ iz H, te izračunajmo ABA^{-1}. Dobivamo da $ABA^{-1} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \notin H$, pa zaključujemo da H nije normalna podrupa od $GL(2, \mathbb{R})$. □

Ako je Y normalna podgrupa, onda se na skupu svih lijevih (odnosno desnih) klasa može definirati množenje:

$$[x_1] \cdot [x_2] = [x_1 \cdot x_2].$$

To je dobro definirana binarna operacija, jer za $x_1' = x_1 y_1$, $x_2' = x_2 y_2$, gdje su $y_1, y_2 \in Y$, imamo $x_1' x_2' = x_1 (y_1 x_2) y_2$. Budući da je $y_1 x_2 \in Y x_2 = x_2 Y$, to postoji $y_1' \in X$ takav da je $y_1 x_2 = x_2 y_1'$. Stoga je $x_1' x_2' = x_1 x_2 y_1' y_2$, pa jer je $y_1' y_2 \in Y$, dobivamo da je $x_1' x_2' \sim x_1 x_2$, što je i trebalo dokazati.

Očito je $[e][x] = [x][e] = [x]$, što znači da je $[e]$ neutralni element za operaciju množenja klasa. Nadalje, $[x^{-1}][x] = [x][x^{-1}] = [e]$, tj. $[x]^{-1} = [x^{-1}]$. Dakle, na skupu $X/\sim$ klasa ekvivalencije dobivamo strukturu grupe. To je *kvocijentna grupa* X/Y. Preslikavanje $q : X \to X/Y$ definirano sa $q(x) = [x]$ naziva se *kvocijentni homomorfizam.*

Napomena 2.5. Ako je grupa X komutativna, onda je i kvocijentna grupa X/Y komutativna.

Primjer 2.14.

1. *Neka je $n \in \mathbb{N}$. Podgrupa $n\mathbb{Z} = \{nz : z \in \mathbb{Z}\} \subseteq \mathbb{Z}$ je normalna (jer je $(\mathbb{Z}, +)$ abelova), pa je definirana kvocijentna grupa $\mathbb{Z}_n = \mathbb{Z}/n\mathbb{Z} = \{[0], [1], \ldots, [n-1]\}$ (klase ostataka modulo n). Za p prost je grupa $\mathbb{Z}_p$ prosta (po Korolaru 2.8).*

2. *Simetrična grupa S_n ima pravu podgrupu A_n (podgrupa svih parnih permutacija, tzv.* alternirajuća grupa*). Iz $[S_n : A_n] = 2$ slijedi da je A_n normalna podgrupa od S_n, što povlači da grupa S_n (za ≥ 3) nije prosta. Može se pokazati da je za $n \geq 5$ grupa A_n prosta.*

Definicija 2.11. *Neka je $f : X \to Y$ homomorfizam grupa.* Jezgra *homomorfizma f je* $\operatorname{Ker} f = f^{-1}(e_Y) \subseteq X$. Slika *homomorfizma f je* $\operatorname{Im} f = f(X) \subseteq Y$.

Lema 2.10. $\operatorname{Ker} f$ *je podgrupa od X;* $\operatorname{Im} f$ *je podgrupa od Y.*

Dokaz:

a) Neka su $x, y \in \operatorname{Ker} f$. Tada je $f(x \cdot y) = f(x) \cdot f(y) = e_Y \cdot e_Y = e_Y$, pa je $x \cdot y \in \operatorname{Ker} f$. Nadalje, za $x \in \operatorname{Ker} f$ je, po Propoziciji 2.4, $e_Y = f(e_X) = f(x^{-1} \cdot x) = f(x^{-1}) \cdot f(x) = f(x^{-1})$, pa je $x^{-1} \in \operatorname{Ker} f$. Time smo dokazali da je $\operatorname{Ker} f$ podgrupa od X.

b) Neka su $y, y' \in \operatorname{Im} f$. Tada postoje $x, x' \in X$ takvi da je $y = f(x)$, $y' = f(x')$. Stoga je $y \cdot y' = f(x) \cdot f(x') = f(x \cdot x') \in \operatorname{Im} f$. Ako je $y \in \operatorname{Im} f$ i $y = f(x)$, onda je, po Propoziciji 2.4, $y^{-1} = f(x^{-1}) \in \operatorname{Im} f$, čime smo dokazali da je $\operatorname{Im} f \leq Y$.

□

Lema 2.11. Ker f *je normalna podrupa od* X.

Dokaz: Prema Propoziciji 2.9, dovoljno je dokazati da za $K = \operatorname{Ker} f$ vrijedi $xKx^{-1} \subseteq K$ za svaki $x \in X$. Neka je $u \in xKx^{-1}$, tj. $u = xkx^{-1}$ za neki $k \in K$. Tada je

$$f(u) = f(xkx^{-1}) = f(x) \cdot f(k) \cdot f(x^{-1}) = f(x) \cdot (f(x))^{-1} = e_Y,$$

što pokazuje da je $u \in K$. □

Napomena 2.6. Slika $\operatorname{Im} f$ općenito ne mora biti normalna podrupa od Y.

Definicija 2.12. *Neka je* $f : X \to Y$ *homomorfizam grupa. Kažemo da je* f

- monomorfizam *ako je* f *injekcija;*
- epimorfizam *ako je* f *surjekcija;*
- izomorfizam *ako je* f *bijekcija.*

Lako se vidi da je izomorfizam grupa relacija ekvivalencije.

Primjer 2.15. *Neka je* $K_4 = \{1, -1, i, -i\}$. *Dokažimo da je preslikavanje* $f : (\mathbb{Z}, +) \to (K_4, \cdot)$, $f(x) = i^x$ *homomorfizam i nađimo mu jezgru.*

Rješenje: Imamo: $f(x + y) = i^{x+y} = i^x \cdot i^y = f(x) \cdot f(y)$, pa je f homomorfizam. Očito je f surjekcija, pa se radi o epimorfizmu. Nadalje,

$$\operatorname{Ker}(f) = \{x \in \mathbb{Z} : f(x) = 1\} = \{x \in \mathbb{Z} : i^x = 1\} = \{4k : k \in \mathbb{Z}\} = 4\mathbb{Z}.$$

♢

Primjer 2.16. *Dokažimo da postoji izomorfizam grupa* $(\mathbb{R}, +)$ *i* $(\mathbb{R}^+, \cdot)$, *ali ne postoji izomorfizam grupa* $(\mathbb{Q}, +)$ *i* $(\mathbb{Q}^+, \cdot)$.

Rješenje:

a) Prelikavanje $f : \mathbb{R} \to \mathbb{R}^+$, $f(x) = e^x$ je bijekcija i homomorfizam, jer vrijedi

$$f(x + y) = e^{x+y} = e^x \cdot e^y = e^x \cdot e^y = f(x) \cdot f(y).$$

b) Pretpostavimo da je $g : \mathbb{Q} \to \mathbb{Q}^+$ izomorfizam. Broj 2 je element od $\mathbb{Q}^+$, pa postoji $q \in \mathbb{Q}$, takav da je $g(q) = 2$. No, sada iz

$$2 = g\left(\frac{q}{2} + \frac{q}{2}\right) = g\left(\frac{q}{2}\right) \cdot g\left(\frac{q}{2}\right) = \left(g\left(\frac{q}{2}\right)\right)^2$$

slijedi da je $g(\frac{q}{2}) = \sqrt{2} \in \mathbb{Q}$, što je kontradikcija.

♢

Lema 2.12. *Homomorfizam $f : X \to Y$ je injekcija (tj. monomorfizam) ako i samo ako je* $\operatorname{Ker} f = \{e_X\}$.

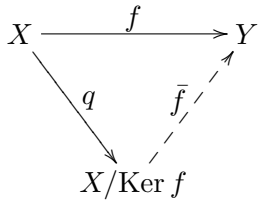
Dokaz: Ako je f injekcija, onda iz $f(e_X) = e_Y$ slijedi da je $\operatorname{Ker} f = \{e_X\}$. Dokažimo obrat. Neka su $x, y \in X$ takvi da je $f(x) = f(y)$. Tada je

$$e_y = (f(x))^{-1} \cdot f(y) = f(x^{-1}) \cdot f(y) = f(x^{-1}y),$$

pa iz $\operatorname{Ker} f = \{e_X\}$ slijedi da je $x^{-1}y = e_X$, tj. $y = x$. Stoga je f injekcija. □

Teorem 2.13 (Teorem o izomorfizmu grupa)**.** *Neka je $f : X \to Y$ homomorfizam grupa. Tada postoji jedinstveni homomorfizam $\bar{f} : X/\operatorname{Ker} f \to Y$ takav da je $\bar{f} \circ q = f$, gdje je $q : X \to X/\operatorname{Ker} f$ kvocijentni homomorfizam. Nadalje,, $\bar{f}$ je monomorfizam. Ukoliko je f epimorfizam, onda je $\bar{f}$ izomorfizam.*

Dokaz:



Neka je $K = \operatorname{Ker} f$, te neka je $[x] = xK \in X/K$. Definirajmo $\bar{f}([x]) = f(x)$. Pokažimo najprije da je definicija dobra. Za $x' \in [x]$ postoji $k \in K$ takav da je $x' = x \cdot k$, pa je zaista

$$f(x') = f(x) \cdot f(k) = f(x).$$

Očito je $\bar{f} \circ q = f$. Pokažimo da je $\bar{f}$ homomorfizam:

$$\bar{f}([x] \cdot [y]) = \bar{f}([xy]) = f(xy) = f(x) \cdot f(y) = f([x]) \cdot f([y]).$$

Da bi provjerili da je $\bar{f}$ monomorfizam, po Lemi 2.12, dovoljno je provjeriti da je $\operatorname{Ker} \bar{f} = \{e_{X/K}\} = \{eK\} = \{K\}$. Neka je $[x] \in X/K$ takav da je $\bar{f}([x]) = e_Y$. Tada je $f(x) = e_Y$, pa je $x \in K$, te je zaista $[x] = xK = K$.

Dokažimo jedinstvenost: ako je $\bar{f} \circ q = f$, onda mora biti $\bar{f}([x]) = \bar{f}(q(x)) = f(x)$, a to je upravo kako smo definirali f.

Konačno, neka je f epimorfizam. Pokažimo da je tada i $\bar{f}$ epimorfizam. Neka je $y \in Y$. Tada postoji $x \in X$ takav da je $f(x) = y$, pa je $\bar{f}([x]) = f(x) = y$, što pokazuje da je $\bar{f}$ surjekcija. □

Primjer 2.17. *Dokažimo da je $(\mathbb{R}, +)/(\mathbb{Z}, +) \simeq (S^1, \cdot)$, gdje je S^1 jednodimenzionalna jedinična sfera (kružnica).*

Rješenje: Definirajmo preslikavanje $f : \mathbb{R} \to S^1$,

$$f(x) = e^{2\pi ix} = \cos(2\pi x) + i\sin(2\pi x).$$

Funkcija f je surjekcija, tj. $f(\mathbb{R}) = S^1$ i f je homomorfizam. Odredimo jezgru od f. Imamo da je $f(x) = \cos(2\pi x) + i\sin(2\pi x) = 1$ ako i samo ako je x cijeli broj. Stoga je $\mathrm{Ker}(f) = \mathbb{Z}$. Po teoremu o izomorfizmu dobivamo da je $\mathbb{R}/\mathbb{Z} \simeq S^1$. ♢

Teorem 2.14 (Cayley)**.** *Svaka grupa X je izomorfna nekoj podgrupi grupe $B(X)$ svih bijekcija skupa X na samog sebe.*

Dokaz: Za $a \in X$ definirajmo preslikavanje $f_a : X \to X$ s $f_a(x) = a \cdot x$. Tvrdimo da je f_a bijekcija. Neka je $f_a(x) = f_a(x')$. Tada je $ax = ax'$, što povlači da je $x = x'$, pa je f_a injekcija. Za svaki $x \in X$ je $a^{-1}x \in X$ i vrijedi $f_a(a^{-1}x) = aa^{-1}x = x$, što pokazuje da je f_a surjekcija.

Definirajmo sada preslikavanje $\phi : X \to B(X)$ sa $\phi(a) = f_a$. Dokazat ćemo da je ϕ monomorfizam grupa. Imamo $\phi(a)(x) = f_a(x) = a \cdot x$, pa je

$$\phi(a \cdot b)(x) = a \cdot b \cdot x = f_a(bx) = f_a(f_b(x)) = (\phi(a) \circ \phi(b))(x), \quad \forall x \in X.$$

Po definicije jednakosti funkcija, ovo povlači da je $\phi(a \cdot b) = \phi(a) \circ \phi(b)$, a to upravo znači da je ϕ homomorfizam grupa.

Dokažimo injektivnost. Neka je $a \neq b$. Tada je $\phi(a)(e) = f_a(e) = a \cdot e = a$, $\phi(b)(e) = f_b(e) = b \cdot e = b$, pa je zaista $\phi(a) \neq \phi(b)$.

Prema Lemi 2.10 je $\phi(X) \subseteq B(X)$ podgrupa od $B(X)$ i $\phi : X \to \phi(X)$ je izomorfizam. □

Definicija 2.13. *Neka su $(G, *)$ i $(H, \cdot)$ grupe, te neka je $G \times H = \{(g, h) : g \in G, h \in H\}$. Tada je $G \times H$ grupa uz binarnu operaciju*

$$(g_1, h_1) \circ (g_2, h_2) = (g_1 * g_2, h_1 \cdot h_2).$$

Grupu $(G \times H, \circ)$ nazivamo direktni produkt *grupa $(G, *)$ i $(H, \cdot)$.*

Primjer 2.18.

a) Jesu li grupe $\mathbb{Z}_2 \times \mathbb{Z}_2$ i $\mathbb{Z}_4$ izomorfne?

b) Jesu li grupe $\mathbb{Z}_2 \times \mathbb{Z}_3$ i $\mathbb{Z}_6$ izomorfne?

Rješenje:

a) Pretpostavimo da postoji izomorfizam $f : \mathbb{Z}_4 \to \mathbb{Z}_2 \times \mathbb{Z}_2$. Neka je $f(1) = (a, b)$. Tada je

$$f(2) = f(1 +_4 1) = f(1) + f(1) = (a +_2 a) + (b +_2 b) = (0, 0).$$

Ali i $f(0) = (0, 0)$, pa f nije injekcija. Drugim riječima, $1 \in \mathbb{Z}_4$ ima red 4, a niti jedan element iz $\mathbb{Z}_2 \times \mathbb{Z}_2$ nema red 4. Dakle, grupe nisu izomorfne.

b) Neka je $f : \mathbb{Z}_2 \times \mathbb{Z}_3 \to \mathbb{Z}_6$, $f(a,b) = 3a +_6 2b$. Očito je f bijekcija. Dokažimo da je homomorfizam:

$$\begin{aligned} f((a,b)(c,d)) &= f(a +_2 c, b +_3 d) = 3(a +_2 c) +_6 2(b +_3 d) \\ &= (3a +_6 2b) +_6 (3c +_6 2d) = f(a,b) +_6 f(c,d), \end{aligned}$$

gdje smo koristili da ako je $x \equiv x' \pmod 2$, onda je $3x \equiv 3x' \pmod 6$, a ako je $x \equiv x' \pmod 3$, onda je $2x \equiv 2x' \pmod 6$. Dakle, grupe $\mathbb{Z}_2 \times \mathbb{Z}_3$ i $\mathbb{Z}_6$ su izomorfne.

Znamo da je $\mathbb{Z}_6$ ciklička grupa s generatorima 1 i 5. Stoga je i grupa $\mathbb{Z}_2 \times \mathbb{Z}_3$ također ciklička. Generatori su joj elementi $(1,1)$ i $(1,2)$.

♢

Primjer 2.19. *Neka je G grupa, te neka su K i H njene izomorfne normalne podgrupe. Mora li biti $G/K \simeq G/H$?*

Rješenje: Odgovor je NE.

Neka je $G = \mathbb{Z}_4 \times \mathbb{Z}_2$, $H = \{(2,0),(0,0)\}$, $K = \{(0,1),(0,0)\}$. Preslikavanje $f : G \to \mathbb{Z}_2 \times \mathbb{Z}_2$, $f(a,b) = (a \bmod 2, b)$ je epimorfizam i $\mathrm{Ker}(f) = H$, pa je $G/H \simeq \mathbb{Z}_2 \times \mathbb{Z}_2$.

Preslikavanje $g : G \to \mathbb{Z}_4$, $f(a,b) = a$ je epimorfizam i $\mathrm{Ker}(f) = K$, pa je $G/k \simeq \mathbb{Z}_4$. No, kao što smo pokazali u prethodnom primjeru, $\mathbb{Z}_2 \times \mathbb{Z}_2 \not\simeq \mathbb{Z}_4$.

♢

2.2 Prsteni i polja

Definicija 2.14. Prsten *je skup R zajedno s dvije binarne operacije $+$ i $\cdot$ za koje vrijedi*

1) $(R,+)$ *je abelova grupa;*

2) $(R,\cdot)$ *je polugrupa;*

3) $x\cdot(y+z)=x\cdot y+x\cdot z, \quad i \quad (x+y)\cdot z=x\cdot z+y\cdot z \quad \forall x,y,z\in R$ *(*distributivnost $\cdot$ *obzirom na* $+$*)*

Napomena 2.7. Neutralni element za zbrajanje označava se s 0. Vrijedi: $0\cdot a=a\cdot a=0$ za svaki $a\in R$. Zaista, $a\cdot 0=a\cdot(0+0)=a\cdot 0+a\cdot 0$, što povlači $a\cdot 0=0$. Neutralni element za množenje (ako postoji) označava se s 1. Ako je množenje komutativno, govori se o *komutativnom prstenu.*

Primjer 2.20.

1) $(\mathbb{Z},+,\cdot)$, $(\mathbb{Q},+,\cdot)$, $(\mathbb{R},+,\cdot)$, $(\mathbb{C},+,\cdot)$ *su prsteni.*

2) M_n *je prsten uz operacije zbrajanja i množenja kvadratnih matrica.*

3) $\mathbb{Z}_m$ *je prsten uz operacije* $+_m$ *i* $\cdot_m$*.*

4) Prsten polinoma

Neka je $(R,+,\cdot)$ neki prsten. *Polinom* u varijabli t nad R je svaki izraz (formalna suma) oblika

$$p(t)=a_nt^n+a_{n-1}t^{n-1}+\cdots+a_1+a_0,$$

gdje su $a_0,a_1,\ldots,a_n\in R$, $n\in\mathbb{N}_0$. Skup svih polinoma nad R označavamo s $R[t]$. U $R[t]$ se uvodi zbrajanje i množenje ovako: ako je $p(t)$ kao gore i $q(t)=b_mt^m+b_{m-1}t^{m-1}+\cdots+b_1+b_0$, onda je

$$p(t)+q(t)=\sum_{k=0}^{\max(m,n)}(a_k+b_k)t^k$$

(pri čemu uzimamo da je $a_k=0$ za $k>n$ i $b_k=0$ za $k>m$),

$$p(t)\cdot q(t)=\sum_{k=0}^{m+n}\left(\sum_{i+j=k}a_ib_j\right)t^k.$$

Skup $R[t]$ uz ove dvije operacije čini prsten. Slično se definira prsten polinoma više varijabli.

Primjer 2.21. *Dokažimo da u prstenu s jedinicom komutativnost zbrajanja slijedi iz ostalih aksioma prstena.*

Rješenje: S jedne strane je

$$(a+1)(b+1) = (a+1)\cdot b + (a+1)\cdot 1 = ab + b + a + 1,$$

a s druge

$$(a+1)(b+1) = a\cdot(b+1) + 1\cdot(b+1) = ab + a + b + 1.$$

Stoga je $b + a = a + b$. ♢

Primjer 2.22. *Dokazati: Ako svi elementi komutativnog prstena P imaju zajednički djelitelj d, onda P ima jedinicu.*

Rješenje: Po pretpostavci, za svaki $x \in P$ postoji $y \in P$ takav da je $x = dy$. Posebno, za $d \in P$ postoji $e \in P$ takav da je $d = de$. Tvrdimo da je e jedinica. Zaista,

$$xe = (ay)e = a(ye) = a(ey) = (ae)y = ay = x,$$

pa je $ex = xe = x$, što se i tvrdilo. ♢

Definicija 2.15. *Neka su $(R, +, \cdot)$ i $(P, +, \cdot)$ dva prstena. Preslikavanje $f : R \to P$ za koje vrijedi $f(x+y) = f(x) + f(y)$ i $f(x\cdot y) = f(x)\cdot f(y)$ zovemo* homomorfizam prstena.

Definicija 2.16. *Neka je $(R, +, \cdot)$ prsten, te $P \subseteq R$. Ako je P prsten obzirom na operacije iz R, onda kažemo da je P* potprsten *od R.* Ideal *I u prstenu R je potprsten sa svojstvom da za svaki $x \in R$ vrijedi $xI \subseteq I$ i $Ix \subseteq I$.*

Neka je $I \subseteq R$ ideal. Tada je $(I, +)$ podgrupa od $(R, +)$, a kako je grupa $(R, +)$ abelova, to je I normalna podrupa. Stoga je dobro definirana kvocijenta grupa R/I, koja je abelova. Elementi su joj klase $x + I$, $x \in R$, a zbrajanje je dano sa

$$(x + I) + (x' + I) = x + x' + I.$$

U R/I se može uvesti i množenje sa

$$(x + I)\cdot(x' + I) = x\cdot x' + I, \quad \text{tj. } [x]\cdot[x'] = [x\cdot x'].$$

Dokažimo da je množenje dobro definirano. Neka je $x_1 \sim x$, $x_1' \sim x'$, tj. postoje $y, y' \in I$ takvi da je $x_1 = x + y$, $x_1' = x' + y'$. Tada je

$$x_1\cdot x_1' = (x+y)\cdot(x'+y') = xx' + yx' + xy' + yy' \in xx' + I$$

(jer je $yx' \in Ix' \subseteq I$, $xy' \in xI \subseteq I$, $yy' \in I$), tj. $x_1x_1' \sim xx'$. Lako se provjeri da je R/I uz ove dvije operacije prsten. Zovemo ga *kvocijentni prsten* od R po idealu I (uočimo da nije dovoljno da I bude samo potprsten).

Primjer 2.23. $\mathbb{Z}_m = \mathbb{Z}/m\mathbb{Z}$

Recimo ovdje nešto o implementaciji operacija u prstenu $\mathbb{Z}_m$. To pitanje je važno za primjene jer je u većini kriptosustava s javnim ključem, šifriranje i dešifriranje je opisano pomoću operacija u prstenu $\mathbb{Z}_m$, za neki veliki prirodni broj m. Zbrajanje u $\mathbb{Z}_m$ je vrlo jednostavno. Naime, za $x, y \in \mathbb{Z}_m$, shvatimo x i y kao nenegativne cijele brojeve manje od n, i tada je

$$x +_m y = \begin{cases} x+y & \text{ako je } x+y < m, \\ x+y-m & \text{ako je } x+y \geq m. \end{cases}$$

S druge strane, množenje u $\mathbb{Z}_m$ nije tako jednostavno. Posebno, ono je bitno kompliciranije od običnog množenja prirodnih brojeva, zato što pored množenja uključuje i (netrivijalnu) modularnu redukciju. Direktna metoda za računanje produkta $x \cdot_m y$ u $\mathbb{Z}_m$ je da izračunamo najprije $x \cdot y$, a potom izračunamo ostatak r pri djeljenju $x \cdot y$ s m. Tada je $x \cdot_m y = r$.

Postoji nekoliko poboljšanja ove metode. Opisat ćemo *Montgomeryjevu redukciju* (iz 1985. godine) čija je glavna ideja izbjegavanje klasične modularne redukcije, tj. dijeljenja. Neka su m, R i T prirodni brojevi takvi da je $R > m$, $(m, R) = 1$ i $0 \leq T < mR$. Ako je m prikazan u bazi b i ima u tom prikazu n znamenaka, onda se obično uzima $R = b^n$. Pokazat ćemo da se $TR^{-1} \bmod m$ može izračunati bez klasičnog dijeljenja. Preciznije, dijeljenje s m zamjenjuje se puno jednostavnijim dijeljenjem s R, koje je zapravo (u slučaju $R = b^n$) jednostavni pomak za n znamenaka.

Lema 2.15. *Neka je* $m' = -m^{-1} \bmod R$, *te* $U = Tm' \bmod R$. *Tada je* $V = (T + Um)/R$ *cijeli broj i* $V \equiv TR^{-1} \pmod{m}$. *Nadalje,* $TR^{-1} \bmod m = V$ *ili* $TR^{-1} \bmod m = V - m$.

Dokaz: Iz definicije brojeva m' i U slijedi da postoje $k, l \in \mathbb{Z}_m$ takvi da je $mm' = -1 + kR$, $U = Tm' + lR$. Sada je

$$\frac{T + Um}{R} = \frac{T + Tmm' + lRm}{R} = \frac{T + T(-1 + kR) + lRm}{R} = kT + lm \in \mathbb{Z}.$$

Očito je $V \equiv (T + Um)R^{-1} \equiv TR^{-1} \pmod{m}$. Konačno, iz $T < mR$ i $U < R$ slijedi $0 \leq V < (mR + mR)/R = 2m$, pa iz $V \equiv TR^{-1} \pmod{m}$ slijedi $V - (TR^{-1} \bmod m) = 0$ ili m. □

Izraz $TR^{-1} \bmod m$ naziva se *Montgomeryjeva redukcija* od T modulo m u odnosu na R, dok se $xR \bmod m$ naziva *Montgomeryjev prikaz* od x. *Montgomeryjev produkt* brojeva x i y je broj $\mathrm{Mont}(x, y) = xyR^{-1} \bmod m$. Ovo je dobro definirano, jer je $xy < m^2 < mR$. Vrijedi:

$$\mathrm{Mont}(xR \bmod m, yR \bmod m) = (xR)(yR)R^{-1} = xyR \bmod m.$$

Dakle, za brojeve u Montgomeryjevom prikazu modularno se množenje može provesti bez modularne redukcije modulo m. Naravno, modularnu redukciju

trebamo da bismo uopće dobili Montgomeryjev prikaz. No, ukoliko više puta koristimo jedan te isti broj, kao što je slučaj kod potenciranja, Montgomeryjeva metoda je znatno efikasnija od obične modularne redukcije.

Primjer ideala $m\mathbb{Z}$ u prstenu $\mathbb{Z}$ se može poopćiti na prirodan način. Ako je R komutativni prsten s jedinicom i $a \in R$, onda je skup $Ra = \{ra : r \in R\}$ ideal u R. To je ujedno i najmanji ideal u R koji sadrži a. Za ideal Ra kažemo da je *generiran* elementom a. Zovemo ga *glavnim idealom* u R i označavamo sa (a).

Definicija 2.17. *Komutativni prsten s jedinicom u kome je svaki ideal glavni ideal, zove se* prsten glavnih ideala*.*

Definicija 2.18. *Neka je R komutativni prsten s jedinicom, te $a_1, a_2, \ldots, a_n \in R$. Sa $(a_1, a_2, \ldots, a_n)$ označavamo najmanji ideal u R koji sadrži elemente $a_1, a_2, \ldots, a_n$. Kažemo da je taj ideal generiran elementima $a_1, \ldots, a_n$.*

Propozicija 2.16.

$$(a_1, a_2, \ldots, a_n) = \{\sum_{i=1}^{n} x_i a_i : x_i \in R\}$$

Dokaz: Označimo skup na desnoj strani s J. J je očito ideal i sadrži $a_1, \ldots, a_n$, pa je $(a_1, \ldots, a_n) \subseteq J$.

Za svaki $x \in R$ je $xa_i \in (a_1, \ldots, a_n)$. Budući da je $(a_1, \ldots, a_n)$ podgrupa, slijedi da je $J \subseteq (a_1, \ldots, a_n)$. □

Primjer 2.24. Neka je $I \subseteq \mathbb{Z}$ ideal, $I \neq \{0\}$, $I \neq \mathbb{Z}$. Neka je $d = \min\{a \in I : a > 0\}$. Po teoremu o dijeljenju s ostatkom, za svaki $n \in \mathbb{Z}$ postoje $q, r \in \mathbb{Z}$, $0 \leq r < d$, takvi da je $n = dq + r$. Ako je $n \in I$, onda je zbog $dq \in I$ i $r \in I$. Ali $r < d$ i $r \in I$ povlači da je $r = 0$. Dakle, svaki $n \in I$ je oblika $n = q \cdot d$, tj. $I = (d)$, pa je $\mathbb{Z}$ prsten glavnih ideala.

Propozicija 2.17. *U prstenu $\mathbb{Z}$ vrijedi $(m, n) = (d)$, gdje je $d = \mathrm{nzd}(m, n)$.*

Dokaz: Neka je $\delta > 0$ takav da je $(m, n) = (\delta)$. Budući da je $(m) \subseteq (m, n) = (\delta)$, to za svaki k postoji l takav da je $km = l\delta$. Specijalno, za $k = 1$ postoji l takav da je $m = l\delta$. Dakle, $\delta | m$. Analogno se pokazuje da $\delta | n$. Stoga $\delta | d$.

Za sve $x, y \in \mathbb{Z}$ postoji $z \in \mathbb{Z}$ takav da je $xm + yn = zd$. Odavde slijedi $(\delta) = (m, n) \subseteq (d)$. Dakle, za svaki k postoji l takav da je $k\delta = ld$. Specijalno, za $k = 1$ postoji l takav da je $\delta = ld$. Stoga $d | \delta$. Budući da $\delta | d$, $d | \delta$ i $d, \delta > 0$, zaključujemo da je $\delta = d$. □

Definicija 2.19. *Neka je R prsten, te $x \in R \setminus \{0\}$. Ako postoji $y \in R \setminus \{0\}$ takav da je $x \cdot y = 0$, onda se kaže da je x* djelitelj nule*. Integralna domena je prsten s jedinicom u kome nema djelitelja nule.*

Teorem 2.18. *$\mathbb{Z}_m$ je integralna domena ako i samo ako je m prost broj.*

Dokaz: Pretpostavimo da m nije prost, tj. da postoje a, b takvi da je $m = ab$ i $1 < a, b < m$. Tada je $[a] \cdot [b] = [ab] = [m] = [0]$, tj. $[a]$ i $[b]$ su djelitelji nule, pa $\mathbb{Z}_m$ nije integralna domena.

Neka je sada m prost. Pretpostavimo da su $[a], [b] \in \mathbb{Z}_m$ takvi da je

$$[a] \cdot [b] = [0], \quad [a] \neq [0], \ [b] \neq [0].$$

No, $[a] \cdot [b] = [ab] = [0]$ povlači da je ab djeljivo s m, što je nemoguće budući da ni a ni b nisu djeljivi s m. Dakle, $\mathbb{Z}_m$ je integralna domena. □

Definicija 2.20. *Neka je $(R, +, \cdot)$ prsten s jedinicom. Za element $a \in R$ kažemo da je* invertibilan *ako postoji $b \in R$ takav da je $a \cdot b = b \cdot a = 1$. Oznaka: $b = a^{-1}$. Neka je R^* skup svih invertibilnih elemenata u R. Tada je $(R^*, \cdot)$ grupa koju nazivamo* grupa jedinica *prstena R.*

Definicija 2.21. *Prsten R s jedinicom u kome je svaki element $x \neq 0$ invertibilan (tj. $(R \setminus \{0\}, \cdot)$ je grupa) zove se* tijelo. *Komutativno tijelo zove se* polje.

Primjer 2.25. *$(\mathbb{Q}, +, \cdot)$, $(\mathbb{R}, +, \cdot)$, $(\mathbb{C}, +, \cdot)$ su polja.*

Propozicija 2.19. *Sljedeća svojstva su invarijantna u odnosu na izomorfizam prstena:*

a) komutativnost;

b) posjedovanje jedinice;

c) biti integralna domena;

d) biti tijelo;

e) biti polje.

Dokaz: Neka je $f : (P_1, +, \cdot) \to (P_2, +, \cdot)$ izomorfizam, te neka P_1 ima promatrano svojstvo.

a) Za $a, b \in P_2$ postoje $x, y \in P_1$ takvi da je $f(x) = a$, $f(y) = b$. Sada je

$$ab = f(x)f(y) = f(xy) = f(yx) = f(y)f(x) = ba.$$

b) Neka je e_1 jedinica u P_1. Tada je $e_2 = f(e_1)$ jedinica u P_2. Zaista,

$$ae_2 = f(x)f(e_1) = f(xe_1) = f(x) = a,$$
$$e_2a = f(e_1)f(x) = f(e_1x) = f(x) = a.$$

c) Pretpostavimo da je $ab = 0$ za $a, b \in P_2$. Iz $0 = f(x)f(y) = f(xy)$ i bijektivnosti funkcije f slijedi da je $xy = 0$, pa je $x = 0$ ili $y = 0$. Propozicija 2.4 sada povlači da je $f(x) = 0$ ili $f(y) = 0$, tj. $a = 0$ ili $b = 0$.

d) Neka je $a = f(x) \in P_2$. Tvrdimo da je $a^{-1} = f(x^{-1})$. Zaista,

$$aa^{-1} = f(x)f(x^{-1}) = f(xx^{-1}) = f(e_1) = e_2,$$
$$a^{-1}a = f(x^{-1})f(x) = f(x^{-1}x) = f(e_1) = e_2.$$

e) Slijedi iz a) i d).

□

Primjer 2.26. *Pokažimo da sve matrice iz* $M_2(\mathbb{R})$ *oblika* $\begin{bmatrix} a & -b \\ b & a \end{bmatrix}$ *čine polje izomorfno polju* $(\mathbb{C}, +, \cdot)$.

Rješenje: Neka je $\mathcal{C} = \left\{ \begin{bmatrix} a & -b \\ b & a \end{bmatrix} : a, b \in \mathbb{R} \right\}$. Lako se provjeri da je $\mathcal{C}$ prsten. Treba provjeriti da je $\mathcal{C}$ zatvoren na razlike i produkte, tj. da je

$$\begin{bmatrix} a & -b \\ b & a \end{bmatrix} - \begin{bmatrix} c & -d \\ d & b \end{bmatrix} = \begin{bmatrix} a-c & -(b-d) \\ b-d & a-c \end{bmatrix} \in \mathcal{C},$$

$$\begin{bmatrix} a & -b \\ b & a \end{bmatrix} \cdot \begin{bmatrix} c & -d \\ d & b \end{bmatrix} = \begin{bmatrix} ac-bd & -(bc+ad) \\ bc+ad & ac-bd \end{bmatrix} \in \mathcal{C}.$$

Definirajmo $f : \mathbb{C} \to \mathcal{C}$ sa $f(a+ib) = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$. Očito je f bijekcija. Dokažimo da je homomorfizam prstena:

$$\begin{aligned} f((a+ib)+(c+id)) &= f(a+c+i(b+d)) = \begin{bmatrix} a+c & -b-d \\ b+d & a+c \end{bmatrix} \\ &= \begin{bmatrix} a & -b \\ b & a \end{bmatrix} + \begin{bmatrix} c & -d \\ d & c \end{bmatrix} = f(a+ib) + f(c+id), \end{aligned}$$

$$\begin{aligned} f((a+ib)\cdot(c+id)) &= f(ac-bd+i(bc+ad)) = \begin{bmatrix} ac-bd & -bc-ad \\ bc+ad & ac-bd \end{bmatrix} \\ &= \begin{bmatrix} a & -b \\ b & a \end{bmatrix} \cdot \begin{bmatrix} c & -d \\ d & c \end{bmatrix} = f(a+ib) \cdot f(c+id). \end{aligned}$$

□

Primjer 2.27. *Dokažimo da brojevi oblika* $a + b\sqrt{2}$, $a, b \in \mathbb{Q}$, *čine polje.*

Rješenje: Označimo sa $P = \{a + b\sqrt{2} : a, b \in \mathbb{Q}\}$, te uzmimo $a + b\sqrt{2}$, $c + d\sqrt{2} \in P$. Iz

$$(a + b\sqrt{2}) - (c + d\sqrt{2}) = (a - c) + (b - d)\sqrt{2} \in P$$

slijedi da je $(P, +)$ abelova grupa.

Pretpostavimo sada da je $c + d\sqrt{2} \neq 0$. Tada je

$$(a + b\sqrt{2}) \cdot (c + d\sqrt{2})^{-1} = \frac{a + b\sqrt{2}}{c + d\sqrt{2}} \cdot \frac{c - d\sqrt{2}}{c - d\sqrt{2}} = \frac{ac - 2bd}{c^2 - 2d^2} + \frac{bc - ad}{c^2 - 2d^2}\sqrt{2} \in P$$

(jer je $c^2 - 2d^2 \neq 0$), pa je $(P \setminus \{0\}, \cdot)$ abelova grupa i $(P, +, \cdot)$ polje. □

Primjer 2.28. *Čine li brojevi oblika $a + b\sqrt[3]{2}$, $a, b \in \mathbb{Q}$, polje?*

Rješenje: Ne, jer $\sqrt[3]{2} \cdot \sqrt[3]{2} = \sqrt[3]{4}$ ne leži u tom skupu. Zaista, ako je $\sqrt[3]{4} = a + b\sqrt[3]{2}$, onda kubiranjem i sređivanjem dobivamo

$$\sqrt[3]{2}(3a^2b + 3ab^3) = 4 - a^3 - 3a^2b^2 - 2b^3.$$

Budući da $\sqrt[3]{2}$ nije racionalan broj, to mora biti $3a^2b + 3ab^3 = 3ab(a + b^2) = 0$. Imamo tri mogućnosti:

1) $a = 0 \quad \Rightarrow \quad \sqrt[3]{4} = b\sqrt[3]{2} \quad \Rightarrow \quad \sqrt[3]{2} = b \in \mathbb{Q}$, kontradikcija;

2) $b = 0 \quad \Rightarrow \quad \sqrt[3]{4} = a \in \mathbb{Q}$, kontradikcija;

3) $a + b^2 = 0 \quad \Rightarrow b^2 - b\sqrt[3]{2} + \sqrt[3]{4} = 0$, što nema realnih rješenja.

□

Napomena 2.8. *a)* U svakom tijelu (pa prema tome i u polju) jednadžbe $ax = b$ i $ya = b$, za $a \neq 0$, imaju uvijek jedinstveno rješenje.

b) Svako tijelo je integralna domena. Zaista, $ax = 0$, $a \neq 0$ povlači da je $x = 0$, zbog jedinstvenosti rješenja jednadžbe.

c) U tijelu nema pravih ideala (ideala različitih od $\{0\}$ i R). Zaista, neka je $I \neq \{0\}$ ideal i neka je $a \in I \setminus \{0\}$. Tada zbog $xI \subseteq I$ za svaki $x \in R$, mora biti i $a^{-1}a = 1 \in I$. Ali tada je za svaki $x \in R$, $x = x \cdot 1 \in I$, pa je $I = R$.

Primjer 2.29. *Primjer tijela koje nije polje:* kvaternioni $\mathbb{H}$ *(hiperkompleksni brojevi).*

Kao skup i kao abelova grupa $\mathbb{H}$ je isto što i $\mathbb{R}^4$:

$$(t, x, y, z) + (t', x', y', z') = (t + t', x + x', y + y', z + z').$$

Uobičajene oznake: $1 = (1, 0, 0, 0)$, $i = (0, 1, 0, 0)$, $j = (0, 0, 1, 0)$, $k = (0, 0, 0, 1)$, $(t, x, y, z) = t + xi + yj + zk$. Dovoljno je definirati tablicu množenja za $1, i, j, k$:

	1	i	j	k
1	1	i	j	k
i	i	-1	k	$-j$
j	j	$-k$	-1	i
k	k	$-j$	$-i$	-1

Očito je da množenje nije komutativno: $i \cdot j \neq j \cdot i$. Lako se provjeri da je inverz dan sa

$$(t + xi + yj + zk)^{-1} = \frac{t - xi - yj - zk}{t^2 + x^2 + y^2 + z^2}.$$

Dakle, $(\mathbb{H}, +, \cdot)$ je tijelo, ali nije polje. $\diamondsuit$

Teorem 2.20. *Svaka konačna komutativna integralna domena je polje.*

Dokaz: Neka je F konačna komutativna integralna domena. Dovoljno je dokazati da svaki $a \in F$, $a \neq 0$, ima inverz. Definirajmo $f : F \to F$ s $f(x) = ax$. Pokažimo da je f injekcija. Neka je $f(x) = f(x')$. To znači da je $ax = ax'$, tj. $a(x - x') = 0$. Budući da F nema djelitelja nule, zaključujemo da je $x - x' = 0$, tj. $x = x'$. No, svaka injekcija s konačnog skupa u samog sebe je i bijekcija. Dakle, postoji $x \in F$ takav da je $f(x) = ax = 1$. Zbog komutativnosti je i $xa = 1$, pa je $x = a^{-1}$. $\square$

Korolar 2.21. *Ako je p prost, onda je $\mathbb{Z}_p$ polje.*

Definicija 2.22. Karakteristika polja F *je najmanji prirodni broj n takav da je*

$$n \cdot 1 = \underbrace{1 + 1 + \cdots + 1}_{n \text{ jedinica}} = 0.$$

Ako takav n ne postoji, onda se kaže da je F polje karakteristike 0.

Napomena 2.9.

1) Karakteristika polja je uvijek prost broj (ukoliko je $\neq 0$). Zaista, pretpostavimo da je $n = a \cdot b$, $1 < a, b < n$, minimalni broj za koji je $n \cdot 1 = 0$. Tada je $n \cdot 1 = (a \cdot 1) \cdot (b \cdot 1) = 0$. Odavde je $a \cdot 1 = 0$ ili $b \cdot 1 = 0$, što je u suprotnosti s minimalnošću od n.

2) Neka je F polje. Svako potpolje od F mora sadržavati 1, pa dakle postoji minimalno potpolje od F (ono generirano skupom $\{1\}$). Ukoliko je F polje karakretistike $p \neq 0$, onda je to najmanje polje izomorfno polju $\mathbb{Z}_p$, a ako je F polje karakteristike 0, onda je to polje izomorfno polju $\mathbb{Q}$.

2.3 Konačna polja

Konačno polje s q elemenata označavat ćemo s $\mathbb{F}_q$ (koristi se još i oznaka $GF(q)$ koja dolazi od "Galoisovog polja"). Konačno polje ne može biti karakteristike 0, stoga neka je p karakteristika od $\mathbb{F}_q$. Tada $\mathbb{F}_q$ sadrži prosto polje $\mathbb{Z}_p = \mathbb{Z}/p\mathbb{Z}$ (označavat ćemo ga i sa $\mathbb{F}_p$). Nadalje, $\mathbb{F}_q$ je konačno dimenzionalan vektorski prostor nad $\mathbb{F}_p$. Neka je n njegova dimenzija, a $\{e_1, \ldots, e_n\}$ baza. Tada se svaki element $a \in \mathbb{F}_q$ može na jednoznačan način prikazati u obliku linearne kombinacije

$$a = \lambda_1 e_1 + \cdots \lambda_n e_n,$$

gdje su $\lambda_i \in \mathbb{Z}_p$. Na taj način svakom $a \in \mathbb{F}_q$ možemo bijektivno pridružiti uređenu n-torku $(\lambda_1, \ldots, \lambda_n) \in (\mathbb{Z}_p)^n$. Stoga je $q = p^n$.

Pokazat ćemo da vrijedi i obrat: za svaku potenciju prostog broja $q = p^n$ postoji polje od q elemenata, i ono je jedinstveno do na izomorfizam.

Elementi polja $\mathbb{F}_q$ različiti od nule tvore abelovu grupu s obzirom na množenje. Tu grupu označavamo sa $\mathbb{F}_q^*$. Iz Lagrangeovog teorema slijedi da red svakog elementa $a \in \mathbb{F}_q^*$ dijeli $q-1$. Analogno Teoremu 1.31 o postojanju primitivnih korijena modulo p, dokazuje se sljedeći teorem.

Teorem 2.22. *Grupa $\mathbb{F}_q^*$ je ciklička. Ako je g generator od $\mathbb{F}_q^*$, onda je g^j također generator ako i samo ako je $\gcd(j, q-1) = 1$. Stoga postoji točno $\varphi(q-1)$ generatora grupe $\mathbb{F}_q^*$.*

Teorem 2.23. *Ako je $\mathbb{F}_q$ polje s $q = p^n$ elemenata, onda svaki element tog polja zadovoljava jednadžbu $X^q - X = 0$, i $\mathbb{F}_q$ je upravo skup svih korijena ove jednadžbe. Obrnuto, za svaku prostu potenciju $q = p^n$, polje razlaganja polinoma $X^q - X$ nad $\mathbb{F}_p$ je polje s q elemenata.*

Dokaz: Neka je $\mathbb{F}_q$ konačno polje. Budući da red svakog nenul elementa od $\mathbb{F}_q$ dijeli $q-1$, slijedi da svi nenul elementi zadovoljavaju jednadžbu $X^{q-1} = 1$. Množeći obje strane sa X, dobivamo da nenul elementi zadovoljavaju jednadžbu $X^q - X = 0$. Očito i element 0 također zadovoljava ovu jednadžbu. Dakle, svih q elemenata od $\mathbb{F}_q$ su korijeni polinoma $X^q - X$, koji je stupnja q. Budući da polinom q-tog stupnja ne može imati više od q korijena, zaključujemo da su njegovi korijeni upravo elementi polja $\mathbb{F}_q$. To znači da je $\mathbb{F}_q$ polje razlaganja polinoma $X^q - X$, tj. najmanje proširenje polja $\mathbb{F}_p$ koje sadrži sve njegove korijene.

Obrnuto, neka je $q = p^n$ potencija prostog broja i neka je $\mathbb{F}$ polje razlaganja nad $\mathbb{F}_p$ polinoma $X^q - X$. Derivacija ovog polinoma je $qX^{q-1} - 1 = -1$ (jer je q višekratnik od p i zato jednak nuli u polju $\mathbb{F}_p$). Odavde slijedi da polinom $X^q - X$ nema višestrukih korijena (jer bi oni morali biti korijeni i od derivacije). Dakle, polje $\mathbb{F}$ mora sadržavati barem q različitih

korijena od $X^q - X$. Ali skup od q korijena već čini polje. Treba provjeriti da je suma i produkt dva korijena ponovo korijen. Zaista, ako su a i b korijeni polinoma $X^q - X$, onda je $a^q = a$ i $b^q = b$. Odavde je $(ab)^q = a^q b^q = ab$, pa je ab također korijen. Iz binomnog poučka slijedi da je $(a+b)^p = a^p + b^p + \sum_{i=1}^{p-1} \binom{p}{i} a^{p-i} b^i = a^p + b^p$ u svakom polju karakteristike p. Sada je $a^{p^2} + b^{p^2} = (a^p + b^p)^p = (a+b)^{p^2}$, te induktivno dobivamo da je $a^q + b^q = (a+b)^q$, što znači da je i $a+b$ također korijen od $X^q - X$. Zaključujemo da je skup od q korijena najmanje polje koje sadrži korijene od $X^q - X$, tj. polje razlaganja ovog polinoma je konačno polje s q elemenata. □

U dokazu prethodnog teorema pokazali smo da potenciranje na p-tu potenciju čuva zbrajanje i množenje. Pokazat ćemo još neka važna svojstva tog preslikavanja.

Propozicija 2.24. *Neka je $\mathbb{F}_q$ konačno polje s $q = p^n$ elemenata i neka je $\sigma(a) = a^p$. Tada je σ automorfizam polja $\mathbb{F}_q$ koji zovemo* Frobeniusov automorfizam. *Elementi od $\mathbb{F}_q$ koji su fiksni na djelovanje od σ su točno elementi prostog polja $\mathbb{F}_p$.*

Dokaz: Pokazali smo već da preslikavanje σ čuva zbrajanje i množenje, što znači da je homomorfizam. Elementi koji ostaju fiksni na djelovanje od σ^j su korijeni od $X^{p^j} - X$. Ako je $j = 1$, to su točno p elemenata prostog polja $\mathbb{F}_p$ (po Teoremu 2.23 za $n = 1$). Ako je $j = n$, dobivamo da su fiksni elementi od σ^n svi elementi polja $\mathbb{F}_q$, tj. da je σ^n identiteta. No, to povlači da je σ bijekcija (inverzno preslikavanje je $a \mapsto a^{p^{n-1}}$). □

Napomena 2.10. *Za $\alpha \in \mathbb{F}_q$, svi konjugati od α (elementi od $\mathbb{F}_q$ koji zadovoljavaju isti normirani ireducibilni polinom s koeficijentima u $\mathbb{F}_p$) su elementi $\sigma^j(\alpha) = \alpha^{p^j}$.*

Postavlja se pitanje kako efektivno realizirati konačno polje s p^n elemenata, te operacije na njemu. Do sada smo takvo nešto pokazali samo za polja $\mathbb{F}_p = \mathbb{Z}_p$.

Primjer 2.30. *Polje $\mathbb{F}_9$.*

Da bi konstruirali polje $\mathbb{F}_9$, uzimamo neki normirani (vodeći koeficijent je 1) ireducibilni (koji se ne može rastaviti na faktore) kvadratni polinom u $\mathbb{F}_3[t]$. Dakle, traži se polinom oblika $t^2 + at + b$, $a, b \in \mathbb{F}_3$, koji nema nultočaka u $\mathbb{F}_3$. Jasno je da mora biti $b \neq 0$. Polinomi $t^2 + 2 = t^2 - 1 = (t-1)(t+1)$, $t^2 \pm 2t + 1 = (t \pm 1)^2$ su očito reducibilni. Tako da konačno dobivamo točno tri ireducibilna polinoma:

$$t^2 + 1, \quad t^2 + t + 2, \quad t^2 + 2t + 2.$$

Sada $\mathbb{F}_9$ možemo realizirati kao $\mathbb{Z}_3[t]/(g(t))$, gdje je $g(t)$ bilo koji od ova tri ireducibilna polinoma. To znači da su elementi od $\mathbb{F}_9$ polinomi u $\mathbb{Z}_3[t]$

stupnja ≤ 1, te da ako dobijemo da je produkt dva elementa stupnja većeg od 1, onda ga zamjenjujemo njegovim ostatkom pri dijeljenju s $g(t)$.

Uzmimo da je $g(t) = t^2 + 1$. Svi elementi od $\mathbb{F}_9$:

$$0,\ 1,\ 2,\ t,\ t+1,\ t+2,\ 2t,\ 2t+1,\ 2t+2.$$

Izračunajmo potencije elementa $a = t + 1$:

$$\begin{gathered} a^0 = 1,\ a^1 = t+1,\ a^2 = t^2 + 2t + 1 = 2t, \\ a^3 = 2t^2 + 2t = 2t - 2 = 2t + 1,\ a^4 = 2t^2 + 3t + 1 = 2t^2 + 1 = -1 = 2, \\ a^5 = 2t + 2,\ a^6 = 2t^2 + 4t + 2 = 2t^2 + t + 2 = t, \\ a^7 = t^2 + t = t - 1 = t + 2,\ a^8 = t^2 + 3t + 2 = 1. \end{gathered}$$

Vidimo da je a generator cikličke grupe $\mathbb{F}_9^*$. ♢

U općem slučaju, polje $\mathbb{F}_q$ za $q = p^n$ realiziramo kao kvocijentni prsten $\mathbb{Z}_p[t]/(g(t))$, gdje je $g(t)$ neki normirani ireducibilni polinom stupnja n u $\mathbb{Z}_p[t]$, a $(g(t))$ označava glavni ideal generiran s $g(t)$ (ovaj prsten je polje zbog toga što je $g(t)$ ireducibilan). Elemente ovog polja se može prikazati kao polinome nad $\mathbb{Z}_p$ stupnja $\leq k - 1$, dok su pripadne operacije zbrajanje i množenje polinoma u $\mathbb{Z}_p[t]$, s time da se nakon množenja računa ostatak pri dijeljenju s polinomom $g(t)$.

Uočimo da su $\mathbb{F}_{p^k}$ i $\mathbb{Z}_{p^k}$ za $k \geq 2$ bitno različite strukture. U $\mathbb{F}_{p^k}$ su svi ne-nul elementi invertibilni, dok u $\mathbb{Z}_{p^k}$ ima točno $\varphi(p^k) = p^k - p^{k-1}$ invertibilnih elemenata.

Na ovom mjestu se možemo pitati kako naći ireducibilni polinom stupnja n nad $\mathbb{Z}_p$ (i imali li uopće takvih polinoma). Pokazuje se da normiranih ireducibilnih polinoma stupnja n nad $\mathbb{Z}_p$ ima približno p^n/n, tj. otprilike svaki n-ti normirani polinom stupnja n nad $\mathbb{Z}_p$ je ireducibilan. Npr. ako je n prost broj, onda postoji točno $\frac{p^n - p}{n}$ različitih normiranih ireducibilnih polinoma stupnja n u $\mathbb{Z}_p[t]$. Testiranje je li konkretni polinom ireducibilan zasniva se na činjenici da je polinom $g(t)$ stupnja n nad $\mathbb{Z}_p$ ireducibilan ako i samo ako je $\mathrm{nzd}(g(t), t^{p^j} - t) = 1$ za $j = 1, 2, \ldots \lfloor n/2 \rfloor$. Posljednji uvjet se provjerava Euklidovim algoritmom za polinome. Da bi operacije u polju $\mathbb{F}_q$ bile što efikasnije, obično se polinom $g(t)$ bira tako da ima što manju težinu W (broj koeficijenata različitih od 0). U slučaju $q = 2^n$, koji je najzanimljiviji za primjene u kriptografiji, čini se da je uvijek moguće postići da je $W = 3$ ili $W = 5$.

Primjer 2.31. *Pomoću realizicije iz Primjera 2.30, provjerimo da svi elementi od* $\mathbb{F}_9$ *zadovoljavaju jednadžbu* $X^9 - X = 0$.

Rješenje: Rastavimo $X^9 - X$ na produkt ireducibilnih polinoma nad $\mathbb{Z}_3$:

$$\begin{aligned} X^9 - X &= X(X^4 - 1)(X^4 + 1) = X(X+1)(X-1)(X^2+1)(X^4+1) \\ &= X(X+1)(X+2)(X^2+1)(X^2+X+2)(X^2+2X+2). \end{aligned}$$

Uzmimo element $a = t + 1$ iz Primjera 2.30. On je korijen polinoma $X^2 + X + 2$. Zaista, $a^2 + a + 2 = t^2 + 3t + 4 = t^2 + 1 = 0$. Prema Napomeni 2.10, isti polinom zadovoljava i element $\sigma(a) = a^3 = 2t + 1$. Analogno se provjeri da su $t + 2$ i $2t + 2$ korijeni polinoma $X^2 + 2X + 2$. Jasno je da su t i $2t$ korijeni polinoma $X^2 + 1$, dok su 0, 1 i 2 korijeni linearnih polinoma X, $X + 2$ i $X + 1$. ♢

Završit ćemo ovo poglavlje s jednom primjenom konačnih polja u teoriji brojeva, tako što ćemo dati dokaz Gaussovog kvadratnog zakona reciprociteta, kojeg smo iskazali i koristili u poglavlju o kvadratnim ostacima, ali ga tamo nismo bili dokazali.

Rješenje jednadžbe $x^n = 1$ zovemo *n-ti korijen jedinice*. Ako je n najmanja potencija za koju je $x^n = 1$, onda kažemo da je x *primitivni* n-ti korijen jedinice.

Propozicija 2.25. *Neka je g generator od $\mathbb{F}_q^*$. Tada je g^j n-ti korijen iz jedinice ako i samo ako je $nj \equiv 0 \pmod{q-1}$. Posebno, $\mathbb{F}_q$ ima primitivni n-ti korijen jedinice ako i samo ako n dijeli $q - 1$.*

Dokaz: Imamo da je $g^{nj} = 1$ ako i samo ako je nj višekratnik reda od g, a to je $q - 1$. Dakle, g^j je n-ti korijen iz jedinice ako i samo ako je

$$nj \equiv 0 \pmod{q-1}. \tag{2.1}$$

Neka je $d = \text{nzd}(n, q-1)$. Tada je kongruencija (2.1) ekvivalentna sa $j \equiv 0 \pmod{\frac{q-1}{d}}$. Ako n ne dijeli $q - 1$, onda je $d < n$, pa iz $dj \equiv 0 \pmod{q-1}$, slijedi da g^j nije primitivni n-ti korijen jedinice. Ako n dijeli $q - 1$, stavimo $\xi = g^{(q-1)/n}$. Tada je $\xi^j = q^{j(q-1)/n} = 1$ ako i samo ako $n|j$, pa je ξ primitivni n-ti korijen jedinice. □

Teorem 2.26 (Gaussov kvadratni zakon reciprociteta)**.** *Ako su p i r različiti neparni prosti brojevi, onda vrijedi*

$$\left(\frac{p}{r}\right)\left(\frac{r}{p}\right) = (-1)^{\frac{p-1}{2}\cdot\frac{r-1}{2}}.$$

Drugim riječima, ako su p i r oba oblika $4k+3$, onda jedna od kongruencija $x^2 \equiv p \pmod{r}$, $x^2 \equiv r \pmod{p}$ ima rješenja, a druga nema. A ako barem jedan od brojeva p i r ima oblik $4k+1$, onda ili obje ove kongruencije imaju rješenja ili obje nemaju rješenja.

Rješenje: Neka je n prirodan broj sa svojstvom $p^n \equiv 1 \pmod{r}$. Na primjer, možemo uzeti $n = r - 1$. Tada, po Propoziciji 2.25, polje $\mathbb{F}_{p^n}$ sadrži primitivni r-ti korijen jedinice, koji označimo sa ξ. Definiramo *Gaussovu sumu*

$$G = \sum_{j=0}^{r-1} \left(\frac{j}{r}\right) \xi^j.$$

Tvrdimo da je

$$G^2 = (-1)^{(r-1)/2} r. \tag{2.2}$$

Uočimo najprije da je za svaki r-ti korijen jedinice $\xi' \neq 1$ vrijedi $S = \sum_{j=0}^{r-1} \xi'^j = 0$. Zaista, $\xi' S = S$, jer množenje sa ξ' samo permutira pribrojnike u sumi S, pa iz $(\xi' - 1)S = 0$ i $\xi' \neq 1$ slijedi $S = 0$. Nadalje, vrijednost $(\frac{j}{r})\xi^j$ ovisi samo o ostatku od j modulo r. Sada imamo:

$$\begin{aligned} G^2 &= \sum_{j=1}^{r-1} \Big(\frac{j}{r}\Big)\xi^j \sum_{k=1}^{r-1} \Big(\frac{-k}{r}\Big)\xi^{-k} = \Big(\frac{-1}{r}\Big) \sum_{j=1}^{r-1}\sum_{k=1}^{r-1} \Big(\frac{jk}{r}\Big)\xi^{j-k} \\ &= (-1)^{(r-1)/2} \sum_{j=1}^{r-1} \Big(\frac{j^2 k}{r}\Big)\xi^{j(1-k)} = (-1)^{(r-1)/2} \sum_{k=1}^{r-1} \Big(\frac{k}{r}\Big) \sum_{j=0}^{r-1} (\xi^{k-1})^j \\ &= (-1)^{(r-1)/2} \Big(\frac{1}{r}\Big) \sum_{j=0}^{r-1} \xi^0 = (-1)^{(r-1)/2} r. \end{aligned}$$

Kvadratni zakon reciprociteta ćemo dobiti tako da na dva različita načina prikažemo G^p. Najprije imamo (koristeći (2.2) i Eulerov kriterij):

$$\begin{aligned} G^p &= (G^2)^{(p-1)/2} G = (-1)^{(r-1)/2 \cdot (p-1)/2} r^{(p-1)/2} G \\ &= (-1)^{(p-1)/2 \cdot (r-1)/2} \Big(\frac{r}{p}\Big) G. \end{aligned}$$

S druge strane, koristeći činjenicu da je $(a+b)^p = a^p + b^p$ u polju $\mathbb{F}_{p^n}$, imamo:

$$\begin{aligned} G^p &= \sum_{j=0}^{r-1} \Big(\frac{j}{r}\Big)^p \xi^{pj} = \sum_{j=0}^{r-1} \Big(\frac{j}{r}\Big)\xi^{pj} \\ &= \Big(\frac{p}{r}\Big) \sum_{j=0}^{r-1} \Big(\frac{pj}{r}\Big)\xi^{pj} = \Big(\frac{p}{r}\Big) G. \end{aligned}$$

Budući da je $G \neq 0$, dijeleći dva dobivena izraza za G^p s G, dobivamo upravo kvadratni zakon reciprociteta. □

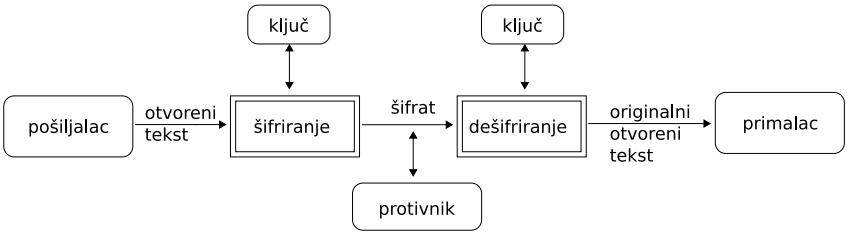
Poglavlje 3

Kriptografija

3.1 Kratki uvod u kriptografiju

Kako uspostaviti sigurnu komunikaciju preko nesigurnog komunikacijskog kanala? Metode za rješavanje ovog problema pročava znanstvena disciplina koja se zove *kriptografija* (ili *tajnopis*). Osnovni zadatak kriptografije je omogućavanje komunikacije dvaju osoba (zovemo ih *pošiljalac* i *primalac* - u kriptografskoj literaturi za njih su rezervirana imena *Alice* i *Bob*) na takav način da treća osoba (njihov *protivnik* - u literaturi se najčešće zove *Eva* ili *Oskar*), koja može nadzirati komunikacijski kanal, ne može razumjeti njihove poruke.

Poruku koju pošiljalac želi poslati primaocu zovemo *otvoreni tekst*. Pošiljalac transformira otvoreni tekst koristeći unaprijed dogovoreni *ključ* K. Taj se postupak zove *šifriranje*, a dobiveni rezultat *šifrat*. Nakon toga pošiljalac pošalje šifrat preko nekog komunikacijskog kanala. Protivnik prisluškujući može saznati sadržaj šifrata, ali kako ne zna ključ, ne može odrediti otvoreni tekst. Za razliku od njega, primalac zna ključ kojim je šifrirana poruka, pa može *dešifrirati* šifrat i odrediti otvoreni tekst.



shema simetrične kriptografije

Ove pojmove ćemo formalizirati u sljedećoj definiciji.

Definicija 3.1. Kriptosustav *je uređena petorka* $(\mathcal{P}, \mathcal{C}, \mathcal{K}, \mathcal{E}, \mathcal{D})$*, gdje je* $\mathcal{P}$ *konačan skup svih otvorenih tekstova,* $\mathcal{C}$ *konačan skup svih šifrata,* $\mathcal{K}$

konačan skup svih mogućih ključeva, $\mathcal{E}$ skup svih funkcija šifriranja i $\mathcal{D}$ skup svih funkcija dešifriranja. Za svaki $K \in \mathcal{K}$ postoji $e_k \in \mathcal{E}$ i odgovarajući $d_K \in \mathcal{D}$. Pritom su $e_K : \mathcal{P} \to \mathcal{C}$ i $d_K : \mathcal{C} \to \mathcal{P}$ funkcije sa svojstvom da je $d_K(e_K(x)) = x$ za svaki $x \in \mathcal{P}$.

Shema koju smo u uvodu opisali predstavlja tzv. *simetrični ili konvencionalni kriptosustav*. Funkcije koje se koriste za šifriranje e_K i dešifriranje d_K ovise o ključu K kojeg Alice i Bob moraju tajno razmjeniti prije same komunikacije. Kako njima nije dostupan siguran komunikacijski kanal, ovo može biti veliki problem.

Godine 1976. Diffie i Hellman su ponudili jedno moguće rješenje problema razmjene ključeva, zasnovano na činjenici da je u nekim grupama potenciranje puno jednostavnije od logaritmiranja.

Diffie i Hellman se smatraju začetnicima *kriptografije javnog ključa*. Ideja javnog ključa se sastoji u tome da se konstruiraju kriptosustavi kod kojih bi iz poznavanja funkcije šifriranja e_K bilo praktički nemoguće (u nekom razumnom vremenu) izračunati funkciju dešifriranja d_K. Tada bi funkcija e_K mogla biti javna. Dakle, u kriptosustavu s javnim ključem svaki korisnik K ima dva ključa: javni e_K i tajni d_K. Ako Alice želji poslati Bobu poruku x, onda je ona šifrira pomoću Bobovog javnog ključa e_B, tj. pošalje Bobu šifrat $y = e_B(x)$. Bob dešifrira šifrat koristeći svoj tajni ključ d_B, $d_B(y) = d_B(e_B(x)) = x$. Uočimo da Bob mora posjedovati neku dodatnu informaciju (tzv. *trapdoor* - skriveni ulaz) o funkciji e_B, da bi samo on mogao izračunati njezin inverz d_B, dok je svima drugima (a posebno Eve) to nemoguće. Takve funkcije čiji je inverz teško izračunati bez poznavanja nekog dodatnog podatka zovu se *osobne jednosmjerne funkcije*.

Napomenimo da su kriptosustavi s javnim ključem puno sporiji od modernih simetričnih kriptosustava (DES, IDEA, AES), pa se stoga u praksi ne koriste za šifriranje poruka, već za šifriranje ključeva, koji se potom koriste u komunikaciji pomoću nekog simetričnog kriptosustava.

Druga važna primjena kriptosustava s javnim ključem dolazi od toga da oni omogućavaju da se poruka "digitalno potpiše". Naime, ako Alice pošalje Bobu šifrat $z = d_A(e_B(x))$, onda Bob može biti siguran da je poruku poslala Alice (jer samo ona zna funkciju d_A), a također jednakost $e_A(z) = e_B(x)$ predstavlja i dokaz da je poruku poslala Alice, pa ona to ne može kasnije zanijekati.

3.2 Data Encryption Standard i Advanced Encryption Standard

Krajem 60-tih i početkom 70-tih godina 20. stoljeća, razvojem financijskih transakcija, kriptografija postaje zanimljiva sve većem broju potencijalnih korisnika. Dotad je glavna primjena kriptografije bila u vojne i diplomatske svrhe, pa je bilo normalno da svaka država (ili čak svaka zainteresirana državna organizacija) koristi svoju šifru za koju je vjerovala da je najbolja. No, tada se pojavila potreba za šifrom koju će moći koristiti korisnici širom svijeta, i u koju će svi oni moći imati povjerenje - dakle, pojavila se potreba uvođenja *standarda* u kriptografiji.

Godine 1972. američki *National Bureau of Standards* (NBS) inicirao je program za zaštitu računalnih i komunikacijskih podataka. Jedan od ciljeva bio je razvijanje jednog standardnog kriptosustava. Godine 1973. NBS je raspisao javni natječaj za takav kriptosustav. Najozbiljnija prijava na natječaj je bila ona koju je poslao IBM-ov tim kriptografa. Algoritam se zasnivao na *Feistelovoj šifri*. Gotovo svi simetrični blokovni algoritmi koji su danas u uporabi koriste ideju koju je uveo voditelj IBM-ovog kriptografskog odjela Horst Feistel 1973. godine. Jedna od glavnih ideja je alternirana uporaba supstitucija i transpozicija kroz više iteracija (tzv. rundi).

Predloženi je algoritam nakon nekih preinaka, u kojima je sudjelovala *National Security Agency* (NSA), prihvaćen kao standard 1976. godine i dobio je ime *Data Encryption Standard* (DES).

DES šifrira otvoreni tekst duljine 64 bita, koristeći ključ K duljine 56 bitova. Tako se dobiva šifrat koji ponovo ima 64 bita. Algoritam se sastoji od 3 etape:

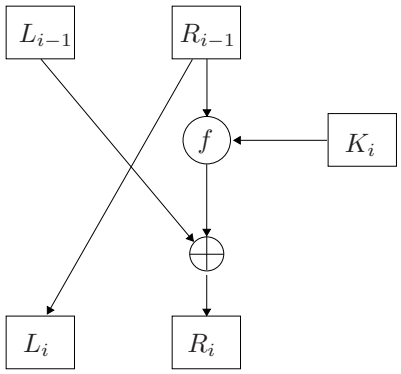
1. Za dani otvoreni tekst x, permutiranjem pomoću fiksne inicijalne permutacije IP dobije se x_0. Zapišemo $x_0 = IP(x)$ u obliku $x_0 = L_0R_0$, gdje L_0 sadrži prva (lijeva) 32 bita, a R_0 zadnja (desna) 32 bita od x_0.

2. Određena funkcija f se 16 puta iterira. Računamo L_iR_i, $i = 1, 2, \ldots, 16$, po sljedećem pravilu:

$$L_i = R_{i-1}, \quad R_i = L_{i-1} \oplus f(R_{i-1}, K_i),$$

   gdje $\oplus$ označava operaciju "ekskluzivno ili" (XOR). Funkciju f ćemo opisati malo kasnije, a $K_1, K_2, \ldots, K_{16}$ su nizovi bitova duljine 48, koji se dobivaju kao permutacije nekih bitova iz K.

3. Primijenimo inverznu permutaciju IP^{-1} na $R_{16}L_{16}$ i tako dobivamo šifrat y. Dakle, $y = IP^{-1}(R_{16}L_{16})$. Uočimo inverzni poredak od L_{16} i R_{16} u ovom zadnjem koraku.



Slika 3.1: Jedna runda DES-a

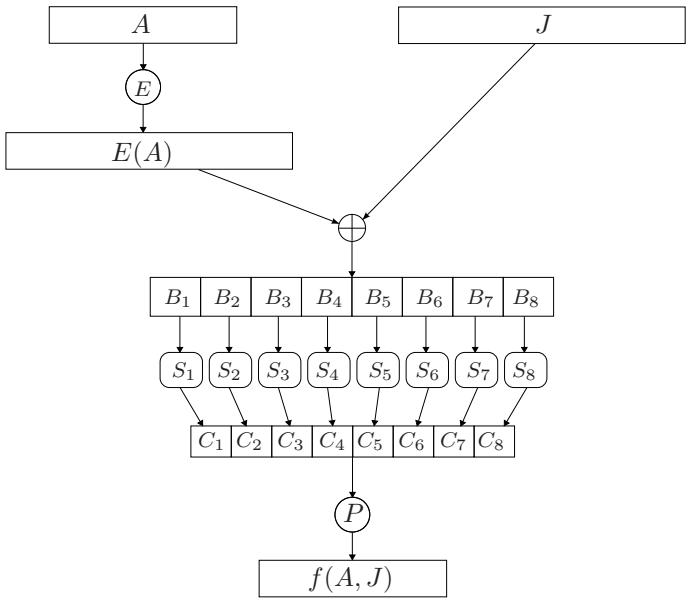
IP								**IP^{-1}**							
58	50	42	34	26	18	10	2	40	8	48	16	56	24	64	32
60	52	44	36	28	20	12	4	39	7	47	15	55	23	63	31
62	54	46	38	30	22	14	6	38	6	46	14	54	22	62	30
64	56	48	40	32	24	16	8	37	5	45	13	53	21	61	29
57	49	41	33	25	17	9	1	36	4	44	12	52	20	60	28
59	51	43	35	27	19	11	3	35	3	43	11	51	19	59	27
61	53	45	37	29	21	13	5	34	2	42	10	50	18	58	26
63	55	47	39	31	23	15	7	33	1	41	9	49	17	57	25

Funkcija f za prvi argument ima niz bitova A duljine 32, a za drugi argument ima niz bitova J duljine 48. Kao rezultat se dobiva niz bitova duljine 32.

Funkcija se računa u sljedeća 4 koraka:

1. Prvi argument A se "proširi" do niza duljine 48 u skladu s fiksnom *funkcijom proširenja* E. Niz $E(A)$ se sastoji od 32 bita iz A, permutiranih na određeni način, s time da se 16 bitova pojavi dvaput.

E					
32	1	2	3	4	5
4	5	6	7	8	9
8	9	10	11	12	13
12	13	14	15	16	17
16	17	18	19	20	21
20	21	22	23	24	25
24	25	26	27	28	29
28	29	30	31	32	1



Slika 3.2: DES funkcija f

2. Izračunamo $E(A) \oplus J$ i rezultat zapišemo kao spoj od osam 6-bitnih nizova
$$B = B_1B_2B_3B_4B_5B_6B_7B_8.$$

3. Sljedeći korak koristi 8 tzv. S-kutija (supstitucijskih kutija) $S_1, S_2, \ldots, S_8$. Svaki S_i je fiksna 4×16 matrica čiji su elementi cijeli brojevi između 0 i 15. Za dani niz bitova duljine 6, recimo $B_j = b_1b_2b_3b_4b_5b_6$, računamo $S_j(B_j)$ na sljedeći način. Dva bita b_1b_6 određuju binarni zapis retka r od S_j ($r = 0, 1, 2, 3$), a četiri bita $b_2b_3b_4b_5$ određuju binarni zapis stupca c od S_j ($c = 0, 1, 2, \ldots, 15$). Sada je $S_j(B_j)$ po definiciji jednako $S_j(r, c)$, zapisano kao binarni broj duljine 4. Na ovaj način izračunamo $C_j = S_j(B_j)$, $j = 1, 2, \ldots, 8$.

4. Niz bitova $C_1C_2C_3C_4C_5C_6C_7C_8$ duljine 32 permutira se pomoću fiksne *završne permutacije* P. Tako se dobije $P(C)$, što je po definiciji upravo $f(A, J)$.

P			
16	7	20	21
29	12	28	17
1	15	23	26
5	18	31	10
2	8	24	14
32	27	3	9
19	13	30	6
22	11	4	25

Konačno, trebamo opisati računanje tablice međuključeva $K_1, K_2, \ldots, K_{16}$ iz ključa K. Ključ K se sastoji od 64 bita, od kojih 56 predstavlja ključ, a preostalih 8 bitova služe za testiranje pariteta. Bitovi na pozicijama 8, 16, $\ldots, 64$ su definirani tako da svaki bajt (8 bitova) sadrži neparan broj jedinica. Ovi se bitovi ignoriraju kod računanja tablice ključeva.

1. Za dani 64-bitni ključ K, ignoriramo paritetne bitove, te permutiramo preostale bitove pomoću fiksne permutacije $PC1$. Zapišemo $PC1(K) = C_0D_0$, gdje C_0 sadrži prvih 28, a D_0 zadnjih 28 bitova od $PC1(K)$.

2. Za $i = 1, 2, \ldots, 16$ računamo:

$$C_i = LS_i(C_{i-1}), \quad D_i = LS_i(D_{i-1}), \quad K_i = PC2(C_iD_i).$$

   LS_i predstavlja ciklički pomak ulijevo za jednu ili dvije pozicije, u ovisnosti od i. Ako je $i = 1, 2, 9$ ili 16, onda je pomak za jednu poziciju, a inače je pomak za dvije pozicije. PC2 je fiksna "kompresijska" permutacija, pomoću koje je opisan izbor i redoslijed 48 od 56 bitova.

Ovim je u potpunosti opisan postupak šifriranja.

PC1							**PC2**					
57	49	41	33	25	17	9	14	17	11	24	1	5
1	58	50	42	34	26	18	3	28	15	6	21	10
10	2	59	51	43	35	27	23	19	12	4	26	8
19	11	3	60	52	44	36	16	7	27	20	13	2
63	55	47	39	31	23	15	41	52	31	37	47	55
7	62	54	46	38	30	22	30	40	51	45	33	48
14	6	61	53	45	37	29	44	49	39	56	34	53
21	13	5	28	20	12	4	46	42	50	36	29	32

Dešifriranje koristi isti algoritam kao šifriranje. Krenemo od šifrata y, ali koristimo tablicu ključeva u obrnutom redoslijedu: $K_{16}, K_{15}, \ldots, K_1$. Kao rezultat dobivamo otvoreni tekst x. Uvjerimo se da ovako definirana funkcija dešifriranja d_K zaista ima traženo svojstvo da je $d_K(y) = x$. Podsjetimo se da smo y dobili kao $y = IP^{-1}(R_{16}L_{16})$. Stoga se primjenom inicijalne permutacije na y dobije $y_0 = R_{16}L_{16}$. Nakon prve runde dešifriranja, lijeva polovica postaje $L_{16} = R_{15}$, a desna $R_{16} \oplus f(L_{16}, K_{16})$. No, iz zadnje runde šifriranja znamo da vrijedi

$$R_{16} = L_{15} \oplus f(R_{15}, K_{16}) = L_{15} \oplus f(L_{16}, K_{16}).$$

Zato je $R_{16} \oplus f(L_{16}, K_{16}) = L_{15}$. Znači, nakon jedne runde dešifriranja dobivamo $R_{15}L_{15}$. Nastavljajući taj postupak, nakon svake sljedeće rudne

dešifriranja dobivat ćemo redom: $R_{14}L_{14}$, $R_{13}L_{13}\ldots$, R_1L_1 i nakon zadnje runde R_0L_0. Preostaje zamijeniti poredak lijeve i desne polovice i primijeniti IP^{-1}. Dakle, na kraju postupka dešifriranja dobivamo $IP^{-1}(L_0R_0)$, a to je upravo otvoreni tekst x, što je i trebalo dokazati.

Vidimo da razlog za zamjenu lijeve i desne polovice prije primjene permutacije IP^{-1} leži upravo u želji da se za dešifriranje može koristiti isti algoritam kao za šifriranje.

Uočimo da su sve operacije u DES-u linearne (sjetimo se da je $\oplus$ zapravo zbrajanje u $\mathbb{Z}_2$), s izuzetkom S-kutija. Stoga su S-kutije izuzetno značajne za sigurnost DES-a. Od objave algoritma, pa sve do danas, S-kutije su obavijene tajnovitošću. Kod nasljednika DES-a upravo će se taj dio promijeniti: S-kutije će biti generirane eksplicitno navedenim algoritmom. Kod DES-a znamo tek neke kriterije koji su korišteni u dizajniranju S-kutija. Ti kriteriji su imali za zadatak povećati tzv. difuziju kriptosustava, tj. postići da na svaki bit šifrata utječe što više bitova otvorenog teksta. Oni su također otežavali i tzv. diferencijalnu kriptoanalizu. Slični su i razlozi zašto u DES-u imamo upravo 16 rundi. Naime, kod izbora broja rundi, ključan je zahtjev da poznati kriptoanalitički napadi ne budu efikasniji od napada "grubom silom".

Originalna IBM-ova ponuda NBS-u je imala 112-bitni ključ. Prva IBM-ova realizacija Feistelove šifre - kriptosustav LUCIFER je imao 128-bitni ključ. Međutim, u verziji DES-a koja je prihvaćena kao standard duljina ključa je smanjena na 56 bitova (da bi ključ stao na tadašnje čipove, ali vjerojatno i pod utjecajem NSA). Mnogi kriptografi su bili protiv tako kratkog ključa jer su smatrali da ne pruža dovoljnu sigurnost protiv napada "grubom silom". Uz 56-bitni ključ imamo $2^{56} \approx 7.2 \cdot 10^{16}$ mogućih ključeva, pa se na prvi pogled napad "grubom silom" čini sasvim nepraktičnim. Međutim, već su 1977. godine Diffie i Hellman ustvrdili da tadašnja tehnologija omogućava konstrukciju računala koje bi otkrivalo ključ za jedan dan, a troškove su procijenili na 20 milijuna dolara. Na osnovu toga su zaključili da je takvo što dostupno samo organizacijama kao što je NSA, ali da će oko 1990. godine DES postati sasvim nesiguran. Godine 1993. Weiner je procijenio da se za 100000 dolara može konstruirati računalo koje bi otkrilo ključ za 35 sati, a za 10 milijuna dolara ono koje bi otkrilo ključ za 20 minuta. Ipak sve su to bili hipotetski dizajni i konačno razbijanje DES-a se dogodilo tek 1998. godine. Tada je *Electronic Frontier Foundation* (EFF) je za 250000 dolara zaista napravila "DES Cracker", koji je razbijao poruke šifrirane DES-om za 56 sati.

Od početka 90-tih godina 20. stoljeća, kad je postalo jasno da je definitivno razbijanje DES-a pitanje trenutka (a i da je već sigurno dostupno organizacijama kao što je NSA), pojavilo se više prijedloga za slične, ali sigurnije kriptosustave. Većini od njih je zajedničko to da imaju veću duljinu ključa od DES-a (obično 128 bitova), te da je generiranje analogona S-kutija

eksplicitnije opisano. Među najpopularnijim takvim kripotsustavima bili su IDEA, CAST i RC5. No, i sam DES je "preživio" u obliku trostrukog DES-a.

Prije nego što kažemo nešto o trostrukom DES-u, odgovorimo na pitanje zašto se ne koristi "dvostruki DES". Kod dvostrukog DES-a bismo svaki blok šifrirali dvaput, s dva različita ključa K i L:

$$y = e_L(e_K(x)), \quad x = d_K(d_L(y)).$$

Poznato je da skup DES permutacija nije podgrupa pripadne grupe permutacija, pa se čini da smo ovako dobili znatno sigurniji kriptosustav. Međutim, postoji nešto što se naziva *napad "susret u sredini"* (engl. meet-in-the-middle), koji je opisao Diffie 1977. godine. Pretpostavimo da je poznat jedan par *otvoreni tekst - šifrat* (x, y). Šifriramo x sa svih 2^{56} mogućih ključeva K. Spremimo rezultate u tablicu i sortiramo ih po vrijednostima od $z = e_K(x)$. Zatim dešifriramo y koristeći svih 2^{56} mogućih ključeva L. Nakon svakog dešifriranja, potražimo rezultat u tablici. (Naime, treba vrijediti $z = d_L(y)$.) Ako ga pronađemo, onda tako dobiveni par (K, L) testiramo na sljedećem poznatom paru *otvoreni tekst - šifrat*. Ako prođu taj test, prihvaćamo ih za korektne ključeve. Vjerojatnost da smo pogriješili je $2^{112-64-64} = 2^{-16}$. Na taj način dobivamo da je za razbijanje dvostrukog DES-a broj operacija reda 2^{56}, što je neznatno više nego za obični DES.

Jedna od najpopularnijih zamjena za DES je *trostruki DES* (koriste se još i nazivi Triple DES i 3DES):

$$y = e_M(d_L(e_K(x))), \quad x = d_K(e_L(d_M(y))).$$

Ovdje je ključ duljine $56 \cdot 3 = 168$ bitova. Često se koristi i verzija u kojoj je $M = K$, pa je u njoj duljina ključa $56 \cdot 2 = 112$, no za nju postoje neki, još uvijek nedovoljno praktični napadi, koji koriste njezinu specifičnu strukturu. Razlog za kombinaciju "*ede*" je kompatibilnost s običnim DES-om: dovoljno je staviti $L = M$ ili $K = L$. Za trostruki DES broj operacija kod napada "susret u sredini" je reda $2^{112} \approx 5 \cdot 10^{33}$, dok je kod diferencijalne kriptoanalize procijenjen na 10^{52}. Možemo reći da je sigurnost kod trostrukog šifriranja upravo onakva kakvu bismo možda naivno očekivali kod dvostrukog. U svakom slučaju, sigurnost 3DES-a je danas i više nego zadovoljavajuća.

Godine 1997. *National Institute of Standards and Technology* (NIST), organizacija koja je naslijedila *National Bureau of Standards*, objavila je natječaj za kriptosustav koji bi trebao kao opće prihvaćeni standard zamijeniti DES. Pobjednik natječaja dobio bi ime *Advanced Encryption Standard* (AES). NIST je postavio sljedeće zahtjeve na kriptosustav:

- mora biti simetričan,
- mora biti blokovni,

- treba raditi sa 128-bitnim blokovima i ključevima s tri duljine: 128, 192 i 256 bitova.

Nekoliko je razloga zbog kojih NIST nije odabrao 3DES kao AES:

- 3DES koristi 48 rundi da bi postigao sigurnost za koju su vjerojatno dovoljne 32 runde,
- softverske implementacije 3DES-a su prespore za neke primjene, posebno za digitalne video podatke,
- 64-bitni blokovi nisu najefikasniji u nekim primjenama.

No, svakako je veliki značaj 3DES-a što je pružio zadovoljavajući privremeni nadomjestak za DES, do izbora novog standarda.

Natječaj je zaključen 15.6.1998. Od 21 pristigle prijave, 15 ih je zadovoljilo NIST-ove kriterije. U kolovozu 1999. NIST je objavio 5 finalista: MARS, RC6, RIJNDAEL, SERPENT i TWOFISH.

Konačno, 2.8.2000. objavljeno je da je RIJNDAEL pobjednik natječaja za AES. RIJNDAEL (čitaj: rejn dol) su razvili belgijski kriptografi Joan Daemen i Vincent Rijmen s *Katholieke Universiteit Leuven*, po kojima je i dobio ime. Razlikuje se od ostalih finalista po tome što se u konstrukciji S-kutija koriste operacije u konačnom polju $\mathbb{F}_{2^8}$. Opisat ćemo kriptosustav RIJNDAEL, tj. *Advanced Encryption Standard* - AES.

Kao što smo već napomenuli, jedna od njegovih specifičnosti je korištenje konačnog polja $\mathbb{F}_{2^8}$ (alternativna je oznaka $GF(2^8)$). Elementi od $\mathbb{F}_{2^8}$ su polinomi oblika $a_7x^7 + a_6x^6 + \cdots + a_1x + a_0$, $a_i \in \{0, 1\}$, a operacije su zbrajanje i množenje polinoma iz $\mathbb{Z}_2[x]$ modulo fiksni ireducibilni polinom $g(x) = x^8 + x^4 + x^3 + x + 1$. Dakle, uzimamo da je $\mathbb{F}_{2^8} = \mathbb{Z}_2[x]/(g(x))$. Elemente od $\mathbb{F}_{2^8}$ možemo prikazati i kao bajtove (nizove od 8 bitova). Npr. polinomu $x^6 + x^4 + x^2 + x + 1$ odgovara bajt `01010111` ili `57` u heksadecimalnom zapisu.

Primjer 3.1. *Pomnožiti polinome* $x^6 + x^4 + x^2 + x + 1$ *i* $x^7 + x + 1$ *iz polja* $\mathbb{F}_{2^8}$.

Rješenje: Najprije pomnožimo ova dva polinoma u prstenu $\mathbb{Z}_2[x]$:

$$(x^6+x^4+x^2+x+1)(x^7+x+1) = x^{13}+x^{11}+x^9+x^8+x^6+x^5+x^4+x^3+1.$$

Potom izračunamo ostatak pri dijeljenju ovog polinoma s $g(x)$. Tako dobijemo rezultat x^7+x^6+1. Heksadecimalno to možemo zapisati kao: $\texttt{57}\cdot\texttt{83} = \texttt{C1}$. ♢

U AES-u se koriste i operacije na polinomima s koeficijentima iz $\mathbb{F}_{2^8}$ stupnja manjeg od 4. Po definiciji, to su polinomi čiji su koeficijenti također polinomi. Svaki takav polinom možemo reprezentirati kao 4-bajtni vektor.

Zbrajanje takvih polinoma se svodi na zbrajanje koeficijenta uz iste potencije (zbrajanje u $\mathbb{F}_{2^8}$ smo već definirali). Da bismo kod množenja kao rezultat dobili polinom stupnja manjeg od 4, moramo produkt reducirati modulo neki polinom četvrtog stupnja. U AES-u je za to izabran polinom x^4+1. Polinom x^4+1 je izabran da bi množenje bilo što jednostavnije za implementirati. Pritom polinom x^4+1 nije ireducibilan nad $\mathbb{Z}_2[x]$, jer u $\mathbb{Z}_2[x]$ vrijedi: $x^4+1=(x+1)^4$. Stoga neće svi polinomi imati inverz. No, to ovdje nije ni nužno. Zapravo, vidjet ćemo da se postojanje inverza zahtijeva samo za jedan konkretan polinom, za kojeg ćemo postojanje inverza provjeriti u sljedećem primjeru.

Uz ovaj izbor polinoma za redukciju, množenje polinoma se može matrično zapisati ovako:

$$\begin{bmatrix} d_0 \\ d_1 \\ d_2 \\ d_3 \end{bmatrix} = \begin{bmatrix} a_0 & a_3 & a_2 & a_1 \\ a_1 & a_0 & a_3 & a_2 \\ a_2 & a_1 & a_0 & a_3 \\ a_3 & a_2 & a_1 & a_0 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix}.$$

Ovdje je polinom $d(x)=d_3x^3+d_2x^2+d_1x+d_0$ rezultat množenja polinoma $a(x)=a_3x^3+a_2x^2+a_1x+a_0$ i $b(x)=b_3x^3+b_2x^2+b_1x+b_0$, u oznaci $d(x)=a(x)\otimes b(x)$.

Primjer 3.2. *Izračunati:*

$$(\texttt{03}x^3+\texttt{01}x^2+\texttt{01}x+\texttt{02})\otimes(\texttt{0B}x^3+\texttt{0D}x^2+\texttt{09}x+\texttt{0E}).$$

Rješenje: Označimo rezultat s $d(x)=d_3x^3+d_2x^2+d_1x+d_0$. Najprije imamo:

$$d_3=\texttt{03}\cdot\texttt{0E}+\texttt{01}\cdot\texttt{09}+\texttt{01}\cdot\texttt{0D}+\texttt{02}\cdot\texttt{0B}.$$

Računamo:

$$\begin{aligned}
\texttt{03}\cdot\texttt{0E} &= (x+1)\cdot(x^3+x^2+x)=x^4+x=\texttt{12},\\
\texttt{01}\cdot\texttt{09} &= \texttt{09},\\
\texttt{01}\cdot\texttt{0D} &= \texttt{0D},\\
\texttt{02}\cdot\texttt{0B} &= x\cdot(x^3+x+1)=x^4+x^2+x=\texttt{16},
\end{aligned}$$

pa je

$$d_3=\texttt{12}+\texttt{09}+\texttt{0D}+\texttt{16}=(x^4+x)+(x^3+1)+(x^3+x^2+1)+(x^4+x^2+x)=\texttt{00}.$$

Na isti način se izračuna $d_2=\texttt{00}$, $d_1=\texttt{00}$, $d_0=\texttt{01}$. Stoga je $d(x)=\texttt{01}$, tj. polinom $b(x)$ je inverz polinoma $a(x)$ s obzirom na operaciju $\otimes$. ♢

AES šifrira blokove od 128 bitova = 16 bajtova. Svaki takav blok se može shvatiti kao 16 elemenata polja, koji se reprezentiraju pomoću 4×4 matrice s elementima iz $\mathbb{F}_{2^8}$. Tu matricu ćemo zvati *AES-blok*. Ključ također ima 128 bitova (postoje i varijante sa 192 i 256 bitova). AES ima 10 rundi (uz 128-bitni ključ). Svaka runda se sastoji od 4 operacije:

`SubBytes`

`ShiftRows`

`MixColumns`

`AddRoundKey`

Prije prve runde vrši se inicijalno dodavanje ključa (`AddRoundKey`), a u posljednjoj se rundi izostavlja transformacija `MixColumns`.

`SubBytes` je jedini nelinearni dio algoritma. Ova transformacija ima 2 koraka:

1. svaki element AES-bloka (matrice) se zamjeni svojim inverzom u $\mathbb{F}_{2^8}$ (s time da element `00`, jedini koji nema inverz, ostaje nepromijenjen),

2. na svaki element $b = b_1 b_2 \cdots b_8$ AES-bloka primijeni se fiksna afina transformacija

$$b'_i = b_i \oplus b_{(i+4) \bmod 8} \oplus b_{(i+5) \bmod 8} \oplus b_{(i+6) \bmod 8} \oplus b_{(i+7) \bmod 8} \oplus c_i,$$

$(i = 1, 2, \ldots, 8)$, gdje je $c = 01100011$.

Ovu transformaciju možemo prikazati pomoću S-kutije. Svaki element AES-bloka zapišemo heksadecimalno. Prvom broju u tom zapisu pridružimo redak, a drugom broju stupac S-kutije. Na presjeku tog retka i stupca nalazi se heksadecimalni zapis transformata promatranog elementa AES-bloka.

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0	63	7C	77	7B	F2	6B	6F	C5	30	01	67	2B	FE	D7	AB	76
1	CA	82	C9	7D	FA	59	47	F0	AD	D4	A2	AF	9C	A4	72	C0
2	B7	FD	93	26	36	3F	F7	CC	34	A5	E5	F1	71	D8	31	15
3	04	C7	23	C3	18	96	05	9A	07	12	80	E2	EB	27	B2	75
4	09	83	2C	1A	1B	6E	5A	A0	52	3B	D6	B3	29	E3	2F	84
5	53	D1	00	ED	20	FC	B1	5B	6A	CB	BE	39	4A	4C	58	CF
6	D0	EF	AA	FB	43	4D	33	85	45	F9	02	7F	50	3C	9F	A8
7	51	A3	40	8F	92	9D	38	F5	BC	B6	DA	21	10	FF	F3	D2
8	CD	0C	13	EC	5F	97	44	17	C4	A7	7E	3D	64	5D	19	73
9	60	81	4F	DC	22	2A	90	88	46	EE	B8	14	DE	5E	0B	DB
A	E0	32	3A	0A	49	06	24	5C	C2	D3	AC	62	91	95	E4	79
B	E7	C8	37	6D	8D	D5	4E	A9	6C	56	F4	EA	65	7A	AE	08
C	BA	78	25	2E	1C	A6	B4	C6	E8	DD	74	1F	4B	BD	8B	8A
D	70	3E	B5	66	48	03	F6	0E	61	35	57	B9	86	C1	1D	9E
E	E1	F8	98	11	69	D9	8E	94	9B	1E	87	E9	CE	55	28	DF
F	8C	A1	89	0D	BF	E6	42	68	41	99	2D	0F	B0	54	BB	16

U implementaciji S-kutije nema bitne razlike u odnosu DES-a. I ovdje vrijednosti iz S-kutije očitavamo iz tablice. Međutim, kod AES-a su kontroverze oko S-kutija izbjegnute navođenjem eksplicitne formule pomoću koje je S-kutija generirana.

`ShiftRows` - ciklički pomiče elemente i-tog retka AES-bloka za i mjesta ulijevo ($i = 0, 1, 2, 3$).

`MixColumns` - stupci matrice se shvate kao polinomi nad $GF(2^8)$. Dakle, stupac $A_i = (a_{0i}, a_{1i}, a_{2i}, a_{3i})$, se shvati kao polinom $a_{3i}x^3 + a_{2i}x^2 + a_{1i}x + a_{0i}$. Sada se ti polinomi pomnože s polinomom $\texttt{03}x^3 + \texttt{01}x^2 + \texttt{01}x + \texttt{02}$ (u skladu s ranije definiranim operacijama nad polinomima, rezultat se računa modulo $x^4 + 1$).

`AddRoundKey` je XOR AES-bloka s međuključem koji odgovara trenutnoj rundi.

S-kutija (operacija `SubBytes`) je dizajnirana tako da bi kriptosustav bio što otporniji na diferencijalnu i linearnu kriptoanalizu, dok su `ShiftRows` i `MixColumns` zaslužni za difuziju.

U konstrukciji međuključeva ključ se najprije proširi korištenjem XOR-a i cikličkog pomaka. Prošireni ključ se sastoji od 44 riječi (4-bajtnih vektora). Međuključevi se biraju iz proširenog ključa tako da se prvi međuključ sastoji od prve četiri riječi, drugi od sljedeće četiri, itd.

Preostalo je još malo detaljnije opisati postupak proširivanja ključa. Prve 4 riječi proširenog ključa predstavljaju zadani ključ. Svaku sljedeću riječ r_i dobijemo iz prethodne riječi r_{i-1} tako da primijenimo XOR s r_{i-4}. Ako je i višekratnik od 4, onda prije operacije XOR izvršimo operacije `RotWord` (rotira riječ jedno mjesto ulijevo, tj. $[r_0 r_1 r_2 r_3]$ promijeni u $[r_1 r_2 r_3 r_0]$), `SubWord` (uzima jedan po jedan bajt iz riječi i na svakog od njih primijeni S-kutiju), te XOR s riječi $[\texttt{02}^{(i-4)/4}, \texttt{00}, \texttt{00}, \texttt{00}]$.

Dešifriranje se odvija po istom algoritmu kao i šifriranje, osim što se umjesto transformacija `SubBytes`, `ShiftRows` i `MixColumns` koriste njihovi inverzi. Primijetimo da smo u Primjeru 3.2 dokazali da polinom $\texttt{03}x^3 + \texttt{01}x^2 + \texttt{01}x + \texttt{02}$ ima inverz $\texttt{0B}x^3 + \texttt{0D}x^2 + \texttt{09}x + \texttt{0E}$ (svi ostali dijelovi algoritma za šifriranje su očito invertibilni).

3.3 RSA kriptosustav

Najpoznatiji kriptosustav s javnim ključem je RSA kriptosustav iz 1977. godine, nazvan po svojim tvorcima Ronaldu Rivestu, Adi Shamiru i Leonardu Adlemanu. Njegova sigurnost je zasnovana na teškoći faktorizacije velikih prirodnih brojeva. Dvadesetak godine kasnije pojavile su se informacije da je vrlo sličan kriptosustav nekoliko godina ranije osmislio Clifford Cocks iz britanske obavještajne agencije "Government Communications Headquarters".

Iako je sigurnost RSA kriptosustava zasnovana na teškoći faktorizacije, u samom se šifriranju i dešifriranju koristi modularno potenciranje, dok se faktorizacija koristi u dobivanju dodatnog podatka ("trapdoora"). Slijedi definicija RSA kriptosustava.

RSA kriptosustav. Neka je $n = pq$, gdje su p i q prosti brojevi. Neka je $\mathcal{P} = \mathcal{C} = \mathbb{Z}_n$, te

$$\mathcal{K} = \{(n, p, q, d, e) : n = pq,\ de \equiv 1 \pmod{\varphi(n)}\}.$$

Za $K \in \mathcal{K}$ definiramo

$$e_K(x) = x^e \bmod n, \qquad d_K(y) = y^d \bmod n, \quad x, y \in \mathbb{Z}_n.$$

Vrijednosti n i e su javne, a vrijednosti p, q i d su tajne, tj. (n, e) je javni, a (p, q, d) je tajni ključ.

Ovdje je $\varphi(n)$ Eulerova funkcija. U našem slučaju je

$$\varphi(n) = \varphi(pq) = \varphi(p)\varphi(q) = (p-1)(q-1) = n - p - q + 1.$$

U dokazu da je d_K inverz od e_K koristimo Eulerov teorem:

$$x^{\varphi(n)} \equiv 1 \pmod{n}, \quad \text{za } \mathrm{nzd}(x, n) = 1.$$

Uvjerimo se da su funkcije e_K i d_K jedna drugoj inverzne.

Imamo: $d_K(e_K(x)) \equiv x^{de} \pmod{n}$. Iz $de \equiv 1 \pmod{\varphi(n)}$ slijedi da postoji prirodan broj k takav da je $de = k\varphi(n) + 1$. Pretpostavimo da je $\mathrm{nzd}(x, n) = 1$. Sada je

$$x^{de} = x^{k\varphi(n)+1} = (x^{\varphi(n)})^k \cdot x \equiv x \pmod{n}.$$

Ako je $\mathrm{nzd}(n, x) = n$, onda je $x^{de} \equiv 0 \equiv x \pmod{n}$. Ako je $\mathrm{nzd}(n, x) = p$, onda je $x^{de} \equiv 0 \equiv x \pmod{p}$. Kako je $(pq, x) = p$, gdje su p i q prosti, slijedi da je $\mathrm{nzd}(q, x) = 1$, pa je prema Eulerovom teoremu $x^{\varphi(q)} = x^{q-1} \equiv 1$

(mod q). Stoga je $x^{de} = (x^{q-1})^{(p-1)k} \cdot x \equiv x \pmod{q}$, pa je $x^{de} \equiv x \pmod{n}$. Slučaj $\text{nzd}(n, x) = q$ je potpuno analogan. Prema tome, zaista je u svakom slučaju $x^{de} \equiv x \pmod{n}$, što znači da je $d_K(e_K(x)) = x$.

Primjer 3.3. *Ilustrirat ćemo šifriranje i dešifriranje u RSA kriptosustavu na sasvim malim parametrima.*

Uzmimo $p = 3$ i $q = 11$. Tada je $n = 33$ i $\varphi(n) = 20$. Eksponent e mora biti relativno prost s 20, pa uzmimo da je $e = 7$. Tada je $d = 3$. Sada je $(n, e) = (33, 7)$ naš javni ključ. Pretpostavimo da nam netko želi poslati poruku $x = 17$. To znači da treba izračunati $e_K(x) = 17^7 \bmod 33$:

$$17^7 = 17 \cdot 17^2 \cdot 17^4 \equiv 17 \cdot 25 \cdot (-2) \equiv -25 \equiv 8 \pmod{33}.$$

Dakle, šifrat je $y = e_K(x) = 8$.

Kad primimo ovaj šifrat, dešifriramo ga pomoću tajnog ključa d:

$$x = d_K(y) \equiv 8^3 \equiv 8 \cdot 8^2 \equiv 8 \cdot (-2) \equiv 17 \pmod{33}.$$

Dakle, $x = 17$. ♢

Sigurnost RSA kriptosustava leži u pretpostavci da je funkcija

$$e_K(x) = x^e \bmod n$$

jednosmjerna. Dodatni podatak ("trapdoor") koji omogućava dešifriranje je poznavanje faktorizacije $n = pq$. Zaista, onaj tko zna faktorizaciju broja n, taj može izračunati $\varphi(n) = (p-1)(q-1)$, te potom dobiti eksponent d rješavajući linearnu kongruenciju

$$de \equiv 1 \pmod{\varphi(n)}$$

(pomoću proširenog Euklidova algoritma).

No, otvoreno je pitanje je li razbijanje RSA kriptosustava, tj. određivanje x iz poznavanja $x^e \bmod n$, ekvivalentno faktorizaciji od n.

Recimo sada nekoliko riječi o izboru parametara u RSA kriptosustavu.

1. Tajno izaberemo dva velika prosta broja p i q slične veličine (barem 100 znamenaka). To radimo tako da najprije generiramo slučajan prirodan broj m s traženim brojem znamenaka, pa zatim pomoću nekog testa prostosti tražimo prvi prost broj veći ili jednak m. (Po teoremu o distribuciji prostih brojeva, možemo očekivati da ćemo trebati testirati približno $\ln m$ brojeva dok ne nađemo prvi prosti broj.) Treba paziti da $n = pq$ bude otporan na metode faktorizacije koje su vrlo efikasne za brojeve specijalnog oblika. Tako bi brojevi $p \pm 1$ i $q \pm 1$ trebali imati barem jedan veliki prosti faktor, jer postoje efikasne metode za faktorizaciju brojeva koji imaju prosti faktor p takav da je jedan od brojeva $p-1$, $p+1$ "gladak", tj. ima samo male proste faktore. Također, p i q ne smiju biti jako blizu jedan drugome, jer ih se onda može naći koristeći činjenicu da su približno jednaki $\sqrt{n}$.

2. Izračunamo $n = pq$ i $\varphi(n) = (p-1)(q-1) = n - p - q + 1$.

3. Izaberemo broj e takav da je $\text{nzd}(e, \varphi(n)) = 1$, te pomoću proširenog Euklidova algoritma izračunamo d takav da je $de \equiv 1 \pmod{\varphi(n)}$. Obično se uzima da je $e < \varphi(n)$. Broj e se može izabrati slučajno, a ima smisla izabrati ga i što manjim, tako da bi šifriranje $x^e \bmod n$ (tzv. modularno potenciranje) bilo što brže. Broj operacija u šifriranju ovisi o veličini broja e, te o broju jedinica u binarnom zapisu od e. Stoga je dugo vremena $e = 3$ bio popularan izbor. No, vidjet ćemo da izbor vrlo malog eksponenta e predstavlja opasnost za sigurnost, te se danas preporuča izbor $e = 2^{16} + 1 = 65537$.

4. Stavimo ključ za šifriranje (n, e) u javni direktorij.

Za efikasnost RSA kriptosustava, važna je činjenica da se modularno potenciranje može izvesti vrlo efikasno. Navedimo ovdje osnovnu metodu za računanje $e_K(x) = x^e \bmod n$, metodu "kvadriraj i množi", a postoje različite njene varijante i poboljšanja. Najprije e prikažemo u bazi 2:

$$e = 2^{s-1} \cdot e_{s-1} + \cdots + 2 \cdot e_1 + e_0,$$

a potom primijenimo sljedeći algoritam:

Kvadriraj i množi

$y = 1$
for $(s - 1 \geq i \geq 0)$ do
    $y = y^2 \bmod n$
    if $(e_i = 1)$ then $y = y \cdot x \bmod n$

Očito je ukupan broj množenja $\leq 2s$, pa je ukupan broj operacija $O(\log e \cdot \log^2 n)$. To znači da je ovaj algoritam polinomijalan.

Jedan očit napad na RSA je faktorizacija od n. Ako napadač faktorizira n, onda može naći $\varphi(n)$ i d. Spomenimo da trenutno najbrži algoritmi za faktorizaciju trebaju

$$e^{O((\log n)^{1/3}(\log\log n)^{2/3})}$$

operacija, tako da su brojevi od preko 250 znamenaka zasad sigurni od ovog napada. Dakle, nije poznat niti jedan polinomijalni algoritam za faktorizaciju. Ovdje ipak treba reći da je u nekim slučajevima n puno lakše faktorizirati, pa takve n-ove treba izbjegavati. Takav je slučaj npr. ako su p i q jako blizu jedan drugoga ili ako $p - 1$ i $q - 1$ imaju samo male proste faktore.

Važno je napomenuti da ako napadač otkrije tajni eksponent d, onda nije dovoljno promijeniti samo eksponent e, već moramo promijeniti i n.

Zaista, ako je poznat broj d takav da je $a^{ed} \equiv a \pmod{n}$ za sve a, $\text{nzd}(a,n) = 1$, onda za broj $m = ed - 1$ vrijedi $a^m \equiv 1 \pmod{n}$ za sve a, $\text{nzd}(a,n) = 1$. Kako je grupa svih reduciranih ostataka modulo p ciklička, ovo je ekvivalentno tome da je m zajednički višekratnik od $p-1$ i $q-1$. Dakle, poznavanje broja m je slabije od poznavanja broja $\varphi(n) = (p-1)(q-1)$. Pokazat ćemo kako ipak pomoću m možemo (s velikom vjerojatnošću) faktorizirati n.

Uvrštavanjem $a = -1$ vidimo da je m paran. Provjerimo zadovoljava li $m/2$ isto svojstvo kao m. Ako postoji a, $\text{nzd}(a,n) = 1$, za koji nije $a^{m/2} \equiv 1 \pmod{n}$, onda takvih a-ova ima barem 50%. Zaista, svakom broju b koji zadovoljava kongruenciju $b^{m/2} \equiv 1 \pmod{n}$, odgovara broj ab koji tu kongruenciju ne zadovoljava. Stoga, ako testiramo nekoliko desetaka a-ova i ako svi oni zadovoljavaju $a^{m/2} \equiv 1 \pmod{n}$, onda s velikom vjerojatnošću možemo m zamijeniti s $m/2$. Nastavljamo ovo dijeljenje s 2 dokle god je to moguće. Na kraju imamo dvije mogućnosti:

1) $m/2$ je višekratnik od $p-1$, a nije od $q-1$ (ili obrnuto). Tada je $a^{m/2} \equiv 1 \pmod{p}$ uvijek, ali je $a^{m/2} \equiv -1 \pmod{q}$ u 50% slučajeva (ako je c primitivni korijen modulo q, onda je $a^{m/2} \equiv 1 \pmod{q}$ za $a = c^{2k}$, dok je $a^{m/2} \equiv -1 \pmod{q}$ za $a = c^{2k+1}$).

2) $m/2$ nije višekratnik ni od $p-1$ ni od $q-1$. Tada je $a^{m/2} \equiv \pm 1 \pmod{p}$, $a^{m/2} \equiv \pm 1 \pmod{q}$ i svaka od četiri mogućnosti nastupa u 25% slučajeva.

Dakle, za proizvoljan a imamo s vjerojatnošću 50% da je $a^{m/2} - 1$ djeljiv s jednim od brojeva p i q, a nije djeljiv s drugim. Uzimajući slučajno nekoliko desetaka a-ova, možemo s vrlo velikom vjerojatnošću očekivati da ćemo pronaći a s gornjim svojstvom. Recimo da je $a^{m/2} - 1$ djeljiv s p, a nije s q. Tada je $\text{nzd}(n, a^{m/2} - 1) = p$ i mi smo uspjeli faktorizirati n. Ovdje treba naglasiti da se najveći zajednički djelitelj od n i $a^{m/2} - 1$ može efikasno izračunati Euklidovim algoritmom.

Upravo opisani algoritam za faktorizaciju je primjer tzv. vjerojatnosnog algoritma, i to iz klase *Las Vegas algoritama*. To su algoritmi koji ne daju uvijek odgovor, ali kada ga daju, onda je odgovor sigurno točan. U slučaju da algoritam ne da odgovor, možemo mu dati novu nezavisnu šansu za nalaženje odgovora. Na taj se način vjerojatnost uspjeha algoritma povećava s količinom raspoloživog vremena. Druga klasa vjerojatnosnih algoritma su *Monte Carlo algoritmi* koji uvijek daju odgovor, ali on može biti netočan. Primjer takvog algoritma je Miller-Rabinov test za testiranje prostosti.

Na prvi se pogled ne čini lošom ideja da izbjegnemo generiranje različitih modula $n = pq$, već da izaberemo jedan "dobar" n jednom za svagda. Taj n bi koristili svi korisnici, a iz jednog povjerljivog centra bi se korisnicima distribuirali parovi e_K, d_K iz kojih bi oni onda formirali svoje javne i tajne ključeve. Međutim, ideja je loša jer bi rezultirala nesigurnim sustavom i to

upravo zbog gore opisanog vjerojatnosnog algoritma. Naime, korisnik A bi pomoću tog algoritma te njegovih eksponenata e_A i d_A mogao faktorizirati n. Nakon toga bi A lako, poznavajući faktore od n i javni ključ e_B, mogao izračunati tajni ključ d_B bilo kojeg drugog korisnika. Ova primjedba pokazuje da jedan RSA modul ne bi smjelo koristiti više korisnika.

Sljedeća, također samo na prvi pogled dobra ideja je da pokušamo izabrati parametre RSA kripotsustava tako da jedan od eksponenata e ili d bude mali. Budući da je broj operacija za modularno potenciranje linearan u broju bitova eksponenta, to bi moglo smanjiti vrijeme potrebno za šifriranje, odnosno dešifriranje. To bi posebno moglo biti od interesa u situacijama kad postoji veliki nesrazmjer u snazi dvaju uređaja koji sudjeluju u komunikaciji, kao što je npr. slučaj kad "pametna kartica" komunicira s centralnim računalom. U toj situaciji bismo možda poželjeli kartici dodijeliti mali tajni eksponent, a računalu mali javni eksponent, da bismo minimizirali onaj dio računanja koje treba provesti kartica. Međutim, vidjet ćemo da takav izbor eksponenata ipak nije dobar. Sljedeći teorem M. Wienera iz 1990. godine pokazuje da u slučaju izbora relativno malog tajnog eksponenta d (u odnosu na n) postoji efikasan algoritam za razbijanje šifre.

Teorem 3.1. *Neka je $n = pq$ i $p < q < 2p$, te neka je $e < \varphi(n)$ i $d < \frac{1}{3}n^{0.25}$. Tada postoji polinomijalni algoritam koji iz poznavanja n i e izračunava d.*

Dokaz: Iz $ed \equiv 1 \pmod{\varphi(n)}$ slijedi da postoji prirodan broj k takav da je $ed - k\varphi(n) = 1$. Odavde je

$$\left|\frac{e}{\varphi(n)} - \frac{k}{d}\right| = \frac{1}{d\varphi(n)}. \tag{3.1}$$

Dakle, $\frac{k}{d}$ je dobra aproksimacija od $\frac{e}{\varphi(n)}$. Međutim, mi ne znamo $\varphi(n)$. Stoga ćemo $\varphi(n)$ aproksimirati s n. Iz $\varphi(n) = n - p - q + 1$ i $p + q - 1 < 3\sqrt{n}$ slijedi $|n - \varphi(n)| < 3\sqrt{n}$. Zamijenimo $\varphi(n)$ s n u (3.1), pa dobivamo:

$$\begin{aligned}\left|\frac{e}{n} - \frac{k}{d}\right| &= \left|\frac{ed - k\varphi(n) - kn + k\varphi(n)}{nd}\right| = \left|\frac{1 - k(n - \varphi(n))}{nd}\right| \\ &\leq \frac{3k\sqrt{n}}{nd} = \frac{3k}{d\sqrt{n}}.\end{aligned}$$

Sada je $k\varphi(n) = ed - 1 < ed$, pa iz $e < \varphi(n)$ (to je standardna pretpostavka u RSA kriptosustavu), slijedi $k < d < \frac{1}{3}n^{0.25}$, te dobivamo

$$\left|\frac{e}{n} - \frac{k}{d}\right| \leq \frac{1}{d\sqrt[4]{n}} < \frac{1}{2d^2}. \tag{3.2}$$

Iz teorije diofantskih aproksimacija (Legendreov teorem) slijedi da relacija (3.2) povlači da je k/d neka konvergenta razvoja u verižni razlomak od e/n.

Iz rekurzija za konvergente p_k/q_k slijedi da je $q_k \geq F_k$, gdje je F_k k-ti Fibonaccijev broj, što znači da nazivnici konvergenti rastu eksponencijano. U našem slučaju dakle slijedi da ima $O(\log n)$ konvergenti od $\frac{e}{n}$. Jedna od njih je $\frac{k}{d}$. Dakle, izračunamo sve konvergente od $\frac{e}{n}$ i testiramo koja od njih zadovoljava uvjet $(x^e)^d \equiv x \pmod{n}$ za slučajno odabran broj x. To daje polinomijalni algoritam za otkrivanje tajnog ključa d.

Drugi način za testiranje točnosti pretpostavke da je neka konkretna konvergenta jednaka $\frac{k}{d}$, jest da se, uz tu pretpostavku, izračuna $\varphi(n) = (p-1)(q-1) = (ed-1)/k$. Tada se može izračunati $\frac{p+q}{2}$ iz identiteta

$$\frac{pq-(p-1)(q-1)+1}{2} = \frac{p+q}{2},$$

te $\frac{q-p}{2}$ iz identiteta $(\frac{p+q}{2})^2 - pq = (\frac{q-p}{2})^2$. Ako se na ovaj način dobije da su brojevi $\frac{p+q}{2}$ i $\frac{q-p}{2}$ cijeli, onda zaključujemo da je promatrana konvergenta stvarno jednaka $\frac{k}{d}$. Tada iz $\frac{p+q}{2}$ i $\frac{q-p}{2}$ možemo lako dobiti i faktorizaciju modula $n = pq$. □

Primjer 3.4. *Pretpostavimo da su u RSA kriptosustavu zadani modul*

$$n = 7978886869909,$$

javni eksponent

$$e = 3594320245477,$$

te da je poznato da tajni eksponent d zadovoljava $d < \frac{1}{3}n^{0.25} < 561$.

Da bismo primijenili Wienerov napad, računamo razvoj broja $\frac{e}{n}$ u verižni razlomak. Dobivamo:

$$[0; 2, 4, 1, 1, 4, 1, 2, 31, 21, 1, 3, 1, 16, 3, 1, 114, 10, 1, 4, 5, 1, 2].$$

Potom računamo pripadne konvergente:

$$0, \frac{1}{2}, \frac{4}{9}, \frac{5}{11}, \frac{9}{20}, \frac{41}{91}, \frac{50}{111}, \frac{141}{313}, \frac{4421}{9814}, \ldots.$$

Konačno, provjeravamo koji od nazivnika 2, 9, 11, 20, 91, 111, 313 zadovoljava kongruenciju $(x^e)^d \equiv x \pmod{n}$ za npr. $x = 2$. Tako dobivamo da je tajni eksponent $d = 313$. ♢

U ovom primjeru smo vidjeli da je prava konvergenta bila upravo zadnja koja je zadovoljavala uvjet za veličinu nazivnika. To nam sugerira da možda uopće nije nužno testirati sve konvergente u zadanom rasponu, već da bi moglo biti moguće karakterizirati pravu konvergentu. Zaista, to se može napraviti preciznijom ocjenom za $\varphi(n)$. Uz razumnu pretpostavku da je $n > 10^8$, dobije se da je $\frac{k}{d}$ jedinstvena konvergenta koja zadovoljava nejednakost

$$\frac{2e}{n\sqrt{n}} < \frac{k}{d} - \frac{e}{n} < \frac{2.122\,e}{n\sqrt{n}}.$$

Postoje i proširenja Wienerovog napada (Verheul - van Tilborg, Dujella) na RSA u kojem je tajni ključ nešto veći od $\sqrt[4]{n}$. U njima se tajni eksponent d traži u obliku $rq_{m+1} \pm sq_m$ za nenegativne cijele brojeve (r, s). Pomoći tih napada se tajni ključ može otkriti ukoliko je $d < 2^{30} n^{0.25}$.

Postoje i napadi koji, umjesto verižnih razlomaka, koriste Coppersmithovu metodu za nalaženje rješenja polinomijalnih kongruencija. Naime, radi se o sljedećem problemu. Neka je zadan polinom $f(x) \in \mathbb{Z}[x]$ stupnja d i neka je poznato da postoji "malo" rješenje kongruencije $f(x) \equiv 0 \pmod{N}$, tj. rješenje x_0 za koje vrijedi $|x_0| < N^{1/d}$. Pitanje je možemo li efikasno naći x_0. Coppersmith je pokazao da je odgovor na ovo pitanje potvrdan. Osnovna ideja je konstruirati novi polinom $h(x) = h_0 + h_1 x + \cdots h_n x^n \in \mathbb{Z}[x]$ za kojeg će također vrijediti $h(x_0) \equiv 0 \pmod{N}$, ali koji će imati male koeficijente. Preciznije, traži se da "norma" $\|h(x)\| := \left(\sum_{i=0}^{n} h_i^2\right)^{1/2}$ bude mala. Tada se može iskoristiti sljedeća jednostavna činjenica: ako za prirodan broj X vrijedi

$$\|h(xX)\| < \frac{N}{\sqrt{n}}$$

i $|x_0| < X$ zadovoljava kongruenciju $h(x_0) \equiv 0 \pmod{N}$, onda je x_0 nultočka polinoma h, tj. vrijedi ne samo kongruencija, već i jednakost $h(x_0) = 0$.

Polinom $h(x)$ s traženim svojstvom može se naći pomoću LLL-algoritma, koji nalazi male vektore u rešetki, a ima brojne primjene u matematici (posebno u kombinatorici i teoriji brojeva), a također i u kriptoanalizi. Neka su $b_1, \ldots, b_m$ linearno nezavisni vektori u $\mathbb{R}^n$. *Rešetka* razapeta ovim vektorima je skup svih njihovih cjelobrojnih linearnih kombinacija:

$$L = \left\{ \sum_{i=1}^{m} n_i \cdot b_i \,:\, n_i \in \mathbb{Z} \right\}.$$

Kaže se da je $B = \{b_1, \ldots, b_m\}$ *baza rešetke* L. Jasno je da jedna rešetka može imati više različitih baza i pitamo se možemo li izabati bazu koja bi imala neko dodatno dobro svojstvo. Ako B shvatimo kao bazu pripadnog vektorskog potprostora, onda znamo da Gram-Schmidtovim postupkom možemo dobiti ortogonalnu bazu za isti vektorski potprostor ($b_i^* = b_i - \sum_{j=1}^{i-1} \mu_{ij} b_j^*$, $i = 1, \ldots, n$, gdje je $\mu_{ij} = \frac{\langle b_i, b_j^* \rangle}{\langle b_j^*, b_j^* \rangle}$). Međutim, ta nova baza ne mora razapinjati istu rešetku kao polazna baza B, i općenito rešetka ne mora imati ortogonalnu bazu. A. K. Lenstra, H. W. Lenstra i L. Lovász uveli su pojam LLL-reducirane baze koja je "gotovo ortogonalna" u smislu da za koeficijente μ_{ij}, $1 \leq j < i \leq n$, vrijedi $|\mu_{i,j}| \leq \frac{1}{2}$ (ovdje $\langle\ ,\ \rangle$ označava skalarni produkt na $\mathbb{R}^n$). Dodatno važno svojstvo ove LLL-reducirane baze je da je prvi vektor u toj bazi vrlo kratak, tj. ima malu normu. Može se dokazati da uvijek vrijedi $\|b_1\| \leq 2^{(m-1)/2} \|x\|$, za sve ne-nul vektore $x \in L$, no, u praksi se vrlo često događa da je $\|b_1\|$ upravo najkraći ne-nul vektor iz L.

Lenstra, Lenstra i Lovász su 1982. godine prikazali su polinomijalni algoritam za konstrukciju LLL-reducirane baze iz proizvoljne baze rešetke (po njima nazvan LLL-algoritam), koji je, kao što smo već rekli, ubrzo našao brojne primjene, od kojih je prva bila u faktorizaciji polinoma s racionalnim koeficijentima. Jedna od tih primjena je i konstrukcija gore spomenutog polinoma $h(x)$ s malim koeficijentima, čiji se koeficijenti mogu dobiti kao komponente prvog vektora LLL-reducirane baze određene rešetke koja se dobije pomoću koeficijenata polaznog polinoma $f(x)$.

Boneh i Durfee su opisali jedan napad na RSA ovakvog tipa koji je primjenjiv u slučaju da je $d < n^{0.292}$. Slično kao kod Weinerova napada, kreće se od jednakosti $ed - k\varphi(n) = 1$, koja se može zapisati i kao

$$ed - k(n + 1 - p - q) = 1.$$

Stavimo $s = p + q$, $a = n + 1$. Sada je nalaženje malog tajnog eksponenta d, recimo $d < n^{\delta}$, ekvivalentno nalaženju malih rješenja k i s kongruencije

$$f(k, s) = k(s - a) \equiv 1 \pmod{e}.$$

Zaista, za k i e imamo sljedeće ocjene:

$$|s| < 3\sqrt{n} \approx e^{0.5}, \quad |k| < \frac{de}{\varphi(n)} \approx e^{\delta}.$$

Dakle, situacija je slična kao kod gore navedenog Coppersmithova rezultata, samo što se ovdje radi o polinomu od dvije varijable, pa se Coppersmithov teorem ne može direktno primijeniti da bi se strogo dokazala korektnost ovog napada. Ipak, pokazalo se da on u praksi radi sasvim zadovoljavajuće.

Savjet je da se izbjegava slučaj kada je $d < \sqrt{n}$, jer je poznato da su svi ovi gore spomenuti napadi sasvim neprimjenjivi ako je $d > \sqrt{n}$.

Također postoje i napadi na RSA uz pretpostavku da je eksponent e mali, pa bi i to trebalo izbjegavati. U ranijim implementacijama RSA kriptosustava, često se uzimalo $e = 3$, da bi se minimiziralo vrijeme potrebno za šifriranje. Pokazat ćemo zašto taj izbor za e nije dobar.

Pretpostavimo da imamo tri korisnika s različitim vrijednostima javnog modula n_1, n_2, n_3, te pretpostavimo da svi oni koriste isti javni eksponent $e = 3$. Nadalje, pretpostavimo da im netko želi poslati identičnu poruku m. Tada njihov protivnik može doznati sljedeće šifrate:

$$c_1 \equiv m^3 \pmod{n_1}, \quad c_2 \equiv m^3 \pmod{n_2}, \quad c_3 \equiv m^3 \pmod{n_3}.$$

Nakon toga, on može, koristeći Kineski teorem o ostatcima naći rješenje sustava linearnih kongruencija

$$x \equiv c_1 \pmod{n_1}, \quad x \equiv c_2 \pmod{n_2}, \quad x \equiv c_3 \pmod{n_3}.$$

Na taj način, dobit će broj x sa svojstvom $x \equiv m^3 \pmod{n_1 n_2 n_3}$. No, kako je $m^3 < n_1 n_2 n_3$, zapravo vrijedi jednakost $x = m^3$, pa protivnik može izračunati originalnu poruku m tako na nađe treći korijen iz x.

Primjer 3.5. *Za ilustraciju ovog napada, pretpostavimo da je* $n_1 = 329$*,* $n_2 = 341$*,* $n_3 = 377$*. Pretpostavimo da je protivnik saznao šifrate* $c_1 = 43$*,* $c_2 = 30$*,* $c_3 = 372$*, te želi saznati zajednički otvoreni tekst* m*.*

Rješenje: Rješavanjem sustava

$$x \equiv 43 \pmod{329}, \quad x \equiv 30 \pmod{341}, \quad x \equiv 372 \pmod{377},$$

pomoću Kineskog teorema o ostatcima, dobiva se da je

$$\begin{aligned} x &\equiv 341 \cdot 377 \cdot 172 + 329 \cdot 377 \cdot 232 + 329 \cdot 341 \cdot 317 \equiv 86451373 \\ &\equiv 1860867 \pmod{329 \cdot 341 \cdot 377}. \end{aligned}$$

To znači da je $x = 1860867$ i $m = \sqrt[3]{x} = 123$. ♢

Upravo opisani napad može se izbjeći tako da se porukama prije šifriranja doda neki "slučajni dodatak" (engl. random pad). Na taj način, nikad nećemo različitim primateljima slati potpuno identične poruke. No, postoje napadi (zasnovani na gore spomenutom Coppersmithovom rezultatu i LLL-algoritmu) koji pokazuju da ni u tom slučaju RSA kriptosustav s vrlo malim eksponentom e nije siguran. Može se preporučiti uporaba eksponenta $e = 65537$, koji je dovoljno velik da bi onemogućio sve poznate napade na RSA s malim eksponentom, a prednost mu je vrlo brzo šifriranje jer ima malo jedinica u binarnom zapisu. Naime, $65537 = 2^{16} + 1$.

Možemo zaključiti da i nakon tri desetljeća intenzivnog proučavanja, još uvijek nije pronađena metoda koja bi razbila RSA kriptosustav. Svi poznati napadi na RSA zapravo samo pokazuju na što treba paziti i što treba izbjegavati kod izbora parametara i implementacije RSA. Zasad se, uz korektnu implementaciju, RSA može smatrati sigurnim kriptosustavom.

3.4 Ostali kriptosustavi s javnim ključem

3.4.1 Rabinov kriptosustav

Usko povezan s problemom faktorizacije je *problem računanja kvadratnog korijena* u $\mathbb{Z}_n$. Neka je $n = pq$, gdje su p, q prosti brojevi. Za $1 \leq a \leq n-1$ treba naći $x \in \mathbb{Z}$ takav da je $x^2 \equiv a \pmod{n}$, uz pretpostavku da takav x postoji, tj. da je a kvadratni ostatak modulo n. Vidjeli smo da postoji efikasan algoritam za rješavanje kongruencije $x^2 \equiv a \pmod{p}$. Algoritam je posebno jednostavan ako je $p \equiv 3 \pmod{4}$. Naime, tada je rješenje $x \equiv \pm a^{(p+1)/4} \pmod{p}$. Kombinirajući dva rješenja $\pm r$ kongruencije $x^2 \equiv a \pmod{p}$ i dva rješenja $\pm s$ kongruencije $x^2 \equiv a \pmod{q}$, po Kineskom teoremu o ostatcima dobivamo četiri rješenja kongruencije $x^2 \equiv a \pmod{pq}$.

Obrnuto, ako znamo riješiti problem kvadratnog korijena, onda znamo riješiti i problem faktorizacije. Neka je dan složen broj n. Odaberimo slučajan broj x takav da je $\mathrm{nzd}(x, n) = 1$ (ako je $\mathrm{nzd}(x, n) > 1$, onda smo našli faktor od n) i izračunajmo $a = x^2 \bmod n$. Primijenimo algoritam (pretpostavka je da takav algoritam postoji) za problem kvadratnog korijena na broj a. Tako dobijemo broj y. Ako je $y \equiv \pm x \pmod{n}$, onda biramo novi x. Ako je $y \not\equiv \pm x \pmod{n}$, onda iz $n|(x-y)(x+y)$ slijedi da je $\mathrm{nzd}(x-y, n)$ netrivijalni faktor od n. Kako postoje četiri kvadratna korijena od a modulo n, vjerojatnost uspjeha ovog algoritma u jednom koraku je $1/2$.

Rabinov kriptosustav iz 1979. godine zasnovan je na teškoći računanja kvadratnog korijena po fiksnom složenom modulu. Štoviše, za njega vrijedi da je njegovo razbijanje ekvivalentno rješavanju problema kvadratnog korijena, pa je, u skladu s gore pokazanim, ekvivalentno i problemu faktorizacije. Ova činjenica pokazuje jednu, barem teoretsku, prednost ovog kriptosustava pred RSA kriptosustavom.

Rabinov kriptosustav. Neka je $n = pq$, gdje su p i q prosti brojevi takvi da je $p \equiv q \equiv 3 \pmod{4}$. Neka je $\mathcal{P} = \mathcal{C} = \mathbb{Z}_n$, te

$$\mathcal{K} = \{(n, p, q) : n = pq\}.$$

Za $K \in \mathcal{K}$ definiramo

$$e_K(x) = x^2 \bmod n, \qquad d_K(y) = \sqrt{y} \bmod n.$$

Vrijednost n je javna, a vrijednosti p i q su tajne.

Ovdje $a = \sqrt{b} \bmod n$ znači da je $a^2 \equiv b \pmod{n}$. Uvjet $p \equiv q \equiv 3 \pmod{4}$ se može izostaviti. No, uz ovaj uvjet je dešifriranje jednostavnije i efikasnije.

Jedan nedostatak Rabinovog kriptosustava je da funkcija e_K nije injekcija, a to je zahtjev koji smo stavljali pred svaki kriptosustav. Naime, postoje četiri kvadratna korijena modulo n, pa dešifriranje nije moguće provesti na jednoznačan način (osim ako je otvoreni tekst neki smisleni tekst, a to nije slučaj kod razmjene ključeva, za što se kriptosustavi s javnim ključem prvenstveno koriste). Jedan način za rješavanje ovog problema je da se u otvoreni tekst na umjetan način ubaci izvjesna pravilnost. To se može napraviti npr. tako da se posljednja 64 bita dupliciraju (ponove). Tada možemo očekivati da će samo jedan od 4 kvadratna korijena dati rezultat koji ima zadanu pravilnost.

Williams je 1980. godine dao jednu modifikaciju Rabinova kriptosustava kojom se također eliminira ovaj nedostatak. U toj modifikaciji se kreće od prostih brojeva p, q sa svojstvom $p \equiv 3 \pmod 8$, $q \equiv 7 \pmod 8$. Tada je Jacobijev simbol $(\frac{2}{pq}) = -1$, pa se svojstva Jacobijeva simbola mogu iskoristiti za identifikaciju "pravog" kvadratnog korijena.

Primjer 3.6. *U Rabinovom kriptosustavu s parametrima*

$$(n, p, q) = (437, 19, 23),$$

dešifrirati šifrat $y = 35$. *Poznato je da je otvoreni tekst prirodan broj* $x < n$ *kojem su zadnja četiri bita u binarnom zapisu međusobno jednaka.*

Rješenje: Najprije trebamo naći kvadratne korijene od 35 modulo 19 i modulo 23. Budući da je $19 \equiv 23 \equiv 3 \pmod 4$, možemo ih naći po formuli $\pm a^{(p+1)/4} \pmod p$. Za $p = 19$, dobivamo $35^5 \equiv 4 \pmod{19}$, a za $p = 23$, $35^6 \equiv 9 \pmod{23}$. Sada koristimo Kineski teorem o ostacima da bi našli kvadratne korijene od 35 modulo $437 = 19 \cdot 23$, rješavajući 4 sustava linearnih kongruencija:

$$x \equiv \pm 4 \pmod{19}, \quad x \equiv \pm 9 \pmod{23}. \tag{3.3}$$

Podsjetimo se što smo ranije rekli o rješavanju sustava od dvije linearne jednadžbe

$$x \equiv a_1 \pmod{m_1}, \quad x \equiv a_2 \pmod{m_2}.$$

Primjenom Euklidovog algoritma dobivamo u, v takve da je $um_1 + vm_2 = 1$. Tada je $x = um_1a_2 + vm_2a_1 \bmod m_1m_2$ rješenje sustava.

U našem slučaju je $u = -6$, $v = 5$, pa su rješenja sustava (3.3) $x \equiv 129, 175, 262, 308 \pmod{437}$. Zapišimo ova četiri broja binarno:

$$\begin{aligned} 129 &= 10000001, \\ 175 &= 10101111, \\ 262 &= 100000110, \\ 308 &= 100110100, \end{aligned}$$

pa vidimo da je traženi otvoreni tekst $x = 175$. $\diamondsuit$

Postoji još jedna slabost Rabinova kriptosustava, koja je, pomalo paradoksalno, usko povezana upravo s njegovom gore navedenom teoretskom prednošću. Radi se o napadu "odabrani šifrat". Pretpostavimo da Eva uspije nekako nagovoriti Boba da joj dešifrira (s njegovim tajnim ključem) šifrat y koji je ona sama izabrala. Sada Eva može primijeniti gore opisani algoritam za faktorizaciju korištenjem kvadratnog korijena, i s vjerojatnošću $\frac{1}{2}$ faktorizirati n. Uspije li Boba nagovoriti na dešifriranje k šifrata, vjerojatnost da će ona nakog toga uspjeti faktorizirati n bit će $1 - \frac{1}{2^k}$.

Napad "odabrani šifrat" predstavlja opasnost i po ostale kriptosustave s javnim ključem. Kod RSA kriptosustava, Eva može dešifrirati šifrat y (dobiven iz njoj nepoznatog otvorenog teksta x po pravilu $y = x^e \bmod n$) tako da izabere slučajan otvoreni tekst x_1, izračuna $y_1 = yx_1^e \bmod n$, te zamoli Boba za joj dešifrira y_1. Rezultat dešifriranja je zapravo xx_1, pa iz njega Eva može izračunati traženi otvoreni tekst x. No, postoji ipak bitna razlika u odnosu na gore opisani napad na Rabinov kriptosustav, jer je ovdje Eva uspjela dešifrirati samo jednu poruku, dok je gore uspjela faktorizirati n i otkriti Bobov tajni ključ.

3.4.2 Kriptosustavi zasnovani na problemu diskretnog logaritma

Neka je G konačna abelova grupa. Da bi bila prikladna za primjene u kriptografiji javnog ključa, grupa G bi trebala imati svojstvo da su operacije množenja i potenciranja u njoj jednostavne, dok je logaritmiranje (inverzna operacija od potenciranja) vrlo teško. Također bi trebalo biti moguće generirati slučajne elemente grupe na gotovo uniforman način. Ipak, centralno pitanje jest koliko je težak tzv. *problem diskretnog logaritma* u grupi G.

Problem diskretnog logaritma: Neka je $(G, *)$ konačna grupa, $g \in G$, $H = \{g^i : i \geq 0\}$ podgrupa od G generirana s g, te $h \in H$. Treba naći najmanji nenegativni cijeli broj x takav da je $h = g^x$, gdje je $g^x = \underbrace{g * g * ... * g}_{x \text{ puta}}$. Taj broj x se zove *diskretni logaritam* i označava se s $\log_g h$.

Činjenicu da postoje grupe u kojima je problem diskretnog logaritma težak, iskoristili su Diffie i Hellman u svom rješenju problema razmjene ključeva.

Pretpostavimo da se Alice i Bob žele dogovoriti o jednom tajnom slučajnom elementu u grupi G, kojeg bi onda poslije mogli koristi kao ključ za šifriranje u nekom simetričnom kriptosustavu. Oni taj svoj dogovor moraju

provesti preko nekog nesigurnog komunikacijskog kanala, bez da su prethodno razmjenili bilo kakvu informaciju. Jedina informacija koju imaju jest grupa G i njezin generator g (pretpostavimo zbog jednostavnosti da je grupa G ciklička).

Slijedi opis Diffie-Hellmanovog protokola. Sa $|G|$ ćemo označavati broj elemenata u grupi G.

Diffie-Hellmanov protokol za razmjenu ključeva:

1. Alice generira slučajan prirodan broj $a \in \{1, 2, ..., |G| - 1\}$. Ona pošalje Bobu element g^a.
2. Bob generira slučajan prirodan broj $b \in \{1, 2, ..., |G| - 1\}$, te pošalje Alice element g^b.
3. Alice izračuna $(g^b)^a = g^{ab}$.
4. Bob izračuna $(g^a)^b = g^{ab}$.

Sada je njihov tajni ključ $K = g^{ab}$.

Njihov protivnik (Eve), koji može prisluškivati njihovu komunikaciju preko nesigurnog komunikacijskog kanala, zna sljedeće podatke: G, g, g^a, g^b. Eve treba iz ovih podataka izračunati g^{ab} (kaže se da Eve treba riješiti *Diffie-Hellmanov problem* (DHP)). Ako Eve iz poznavanja g i g^a može izračunati a (tj. ako može riješiti problem diskretnog logaritma (DLP)), onda i ona može pomoću a i g^b izračunati g^{ab}. Vjeruje se da su za većinu grupa koje se koriste u kriptografiji ova dva problema, DHP i DLP, ekvivalentni (tj. da postoje polinomijalni algoritmi koji svode jedan problem na drugi).

U originalnoj definiciji Diffie-Hellmanovog protokola za grupu G se uzima multiplikativna grupa $\mathbb{Z}_p^*$ svih ne-nul ostataka modulo p, gdje je p dovoljno velik prost broj. Poznato je da je grupa $\mathbb{Z}_p^*$ ciklička. Generator ove grupe se naziva *primitivni korijen* modulo p. Broj $g \in \{1, 2, ..., p - 1\}$ je primitivni korijen modulo p ako je g^{p-1} najmanja potencija broja g koja daje ostatak 1 pri djeljenju s p.

Sada ćemo opisati *ElGamalov kriptosustav* iz 1985. godine, koji zasnovan na teškoći računanja diskretnog logaritma u u grupi $(\mathbb{Z}_p^*, \cdot_p)$.

Pokazuje se da je ovaj problem približno iste težine kao problem faktorizacije složenog broja n (ako su p i n istog reda veličine), a i neke od metoda koje koriste u najboljim poznatim algoritmima za rješavanje tih problema su vrlo slične.

ElGamalov kriptosustav: Neka je p prost broj i $\alpha \in \mathbb{Z}_p^*$ primitivni korijen modulo p. Neka je $\mathcal{P} = \mathbb{Z}_p^*$, $\mathcal{C} = \mathbb{Z}_p^* \times \mathbb{Z}_p^*$ i

$$\mathcal{K} = \{(p, \alpha, a, \beta) \,:\, \beta \equiv \alpha^a \pmod{p}\}.$$

Vrijednosti p, α, β su javne, a vrijednost a je tajna.
Za $K \in \mathcal{K}$ i tajni slučajni broj $k \in \{0, 1, \ldots, p-1\}$ definiramo

$$e_K(x, k) = (\alpha^k \bmod p,\ x\beta^k \bmod p).$$

Za $y_1, y_2 \in \mathbb{Z}_p^*$ definiramo

$$d_K(y_1, y_2) = y_2 (y_1^a)^{-1} \bmod p.$$

Mogli bismo reći da se otvoreni tekst x "zamaskira" množeći s β^k. Onaj tko poznaje tajni eksponent a može iz α^k izračunati β^k i "ukloniti masku".

Da bi eksponent a stvarno bio tajan, prost broj p mora biti dovoljno velik da bi u $\mathbb{Z}_p^*$ problem diskretnog logaritma bio praktički nerješiv. Stoga se danas preporuča korištenje prostih brojeva od oko 1024 bita. Također bi red grupe, tj. broj $p-1$, trebao imati barem jedan veliki prosti faktor (od barem 160 bitova).

Primjer 3.7. *Neka je u ElGamalovom kriptosustavu* $p = 23$, $\alpha = 5$, $a = 17$, $\beta = 15$. *Dešifrirati šifrat* $(y_1, y_2) = (17, 6)$.

Rješenje: Računamo $y_1^a = 17^{17} \equiv 11 \pmod{23}$. Potom nađemo inverz od 11 modulo 23. Dobivamo da je inverz $-2 \equiv 21 \pmod{23}$. Konačno izračunamo $6 \cdot 21 \bmod 23$ i dobivamo otvoreni tekst $x = 11$. ♢

No, nije $\mathbb{Z}_p^*$ jedina grupa kod koje je potenciranje puno lakše od logaritmiranja. ElGamalov kriptosustav se može na sasvim isti način definirati i na multiplikativnoj grupi konačnog polja karakteristike 2, tj. $\mathbb{F}_{2^n}$. Najbolji poznati algoritmi za problem diskretnog logaritma u $\mathbb{F}_{2^n}$ trebaju

$$e^{O(n^{1/3}(\log n)^{2/3})}$$

operacija. Tako se smatra da je za $n > 1000$ odgovarajući kriptosustav siguran ukoliko $2^n - 1$ ima barem jedan veliki prosti faktor.

Dapače, ima grupa, poput grupe eliptičke krivulje nad konačnim poljem, kod kojih je razlika u težini ova dva problema (potenciranja i logaritmiranja) još veća.

Ideju o tome da bi eliptičke krivulje mogle biti korisne u konstrukciji kriptosustava s javnim ključem prvi su javno iznijeli Koblitz i Miller 1985. godine.

Definicija 3.2. *Neka je K polje karakteristike različite od 2 i 3.* Eliptička krivulja *nad poljem K je skup svih uređenih parova $(x, y) \in K \times K$ koji zadovoljavaju jednadžbu*

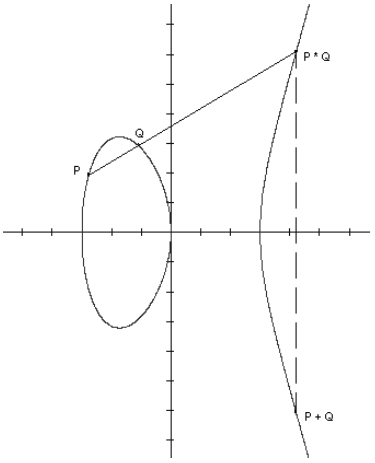
$$E: \qquad y^2 = f(x) = x^3 + ax + b,$$

gdje su $a, b \in K$ i polinom $f(x)$ nema višestrukih korijena, zajedno s "točkom u beskonačnosti" koju ćemo označavati sa $\mathcal{O}$. Taj skup označavamo s $E(K)$.

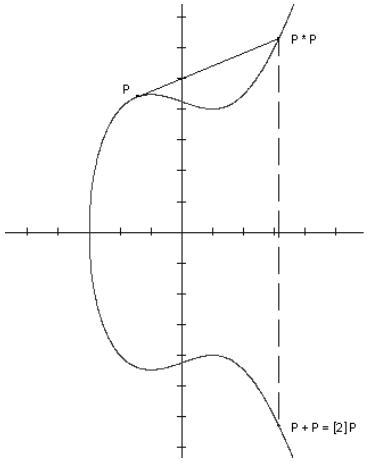
Vrlo slično se definira eliptička krivulja i nad poljima karakteristike 2 ili 3.

Jedno od najvažnijih svojstava eliptičkih krivulja jest da se na njima može, na prirodan način, uvesti operacija uz koju one postaju abelove grupe. Da bi to objasnili, uzmimo da je $K = \mathbb{R}$ polje realnih brojeva. Tada eliptičku krivulju $E(\mathbb{R})$ (bez točke u beskonačnosti) možemo prikazati kao podskup ravnine.

Definirat ćemo operaciju zbrajanja na $E(\mathbb{R})$. Neka su $P, Q \in E(\mathbb{R})$. Povucimo pravac kroz točke P i Q. On siječe krivulju E u tri točke. Treću točku označimo s $P * Q$. Sada definiramo da je $P + Q$ osnosimetrična točka točki $P*Q$ s obzirom na os x. Ako je $P = Q$, onda umjesto sekante povlačimo tangentu kroz točku P. Po definiciji stavljamo da je $P + \mathcal{O} = \mathcal{O} + P = P$ za svaki $P \in E(\mathbb{R})$.



sekanta



tangenta

Pokazuje se da skup $E(\mathbb{R})$ uz ovako definiranu operaciju zbrajanja postaje abelova grupa. Očito je $\mathcal{O}$ neutralni element, dok je $-P$ osnosimetrična točka točki P u odnosu na os x. Možemo zamišljati da je $\mathcal{O}$ treća točka presjeka od E s (vertikalnim) pravcem kroz P i $-P$. Komutativnost je također očita, a najteže je provjeriti asocijativnost.

Analitička geometrija nam omogućava da operaciju zbrajanja, koju smo definirali geometrijski, zapišemo pomoću algebarskih formula. Te formule nam omogućavaju da definiramo zbrajanje točaka na eliptičkoj krivulji nad

proizvoljnim poljem K (uz malu modifikaciju za slučaj polja s karakteristikom 2 i 3). Ponovo je skup $E(K)$, uz tako definirano zbrajanje, abelova grupa.

Za primjene eliptičkih krivulja u kriptografiji posebno je važan slučaj kada je polje $K = \mathbb{Z}_p$, ili općenitije kada je K konačno polje. Među konačnim poljima, pored polja $\mathbb{Z}_p$, najvažnija su polja karakteristike 2. Eliptičke krivulje imaju važnu primjenu i kod faktorizacije i dokazivanja prostosti.

Budući da je polovica elemenata iz $\mathbb{Z}_p^*$ jednaka kvadratu nekog elementa iz $\mathbb{Z}_p^*$, za očekivati je da $E(\mathbb{Z}_p)$ ima približno p točaka (ako je točka (x, y) u $E(\mathbb{Z}_p)$, onda je također i točka $(x, -y)$ u $E(\mathbb{Z}_p)$). Preciznije, po Hasseovom teoremu, vrijedi:

$$p + 1 - 2\sqrt{p} < |E(\mathbb{Z}_p)| < p + 1 + 2\sqrt{p}.$$

Nadalje, $E(\mathbb{Z}_p) \cong \mathbb{Z}_m \times \mathbb{Z}_k$ i vrijedi $k|m$ i $k|p-1$. Dakle, $E(\mathbb{Z}_p)$ je produkt dvije cikličke grupe.

Svi kriptosustavi koji u svojoj originalnoj definiciji koriste grupu $\mathbb{Z}_p^*$, kao što je npr. ElGamalov, mogu se vrlo lako modificirati tako da koriste grupu $E(\mathbb{Z}_p)$. No, doslovno prevođenje ElGamalovog kriptosustava u eliptičke krivulje ima nekoliko nedostataka.

Prvi je da prije šifriranja moramo elemente otvorenog teksta prebaciti u točke na eliptičkoj krivulji. Za to ne postoji zadovoljavajući deterministički algoritam. No, postoji vjerojatnosni algoritam, koji koristi činjenicu da kvadrati u konačnom polju predstavljaju 50% svih elemenata. To znači da s približnom vjerojatnošću $1-\frac{1}{2^k}$ možemo očekivati da ćemo iz k pokušaja pronaći broj x takav da je $x^3 + ax + b$ kvadrat u $\mathbb{Z}_p$. Za $k = 30$ to je sasvim zadovoljavajuća vjerojatnost. Pretpostavimo sada da su nam osnovne jedinice otvorenog teksta cijeli brojevi između 0 i M. Pretpostavimo nadalje da je $p > Mk$. Sada otvorenom tekstu m pridružujemo točku na eliptičkoj krivulji $E(\mathbb{Z}_p)$ na sljedeći način. Za brojeve x oblika $mk + j$, $j = 1, 2, \ldots, k$ provjeravamo je li $x^3 + ax + b$ kvadrat u $\mathbb{Z}_p$. Kad nađemo takav broj, izračunamo y koji zadovoljava da je $y^2 \equiv x^3 + ax + b \pmod{p}$, te broju m pridružimo točku (x, y) na $E(\mathbb{Z}_p)$. Obrnuto, iz točke (x, y) pripadni otvoreni tekst m možemo dobiti po formuli

$$m = \left\lfloor \frac{x-1}{k} \right\rfloor.$$

Drugi problem je da se šifrat jednog elementa otvorenog teksta kod ove varijante ElGamalovog kriptosustava sastoji od uređenog para točaka na eliptičkoj krivulji. To znači da, prilikom šifriranja, poruka postane otprilike 4 puta dulja.

Navest ćemo jednu varijantu ElGamalovog kriptosustava koja koristi eliptičke krivulje. Zove se *Menezes-Vanstoneov kriptosustav*. U njemu se eliptičke krivulje koriste samo za "maskiranje", dok su otvoreni tekstovi

i šifrati proizvoljni uređeni parovi elemenata iz polja (a ne nužno parovi koji odgovaraju točkama na eliptičkoj krivulji). Kod ovog kriptosustava, šifrirana poruka je (samo) 2 puta dulja od originalne poruke.

Menezes-Vanstoneov kriptosustav: Neka je E eliptička krivulja nad $\mathbb{Z}_p$ ($p > 3$ prost), te H ciklička podgrupa od E generirana s α. Neka je $\mathcal{P} = \mathbb{Z}_p^* \times \mathbb{Z}_p^*$, $\mathcal{C} = E \times \mathbb{Z}_p^* \times \mathbb{Z}_p^*$ i

$$\mathcal{K} = \{(E, \alpha, a, \beta) \,:\, \beta = a\alpha\},$$

gdje $a\alpha$ označava $\alpha + \alpha + \cdots + \alpha$ (a puta), a $+$ je zbrajanje točaka na eliptičkoj krivulji.
Vrijednosti E, α, β su javne, a vrijednost a je tajna.
Za $K \in \mathcal{K}$ i tajni slučajni broj $k \in \{0, 1, \ldots, |H| - 1\}$, te za $x = (x_1, x_2) \in \mathbb{Z}_p^* \times \mathbb{Z}_p^*$ definiramo

$$e_K(x, k) = (y_0, y_1, y_2),$$

gdje je $y_0 = k\alpha$, $(c_1, c_2) = k\beta$, $y_1 = c_1 x_1 \bmod p$, $y_2 = c_2 x_2 \bmod p$.
Za šifrat $y = (y_0, y_1, y_2)$ definiramo

$$d_K(y) = (y_1 (c_1)^{-1} \bmod p,\, y_2 (c_2)^{-1} \bmod p),$$

gdje je $a y_0 = (c_1, c_2)$.

Kao što smo već spomenuli, glavni razlog za uvođenje eliptičkih krivulja u kriptografiju javnog ključa jest taj da je problem diskretnog logaritma u grupi $E(\mathbb{Z}_p)$ još teži od problema diskretnog logaritma u grupi $\mathbb{Z}_p^*$.

To pak znači da se ista sigurnost može postići s manjim ključem. Tako je npr. umjesto ključa duljine duljine 1024 bita, dovoljan ključ duljine 160 bitova. To je osobito važno kod onih primjena (kao što su npr. čip-kartice) kod kojih je prostor za pohranu ključeva vrlo ograničen.

3.4.3 Merkle-Hellmanov kriptosustav

Merkle-Hellmanov kriptosustav iz 1978. godine za osnovu ima *problem ruksaka*. Pretpostavimo da imamo n predmeta s volumenima v_1, v_2, ... , v_n kojima želimo napuniti ruksak volumena V. Dakle, želimo naći $J \subseteq \{1, 2, \ldots, n\}$ tako da je $\sum_{j \in J} v_j = V$ (ako takav podskup postoji). Ekvivalentna formulacija je:

Problem ruksaka. Za dani skup $\{v_1, v_2, \ldots, v_n\}$ od n prirodnih brojeva i prirodan broj V, naći niz $m = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)$ od n binarnih znamenaka ($\varepsilon_i \in \{0, 1\}$) tako da je

$$\varepsilon_1 v_1 + \varepsilon_2 v_2 + \cdots + \varepsilon_n v_n = V,$$

ako takav m postoji.

Poznato je da je ovaj opći problem ruksaka vrlo težak. On spada u tzv. NP-potpune probleme. To, pored ostalog, znači da nije poznat polinomijalni algoritam za njegovo rješavanje. Međutim, jedan njegov specijalni slučaj, *superrastući problem ruksaka*, je puno lakši. To je slučaj kad je niz v_1, v_2, ..., v_n rastući i vrijedi

$$v_j > v_1 + v_2 + \cdots v_{j-1} \quad \text{za } j = 2, 3, ..., n.$$

Primjer superrastućeg niza je niz $v_i = 2^{i-1}$. Tada su ε_i-ovi upravo binarne znamenke broja V. Jasno je da u slučaju superrastućeg niza, u svakom koraku u ruksak moramo staviti najveći predmet koji u njega stane. To vodi do sljedećeg algoritma:

Algoritam za superrastući problem ruksaka za $\{v_1, \ldots, v_n, V\}$**:**

for $i = n, \ldots, 1$ do
    if $(V \geq v_i)$ then $V = V - v_i$; $\varepsilon_i = 1$
        else $\varepsilon_i = 0$
if $(V = 0)$ then return "$(\varepsilon_1, \ldots, \varepsilon_n)$ je rješenje"
    else return "nema rješenja"

Ideja Merkle-Hellmanovog kriptosustava je "zamaskirati" superrastući niz tako da izgleda kao sasvim slučajan niz. Onaj kome je poruka namjenjena (Bob) zna kako ukloniti masku, pa može pročitati poruku rješavajući supperrastući problem ruksaka. Svi drugi moraju rješavati, puno teži, opći problem ruksaka, pa ne mogu pročitati poruku. "Maskiranje" se provodi modularnim množenjem.

Merkle-Hellmanov kriptosustav. Neka je $v = (v_1, v_2, \ldots, v_n)$ superrastući niz prirodnih brojeva, te neka je $p > v_1 + v_2 + \cdots + v_n$ prost broj i $1 \leq a \leq p-1$. Za $i = 1, 2, \ldots n$, definiramo

$$t_i = av_i \bmod p$$

i označimo $t = (t_1, t_2, \ldots, t_n)$. Neka je

$$\mathcal{P} = \{0,1\}^n, \quad \mathcal{C} = \{0, 1, \ldots, n(p-1)\} \quad \text{i} \quad \mathcal{K} = \{(v, p, a, t)\},$$

gdje su v, p, a i t konstruirani na gore opisani način.
Za $K \in \mathcal{K}$ definiramo

$$e_K(x_1, x_2, \ldots, x_n) = x_1t_1 + x_2t_2 + \cdots + x_nt_n.$$

Za $0 \leq y \leq n(p-1)$ definiramo $z = a^{-1}y \bmod p$, riješimo (superrastući) problem ruksaka za skup $\{v_1, v_2, \ldots, v_n, z\}$ i tako dobivamo

$$d_K(y) = (x_1, x_2, \ldots, x_n).$$

Vrijednost t je javna, dok su vrijednosti p, a i v tajne.

Primjer 3.8. *Zadan je Merkle-Hellmanov kriptosustav s parametrima*

$$v = (2, 5, 11, 23, 45, 91), \;\; p = 181, \;\; a = 111,$$
$$t = (41, 12, 135, 19, 108, 146).$$

Dešifrirati šifrat $y = 296$.

Rješenje: Najprije izračunamo inverz od a modulo p (Euklidovim algoritmom). Dobivamo da je inverz 106. Zatim računamo $z = 106 \cdot 296 \bmod 181 = 63$. Sada riješimo superrastući problem ruksaka za v i z. Dobivamo da je $63 = 45 + 11 + 5 + 2$, pa je otvoreni tekst

$$(x_1, x_2, x_3, x_4, x_5, x_6) = (1, 1, 1, 0, 1, 0).$$

Provjera: $e_K(x_1, x_2, x_3, x_4, x_5, x_6) = t_1 + t_2 + t_3 + t_5 = 296$. ♢

Merkle-Hellmanov kriptosustav ima jednu vrlo veliku prednost u odnosu na ostale kriptosustave s javnim ključem. Naime, šifranje njime je znatno brže, te je on po brzini bio usporediv s najboljim simetričnim kriptosustavima. No, godine 1982. je uslijedilo razočaranje, kad je Shamir pronašao polinomijalni algoritam za razbijanje Merkle-Hellmanovog kriptosustava. Pokazalo se da se ovako jednostavnim maskiranjem vrlo specijanog niza ipak

ne dobiva sasvim slučajan niz. U razbijanju se koriste algoritmi za diofantske aproksimacije (verižni razlomci i LLL-algoritam).

Ideja je promotriti rešetku L u prostoru $\mathbb{R}^{n+1}$ generiranu vektorima $b_1 = (1, 0, 0, \ldots, 0, -t_1)$, $b_2 = (0, 1, 0, \ldots, 0, -t_2)$, $\ldots$, $b_n = (0, 0, \ldots, 0, 1, -t_n)$, $b_{n+1} = (0, 0, 0, \ldots, 0, y)$. Ako je $x_1 t_1 + \cdots x_n t_n = y$ (a to je problem koji treba riješiti kod dekriptiranja), onda je

$$x = (x_1, x_2, \ldots, x_n, 0) = \sum_{i=1}^{n} x_i \cdot b_i + b_{n+1}.$$

Dakle, x je element rešetke L, a budući da su mu sve komponente jednake 0 ili 1, on je "kratak". Zato se može očekivati da ćemo ga moći dobiti kao jedan od elemenata LLL-reducirane baze, dobivene LLL-algoritmom. Više o kriptoanalizi Merkle-Hellmanovog, a i nekih drugih kriptosustava zasnovanih na problemu ruksaka, može se naći u članku A. M. Odlyzka iz 1990. godine, znakovita naslova "The rise and fall of knapsack cryptosystems".

Prema tome se Merkle-Hellmanov kriptosustav ne može više smatrati sigurnim kriptosustavom. Ipak, ideja na kojoj je zasnovan je vrlo zanimljiva. Ta ideja je korištenje jednostavnog specijalnog slučaja nekog teškog (NP-potpunog) problema, s time da se taj specijalni slučaj prikrije tako da izgleda kao opći.

3.4.4 McElieceov kriptosustav

Slična ideja onoj, koju smo vidjeli kod Merkle-Hellmanovog kriptosustava, koristi se i u McElieceovom kriptosustavu. Ovdje je pripadni NP-potpuni problem dekodiranje općih linearnih kodova za ispravljanje grešaka. Kao osnova u ovom kriptosustavu koristi se specijalna klasa *Goppinih kodova* za koje postoji polinomijalni algoritam za dekodiranje.

Neka su k i n prirodni brojevi, $k \leq n$. *Binarni linearni* $[n,k]$*-kôd* C je k-dimenzionalni potprostor vektorskog prostora $\mathbb{Z}_2^n$. Generirajuća matrica za linearni kôd C je $k \times n$ matrica G čiji redci tvore bazu za C.

Za $x = (x_1, \ldots, x_n)$, $y = (y_1, \ldots, y_n) \in \mathbb{Z}_2^n$, definiramo *Hammingovu udaljenost* s

$$d(x,y) = |\{i \,:\, 1 \leq i \leq n, x_i \neq y_i\}|,$$

tj. kao broj koordinata u kojima se x i y razlikuju. Neka je C neki linearni $[n,k]$-kôd. Definiramo *minimalnu udaljenost* od C s

$$d = d(C) = \min\{d(x,y) \,:\, x,y \in C, x \neq y\}.$$

Tada za C kažemo da je $[n,k,d]$-kôd.

Primjer jednog $[7,4,3]$-koda dan je sljedećom generirajućom matricom:

$$G = \begin{bmatrix} 1&0&0&0&1&1&0 \\ 0&1&0&0&1&0&1 \\ 0&0&1&0&0&1&1 \\ 0&0&0&1&1&1&1 \end{bmatrix}$$

Svrha kodova za ispravljanje grešaka jest da isprave slučajne greške koje mogu nastati prilikom prenošenja binarnih podataka preko kanala sa "šumom" (engl. chanell with noise). Neka je G generirajuća matrica za $[n,k,d]$-kôd C. Pretpostavimo da je x binarna k-torka koju Alice želi prenijeti Bobu preko kanala sa šumom. Tada Alice *kodira* x kao n-torku $y = xG$, te pošalje y preko kanala. Bob primi n-torku r koja se, budući da kanal ima šum, ne mora podudarati s y. Da bi *dekodirao* r, Bob traži element ("kodnu riječ") $y' \in C$ koja ima najmanju Hammingovu udaljenost od r, te potom izračuna k-torku x' takvu da je $y' = x'G$. Tada Bob može očekivati da je $y' = y$, pa onda i $x' = x$, tj. da je uspio ispraviti sve greške nastale prilikom prijenosa. Nije teško vidjeti da ukoliko broj grešaka nije veći od $(d-1)/2$, onda se na ovaj način mogu "ispraviti sve greške", tj. može se jednoznačno rekonstruirati poslana poruka.

Pokazuje se da je problem nalaženja najbliže kodne riječi vrlo težak problem za opće linearne kodove. No, postoje kodovi za koje postoje efikasni algoritmi za dekodiranje. McEliece je 1978. godine predložio da se jedna klasa takvih kodova, tzv. Goppini kodovi, iskoristi u konstrukciji kriptosustava s javnim ključem. Parametri Goppinih kodova imaju oblik $n = 2^m$, $d = 2t+1$, $k = n - mt$.

McElieceov kriptosustav. Neka je G generirajuća matrica za $[n,k,d]$-kôd C. Neka je S invertibilna $k \times k$ matrica nad $\mathbb{Z}_2$, te P $n \times n$ permutacijska matrica (u svakom retku i svakom stupcu ima točno jednu jedinicu, a svi ostali elementi su nule). Stavimo $G' = SGP$. Sada definiramo $\mathcal{P} = \mathbb{Z}_2^k$, $\mathcal{C} = \mathbb{Z}_2^n$, $\mathcal{K} = \{(G,S,P,G')\}$, gdje su G,S,P,G' konstruirani na prije opisani način. Ovdje je G' javni, a G,S,P tajni ključ.
Za $K = (G,S,P,G') \in \mathcal{K}$, definiramo

$$e_K(x,e) = xG' + e,$$

gdje je $e \in \mathbb{Z}_2^n$ slučajno generirani vektor težine $t = (d-1)/2$ (ima t jedinica, a ostalo nule).
Šifrat $y \in \mathbb{Z}_2^n$ se dešifrira na sljedeći način. Najprije se izračuna $y_1 = yP^{-1}$. Potom se dekodira y_1, i tako se dobije kodna riječ $x_1 \in C$ koja je najbliža y_1 (budući da je P permutacijska matrica, imamo da je $y_1 = xSG + e_1$, gdje je $e_1 = eP^{-1}$ binarni niz težine $t = (d-1)/2$, pa znamo da mora vrijediti da je $x_1 = xSG$). Izračuna se $x_0 \in \mathbb{Z}_2^k$, takav da je $x_0G = x_1$. Konačno se izračuna $x = x_0S^{-1}$.

Uočimo da y_1 možemo dekodirati jer je kodiran Goppinim kodom G, a y ne možemo jer je kodiran s G' za koji nemamo efikasan algoritam za dekodiranje. McEliece je predložio korištenje Goppinog koda s parametrima $[1024, 524, 101]$. Otvoreni tekst je binarna 524-torka, a odgovarajući šifrat je binarna 1024-torka. Javni ključ je 524×1024 binarna matrica. Ovaj kriptosustav nije korišten u praksi, prvenstveno zbog ogromne veličine javnog ključa.

3.4.5 NTRU kriptosustav

Jedan od najzanimljivijih novijih kriptosustava, koji je još uvijek predmet intenzivnog proučavanja, je *NTRU kriptosustav*, koji su 1997. godine predložili Hoffstein, Pipher i Silverman s *Brown University* (ima više objašnjenja za skraćenicu NTRU: *Number Theory Research Unit* ili *Number Theorists aRe Us* ili *N-th degree* tru*ncated polynomial ring*). U ovom se kriptosustavu kod šifriranja koriste polinomi. Preciznije, koristi se prsten $R = \mathbb{Z}[X]/(X^n - 1)$. Na elementima od R definira se operacija *cikličke konvolucije*, tj. za $F = \sum_{i=0}^{n-1} F_i x^i$, $G = \sum_{i=0}^{n-1} G_i x^i$, definira se $H = F \circledast G = \sum_{k=0}^{n-1} H_k x^k \in R$, s

$$H_k = \sum_{i+j \equiv k \pmod n} F_i G_j.$$

Pored toga se koristi redukcija ovako dobivenih polinoma po dvama relativno prostim modulima p i q. Sigurnost ovog kriptosustava se upravo zasniva na nezavisnosti tih dviju redukcija.

Da bi kreirao svoje NTRU ključeve, Bob na slučajan način izabire dva polinoma $f, g \in R$. Polinom f mora zadovoljavati uvjet da ima inverze modulo p i modulo q. Označimo te inverze s F_p i F_q, tj.

$$F_p \circledast f \equiv 1 \pmod{p}, \quad F_q \circledast f \equiv 1 \pmod{q}.$$

Bob zatim izračuna polinom h takav da je

$$h \equiv F_q \circledast g \pmod{q}.$$

Sada je h Bobov javni, a f tajni ključ.

Pretpostavimo da Alice želi poslati Bobu poruku m, za koju ćemo pretpostaviti da je također element skupa R. Ona izabire slučajan polinom $r \in R$ i koristi Bobov javni ključ da izračuna $e \equiv p \cdot r \circledast h + m \pmod{q}$. Alice pošalje Bobu šifrat e.

Kad Bob primi šifrat e, najprije pomoću svog tajnog ključa f izračuna

$$a \equiv f \circledast e \pmod{q},$$

gdje su koeficijenti polinoma a izabrani iz sustava najmanjih ostaka po apsolutnoj vrijednosti, tj. iz segmenta $[-\frac{q}{2}, \frac{q}{2}]$. Sada Bob računa originalni otvoreni tekst m iz

$$m \equiv F_p \circledast a \pmod{p}.$$

U praktičnoj implementaciji, polinomi f, g, r i m se biraju iz fiksiranih podskupova R_f, R_g, R_r, R_m od R, koji se sastoje od polinoma s koeficijentima iz skupa $\{-1, 0, 1\}$ sa zadanim brojem ne-nul koeficijenata. Ove skupove također smatramo parametrima kriptosustava. Tipičan izbor za parametre p i q je $p = 3$, $q = 128$ (i naravno, ti parametri su javni).

Provjerimo hoće li Bob na ovaj način zaista dešifrirati šifrat. Imamo:

$$\begin{aligned} a &\equiv f \circledast e \equiv f \circledast (p \cdot r) \circledast h + f \circledast m \equiv f \circledast (p \cdot r) \circledast F_q \circledast g + f \circledast m \\ &\equiv p \cdot r \circledast g + f \circledast m \pmod{q}. \end{aligned}$$

Izbor parametara nam osigurava da će (skoro uvijek) ovaj posljednji polinom imati koeficijente iz segmenta $[-\frac{q}{2}, \frac{q}{2}]$. To znači da vrijedi jednakost

$$a = p \cdot r \circledast g + f \circledast m \quad \text{u } R.$$

Zato je

$$F_p \circledast a \equiv F_p \circledast f \circledast m \equiv m \pmod{p}.$$

Šifriranje i dešifriranje kod NTRU kriptosustava je brže nego kod npr. RSA kriptosustava, što svakako predstavlja jednu njegovu prednost. No,

poznato je nekoliko mogućih napada na NTRU koji koriste LLL-algoritam za nalaženje najkraćeg elementa u rešetki. Također, mogući su i napadi koji koriste činjenicu da u NTRU-u postoje šifrati koje je nemoguće dešifrirati (postoji vrlo mala vjerojatnost da će neki koeficijent polinoma $p \cdot r \circledast g + f \circledast m$ biti izvan segmenta $[-\frac{q}{2}, \frac{q}{2}]$). Zasad nije sasvim jasno koliko su ti napadi ozbiljna prijetnja za sigurnost ovog kriptosustava. Svakako će na to pitanje trebati odgovoriti, prije nego što dođe do eventualnog ulaska ovog kriptosustava u najširu uporabu.

Jedna potencijalna, za sada samo teoretska, prednost NTRU kriptosustava u odnosu na RSA je vezana uz pitanje što bi se dogodilo s njihovom sigurnošću ako bi se uspjelo konstruirati kvantna računala. Za razliku od klasičnih računala kod kojih je osnovna jedinica informacije jedan bit (koji može biti 0 ili 1), kvantna računala bi koristila ideje iz kvantne mehanike, te bi kod njih osnovna jedinica informacije - qubit - nosila puno više informacija. Takva računala još nisu praktično realizirana u obliku koji bi bio konkurencija klasičnim računalima, ali takva realizacija nije isključena u doglednoj budućnosti. Stoga se u posljednjih 15-tak godina radi i na algoritmima specijalno dizajniranim baš za takva računala.

Od poznatih algoritama za kvantna računala, dva su vrlo relevantna za kriptoanalizu: Groverov i Shorov. Groverov algoritam ubrzava pretraživanja nesortiranih datoteka, te pomoću njega napadi "grubom silom" postaju puno učinkovitiji (kod simetričnih kriptosustava to bi značilo da će biti potreban dvostruko duži ključ da se zadrži sadašnja razina sigurnosti). Shorov algoritam koristi činjenicu da kvantne metode omogućavaju vrlo brzo računanje perioda periodičnih funkcija. To daje polinomijalne kvantne algoritme za probleme faktorizacije i diskretnog algoritma. Znači da bi efektivna konstrukcija dovoljno snažnih kvantnih računala učinila neupotrebljivim kriptosustave javnog ključa zasnovane na faktorizaciji (RSA, Rabin) i problemu diskretnog logaritma (ElGamal, ECC). Od kriptosustava koje smo mi spominjali, čini se da bi McElieceov i NTRU možda mogli biti sigurni i u eri kvantnih računala. Pitanja tzv. post-kvantne kriptografije, tj. kriptosustava dizajniranih da budu otporni na napade koje bi realizirala kvantna računala, su predmet intenzivnih istraživanja.