Kódování

Cílem kódování je přenášet informace tak, abychom postřehli, pokud by došlo k jejich zkreslení náhodnou chybou, v ideálním případě dokonce takto vzniklou chybu nejen odhalit, ale i opravit.

Použití: ukládání dat (paměť počítače, CD, DVD, Blue-ray, čárové kódy, QR-kódy), komunikace (digitální TV, vesmírné sondy), atd.

Přenášená informace bude zapsána pomocí abecedy, která má p písmen (tj. jakýchsi symbolů), kde p je pevně zvolené prvočíslo. Tuto abecedu tedy můžeme ztotožnit s množinou $\mathbb{Z}_p$ všech zbytkových tříd modulo p. Přenášet budeme slova délky n, každé takové kódové slovo $a_1 a_2 a_3 \dots a_{n-1} a_n$ lze tedy chápat jako polynom

$$a(x) = a_1 x^{n-1} + a_2 x^{n-2} + \cdots + a_{n-1} x + a_n \in \mathbb{Z}_p[x]$$

stupně $\mathrm{st}(a(x)) < n$. Číslo n nazýváme délka kódu.

Kdyby každý polynom stupně menšího než n bylo některé z kódových slov, tak bychom nemohli postřehnout, že při přenosu došlo k nějaké náhodné chybě.

Polynomiální kód délky n daný polynomem $g(x) \in \mathbb{Z}_p[x]$

Máme tedy pevně zvolené prvočíslo p a přirozené číslo n.

Zafixujme také polynom $g(x) \in \mathbb{Z}_p[x]$ stupně $\mathrm{st}(g(x)) = k < n$.

Kódová slova budou ty polynomy $a(x) \in \mathbb{Z}_p[x]$ stupně $\mathrm{st}(a(x)) < n$, které jsou dělitelné polynomem $g(x)$ v okruhu $\mathbb{Z}_p[x]$, jsou to tedy polynomy $a(x) = g(x) \cdot h(x)$, kde $h(x) \in \mathbb{Z}_p[x]$ je libovolný polynom stupně $\mathrm{st}(h(x)) < n - k$.

Pokud chceme odeslat zprávu zapsanou pomocí $n - k$ písmen, tj. polynom $b(x) = b_1 x^{n-k-1} + \cdots + b_{n-k-1}x + b_{n-k} \in \mathbb{Z}_p[x]$ stupně $\mathrm{st}(b(x)) < n - k$, vydělíme polynom $x^k \cdot b(x)$ polynomem $g(x)$ se zbytkem a dostaneme polynomy $q(x)$, $r(x) \in \mathbb{Z}_p[x]$ tak, že $x^k \cdot b(x) = g(x) \cdot q(x) + r(x)$, kde $\mathrm{st}(r(x)) < k$.

Odešleme pak polynom $g(x) \cdot q(x) = x^k \cdot b(x) - r(x)$.

Každé kódové slovo se tedy skládá z $n - k$ významových písmen (daných polynomem $b(x)$) následovaných k kontrolními písmeny (daných polynomem $-r(x)$). Je však nutné vhodně zvolit polynom $g(x)$. Určitě by nebyla vhodná volba $g(x) = x^k$, protože pak bychom každou zprávu $b(x)$ doplnili nulovým polynomem.

Příklad pro prvočíslo $p = 2$, tedy $\mathbb{Z}_2 = \{0, 1\}$

Zvolme polynom $g(x) = x + 1 \in \mathbb{Z}_2[x]$, tedy $k = \mathrm{st}(g(x)) = 1$.
Zvolme libovolnou délku kódu $n > 1$.
Odesílanou zprávou je nějaký polynom $b(x) \in \mathbb{Z}_2[x]$ stupně $\mathrm{st}(b(x)) < n - 1$.
Naše kódová slova se tedy budou skládat z $n - 1$ významových písmen (to jsou koeficienty polynomu $b(x)$) následovaných jedním kontrolním písmenem. Kontrolní písmeno určujeme takto: polynom $x \cdot b(x)$ vydělíme polynomem $g(x) = x + 1$ se zbytkem a dostaneme polynomy $q(x)$, $r(x) \in \mathbb{Z}_2[x]$ tak, že

$$x^k \cdot b(x) = g(x) \cdot q(x) + r(x) = (x + 1) \cdot q(x) + r(x),$$

kde $\mathrm{st}(r(x)) < k = 1$.
Tedy polynom $r(x) \in \mathbb{Z}_2[x]$ je konstantní, tj. $r(x) \in \{0, 1\}$.
Dosazením $x = 1$ dostaneme $r(1) = b(1)$, a proto $r(x) = b(1)$.
Protože $b(1) = 0$, má-li odesílaná zpráva sudý počet jedniček, a $b(1) = 1$, má-li odesílaná zpráva lichý počet jedniček, doplňujeme zprávu jedním písmenem tak, aby celkový počet jedniček byl sudý.
Kód tedy pozná, že došlo k jedné chybě, opravit ji neumí. Pokud došlo ke dvěma chybám, nic nepozná.

Metrický prostor, Hammingova vzdálenost kódových slov

Metrickým prostorem rozumíme nějakou neprázdnou množinu M (její prvkům říkáme body) spolu s metrikou na množině M, což je zobrazení $\rho : M \times M \to \mathbb{R}_0^+$, kde $\mathbb{R}_0^+$ značí množinu nezáporných reálných čísel, splňující pro každé $x, y, z \in M$

- $\rho(x, y) = 0 \iff x = y$,
- $\rho(x, y) = \rho(y, x)$,
- $\rho(x, y) + \rho(y, z) \geq \rho(x, z)$.

Při práci v metrickém prostoru můžeme používat svou geometrickou intuici, například mluvit o koulích s daným středem a daným poloměrem (jde o množinu bodů, jejichž vzdálenost od daného středu je menší než daný poloměr).

Pro každé dva polynomy $a(x)$, $b(x) \in \mathbb{Z}_p[x]$ stupnů $\mathrm{st}(a(x)) < n$, $\mathrm{st}(b(x)) < n$, definujeme jejich Hammingovu vzdálenost jako počet nenulových koeficientů rozdílu $a(x) - b(x)$, tj. počet koeficientů, v nichž se oba polynomy liší. Uvědomte si, že jde o metrický prostor.

Užití Hammingovy vzdálenosti kódových slov

Máme-li být schopni postřehnout, že došlo k chybě, pokud při přenosu bylo právě na jedné pozici přenášené písmeno náhodné změněno, je třeba, aby vzdálenost libovolných dvou různých kódových slov byla alespoň 2.

Pokud tato vzdálenost libovolných dvou různých kódových slov bude alespoň 3, budeme dokonce schopni takovou chybu i opravit.

Obecněji, je-li pro nějaké $t \in \mathbb{N}$ vzdálenost libovolných dvou různých kódových slov alespoň $t + 1$, pak lze chybu detekovat, když došlo při přenosu ke změně na nejvýše t pozicích. Je-li tato vzdálenost alespoň $2t + 1$, pak takovou chybu lze dokonce i opravit.

Protože u polynomiálního kódu je rozdíl libovolných dvou kódových slov opět kódové slovo, lze místo o nejmenší vzdálenosti dvou různých kódových slov hovořit o nejmenší vzdálenosti nenulového kódového slova od nuly.

Příklad

Zvolme $p = 2$, tedy písmena jsou 0 a 1. Dále položme $n = 5$, $g(x) = x^2 + x + 1$. Pak kódová slova jsou polynomy

$$0, \quad x^2 + x + 1, \quad x^3 + x^2 + x, \quad x^3 + 1,$$
$$x^4 + x^3 + x^2, \quad x^4 + x^3 + x + 1, \quad x^4 + x, \quad x^4 + x^2 + 1.$$

Budeme-li polynomy psát jako posloupnosti koeficientů, budeme kódovat takto:

$$\begin{aligned} 000 &\mapsto 00000, & 100 &\mapsto 10010, \\ 001 &\mapsto 00111, & 101 &\mapsto 10101, \\ 010 &\mapsto 01001, & 110 &\mapsto 11011, \\ 011 &\mapsto 01110, & 111 &\mapsto 11100. \end{aligned}$$

Je ihned vidět, že nejmenší vzdálenost nenulového kódového slova od nuly je 2, jsme tedy schopni detekovat chybu na jedné pozici. Opravit tuto chybu nejsme obecně schopni, například posloupnost 01000 by mohla vzniknout jednou chybou na druhé pozici anebo jednou chybou na páté pozici.

Kód opravující chybu na jedné pozici

Věta 1. Zvolme prvočíslo p a přirozené číslo $m > 1$. Nechť K je těleso mající právě p^m prvků, nechť $\alpha \in K$ je libovolný generátor multiplikativní grupy $K^\times$ tělesa K a $g(x)$ je minimální polynom prvku α nad tělesem $\mathbb{Z}_p$. Pak polynomiální kód délky $n = \frac{p^m-1}{p-1}$ daný polynomem $g(x)$ je schopen opravit chybu na jedné pozici.

Důkaz. Platí $\mathrm{st}(g(x)) = m < 1 + p + \cdots + p^{m-1} = n$. Stačí tedy ukázat, že žádné kódové slovo nemá vzdálenost 1 nebo 2 od nuly. Předpokládejme naopak, že takové kódové slovo existuje, tj. pro nějaké $i \in \{0, 1, \ldots, n-1\}$, $a \in \mathbb{Z}_p^\times$, je polynom ax^i kódové slovo, anebo pro nějaká $0 \leq i < j < n$, $a, b \in \mathbb{Z}_p^\times$, je polynom $ax^j + bx^i$ kódové slovo. Protože polynom $g(x)$ je nesoudělný s polynomem x, první případ $g(x) \mid ax^i$ vede ke sporu. Druhý případ $g(x) \mid ax^j + bx^i = a(x^{j-i} + ba^{-1})x^i$ dává $g(x) \mid x^{j-i} + ba^{-1}$, tedy $\alpha^{j-i} = -ba^{-1} \in \mathbb{Z}_p^\times$, odkud $\alpha^{(j-i)(p-1)} = 1$. Protože řád prvku α v grupě $K^\times$ je $p^m - 1$, dostáváme $p^m - 1 \mid (j-i)(p-1)$, tj. $n \mid j - i$, což je ve sporu s tím, že $0 < j - i < n$.

Užití kódu z věty 1 - kódování

Parametry kódu: p prvočíslo, $m \in \mathbb{N}$, $m > 1$,
K těleso, $|K| = p^m$, $K^\times = \langle \alpha \rangle$,
$g(x)$ minimální polynom α nad $\mathbb{Z}_p$,
$n = \frac{p^m - 1}{p-1} = p^{m-1} + p^{m-2} + \cdots + p + 1$.

Nutný předpoklad pro správnou funkci kódu: K chybám při přenosu dochází tak zřídka, že lze očekávat, že při odeslání n písmen (tj. koeficientů odesílaného polynomu) bude nejvýše jedno písmeno přijato chybně.

Zpráva: polynom $b(x) \in \mathbb{Z}_p[x]$, $\mathrm{st}(b(x)) < n - m$,
dělením se zbytkem:
$x^m \cdot b(x) = g(x) \cdot q(x) + r(x)$, kde $\mathrm{st}(r(x)) < m$.

Odeslaná informace:
$g(x) \cdot q(x) = x^m \cdot b(x) - r(x)$.

Užití kódu z věty 1 - dekódování

Parametry kódu: těleso K, $|K| = p^m$, p prvočíslo, $m > 1$, $K^\times = \langle\alpha\rangle$, $g(x)$ minimální polynom α nad $\mathbb{Z}_p$, $n = \frac{p^m-1}{p-1}$.

Přijatá informace: polynom $h(x) \in \mathbb{Z}_p[x]$, $\mathrm{st}(h(x)) < n$.

Pokud nedošlo při přenosu **k žádné chybě**, byl polynom $h(x)$ odeslán, proto $g(x) \mid h(x)$, tj. $h(\alpha) = 0$.

Pokud došlo při přenosu **k jediné chybě**, byl odeslán polynom $h(x) - cx^j$, kde $c \in \mathbb{Z}_p^\times$, $0 \leq j < n$. Potřebujeme určit c, j. Platí $g(x) \mid h(x) - cx^j$, tj. $h(\alpha) = c\alpha^j \neq 0$. Protože $h(\alpha) \in K^\times = \langle\alpha\rangle$, existuje jediné $t \in \mathbb{Z}$, $0 \leq t < p^m - 1$ splňující $h(\alpha) = \alpha^t$. Toto t nalezneme, pak $c = \alpha^{t-j} \in \mathbb{Z}_p^\times$. Z malé Fermatovy věty $1 = c^{p-1} = \alpha^{(t-j)(p-1)}$. Protože řád α je $p^m - 1$, platí $p^m - 1 \mid (t-j)(p-1)$, tj. $n = \frac{p^m-1}{p-1} \mid t - j$. Číslo j tedy nalezneme jako zbytek po dělení čísla t číslem n a víme, že $c = \alpha^{t-j} \in \mathbb{Z}_p^\times$.

Odeslaný polynom: je-li $h(\alpha) = 0$, byl odeslán $h(x)$; je-li $h(\alpha) \neq 0$, byl odeslán $h(x) - cx^j$ (pro výše určené c, j).

Kód Reed–Solomon opravující chyby na více pozicích

Věta 2. *Zvolme prvočíslo p a přirozené číslo m. Nechť K je těleso mající právě p^m prvků, nechť $\alpha \in K$ je libovolný generátor multiplikativní grupy $K^\times$ tělesa K. Nechť $r \geq -1$, $t > 0$ jsou celá čísla. Předpokládejme, že polynom $g(x)$ je dělitelný minimálním polynomem prvku α^{r+j} pro každé $j = 1, 2, \ldots, 2t$ a platí $\mathrm{st}(g(x)) < p^m - 1$. Pak polynomiální kód délky $n = p^m - 1$ daný polynomem $g(x)$ je schopen opravit chybu na t pozicích.*

Důkaz. Protože $\mathrm{st}(g(x)) < p^m - 1$, existuje v $K^\times$ prvek, který není kořen $g(x)$, a tedy $2t < n$. Předpokládejme, že existuje nenulové kódové slovo, jehož vzdálenost od nuly nepřevyšuje $2t$. Existují tedy $b_1, \ldots, b_{2t} \in \mathbb{Z}_p$, ne všechny nuly, a $0 \leq k_1 < k_2 < \cdots < k_{2t} < n$ tak, že polynom $f(x) = \sum_{i=1}^{2t} b_i x^{k_i}$ je kódové slovo, tj. $g(x) \mid f(x)$ a $f(\alpha^{r+j}) = 0$ pro každé $j = 1, 2, \ldots, 2t$. Pak $(\alpha^{(r+j)k_i})_{j,i=1,2,\ldots,2t}$ je matice s lineárně závislými sloupci. Ovšem pro $k = \sum_{i=1}^{2t} k_i$ platí $0 = \det(\alpha^{(r+j)k_i})_{j,i=1,2,\ldots,2t} = \alpha^{(r+1)k} \cdot \prod_{1 \leq i < j \leq 2t} (\alpha^{k_j} - \alpha^{k_i})$ užitím vzorce pro Vandermondův determinant. Ale to je součin mající pouze nenulové činitele, neboť α má řád n, spor.

Příklad

Nechť $K = \mathbb{Z}_2[y]/(y^4 + y + 1)$, označme $\alpha = [y]_{y^4+y+1}$ třídu obsahující polynom y.

Minimální polynom prvků $\alpha, \alpha^2, \alpha^4, \alpha^8$ je $x^4 + x + 1$.

Minimální polynom prvků $\alpha^3, \alpha^6, \alpha^{12}, \alpha^9$ je $x^4 + x^3 + x^2 + x + 1$.

Minimální polynom prvků α^5, α^{10} je $x^2 + x + 1$.

Proto předpoklady předchozí věty pro $p = 2$, $m = 4$, $n = 15$, $r = 0$, $t = 3$ splňuje polynom

$$g(x) = (x^4 + x + 1) \cdot (x^4 + x^3 + x^2 + x + 1) \cdot (x^2 + x + 1) = \\ = x^{10} + x^8 + x^5 + x^4 + x^2 + x + 1.$$

Odpovídající kód délky 15 má 5 významových a 10 kontrolních písmen. Je schopen opravit chyby až na třech pozicích.