1 Co jsou lineární kódy

Žádný záznam informace a žádný přenos dat není absolutně odolný vůči chybám. Někdy je riziko poškození zanedbatelné, v mnoha případech je však zaznamenaná a přenášená informace jištěna přidáním dat, která se jeví jako redundantní (nadbytečná) v případě, že nedojde k žádné chybě, která však při výskytu chyby mohou pomoci

- signalizovat, že k chybě došlo,
- případně umožnit její opravu.

Typickým příkladem je přidání paritního bitu. Předpokládejme, že chceme ochránit data složená z hexadecimálních cifer. Každou cifru reprezentujeme čtyřmi bity. Předpokládejme, že data posíláme po kanálu zatíženým značným šumem, a proto přidáme ke každé cifře paritní bit.

Například hexadecimální AF01, což je bitově

$$1010 \mid 1111 \mid 0000 \mid 0001$$

se vyšle jako

$$10100 \mid 11110 \mid 00000 \mid 00011.$$

Takový kód se nazývá paritní a umožní odhalit nejvýše jednu chybu v bloku.

Pokud například při přenosu bloku 10100 dojde při různých přenosech postupně k chybě na pozicích 1, 2, 3, 4 a 5, dostaneme slova 00100, 11100, 10000, 10110, 10101. Pokaždé sice odhalíme, že data byla poškozena, neboť přijatý blok nevyhovuje paritní kontrole – má lichý počet bitů, ale nemáme žádný nástroj, jak určit, na které pozici k chybě došlo. Dvojnásobné chyby neodhalíme vůbec.

Existují lepši zabezpečení? Zdá se, že opravdu bezpečné, i když co do využití místa nepříliš úsporné, by bylo zdvojení každé pozice. Naše zpráva AF01 by pak byla vyslána jako

$$11001100 \mid 11111111 \mid 00000000 \mid 00000011.$$

Nicméně po krátné úvaze uvidíme, že jsme si příliš nepolepšili. Některé dvojnásobné chyby náš kód odhalí, ale pokud k nim dojde na bitech bezprostředně po sobě následujících, tak ji odhalit nemusí. K tomu bychom potřebovali kód, který každou pozici ztrojnásobí. Trojnásobný kód

- signalizuje výskyt dvojnásobných chyb, a
- umožňuje opravit chyby jednonásobné.

Takový kód je ovšem současně značně neúsporný. Čtveřici bitů kóduje dvanácti bity. Ukážeme, že téhož efektu lze dosáhnout kódem, který čtveřici bitů kóduje

pomocí sedmi bitů. Takovému kódu se říká Hammingův a lze jej vytvořit pomocí Fanovy roviny.

Na množině $\{1,\ldots,7\}$ je dáno 7 bloků: $\{1,2,3\}$, $\{3,4,5\}$, $\{1,5,6\}$, $\{1,4,7\}$, $\{2,5,7\}$, $\{3,6,7\}$, $\{2,4,6\}$. Vidíme, že

- Každé dva různé bloky se protínají právě v jednom bodě,
- každými dvěma různými body prochází právě jeden blok,
- existují čtyři body, z nichž žádné tři neleží v jednom bloku.

Systém podmnožin (bloků), který splňuje výše uvedené podmínky, se nazývá projektivní rovina. Blokům se pak obvykle říká přímky. Pokud projektivní rovinu konstruujeme z nějakého systému bodů a přímek, tak pro rozlišení hovoříme o projektivních bodech a projektivních přímkách.

Dříve, než přistoupíme k definici Hammingova kódu, tak poukážeme na roli, kterou v našich úvahách hrají vektorové prostory nad dvouprvkovým tělesem $F = \{0,1\}$.

Zabýváme se **blokovými kódy** určité délky, kterou zpravidla budeme značit k. Vstupem jsou tedy prvky F^k, což pro nás bude standardní vektorový prostor dimenze k. Každý vektor délky k je pak transformován na vektor délky n. Při transformaci ovšem nedostaneme všechny vektory délky n, ale jenom některé. Každé slovo, které může transformací vzniknout se nazývá **kódové slovo** a jejich sjednocení je **kód**, označíme ho C. Je-li C vektorovým podprostorem F^n, mluvíme o lineárním $[n,k]$ kódu. Protože za F jsme volili těleso $F_2 = \{0,1\}$, jde o **kód binární**. Stejnou definici je možno použít i pro $F = \mathbb{F}_3$, kdy mluvíme o **kódu ternárním**. Obecně pak **q-ární lineární $[n,k]$ kód** je každý podprostor $\mathbb{F}_q^n$, který má dimenzi k. Parametr n nazýváme **délka** a parametr k nazýváme **dimenze**.

Pokud v dalším textu bez dalšího vysvětlení zmíníme $[n,k]$ kód, bude se tím rozumět lineární binární kód. Chceme-li stručně naznačit, že jde o lineární q-ární kód, píšeme $[n,k]_q$.

Je paritní kód hexadecimálních číslic lineární? Množinu kódových slov můžeme zapsat jako

$$\{(\epsilon_1,\ldots,\epsilon_5);\ \epsilon_1+\ldots+\epsilon_5=0\},$$

a to jistě je vektorový podprostor $\mathbb{F}_2^5$. Jde tedy o binární kód s parametry $[5,4]$.

Každý $[n,k]$ kód C je plně určen k-ticí vektorů délky n, které udávají jeho generátory. Matice $k\times n$, jejíž řádky generují C, se nazývá **generující matice**. Paritní kód délky 5 má generující matici rovnou

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix}.$$

Tato matice je tvaru $(I\ A)$, kde I je jednotková matice řádu k (zde řádu 4). Takové generující matici se říká **systematická**, nebo také, že je ve **standardním tvaru**.

Ať $G = (I\ A)$ je generující matice nějakého $[n,k]_q$ kódu. Matice A je tedy rozměrů $k \times k'$, kde $k' = n-k$. Ať $g_i = (0,\ldots,0,1,0,\ldots,0,a_{i1},\ldots,a_{ik'})$ je i-tý řádek G. Je přirozené předpokládat, že vstupní vektor $e_i = (0,\ldots,0,1,0\ldots,0)$, který má 1 v i-té pozici, zakódujeme jako g_i. Vstupní vektor $u = (\lambda_1,\ldots,\lambda_k)$ zakódujeme tudíž jako $\sum \lambda_i g_i$, což je rovno $u \cdot G$.

Generující matici lze tedy použít pro kódování. Například hexadecimální A se kóduje v paritním kódu jako

$$\begin{pmatrix} 1 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 & 0 & 0 \end{pmatrix}.$$

Kódování lineárních kódů lze tedy realizovat násobením matice a vektoru. Jak realizovat dekódování?

Prvním krokem je určení takzvané **paritní** neboli **kontrolní** neboli **prověrkové** matice H. Pro q-ární $[n,k]$ kód C je H maticí rozměrů $k' \times n$, kde $k' = n-k$, jež splňuje $u \in C \Leftrightarrow Hu^T = 0$.

Následující tvrzení bude formálně dokázáno až v další kapitole. Jde o standardní fakt, který bývá uváděn i v základních kurzech lineární algebry.

Tvrzení 1.1. *Ať $(I\ A)$ je generující maticí ve standardním tvaru. Potom $(-A^T\ I)$ je prověrkovou maticí.*

□

Prověrkovou maticí paritního $[5,4]$ kódu je tedy matice 1×5 tvaru

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 1 \end{pmatrix}.$$

Jestliže u je vyslané slovo a w je slovo přijaté, tak z $Hw^{\mathsf{T}} \neq 0$ vyplývá, že při přenosu došlo k chybě. Slovo Hw^{T} se nazývá syndrom a časem ukážeme, jak jej lze využít i pro případnou opravu.

Nyní však potřebujeme vyjasnit, co to přesně znamená, že kód C zachycuje (detekuje) až r-násobné chyby a opravuje až s-násobné chyby. Pro $u, v \in F^n$, kde F je těleso, položme

$$d(u,v) = |\{i;\ 1 \leq i \leq n, u_i \neq v_i\}|.$$

Pro $u \in F^n$, ať

$$|u| = |\{i;\ 1 \leq i \leq n, u_i \neq 0\}|.$$

Vidíme, že $d(u,v) = |u-v|$. Hodnota $d(u,v)$ se nazývá vzdáleností nebo **Hammingovou vzdáleností** vektorů u a v. Hodnota $|u|$ se nazývá **vahou** vektoru u.

Lemma 1.2. *Ať $u, v, w \in F^n$. Pak $d(u,v) + d(v,w) \geq d(u,w)$.*

Důkaz. Ať R, S a T jsou po řadě množiny indexů i, kde $u_i \neq v_i$, $v_i \neq w_i$ a $u_i \neq w_i$. Jistě $T \subseteq R \cup S$, takže $|T| \leq |R| + |S|$. Přitom $|T| = d(u,w)$, $|R| = d(u,v)$ a $|S| = d(v,w)$. □

Vidíme, že Hammingova vzdálenost vytváří na F^n metriku. Má tedy smysl definovat koule $S(u,r) = \{v \in F^n;\ d(u,v) \leq r\}$, kde $u \in F^n$ je střed a r je poloměr.

Minimální vzdáleností kódu C se rozumí $\min\{d(u,v);\ u,v \in C,\ u \neq v\}$.

Tvrzení 1.3. *Ať C je kód s minimální vzdáleností d. Pak C detekuje všechny r-násobné chyby právě když $r < d$, a C opravuje všechny r-násobné chyby právě když $2r < d$.*

Důkaz. Ať je vysláno slovo v a přijato slovo w. K r-násobné chybě dojde právě když $d(v,w) = r$. Jestliže r je menší než d, tak nemůže dojít k tomu, aby w bylo považováno za kódové slovo, které by bylo odlišné od v. Proto z $r < d$ plyne, že C detekuje všechny r-násobné chyby.

Opravit chybu znamená jednoznačně určit nejbližší kódové slovo. Pokud by za nejbližší kódové slovo bylo možné vzít $v' \neq v$, tak máme $r = d(v,w) \geq d(v',w)$ a $2r \geq d(v,w) + d(v',w) \geq d(v,v') \geq d$. Toto nemůže nastat, je-li $2r < d$, takže v takovém případě lze r-násobné chyby jednoznačně opravit. Je-li naopak $d \geq 2r$ a $r < d$, tak lze zvolit $v, v' \in C$ a w taková, aby platilo $d = (v,v') = d(v,w) + d(v',w)$. Pak $d - r = d(v',w) \geq d(v,w) \geq r$ a v není možno odvodit z w jako jednoznačně určené nejbližší kódové slovo, což znamená, že chybu nelze jednoznačně opravit. □

Pro nenulový lineární $[n,k]_q$ kód C definujeme jeho minimální váhu jako $\min\{|u|;\ u \in C,\ u \neq 0\}$. Z $d(u,v) = |u - v|$ okamžitě plyne

Tvrzení 1.4. *Minimální vzdálenost nenulového lineárního kódu je rovna jeho minimální váze.*

□

Formálně se minimální váha nulového kódu délky n definuje jako $n + 1$. Má-li $[n,k]_q$ kód minimální váhu d, hovoříme o $[n,k,d]_q$ kódu. Výše jsme slíbili zkonstruovat binární kód o 16 prvcích, který by opravoval jednonásobné chyby a detekoval až dvojnásobné chyby. Z Tvrzení 1.3 vidíme, že zadání vyhoví každý $[7,4,3]$ kód.

Vraťme se nyní k Fanově rovině. Je-li dán nějaký blok $B \subseteq \{1,\ldots,n\}$, tak jeho incidenční vektor i_B je roven $(\epsilon_1,\ldots,\epsilon_n)$, kde $\epsilon_i = 0$, pokud $i \notin B$ a $\epsilon_i = 1$, pokud $i \in B$.

Ať w je na chvíli binární vektor délky 7, který je tvořen samými jedničkami. Ať C je množina vektorů složená z nulového vektoru, jedničkového vektoru w a vektorů i_B a $w - i_B$, kde B probíhá množinu všech bloků Fanovy roviny. Bloky označme: $B_1 = \{1, 2, 3\}$, $B_2 = \{3, 4, 5\}$, $B_3 = \{1, 5, 6\}$, $B_4 = \{1, 4, 7\}$, $B_5 = \{2, 5, 7\}$, $B_6 = \{3, 6, 7\}$, $B_6 = \{2, 4, 6\}$. Incidenční vektor bloku B_i označme u_i. Ať W je podprostor $\mathbb{F}_2^n$ generovaný vektory u_3, u_5, u_6 a $w - u_1$, tedy lineární podprostor generovaný řádky matice

$$G = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}.$$

Lemma 1.5. *Vektorový prostor W se shoduje s množinou C.*

Důkaz. Množiny W a C mají stejný počet prvků, neboť $\dim W = 4$. Přitom $w \in W$, což lze ověřit součtem všech řádků matice G. Zbývá dokázat $\{u_2, u_4, u_7\} \subseteq W$. Ovšem

$$\begin{aligned} u_2 &= (0, 0, 1, 1, 1, 0, 0) = u_6 + (w - u_1), \\ u_4 &= (1, 0, 0, 1, 0, 0, 1) = u_3 + (w - u_1) \text{ a} \\ u_7 &= (0, 1, 0, 1, 0, 1, 0) = u_5 + (w - u_1). \end{aligned}$$

□

S využitím Tvrzení 1.1 můžeme tedy vyslovit

Tvrzení 1.6. *Kód C s generující maticí G je binární $[7, 4, 3]$ kód. Jeho prověrková matice je rovna*

$$H = \begin{pmatrix} 1 & 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 1 \end{pmatrix}.$$

□

Všimněme si, že ve sloupcích H se vystřídají právě všechny nenulové binární vektory délky 3. Binární Hammingův kód se nazývá každý binární kód, ke kterému lze najít prověrkovou matici s obdobnou vlastností. Jde-li o vektory délky $\ell \geq 2$, dostáváme $[2^\ell - 1, 2^\ell - \ell - 1, 3]$ kód, což v obecnosti dokážeme až v další kapitole. Výše zkonstruovaný kód C odpovídá případu $\ell = 3$; budeme hovořit o základním Hammingově kódu.

Při definici Hammingova kódu neurčujeme pořadí sloupců, jde jen o to, aby se vystřídaly všechny vektory. Obecně se kódy často definují až na pořadí souřadnic. Říká se, že kódy C_1 a C_2 délky n jsou permutačně ekvivalentní, jestliže existuje $\sigma \in S_n$, že $(u_1, \ldots, u_n) \in C_1 \Leftrightarrow (u_{\sigma(1)}, \ldots, u_{\sigma(n)}) \in C_2$.