Cvičení 6. (Kapitola 6.)

Kódování a dekódování pomocí lin. kódů

Pravděpodobnosti P_{spr}, $P_{šp}$

Kódování:

- Vezmeme C - $[n,k]$-kód nad $GF(q)$ s generující maticí G
- $|C| = q^k$, G má rozměr $k \times n$
- C lze použít pro přenos q^k různých slov (zpráv), kterým přiřadíme q^k různých k-tic z q symbolů, tj. celý $V(k,q)$

Délka kód. slova je n, tj. $n-k$ symbolů jsou kontrolní.

Vektor zprávy $\vec{u} \in V(k,q)$ zakódujeme vynásobením maticí G

$$\vec{x} = \vec{u} \cdot G = \sum_{i=1}^{k} u_i \cdot \vec{r}_i \quad \text{a} \quad \vec{x} \in C \leftarrow \text{je kódové slovo}$$

Kódování s maticí ve standardním tvaru

Pro $G = [I_k, A]$, kde A je $k \times (n-k)$ matice

pro $\vec{x} = \vec{u} G$ platí, že $x_i = u_i$ pro $1 \le i \le k$ ← symboly zprávy

$$x_{k+i} = \sum_{j=1}^{k} a_{j,i} \cdot u_j \quad 1 \le i \le n-k \leftarrow \text{redundance (kontrolní symboly)}$$

Dekódování:

$\vec{x} = x_1, \dots, x_n$ vyslané slovo

$\vec{y} = y_1, \dots, y_n$ přijaté slovo

$\vec{e} = \vec{y} - \vec{x}$ ← chybový vektor

Def: C, $[n,k]$-kód nad $GF(q)$ a $\vec{a}$ je jakýkoliv vektor z $V(n,q)$.

Množina $\vec{a} + C = \{\vec{a} + \vec{x} \mid \vec{x} \in C\}$ se nazývá koset.

Vlastnosti kosetů:

- je-li $\vec{a}+C$ koset a $\vec{b} \in \vec{a}+C \Rightarrow \vec{a}+C = \vec{b}+C$.

- Lagrangeův theorem (rozklad grupy podle podgrupy)

  C – $[n,k]$-kód nad $GF(q)$ – pak

  (i) každé slovo z $V(n,q)$ je v nějakém kosetu & C.

  (ii) každý koset obsahuje q^k slov

  (iii) každé 2 kosety jsou buď totožné, nebo disjunktní.

Příklad: $[4,2]$ - kód - sestavit kosety.

Def: „Coset leader" Reprezentant kosetu - slovo s nejmenší váhou v kosetu.
Nemusí být určeno jednoznačně.

Platí: $V(n,q) = (0+C) \cup (a_1+C) \cup \dots \cup (a_s+C)$, kde $s = q^{n-k} - 1$

Standardní Slepianovo Rozmístění (Standard Array) „SR".
Způsob sestavení SR → viz přednáška.

Dekódování založeno na kosetech - pomocí SR. (později syndromové dek.)

Míra efektivnosti kódu - pravděpodobnost, že přijaté slovo bude dekódováno na slovo odeslané $P_{spr} = 1 - P_{šp}$

Pro binární kód a symetrický kanál s pravděp. chyby symbolu p
pravděpodobnost, že chybový vektor je daný vektor váhy i, je: $\underline{p^i(1-p)^{n-i}}$

Pro binární $[n,k]$- kód a α_i počet kosetů reprezentantů váhy i

$$P_{spr}(C) = \sum_{i=0}^{n} \alpha_i p^i (1-p)^{n-i}$$

← Pravděp, že slovo dekódované pomocí SR je slovo odeslané.

$P_{šp}(C) = 1 - P_{spr}(C)$ = "Word error rate"

Pravděpodobnost, že přijaté slovo bylo dekódováno na jiné slovo, než bylo odesláno.

Pozn: α_i není lehké určit pro $i > t$, kde $d(C) = 2t+1$

nebo $d(C) = 2t+2$ pro (n, k, d)-kód.

Pro perfektní $[n, k, 2t+1]$-kódy $\alpha_i = \binom{n}{i}$ pro $0 \le i \le t$

$\alpha_i = 0$ pro $i > t$

Př:

(1) Binární $[7, 4, 3]$-kód.

- $|C| = 2^4 = 16$ ← podprostor ve $V(7,2)$, $|V(7,2)| = 2^7$, $|V(7,2)| = 128$
- Je perfektní? $M\left(\binom{7}{0} + \binom{7}{1}\right) \le 2^7$, $1 + 7$, $M \le \frac{2^7}{2^3} = 2^4 = 16$ ✓ => je perfektní
- Gener. matice 4×7, $[I_4 | A]$

$$\left[\begin{array}{cccc:ccc} 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{array}\right]$$ ← například takto

$d(C) = w(C) = 3$

Celý kód

0000	1000	0100	0010	0001	1100	1010	1001
0000000	1000011	0100101	0010110	0001111	1100110	1010101	1001100
0110011	0101010	0011001	1110000	1101001	1011010	0111100	1111111
0110	0101	0011	1110	1101	1011	0111	1111

- SR má rozměr $q^{n-k} \times q^k = 2^{7-4} \times 2^4 = 8 \times 16$

tj. 7 kosetů reprezentantů (nenulových)

slova $\vec{a}_i$ pro $i = 1 \dots 7$, kde $\vec{a}_i = [a_{ij}]$ $a_{ij} = 1$ pro $i = j$, $a_{ij} = 0$, $i, j = 1 \dots 7$

- Určete $P_{šp}(C)$ pro $p = 0{,}1$

Protože C je perfektní $\alpha_0 = 1$, $\alpha_1 = \binom{7}{1} = 7$

$\alpha_i = \binom{n}{i}$ pro $0 \le i \le t$ a $\alpha_i = 0$ pro $i > t$

$$P_{spr}(C) = \sum_{i=0}^{n} \alpha_i p^i (1-p)^{n-i} = (1-p)^7 + 7 \cdot p(1-p)^6 \doteq \underline{0{,}998}$$

$$P_{šp} = 1 - P_{spr}(C) = 1 - (1-p)^7 - 7 \cdot p(1-p)^6 \doteq \underline{0{,}002}$$

Př. 2. Zkonstruujte standard arrays lineárních kódů,

(i) které mají generující matice.

1) $G_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, 2) $G_2 = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$, 3.) $G_3 = \begin{bmatrix} 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 \end{bmatrix}$

od 1) C_1 je $[2,2]$-kód ve $V(2,2)$ nad $GF(2)$

tj C_1 je celý prostor $V(2,2)$

Standard Array bude obsahovat jen jeden řádek odpovídající kódu C_1

00	10	01	11

od 2.) C_2 je $[3,2]$-kód ve $V(3,2)$ nad $GF(2)$

Rozměr SA je $q^{n-k} \times q^k = 2^1 \times 2^2 = 2 \times 4$

C_2

000	101	011	110
100	001	111	010

$G_2 = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$

od 3.) C_3 je $[5,2]$-kód ve $V(5,2)$ nad $GF(2)$ $G_3 = \begin{bmatrix} 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 \end{bmatrix}$

Rozměr SA je $q^{n-k} \times q^k = 2^3 \times 2^2 = 8 \times 4$

00000	10110	01011	11101
10000	00110	11011	01101
01000	11110	00011	10101
00100	10010	01111	11001
00010	10100	01001	11111
00001	10111	01010	11100
11000	01110	10011	00101
10001	00111	11010	01100

$d(C_3) = 3$

(i) Pomocí standard array pro C_3 dekódujte přijatá slova

$\vec{y}_1 = 11111$, $\vec{y}_2 = 01011$

$\vec{y}_1 \rightarrow$ dekodujeme na $\vec{x}_1 = 11101$ $d(\vec{y}_1, \vec{x}_1) = 1$ $\vec{x}_1 = \vec{y}_1 + \vec{e}_1 = 11111 + 00010$

$\vec{y}_2 \rightarrow$ — " — na $\vec{x}_2 = 01011$

↑ je kódové slovo

(ii) Uveďte příklady a) kdy dojde ke 2 chybám a slovo bude opraveno

odesláno $\vec{x} = 10110$ dvě chyby (10001) → přijato $\vec{y} = 00111$ opraveno na $\vec{x} = 10110$

b) kdy dojde ke 2 chybám a slovo nebude opraveno

odesláno $\vec{x} = 10110$ dvě chyby (00011) → přijato $\vec{y} = 10101$ → dekódováno chybně na 11101

Př:

(3.) Jestliže předpokládáme binární symetrický kanál s pravděpod. chyby v symbolu p, určete $P_{spr}(C)$ pro kódy C_1, C_2 a C_3 z předchozího příkladu, Tj. určete pravděpodobnost, že přijaté slovo bude správně dekódováno na slovo odeslané. pro $p = 0.01$

(i) C_1 je $[2,2]$-kód, $G_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$,

00	10	01	11

$P_{spr}(C_1) = (1-p)^2 = (0,99)^2 = 0,9801$ $\quad P_{šp} = 0,0199$

(ii) C_2 je $[3,2]$-kód, $G_2 = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$,

000	101	011	110
100	001	111	010

$P_{spr}(C_2) = (1-p)^3 + p(1-p)^2 = (1-p)^2 = 0,9801$ $\quad P_{šp} = 0,0199$

(iii) C_3 je $[5,2]$-kód, $G_3 = \begin{bmatrix} 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 \end{bmatrix}$,

00000	10110	01011	11101
10000	00110	11011	01101
01000	11110	00011	10101
00100	10010	01111	11001
00010	10100	01001	11111
00001	10111	01010	11100
11000	01110	10011	00101
10001	00111	11010	01100

$P_{spr}(C_3) = (1-p)^5 + 5 \cdot p \cdot (1-p)^4 + 2 \cdot p^2 (1-p)^3$

$= (1-p)^3 (1 - 2p + p^2 + 5p - 5p^2 + 2p^2)$

$= (1-p)^3 (1 + 3p - 2p^2) = 0,9992$

$P_{šp} = 0,0008$

$P_{šp}\, C_2 : P_{šp}\, C_3 \approx 25 : 1$

Šance na neopravenou chybu v C_2 je 25x větší než v C_3.

Pravděpodobnost chyby je-li kód použit pouze k detekci.

Příjímač nepozná, že došlo k chybě, jestliže slovo přijaté je jedno z kódových slov - různé od slova vysílaného.

Tj. když chybový vektor $\bar{e} = \bar{y} - \bar{x}$ je také kódovým slovem.

Pravděp, že nebude detekována chyba tedy nezávisí na vysílaném slově a je dána vztahem.:

$$P_{undetect}(C) = \sum_{i=1}^{n} A_i p^i (1-p)^{n-1}$$

kde A_i počet slov váhy i v kódu C, který je binární $[n, k]$-kód

Př: Jestliže kódy C_1, C_2, C_3 jsou použity pouze pro detekci chyb. vypočítejte pravděp., že chyba v přijatém slově nebude detekována.

Pro $p = 0.01$

$P_{undetect}(C_1) = p^2 + 2p(1-p) = 0.0199$

$P_{undetect}(C_2) = 3p^2(1-p) = 0,000297$

$P_{undetect}(C_3) = 2 \cdot p^3(1-p)^2 + p^4(1-p) = 0.0000197 \doteq \frac{2}{1000000} = \frac{1}{500000}$

U C_3 cca v 1 slova z půl milionu slov nebude detekována chyba

Další míra efektivity kódu.

Def: „RATE" kódu. „Symbol Rate"

Lin. kód přenáší zprávu o k-symbolech kódovými slovy délky n.

Říkáme, že RATE $[n, k, d]$-kódu je $R(C) = \frac{k}{n}$.

Př. (4.) Předpokládejme, že chceme předat informaci o cestě pomocí pokynů odpovídajících 4 směrům S, V, J, Z.

Bez možnosti opravovat ani detekovat chybu lze použít binární $[2,2]$-kód $C = \begin{cases} 00 & -s \\ 10 & -v \\ 01 & -j \\ 11 & -z \end{cases}$ $R(C) = 1$

Jaký kód použít, chceme-li mít alespoň nějakou ochranu proti chybám a zachovat podmínku, aby $R(C) \geq \frac{1}{2}$?

1. možnost. - Použít $[4, 2, 2]$-kód

SR

0000	1011	0101	1110
1000	0011	1101	0110
0100	1111	0001	1010
0010	1001	0111	1100

$P_{spr} = (1-p)^4 + 3 \cdot p(1-p)^3 =$

$= (1-p)^3(1+2p) = 0,9897$ pro $p = 0.01$

Pak $R(C) = \frac{1}{2}$ a $P_{šp} = 0,0103$

kód opraví jen některé 1 chyby

Pokud budeme chtít zachovat, že odesíláme jen 4 různé zprávy, pak efektivnější kód nenajdeme.

2. možnost - připustíme-li, že různých zpráv může být více - nasl. strategie umožní efektivnější kódování (dekódování)

Cestu např. SSZSVV... bychom nahradili řetězcem 0000110010100... a ten rozdělili na bloky délky 4

Celkem 2^4 různých zpráv bychom zakódovali pomocí binárního - $[7, 4, 3]$-kódu (Viz příklad (1))

Pak $R(C) = \frac{4}{7} > \frac{1}{2}$ ✓ kód opraví každou 1 chybu

a $P_{spr} = 0,998$ a $P_{šp} = 0,002$

Oproti $[4,2]$-kódu je počet špatně dekódovaných slov $[7,4]$-kódem snížen cca na pětinu.