Teorija informacij in sistemov, predavanje 6

Uroš Lotrič

Univerza v Ljubljani,
Fakulteta za računalništvo in informatiko

5. Varno kodiranje

- Omejili se bomo na enostavne linearne bločne kode za binarni simetrični kanal
- dolžina bloka je k znakov, abeceda znakov je enaka abecedi kanala, torej imamo $M = 2^k$ blokov $x_1 \dots x_k$, $x_i \in \{0, 1\}$
- za potrebe varovanja dodamo še nekaj varnostnih znakov, celotna dolžina vsake od M kodnih zamenjav je potem n.
- pri konstrukciji koda (n, k) moramo torej med 2^n možnostmi moramo izbrati M najbolj primernih zamenjav

5.1 Ponavljajoče kode

- namesto enega pošljemo n enakih znakov
- primer za $n = 3$: verjetnost za napako za $n = 3$ in $p = 0.01$ je $p_e = 3\,10^{-4}$, hitrost pade iz 1 na 1/3
- Boljši pristop je, da naredimo kode, kjer se povečujeta n in k hitreje od razlike $n - k$.

5.1 Kontrolne vsote 1

- podatkovnim bitom dodamo nekaj bitov za preverjanje parnosti ali **paritetnih bitov**
- paritetni biti so nastavljeni tako, da je vsota bitov v aritmetiki po modulu 2 fiksna vrednost (0 ali 1)
- modulo 2 aritmetika:

+/-/XOR	0	1
0	0	1
1	1	0

×/AND	0	1
0	0	0
1	0	1

- primer: bloku lahko dodamo še varnostni bit, tako, da je vsota vedno soda (redkeje liha): 00|0, 01|1, 10|1, 11|1. Ta kod zna detektirati eno napako.
- ISBN-10: razširitev te ideje na 10 znakov 0, ..., 9:

5.1 Pravokotni in trikotni kodi

- Pravokotne kode: kodo zapišemo v obliki pravokotnika, sodost po vrsticah in po stolpcih.
- Odkrivanje napak

1	0	0	1	1	1	0	0
0	0	0	1	1	1	1	0
0	0	1	1	0	0	1	1
1	0	0	1	0	0	1	1
0	1	1	1	1	1	0	1
1	0	1	0	1	1	1	1
0	0	1	1	0	0	1	1
1	1	0	0	0	0	1	1

1	0	1	1	1
1	1	0	1	
1	0	0		
0	1			
1				

- Trikotne kode: vsota elementov v stolpcu in vrstici s paritetnim bitom vred mora biti soda.
- Odkrivanje napak

5.2 Hammingova razdalja 1

- **Hammingova razdalja med kodnima zamenjavama** nam pove število znakov, na katerih se razlikujeta.
- primer: $x_a = 11000010$, $x_b = 10010010$, $d(x_a, x_b) = 2$
- kodni zamenjavi sta enaki, če je razdalja 0, razdalja med različnimi kodnimi zamenjavami mora biti vsaj 1, drugače je kod singualaren
- **Hammingova razdalja koda** je podana kot minimalna Hammingova razdalja med dvema kodnima zamenjavama.
- Ideja popravljanja na kocki za $n = 3$

5.2 Hammingova razdalja 2

- Hammingova razdalja koda definira odpornost koda na napake.
- Število napak, ki jih kod zazna:
  $d \geq e + 1 \rightarrow e_{\mathrm{max}} = d - 1$ (nariši na sliki)
- Število napak, ki jih kod lahko popravi:
  $d \geq 2f + 1 \rightarrow f_{\mathrm{max}} = \lfloor \frac{d-1}{2} \rfloor$
- Primer: $d = 3$: detekcija 2 in popravljanje 1 napake (kodni zamenjavi $\{000, 111\}$ na kocki

5.3 Hammingov pogoj

- Za blok dolžine n lahko zgradimo 2^n različnih kodnih zamenjav. Če želimo zagotoviti odpornost na napake, mora biti razdalja d več od 1. Uporabni kodi imajo število kodnih zamenjav $M = 2^k < 2^n$.
- Hammingov pogoj: da bi lahko pravilno dekodirali vse kodne zamenjave pri katerih je prišlo do e ali manj napak mora veljati

$$M \leq \frac{2^n}{\sum_{i=0}^{e} \binom{n}{i}}$$

5.4 Linearni bločni kodi 1

- Kode označimo kot (n, k).
  k - podatkovni biti, m - varnostni biti
- O linearnih bločnih kodih govorimo kadar
  - je vsota vsakega para kodnih zamenjav spet kodna zamenjava.
  - da produkt kodne zamenjave z 1 in 0 spet kodno zamenjavo
- Označimo jih z $L(n, k)$
- vedno obstaja kodna zamenjava s samimi ničlami
- Hammingova razdalja linearnega koda je enaka številu enic v kodni zamenjavi z najmanj enicami.
  Dokaz!

5.4 Linearni bločni kodi: generatorska matrika 1

- Napišimo enačbe za trikotni kod. Podatkovni biti naj bodo označeni kot z_1, z_2 in z_3, varnostni pa kot s_1, s_2 in s_3.

$$\begin{array}{ccc} z_1 & z_2 & s_1 \\ z_3 & s_2 & \\ s_3 & & \end{array} \qquad \text{Velja} \qquad \begin{aligned} z_1 + z_2 + s_1 &= 0 \\ z_3 + s_2 + z_2 &= 0 \\ s_3 + z_3 + z_1 &= 0 \end{aligned}$$

- sestavimo kodno zamenjavo v obliki $\mathbf{x} = (x_1, x_2, x_3, x_4, x_5, x_6) = (z_1, z_2, z_3, s_1, s_2, s_3)$.
- Enačbe, zapisane z x_i:

$$\begin{aligned} x_1 + x_2 + x_4 &= 0 \\ x_2 + x_3 + x_5 &= 0 \\ x_1 + x_3 + x_6 &= 0 \end{aligned}$$

- Generiranje kodne zamenjave lahko opišemo z **generatorsko matriko**

$$\mathbf{x} = \mathbf{zG} = (z_1, z_2, z_3) \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{pmatrix}$$

5.4 Linearni bločni kodi: generatorska matrika 2

- Na področju kodiranje je običajno, da se vektorji pišejo vodoravno in ne navpično.
- V splošnem podatkovni vektor $1 \times k$ množimo z generatorsko matriko $k \times n$, da dobimo kodno zamenjavo $1 \times n$.
- Matrika mora imeti linearno neodvisne vrstice
- Za diskretne kanale brez spomnina jo vedno lahko zapišemo v obliki $\mathbf{G} = (\mathbf{I_k}|\mathbf{A})$
- Kod, čigar generatorska matrika ima to obliko, je **sistematični kod** - prvih k znakov koda je enakih sporočilu (podatkovnim bitom). Ostlih $n - k$ znakov so paritetni biti.

5.4 Matrika za preverjanje sodosti 1

- Prej napisane enačbe lako zapišemo z **matriko za preverjanje sodosti**

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

- potem mora veljati $\mathbf{x}\mathbf{H}^T = \mathbf{0}$
- Pokaži $\mathbf{GH^T} = \mathbf{0}$.
- Če je $\mathbf{G} = (\mathbf{I_k}|\mathbf{A})$, pokaži da je $\mathbf{H} = (\mathbf{A^T}|\mathbf{I_{n-k}})$
- Pokaži: vsota dveh kodnih zamenjav je nova kodna zamenjava