Programerska domača naloga - varianta 4
pri predmetu TEORIJA KODIRANJA IN KRIPTOGRAFIJA

Ljubljana, 10. maj 2016

Naj bo $\mathcal{C}$ popolni Hammingov kod dolžine 31 nad $\mathbb{Z}_2$, torej je $n = 31$ in $k = 26$. Nadzorna matrika H je matrika dimenzije 5×31, katere stolpci so vsi neničelni 5-bitni nizi. Uredimo jih leksikografsko, z identiteto na koncu:

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

Znake kodiramo v 5-bitne nize po naslednji tabeli:

A	00000	B	00001	C	00010	D	00011	E	00100
F	00101	G	00110	H	00111	I	01000	J	01001
K	01010	L	01011	M	01100	N	01101	O	01110
P	01111	Q	10000	R	10001	S	10010	T	10011
U	10100	V	10101	W	10110	X	10111	Y	11000
Z	11001	.	11010	,	11011	¶	11100	?	11101
’	11110	_	11111						

Znak _ označuje presledek, ¶ pa znak za novo vrstico. Besedilo kodiramo tako, da združimo po pet črk/znakov skupaj, nato jih po zgornji tabeli prekodiramo v 25 bitov, dodamo še eno ničlo, da dobimo 26 bitov, nato pa to zakodiramo v ustrezno kodno besedo dolžine 31 bitov s pomočjo matrike $G = [I|A]$, kjer je $H = [-A^T|I]$. Matrika G je velikosti $k \times n$, torej 26×31. Tako dobljeni niz ničel in enk shranimo v datoteko.

Primer:

$$\begin{aligned} \text{PETEK} &\to 01111\ 00100\ 10011\ 00100\ 01010\ 0 \to 0111100100100110010001010 0 \\ &\overset{G}{\to} 0111100100100110010001010000011 \end{aligned}$$

1. Sestavite program, ki tekstovno datoteko zakodira po gornjem postopku. Izhodna datoteka mora biti tudi tekstovna (vsebuje samo ničle in enice). Nato z vašim programom zakodirajte vaše ime in priimek, tako da v datoteko `ime.txt` shranite vaše ime in priimek v eni vrstici, z natanko enim presledkom med imenom in priimkom. Rezultat kodiranja shranite v datoteko `kodiranoime.txt`.

2. Sestavite program, ki s pomočjo sindromov dekodira datoteko (ki je bila kodirana po gornjem postopku in “poslana”po kanalu s šumom) in rezultat shranite kot tekstovno datoteko. S tem programom dekodirajte datoteki `kodirano1.txt` in `kodirano2.txt`, ki ju dobite na spletni učilnici in ju shranite v datoteki `besedilo1.txt` in `besedilo2.txt`. Obakrat je bilo zakodirano in “poslano” isto besedilo, pri čemer je bila pri enem besedilu verjetnost napake pri prenosu posameznega bita 0,02, pri drugem pa 0,005. Oglejte si rezultate dekodiranja in ocenite, pri katerem je bila verjetnost napake posameznega bita večja. Ker je razmaknjenost Hammingovih kodov enaka 3, lahko ti kodi popravijo 1 napako. Kaj opazite, če sta v neki kodni besedi dve napaki? Koliko napačnih črk v tem primeru pričakujemo po dekodiranju take kodne besede? Ali je izbrani kod primeren za kanal, kjer je verjetnost napake posameznega bita 0,02 ali več?

Na spletni učilnici oddajte datoteko ImePriimek.zip, ki vsebuje:

- Poročilo v obliki PDF o tem, kako ste se lotili problema ter odgovore na vprašanja, ki so zastavljena zgoraj. Ocenite tudi časovno in prostorsko zahtevnost vašega algoritma.
- Kodo za oba programa (lahko tudi v eni datoteki). Koda naj ima dovolj komentarjev, razloženo mora biti tudi, kako program pognati.
- Datoteki `ime.txt` in `kodiranoime.txt`.
- Datoteki `besedilo1.txt` in `besedilo2.txt`.