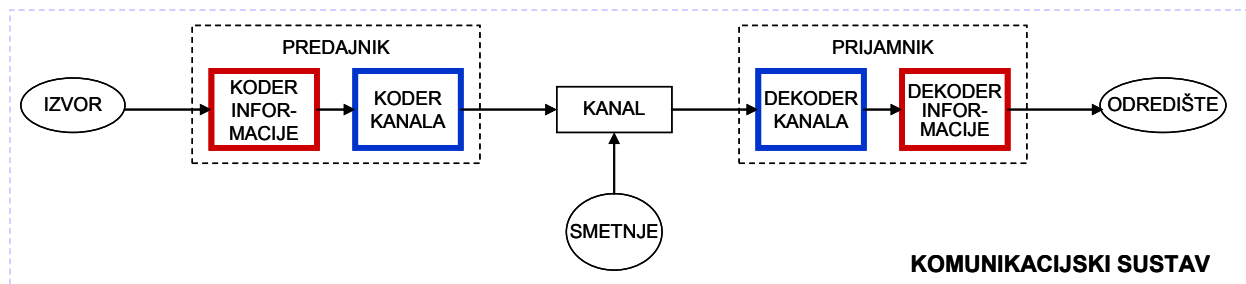


SVEUČILIŠTE U ZAGREBU
Fakultet elektrotehnike i računarstva

Predmet: Teorija informacije (34315)
Ak. godina: 2011./2012.

Projektni zadatak za laboratorijske vježbe

Zadatak



Slika 1. Model komunikacijskog sustava i prijenosni kanal

Na sl. 1 zadan je komunikacijski sustav koji se sastoji od izvora informacije, koda informacije, koda kanala, kanala na koji utječu smetnje, dekoda kanala, dekoda informacije i odredišta. Izvorište generira simbole sa zadanim vjerojatnostima pojavljivanja. Te simbole potrebno je kodirati nekom od metoda entropijskog kodiranja. Nakon toga, novonastale sljedove simbola (tzv. kodirana poruka) potrebno je kodirati nekom od metoda zaštitnog kodiranja (dobivamo zaštitno kodiranu poruku). Zaštitno kodirana poruka prenosi se komunikacijskim kanalom u kojem:

- i) u jednom slučaju nema smetnji, dok
- ii) u drugom slučaju smetnje djeluju na poslane podatke.

Podatke koji su prošli kroz komunikacijski kanal potrebno je dekodirati na odredištu.

U zadatku je potrebno slijed od 10000 simbola prenijeti kroz zadani komunikacijski sustav. Definirana su izvorišta sa simbolima, metode entropijskog kodiranja, zaštitnog kodiranja, te svojstva komunikacijskog kanala.

Izvorišta (datoteke se nalaze u repozitoriju na stranicama predmeta):

1. *Izvorište_1.txt*
2. *Izvorište_2.txt*
3. *Izvorište_3.txt*

Metode entropijskog kodiranja:

4. Shannon-Fanoovo kodiranje
5. Huffmanovo kodiranje

Metode zaštitnog kodiranja:

1. Hammingov kôd $[n, k] = [7, 4]$
2. Ciklični kôd, duljina kodne riječi je 7 bitova i generirajući polinom $g(x) = x^3 + x + 1$

Kanal:

1. Vjerojatnost pogreške bita $p = 1/1000$
2. Vjerojatnost pogreške bita $p = 1/2000$

Pravila za izvođenje zadataka

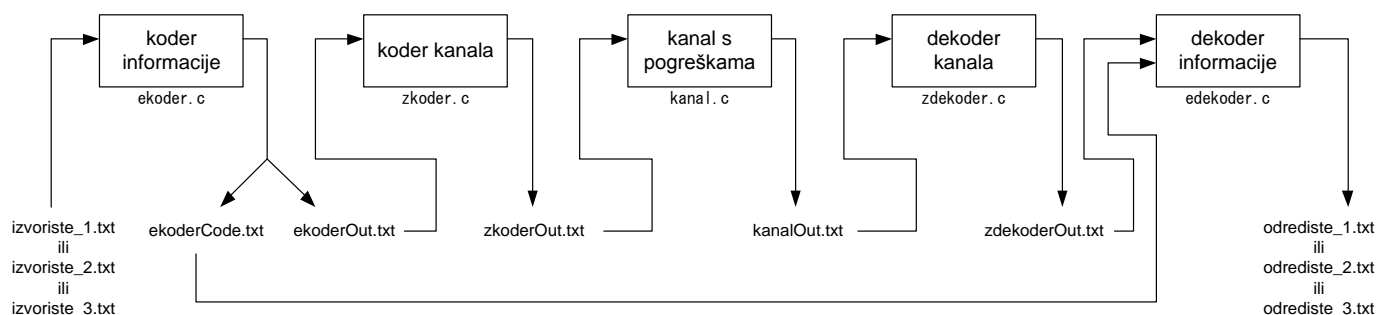
Studenti se trebaju držati pravila za izvođenje i predaju zadatka. Bilo kakva odstupanja od definiranih pravila neće se tolerirati, tj. dodjeljivat će se 0 bodova.

Zadatak se radi u grupi studenata (popis podgrupa za svakog predavača nalazi se na stranicama kolegija). Svaka grupa ima svoj redni broj, npr. redni broj grupe s nazivom „zi13” je broj 13. Ovisno o rednom broju (u nastavku *RedniBroj*), svakoj grupi dodjeljuje se pojedini segment laboratorijskog zadatka prema navedenim formulama:

- a) $brojIzvorista = (RedniBroj \bmod 3) + 1$
- b) $tipEntropijskogKodiranja = (RedniBroj \bmod 2) + 1$
- c) $tipZaštitnogKodiranja = (RedniBroj \bmod 2) + 1$
- d) $tipKanala = (RedniBroj \bmod 2) + 1$

Zadatak se sastoji od dva dijela:

1. **Implementacija komunikacijskog sustava.** Ovisno o dodijeljenim segmentima zadatka, svaka grupa treba implementirati komponente komunikacijskog sustava, odnosno pripadne kodere, zatim kanal, te dekodere.



Slika 2. Shema komunikacijskog sustava (i redoslijeda programskih komponenti) kojeg je potrebno implementirati za laboratorijski zadatak. Strelice označuju ulazne/izlazne datoteke koje se predaju kao argumenti *main* funkcije za određenu programsku komponentu. Programske komponente su označene kućicama.

Primjer odabira komponenti sustava:

grupa s nazivom „zi13” ima za zadatak sljedeće vrijednosti:

$brojIzvorista=2$

$tipEntropijskogKodiranja = 2$

$tipZaštitnogKodiranja = 2$

$tipKanala = 2,$

odnosno grupa „zi13“ za zadatak treba implementirati Huffmanov entropijski koder, ciklični koder i kanal s pogreškom $p = 1/2000$, te pripadne dekodere, te poslati slijed simbola iz izvorišne datoteke (*Izvoriste_2.txt*) kroz implementirani komunikacijski kanal.

Slika 2 ilustrira programske zahtjeve ovog zadatka:

Komunikacijski sustav se može podijeliti u pet cjelina, odnosno pet programskih komponenti koje je potrebno implementirati, a to su: dva kodera, kanal i dva dekodera. Komponente primaju ulaze i generiraju izlaze, onako kako je ilustrirano na slici 2.

Kao izvorišta koriste se datoteke s nizom simbola (*Izvoriste_1.txt*, *Izvoriste_2.txt*, *Izvoriste_3.txt* - nalaze se u repozitoriju na stranicama kolegija). Entropijski koder obradom i kodiranjem izvorišta generira dvije izlazne datoteke:

-izlaznu datoteku koja sadrži kodirani niz simbola iz izvorišta (odnosno slijed 0 i 1), a koja se prenosi do kodera kanala, odnosno zaštitnog kodera, te
-izlaznu datoteku koja sadrži tablicu simbola, tj. za svaki simbol ekvivalentni binarni kod. Ova datoteka služi entropijskom dekoderu da dekodira primljeni niz. Datoteka mora imati zapis oblika **simbol=kod**.

Primjer zapisa tablice simbola:

Izvorište sadrži simbole {*a*, *b*, *c*, *d*}. Entropijski koder je za svaki simbol odredio sljedeći binarni zapis. *d*= 001, *c*=01, *a*=1, *b*=10. Izlazna datoteka *ekoderCode.txt* ima sljedeći zapis:

```
d= 001
c=01
a=1
b=10
```

U praksi nije slučaj da entropijski koder šalje entropijskom dekoderu ovakav zapis – umjesto toga on frekvencije pojavljivanja svih simbola (iz npr. ASCII tablice simbola) u binarnom zapisu šalje kroz cijeli komunikacijski sustav do entropijskog dekodera, koji zatim uz pomoć tih frekvencija prvo rekonstruira tablicu kodiranja, a onda dekodira niz 0 i 1 u odgovarajući niz simbola.

Komponente se rade isključivo u programskim jeziku „c“ i imenuju se onako kako je to ilustrirano na slici 2 .

Prema slici 2, svaka komponenta, osim entropijskog kodera i dekodera, ima po jedan ulaz i jedan izlaz. U programska implementaciji svaka komponenta **mora** primiti putanju ulazne i izlazne datoteke (**isključivo u tom redoslijedu!**) preko argumenata koji se predaju preko komandne linije (***int main (int argc, char **argv)***).

Entropijski koder kao **treći argument prima i putanju do datoteke u koju će zapisati entropijski kod**. Entropijski dekodeer kao **drugi argument prima putanju do datoteke u kojoj se nalazi entropijski kod**, a kao treći argument putanju do datoteke u koju će zapisati odredište.

Primjeri pozivanja izvršne datoteke:

Izvršna datoteka *ekoder.exe* (dobivena iz datoteke *ekoder.c*) se iz komandne linije pokreće na sljedeći način:

ekoder *putanja_do_izvorista/izvoriste1.txt* *putanja_do_izlaza/ekoderOut.txt*
putanja_do_koda/ekoderCode.txt

Izvršna datoteka *zkoder.exe* (dobivena iz datoteke *zkoder.c*) se iz komandne linije pokreće na sljedeći način:

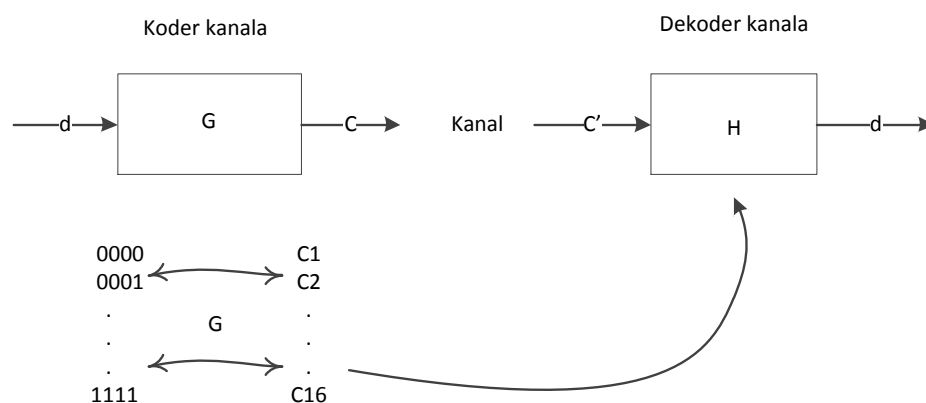
zkoder *putanja_do_ulaza / ekoderOut.txt* *putanja_do_izlaza /zkoderOut.txt*

Izvršna datoteka *edekoder.exe* (dobivena iz datoteke *edekoder.c*) se iz komandne linije pokreće na sljedeći način:

edekoder *putanja_do_ulaza/zdekoderOut.txt* *putanja_do_koda/ekoderCode.txt*
putanja_do_odredišta/odrediste_1.txt

Napomene uz izvođenje ovog dijela zadatka:

1. Entropijsko kodiranje: simbolu s manjom vjerojatnosti (ili združenom ako su vjerojatnosti iste) dodjeljuje se 0, a onom s većom 1 (kao što je napravljeno u knjizi na stranicama 257. i 258.)
2. Ukoliko ukupni slijed bitova koji izlazi iz koder informacije nije odgovarajuće duljine (tj. djeljiv s 4) potrebno ga je nadopuniti nulama jer je to bitno za koder kanala (tj. zaštitno kodiranje).
3. Hammingov koder zaštitno kodira bitove po redu. Npr. ako je slijed bitova u *ekoderOut.txt* datoteci 01111100, onda će prva kodna riječ u *zkoderOut.txt* datoteke izgledati ovako: $p_{11}p_{12}0p_{13}111$ (p_{ij} su zaštitni bitovi)....
4. Ciklički koder koristi generirajuću matricu cikličkog koda koja je opisana u knjizi na str. 172 . Dekoder cikličkog koda koristi matricu provjere pariteta \mathbf{H} koja je dana u knjizi na str 174. Ista nije u obliku $\mathbf{H}=[\mathbf{A}^T|\mathbf{I}]$. U cilju provedbe dekodiranja u dekoder kanala tj. u *zdekoder.c* potrebno je upisati tablicu preslikavanja poruke \mathbf{d} u kodnu riječ \mathbf{C} kako bi se moglo jednoznačno provesti dekodiranje.



5. Ukoliko entropijski dekoder naiđe na niz bitova koji ne postoje u tablici simbola (*ekoderCode.txt*), onda ih mijenja sa znakom *hash* (#).

Primjer za slučaj višestruke pogreške u jednoj kodnoj riječi:

Izvorište generira simbole iz abecede $\{a, b, c\}$. Neka izvorište generira 4 simbola sljedećim redoslijedom: *abac*. Neka koder informacije za svaki simbol određuje sljedeći binarni zapis: $a = 1$, $b = 01$, $c = 00$. Entropijskim kodiranjem se stvaraju dvije datoteke: 1) *ekoderCode.txt*, s tablicom simbola, 2) *ekoderOut.txt*, u kojoj se nalaze kodirani simboli. Neka je sadržaj *ekoderOut.txt* datoteke sljedeći:

101100

Koder kanala ovom nizu dodaje dvije 0, kako bi niz bitova bio odgovarajuće duljine prije zaštitnog kodiranja (tj. djeljiv s 4). Zaštitnim kodiranjem se u datoteku *zkoderOut.txt* zapisuje sljedeći niz:

$p_{11}p_{12}1p_{13}011p_{21}p_{22}0p_{23}000$, gdje su p_{ij} zaštitni bitovi.

Pretpostavimo da se u kanalu dogodi dvostruka (tj. višestruka) pogreška na prvoj kodnoj riječi, tako da u dekoder kanala ulazi sljedeći niz:

$p_{11}p_{12}0p_{13}010p_{21}p_{22}0p_{23}000$

Nadalje, neke se nakon dekodiranja (dekoder kanala) na ulazu dekodera informacije pojavljuje sljedeći slijed bitova (Pretpostavka je da dekoder kanala kodnu riječ s dvostrukom pogreškom dekodira kao 1010.):

10100000

Dekoder informacije će dani slijed bitova (1 01 00 00 0) dekodirati kao:

abcc#

2. **Izrada izvještaja.** Uz implementaciju programskog dijela zadatka, potrebno je popuniti izvještaj koji se nalazi u repozitoriju na stranicama kolegija. Izvještaj se sastoji od teorijskog dijela pitanja koji su vezani uz dodijeljene segmente laboratorijskog zadatka za svaku grupu.

Primjer odabira izvorišta za analizu:

Grupa s nazivom “zi13” za pitanja o izvorištu simbola analizira datoteku pod rednim brojem 2, odnosno datoteku *Izvoriste_2.txt* ($brojIzvorista = (RedniBroj \bmod 3) + 1 = 2$)

Pravila za predaju zadatka

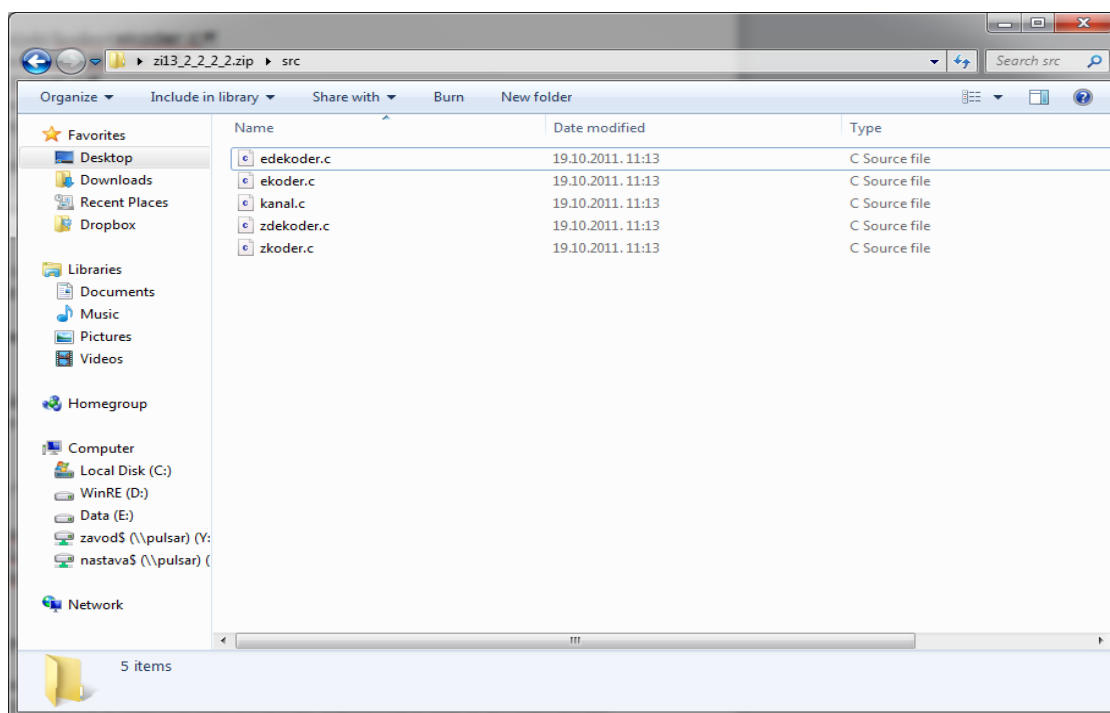
1. Rješenje laboratorijskog zadatka mora sadržavati:

a. Dobro komentirani izvorni kôd. Izvorni kôd laboratorijskog zadatka potrebno je podijeliti na cjeline (imena datoteka u zagradama):

- entropijski koder (ekoder.c)
- zaštitni koder (zkoder.c)
- kanal (kanal.c)
- zaštitni dekodekoder (zdekoder.c)
- entropijski dekodekoder (edekoder.c)

Sve datoteke iz potrebno je pohraniti u direktorij imena “src” (bez navodnika).

b. Izvještaj sa odgovorima na teorijska pitanja u .pdf formatu.



Slika 3. Izgled „src“ direktorija u kojem se nalaze datoteke implementiranih komponenti

2. Direktorij „src“ i izvještaj, odnosno rješenje laboratorijskog zadatka, potrebno je zapakirati u .zip (**Isključivo!** Za predanu .rar arhivu dodjeljivat će se 0 bodova.) arhivu imena:

“grupa_brojIzvorista_tipEntropijskogKodiranja_tipZastitnogKodiranja_tipKanal.zip”,

gdje varijable nakon donjih crtica označavaju odabrane segmente laboratorijskog zadatka dobivene formulama iznad.

Primjer imenovanja arhive za predaju:

grupa s nazivom "zi13" ima za zadatak sljedeće vrijednosti:

brojIzvorišta = 2

tipEntropijskogKodiranja = 2

tipZaštitnogKodiranja = 2

tipKanala = 2

.zip arhiva te grupe treba biti sljedećeg imena: "zi13_2_2_2_2.zip" (**Bez navodnika!**).

1. **Predaju zadatka potrebno je napraviti do 16. siječnja 2012. u 23:59:59. Kašnjenje s predajom laboratorijskog zadatka neće se tolerirati.**
2. **Zadaci se predaju putem sustava Moodle. Na sustavu će pojedinačno, za svakog predavača posebno, biti otvoren resurs unutar kojeg će grupe studenata predavati rješenja zadatka. Dovoljno je da jedan student iz grupe preda rješenje zadatka!**
3. **Predani izvorni kôd bit će testiran na drugim datotekama, tj. izvorštima sa drugim vjerojatnostima pojavljivanja simbola.**

Sve nejasnoće u vezi laboratorijskih vježbi rješavaju se slanjem elektroničke pošte na adresu: aleksandra.cerekovic@fer.hr ili dolaskom na konzultacije koje se moraju prethodno najaviti putem elektroničke pošte (aleksandra.cerekovic@fer.hr).

U Zagrebu, 20. listopada 2011.

Za predmet Teorija informacije

dr.sc. Aleksandra Čereković