TARTU ÜLIKOOL
Matemaatika-informaatika teaduskond
Matemaatika instituut
Matemaatika eriala

Kristo Visk

Permutatsioonikoodid ja allikakodeerimine

Bakalaureusetöö

Juhendaja: Dr. Ago-Erik Riet

Tartu 2015

Permutatsioonikoodid ja allikakodeerimine

Bakalaureusetöö
Kristo Visk

Lühikokkuvõte. Bakalaureusetöö eesmärk on uurida permutatsioonikoodide ühest dekodeeritavust ja prefiksivabadust. Toome sisse permutatsioonikonstandi, permutatsiooni mustri ja kindlat tüüpi permutatsioonikoodide mustrivabaduse mõisted. Näitame, et permutatsioonikoodi prefiksivabadusest järeldub tema ühene dekodeeritavus ning uurime seoseid permutatsioonikonstandi ja prefiksivabaduse vahel. Samuti veendume, et mustrivabad permutatsioonikoodid on prefiksivabad.

Märksõnad. Diskreetne matemaatika, koodid, permutatsioonikoodid, ühene dekodeeritavus, prefiksivabadus, välkmälu.

Permutation codes and source coding

Bachelor's thesis
Kristo Visk

Abstract. The purpose of this Bachelor's thesis is to explore unique decodability and prefix-freedom of permutation codes. We introduce four notions: permutation code, permutation constant, permutation pattern and pattern-freedom of certain permutation codes. In this thesis it is shown that prefix-free permutation codes are uniquely decodable. We also investigate relations between the permutation constant and prefix freedom. At the end of the thesis it is shown that pattern-free permutation codes are prefix-free.

Keywords. Discrete mathematics, codes, permutation codes, unique decodability, prefix-freedom, flash memory.

Sisukord

Sissejuhatus

Käesoleva bakalaureusetöö eesmärk on uurida permutatsioonikoode ja nende omadusi: ühest dekodeeritavust ja prefiksivabadust. Koodi ühene dekodeeritavus ja prefiksivabadus on defineeritud raamatu [NB] eeskujul. Permutatsioonikoodi mõiste ning seosed vastavate omaduste vahel on iseseisva teoreetilise uurimuse tulemus.

Koode kasutatakse igapäevaelus erinevatel eesmärkidel. Tuntumad neist on näiteks kodeerimine andmete turvalisuse tagamiseks, kodeerimine seadmete töökindluse ja võimekuse suurendamiseks, kodeerimine andmete efektiivsemaks haldamiseks.

Allikakodeerimine ehk andmete pakkimine on lähteandmete kodeerimine andmevahetuse efektiivsemaks muutmise eesmärgil. Me defineerime permutatsioonikoodid allikakodeerimises. Analoogiline pakkimine võib järgumodulatsiooni kontekstis anda viisi andmete pakkimiseks välkmäludesse salvestamisel.

Koodi ühene dekodeeritavus tagab selle, et igale lõpliku pikkusega koodsõnade jadale vastab parajasti üks lähtekood. Koodi prefiksivabadus võimaldab koodsõnade jadasid vasakult paremale lugedes üheselt dekodeerida.

Töö esimeses peatükis kirjeldatakse artiklite [JMSB] ja [MBZ] põhjal permutatsioonikoodide seost välkmäluga ning seeläbi nende ühte võimalikku rakendusvaldkonda.

Teine peatükk koosneb viiest alajaotusest. Esimeses alajaotuses antakse vajalikud põhimõisted edasise teksti mõistmiseks. Teises alajaotuses esitatakse koodi ja tema parameetrite definitsioonid ning uuritakse koodide ühest dekoderitavust. Kolmandas alajaotuses antakse permutatsioonikoodi definitsioon. Neljandas alajaotuses defineeritakse permutatsioonikoodi prefiksivabadus ning uuritakse seoseid prefiksivabaduse ja ühese dekodeeritavuse

vahel. Viiendas alajaotuses tuuakse sisse permutatsiooni mustri mõiste ning mustrite seos prefiksivabaduse ja ühese dekodeeritavusega.

1. Permutatsioonikoodide seos välkmäluga

Säilmälu on arvutimälu osa, mis säilitab informatsiooni ka vooluvõrgust eemaldatuna. Välkmälu on elektrooniliselt programmeeritav säilmälu seade. Magnetmäluga võrreldes on välkmälu odavam, töökindlam ja suurema säilitamistihedusega. Andmed salvestatakse välkmällu erineva suurusega laengute andmisel mälupesadesse. Mällu salvestamise protseduur on asümmeetriline: konkreetse mälupesa laengutaset on laengu andmise kaudu võimalik suurendada, aga laengutaseme vähendamiseks tuleb kustutada ja üle kirjutada terve antud mälupesa sisaldav plokk.

Seadme vananedes (või teatud tüüpi kahjustuste korral) võib laengut mälupesadest lekkima hakata. Leke võib aga toimuda erinevates pesades erineva kiirusega, mistõttu muutub vajalike laengutasemete säilitamine keeruliseks. Tekkinud vigade korrigeerimiseks tuleb jälgida laengutaset igas pesas eraldi ning vajadusel seda suurendada. Laengu andmisel konkreetsesse mälupessa võib antud pesa saada liiga suure laengu – tekib ületäitmisviga. Kuna laengu vähendamine mälupesas on tülikas, väheneb seadme vananedes andmete haldamise kiirus ning võivad tekkida vead andmete lugemisel.

Laengutasemete efektiivsemaks haldamiseks on hakatud välja töötama mitmetasandilisi mälupesasid. Seeläbi suureneb pesades säilitatavate bittide arv. Ühebitiste seadmete korral on iga mälupesa kas tühi või täidetud. Mitmetasandilisel juhul on laengutase täidetud pesas jaotatud lõplikuks arvuks osadeks. Seeläbi saab iga mälupesa laengutaset täpsemini määrata. Mingis konkreetses plokis sisalduvate pesade laengutasemeid on võimalik järjestada permutatsioonide abil. Permutatsioon vastab mälupesade plokile järgmisel viisil. Oletame, et mälupesad plokis on nummerdatud naturaalarvudega $1, 2, 3$. Siis permutatsioon $(2, 3, 1)$ on mälupesade järjestus laengu suuruse järgi: pesas 2 on suurim laeng, sellele järgneb pesa 3 ja kõige väiksem laeng on pesas 1. Laengu lekke korral muutub ka antud permutatsioon.

Artiklis [JMSB] tutvustatakse järgumodulatsiooni skeemi, mis tugineb mitmetasandiliste mälupesade kodeerimisele. Kodeerides plokis sisalduvaid mälupesasid permutatsioonideks ning hallates antavaid laenguid sobivalt jaotatud laengutasemete põhjal, on võimalik järgumodulatsiooni kasutades ületäitmisvigu vältida. Koodide abil saab veaparandusvõimet suurendada.

2. Koodid

2.1 Põhimõisted

Olgu $k, l, n \in \mathbb{N} = \{1, 2, \dots\}$.

Defineerime hulga $[k] := \{1, \dots, k\} \subseteq \mathbb{N}$. *Permutatsiooniks* hulgal $[k]$ nimetatakse bijektiivset kujutust

$$\sigma : [k] \to [k].$$

Kõikide permutatsioonide kogumit hulgal $[k]$ tähistame edaspidi

$$\mathbb{S}_k := \{\sigma \mid \sigma : [k] \to [k] \text{ on bijektsioon}\}.$$

Ilmselt $|\mathbb{S}_k| = k!$. Permutatsiooni $\sigma : l \mapsto b_l$ $(l = 1, \dots, k)$ tähistamiseks kasutame ka vektorkuju $(b_1, \dots, b_k)$ ja järjendit $b_1 \dots b_k$.

Tähestikuks nimetame mingit lõplikku hulka $S := \{s_i \mid i \in 1, \dots, n\} \neq \varnothing$ võimsusega n. Hulga S elemente s_i kutsume *sümboliteks*, lõpliku pikkusega sümbolijada $s_1 \dots s_k$ *sõneks*. Pikkusega l sõnede hulka tähistame

$$S^l := \{s_1 \dots s_l \mid s_j \in S,\ j = 1, \dots, l\},$$

kõigi lõpliku pikkusega sõnede kogumit

$$S^* := \bigcup_{j \in \mathbb{N} \cup \{0\}} S^j = S^0 \cup S^1 \cup S^2 \cup \dots\ ,$$

kus $S^0 = \{\epsilon\}$ ning ϵ tähistab tühja sõnet.

Lisaks tähistame

$$T_k := \{(n_1, \dots, n_l) \mid n_j \in [k],\ j = 1, \dots, l\ , n_j \neq n_i\ (j \neq i),\ 1 \leq l \leq k\}$$

ning

$$\mathbb{T}_k := \{(n_1, \dots, n_l) \mid n_j \in [l],\ j = 1, \dots, l\ , n_j \neq n_i\ (j \neq i),\ 1 \leq l \leq k\}.$$

2.2 Koodi mõiste ja ühene dekodeeritavus

Definitsioon 2.1 ([NB]). Olgu S_1 ja S_2 mingid tähestikud. *Koodiks* tähestikul S_1 nimetatakse injektiivset kujutust

$$c : S_1 \to S_2^*.$$

Kujutist $c(s)$ nimetame *koodsõnaks* sümbolist $s \in S$. Praktikas on eelpool esitatud definitsioonist siiski vähe kasu. Tihti ei piisa üksikutele sümbolitele koodsõnade määramisest - tarvis on kodeerida sõnesid. Jätkame nüüd koodi c tähestikult S sõnede kogumile S^*.

Olgu $c : S_1 \to S_2^*$ kood. Iga $s_1 \dots s_n \in S_1^*$ $(n \in \mathbb{N})$ korral defineerime

$$c(s_1 \dots s_n) := c(s_1) \dots c(s_n),$$

kus $c(s_1) \dots c(s_n)$ tähistab koodsõnade järjest kirjutamist. Selline jätkamine annab võimaluse tuua sisse ühese dekodeeritavuse mõiste, mis osutub koodi praktiliste rakenduste seisukohalt oluliseks tunnuseks.

Definitsioon 2.2 ([NB]). $c : S_1 \to S_2^*$ on *üheselt dekodeeritav*, kui tema jätk $c : S_1^* \to S_2^*$ on injektsioon.

Vahetult ühese dekodeeritavuse definitsioonist järeldub järgmine lause.

Lause 2.3 ([NB]). *Olgu $c : S_1 \to S_2^*$ üheselt dekodeeritav kood ning olgu $s_1 \dots s_n$ ja $s_1' \dots s_m'$ mingid kaks sõnet hulgast S_1^* $(n, m \in \mathbb{N})$. Siis*

$$c(s_1 \dots s_n) = c(s_1' \dots s_m') \Leftrightarrow m = n \ \wedge \ s_j = s_j' \ (j \in \{1, \dots, n\}).$$

Näide 2.4. Võtame $S_1 := \{x, y, z\}$ ja $S_2 := [3] = \{1, 2, 3\}$. Olgu $c : S_1 \to S_2^*$ kood, kus $c(x) = 3$, $c(y) = 12$, $c(z) = 123$. Ilmselt pole tegu üheselt dekodeeritava koodiga, kuna $c(yx) = c(z)$ ning $yx \neq z$. Seega dekodeerides koodsõna 123 ei saa kindel olla, kas lähtekoodiks on sõne yx või sõne z.

Näide 2.5. Võtame $S_1 := \{x, y, z\}$ ja $S_2 := [3]$. Olgu $c : S_1 \to S_2^*$ kood, kus $c(x) = 1$, $c(y) = 12$, $c(z) = 123$. Siis c on üheselt dekodeeritav. Selle näitamiseks piisab vaadelda kahte suvalist sõne $s_1 \dots s_n$ ja $s_1' \dots s_m'$ hulgast S_1^*, mille korral $c(s_1 \dots s_n) = c(s_1' \dots s_m')$, ehk $c(s_1) \dots c(s_n) = c(s_1') \dots c(s_m')$. Et kodeeritud sõnumid on võrdsed, peavad ka nende algused olema võrdsed. Kui nüüd oletada, et $c(s_1) \neq c(s_1')$, siis lähtudes meie koodi valikust, tekib järgmise koodsõna algusesse sümbol 1, mistõttu meie kodeeritud sõnumite algused erinevad teineteisest. Seega $c(s_1) \dots c(s_n) \neq c(s_1') \dots c(s_m')$, mis on vastuolus eeldusega. Analoogiliselt jätkates saab näidata, et $s_j = s_j'$ ($j \in \{1, \dots, n\}$) ja $m = n$. Seega lause 2.3 põhjal on c üheselt dekodeeritav.

Toome sisse veel ühe koodile iseloomuliku karakteristiku mõiste.

Definitsioon 2.6 ([NB])**.** Olgu $c : S_1 \to S_2^*$ kood ning olgu tema kõige pikema koodsõna pikkus n. Koodi c *parameetriteks* nimetatakse naturaalarve $a_1, \dots, a_n$, kus

$$a_j = |\{s \in S_1 \mid c(s) \in S_2^j\}| \quad (j \in \{1, \dots, n\}).$$

Lihtsamalt öeldes tähistab a_j pikkusega j koodsõnadeks kodeeruvate sümbolite koguarvu.

Näidetes 2.4 ja 2.5 kirjeldatud koodide parameetrid on $a_1 = a_2 = a_3 = 1$.

Näide 2.7. Olgu $S_1 = \{x_j \mid (j \in \{1, \dots, 7\})\}$ ja $S_2 = [3]$. Defineerime koodi c järgneva eeskirja alusel:

$$\begin{cases} c(x_1) = 1 \\ c(x_2) = 2 \\ c(x_3) = 3 \\ c(x_4) = 11 \\ c(x_5) = 12 \\ c(x_6) = 32 \\ c(x_7) = 1322. \end{cases}$$

Siis koodi c parameetrid on $a_1 = a_2 = 3, a_3 = 0, a_4 = 1$.

2.3 Permutatsioonikoodid

Definitsioon 2.8. Olgu S mingi tähestik ja $k \in \mathbb{N}$. *Permutatsioonikoodiks* tähestikul S nimetatakse injektiivset kujutust

$$c : S \to T_k.$$

Edasises kutsume permutatsioonikoode vahel ka lihtsalt koodideks. Paneme tähele, et

$$|T_k| = |\{(n_1, \ldots, n_l) : n_j \in [k],\ n_j \neq n_i\ (j \neq i),\ 1 \leq l \leq k\}| = \sum_{l=1}^{k} \binom{k}{l} l! = \sum_{l=1}^{k} \prod_{i=k-l+1}^{k} i,$$

sest

$$\sum_{l=1}^{k} \binom{k}{l} l! = \sum_{l=1}^{k} \frac{k!}{(k-l)!l!} l! = \sum_{l=1}^{k} \frac{k!}{(k-l)!} = \sum_{l=1}^{k} k \cdot \cdots \cdot (k-l+1) = \sum_{l=1}^{k} \prod_{i=k-l+1}^{k} i.$$

Et permutatsioonikood c peab olema injektiivne, seab parameeter k piirangu tähestiku S suurusele. Näiteks $k = 3$ korral saame, et $|T_k| = 15$, millest järeldub $|S| \leq 15$.

Järeldus 2.9. *Olgu $c : S \to T_k$ permutatsioonikood. Siis*

$$|S| \leq \sum_{l=1}^{k} \prod_{i=k-l+1}^{k} i.$$

Kuna permutatsioonikood on kujutus mingisse lõplikku hulka T_k, saab iga koodsõna pikkus olla ülimalt k sümbolit (sümboliteks on siin hulga $[k]$ elemendid). Tähistame

$$T_k^l := \{(n_1, \ldots, n_l) : n_j \in [k],\ n_j \neq n_i\ (j \neq i)\}.$$

Definitsiooni 2.6 kohaselt on permutatsioonikoodi $c : S \to T_k$ parameetriteks naturaalarvud $a_1, \ldots, a_k$, kus

$$a_j = |\{s \in S \mid c(s) \in T_k^j\}| \quad (j \in \{1, \ldots, k\}).$$

2.4 Prefiksivabad permutatsioonikoodid

Definitsioon 2.10 ([NB])**.** Olgu S tähestik ning $s_1 \ldots s_n$ ja $s_1' \ldots s_m'$ mingid kaks sõne hulgast S^*. Öeldakse, et sõne $s_1 \ldots s_n$ on *prefiksiks* sõnele $s_1' \ldots s_m'$, kui leidub mingi mittetühi sõne $r \in S^*$ nii, et

$$s_1' \ldots s_m' = s_1 \ldots s_n r.$$

Definitsioon 2.11 ([NB])**.** Permutatsioonikoodi $c : S \to T_k$ nimetatakse *prefiksivabaks*, kui ei leidu sümbolipaari $s, s' \in S$ nii, et

$$c(s) = c(s')r$$

mingi mittetühja $r \in T_k$ korral.

Näide 2.12. Olgu $S = \{x, y, z\}$ ning $c : S \to T_3$ permutatsioonikood, kus $c(x) = 23$, $c(y) = 13$ ja $c(z) = 231$. Siis c ei ole prefiksivaba, kuna võttes $s := z$, $s' := x$ ning $r := 1$ saame, et $c(s) = c(z) = 231 = c(x)1 = c(s')r$.

Teoreem 2.13 ([NB]). *Olgu $c : S \to T_k$ prefiksivaba permutatsioonikood. Siis on c üheselt dekodeeritav.*

Tõestus. Olgu $s_1 \dots s_n$ ja $s_1' \dots s_m'$ mingid kaks sõnet hulgast S^*, mille korral $c(s_1 \dots s_n) = c(s_1' \dots s_m')$, ehk $c(s_1) \dots c(s_n) = c(s_1') \dots c(s_m')$. Lause 2.3 põhjal piisab c üheseks dekodeeritavuseks näidata, et $s_j = s_j' \ \forall j \in \{1, \dots, n\}$ ja $m = n$. Oletades nüüd väite vastaselt, et $s_1 \neq s_1'$, saame koodi c injektiivsuse tõttu, et $c(s_1) \neq c(s_1')$. Et aga eelduse kohaselt $c(s_1) \dots c(s_n) = c(s_1') \dots c(s_m')$, peab üks koodsõnadest $c(s_1)$, $c(s_1')$ olema teisele prefiksiks. See on vastuolus eeldusega, et c on prefiksivaba. Seega $s_1 = s_1'$. Matemaatilist induktsiooni kasutades saab näidata, et

$$s_j = s_j' \ \forall j \in \{1, \dots, n\} \wedge m = n.$$

Seega on c üheselt dekodeeritav. □

Osutub, et üheselt dekodeeritav permutatsioonikood ei pruugi olla prefiksivaba.

Näide 2.14. Võtame $S := \{x, y, z\}$. Olgu $c : S \to T_3$ permutatsioonikood, kus $c(x) = 1$, $c(y) = 12$, $c(z) = 123$. Näite 2.5 põhjal on c üheselt dekodeeritav. Samas aga $c(y) = 12 = c(x)2$, mistõttu pole c prefiksivaba.

Järgnevalt toome sisse permutatsioonikonstandi mõiste.

Definitsioon 2.15. Olgu $k \in \mathbb{N}$ ja $a_1, \dots, a_k \in \mathbb{N} \cup \{0\}$. Parameetritele $a_1, \dots, a_k$ vastavaks *permutatsioonikonstandiks* nimetatakse mittenegatiivset reaalarvu

$$P := \frac{a_1}{\binom{k}{1} 1!} + \dots + \frac{a_k}{\binom{k}{k} k!} = \sum_{l=1}^{k} \frac{a_l}{\binom{k}{l} l!} = \sum_{l=1}^{k} \frac{a_l \cdot (k-l)!}{k!}.$$

Paneme tähele, et kui $c : S \to T_k$ on permutatsioonikood parameetritega $a_1, \dots, a_k$ ning P_c tema parameetritele vastav permutatsioonikonstant, siis

$$P_c = \sum_{l=1}^{k} \frac{a_l}{|T_k^l|}.$$

Seega liidetakse permutatsioonikonstandi arvutamisel proportsioonid, mille parameeter a_i moodustab vastava T_k^i võimsusest $(1 \le i \le k)$.

Permutatsioonikonstandi analoogiks klassikalises allikakodeerimises, kus erinevus on selles, et koodsõna sümbolid ei pea olema erinevad, on Kraft-McMillani arv. Raamatus [NB] on tõestatud järgnev teoreem klassikaliste allikakoodide ja Kraft-McMillani arvu kohta. Meie tõestame Kraft-McMillani teoreemile ehk Krafti võrratusele analoogilise võrratuse permutatsioonikoodide kohta.

Teoreem 2.16. *Olgu $k \in \mathbb{N}$ ja $a_1, \ldots, a_k \in \mathbb{N} \cup \{0\}$. Kui*

$$0 < P := \sum_{l=1}^{k} \frac{a_l \cdot (k-l)!}{k!} \le 1,$$

siis leidub prefiksivaba permutatsioonikood $c : S \to T_k$ parameetritega $a_1, \ldots, a_k$.

Tõestus. Kuna eelduse kohaselt $0 < P$ ning iga liidetav

$$\frac{a_l \cdot (k-l)!}{k!}$$

on mittenegatiivne ($l \in \{1, \ldots, k\}$), siis leidub vähemalt üks nullist erinev parameeter a_l. Tähestikuks valime mingi lõpliku hulga S, mis rahuldab tingimust

$$|S| = \sum_{l=1}^{k} a_l.$$

$k = 1$ korral sobib tähestikuks mingi üheelemendiline hulk $S = \{s\}$ ning ainsaks võimalikuks koodiks kujutus $c : S \to T_1$, mille korral $c(s) = 1$. Seega $a_1 = 1$, mistõttu ka $P = 1$. Ilmselt on c prefiksivaba. Oletame nüüd, et $k \ge 2$. Et $P \le 1$, saab ka iga osasumma

$$\sum_{l=1}^{m} \frac{a_l \cdot (k-l)!}{k!} \quad (1 \le m \le k)$$

olla ülimalt võrdne ühega. Seega

$$\frac{a_1 \cdot (k-1)!}{k!} \le 1, \text{ millest } a_1 \le k = \binom{k}{1} 1!,$$

$$\frac{a_1 \cdot (k-1)!}{k!} + \frac{a_2 \cdot (k-2)!}{k!} \le 1, \text{ millest } a_2 \le \frac{k!}{(k-2)!} \cdot \left(1 - \frac{a_1}{k}\right) = \binom{k}{2} 2! - a_1 \binom{k-1}{1}$$

ning

$$a_i \leq \binom{k}{i} i! - \sum_{j=1}^{i-1} a_j \binom{k-j}{i-j} (i-j)!,$$

kus $1 \leq i \leq k$.
Kuna $a_1 \leq k$, saame valida a_1 koodsõna pikkusega 1. Et kood c peab olema prefiksivaba, jääb pikkusega 2 koodsõnade valikuks võimalusi

$$\binom{k}{2} 2! - a_1 \binom{k-1}{1},$$

sest kõikide pikkusega 2 koodsõnade hulgast tuleb välja jätta need, mille esimene sümbol on juba koodsõnana kasutusel. Et aga

$$a_2 \leq \binom{k}{2} 2! - a_1 \binom{k-1}{1},$$

siis on võimalik valida a_2 koodsõna pikkusega 2, säilitades koodi prefiksivabadust. Oletame nüüd, et meil on iga $h \in \{1, \ldots, i-1\} \subset \mathbb{N}$ korral õnnestunud prefiksivabadust säilitades valida a_h koodsõna pikkusega h. Pikkusega i koodsõnade valikuks jääb seega alles

$$\binom{k}{i} i! - \sum_{j=1}^{i-1} a_j \binom{k-j}{i-j} (i-j)!$$

võimalust ning kuna eelneva põhjal

$$a_i \leq \binom{k}{i} i! - \sum_{j=1}^{i-1} a_j \binom{k-j}{i-j} (i-j)!,$$

siis õnnestub meil valida a_i koodsõna pikkusega i, rikkumata prefiksivabaduse tingimust. Matemaatilise induktsiooni põhjal kehtib see suvalise $1 \leq i \leq k$ korral.

Seega kui $P \leq 1$, leidub alati piisavalt koodsõnu prefiksivaba koodi konstrueerimiseks. □

Järeldus 2.17. *Olgu $k \in \mathbb{N}$ ja $a_1, \ldots, a_k \in \mathbb{N} \cup \{0\}$. Kui*

$$P := \sum_{l=1}^{k} \frac{a_l \cdot (k-l)!}{k!} \leq 1,$$

siis leidub prefiksivaba üheselt dekodeeritav permutatsioonikood parameetritega $a_1, \ldots, a_k$.

Näide 2.18. Olgu $k = 3, a_1 = 1, a_2 = 3, a_3 = 1$. Siis

$$P = \sum_{l=1}^{3} \frac{a_l \cdot (3-l)!}{3!} = \frac{1 \cdot (3-1)!}{3!} + \frac{3 \cdot (3-2)!}{3!} + \frac{1 \cdot (3-3)!}{3!} = 1.$$

Seega teoreemi 2.16 kohaselt leidub prefiksivaba permutatsioonikood $c : S \to T_3$ parameetritega a_1, a_2, a_3. Tähestikuks S võtame mingi viie-elemendilise hulga $S = \{x_1, \ldots, x_5\}$. Üheks võimalikuks prefiksivabaks koodiks sobib näiteks kujutus c', mis on antud eeskirjaga

$$\begin{cases} c'(x_1) = 1 \\ c'(x_2) = 21 \\ c'(x_3) = 31 \\ c'(x_4) = 23 \\ c'(x_5) = 321. \end{cases}$$

Ilmneb, et kui kood on prefiksivaba, siis talle vastav permutatsioonikonstant P on ülimalt võrdne ühega. Selle näitamiseks paneme esmalt tähele, et iga $r \in T_k^l$ $(0 < l < k)$ korral saab sõnest r konstrueerida $(k-l)$ sõnet $r' \in T_k^{l+1}$, lisades talle lõppu mingi seni kasutamata sümboli hulgast $[k]$.

Definitsioon 2.19. Olgu $c : S \to T_k$ prefiksivaba permutatsioonikood parameetritega $a_1, \ldots, a_k$ ning olgu $i = \min\{1 \le l \le k \mid a_l > 0\}$. Koodi c *laiendiks* nimetatakse koodi $c^* : S' \to T_k$, mis on saadud koodist c pikkusega i koodsõnade asendamisel koodsõnadega pikkusega $i+1$, lisades algsete koodsõnade lõppu kõikvõimalikke vabu sümboleid. Kui $i = k$, siis $c^* = c$.

Näide 2.20. Olgu $S = \{x_1, \ldots, x_5\}$ ning $c : S \to T_3$ antud võrdustega

$$\begin{cases} c(x_1) = 1 \\ c(x_2) = 21 \\ c(x_3) = 31 \\ c(x_4) = 23 \\ c(x_5) = 321. \end{cases}$$

Siis tema laiendiks on kood $c^* : S' \to T_k$, kus $S' = \{x_1^1, x_1^2, x_2 \ldots, x_5\}$ ning

$$\begin{cases} c(x_1^1) = 12 \\ c(x_1^2) = 13 \\ c(x_2) = 21 \\ c(x_3) = 31 \\ c(x_4) = 23 \\ c(x_5) = 321. \end{cases}$$

Paneme tähele, et kui c^* on koodi c laiend, siis $|S'| = |S| + a_i \cdot (k-i-1)$. Koodi c prefiksivabadus tagab ka meie definitsiooni korrektsuse, kuna lisanduvad koodsõnad ei saa juba kasutusel olla.

Lause 2.21. *Kui kood c^* on koodi c laiend, siis c^* on prefiksivaba, kusjuures nende koodide parameetritele vastavad permutatsioonikonstandid on võrdsed.*

Tõestus. Juhul $c^* = c$ on väite kehtivus ilmne. Oletame nüüd, et $c^* \neq c$.
1) Näitame, et c^* on prefiksivaba. Eelduse kohaselt on c prefiksivaba. Väite tõestuseks piisab näidata, et ükski uutest koodsõnadest ei ole prefiksiks juba kasutusel olevatele koodsõnadele. Oletades nüüd vastuväiteliselt, et üks uutest koodsõnadest r^* rikub koodi prefiksivabaduse, oleks pidanud ka talle vastav algne koodsõna r olema prefiksiks mingile teisele koodsõnale. See on vastuolus koodi c prefiksivabadusega. Seega on ka c^* prefiksivaba.
2) Näitame, et koodide c ja c^* parameetritele vastavad permutatsioonikonstandid on võrdsed. Olgu $a_1, \ldots, a_k$ ja $a_1^*, \ldots, a_k^*$ vastavalt koodide c ja c^* parameetrid ning $i = \min\{1 \leq l < k \mid a_l > 0\}$. Ilmselt $a_j = a_j^*$, kui $j > i+1$ ning $a_j = a_j^* = 0$, kui $0 < j < i$. Koodi laiendi definitsiooni kohaselt $a_i^* = 0$. Et

$$a_{i+1}^* = a_{i+1} + a_i \cdot (k-i),$$

siis

$$\frac{a_{i+1}^* \cdot (k-i-1)!}{k!} + \frac{a_i^* \cdot (k-i)!}{k!} = \frac{a_{i+1} \cdot (k-i-1)!}{k!} + \frac{a_i \cdot (k-i)!}{k!}.$$

Seega kuna ülejäänud liikmed permutatsioonikonstandi arvutamise valemis on samuti võrdsed, on võrdsed ka vastavad permutatsioonikonstandid. □

Teoreem 2.22. *Olgu c prefiksivaba permutatsioonikood. Siis $P_c \leq 1$, kus P_c tähistab koodi c parameetritele vastavat permutatsioonikonstanti.*

Tõestus. Oletame väite vastaselt, et $P_c > 1$. Olgu $a_1, \ldots, a_k$ koodi c parameetrid ning $i = \min\{1 \leq l \leq k \mid a_l > 0\}$. Kui $i < k$, siis laiendame koodi c koodiks c_i^*. Lause 2.21 kohaselt on kood c_i^* prefiksivaba ning $P_{c_i^*} = P_c > 1$. Kui nüüd $i+1 < k$, siis laiendame koodi c_i^* koodiks c_{i+1}^*. Lause 2.21 kohaselt on kood c_{i+1}^* prefiksivaba ning $P_{c_{i+1}^*} = P_{c_i^*} = P_c > 1$. Analoogiliselt jätkates jõuame lõpuks koodini c^*, kus iga koodsõna on pikkusega k. Eelneva põhjal on c^* prefiksivaba ning $P_{c^*} > 1$, mistõttu

$$\frac{a_k \cdot (k-k)!}{k!} = \frac{a_k}{k!} > 1.$$

Samas aga $a_k \leq k!$, kuna $|T_k^k| = k!$. Seega

$$\frac{a_k}{k!} \leq 1.$$

Oleme jõudnud vastuoluni. □

2.5 Prefiksivabadus permutatsiooni mustrite kaudu

Olgu $n \in \mathbb{N}$. Defineerime hulgal $\mathbb{S}_n$ tehte $\circ$ võrdusega

$$(\sigma \circ \tau)(i) := \sigma(\tau(i)) \quad (\sigma, \tau \in \mathbb{S}_n,\ i = 1, \ldots, n).$$

Lause 2.23. $(\mathbb{S}_n, \circ)$ *on rühm.*

Tõestus.
G0) Näitame, et$(\mathbb{S}_n, \circ)$ on kinnine. Olgu $\sigma_1, \sigma_2 \in \mathbb{S}_n$ ning $i, j \in [n]$ $(i \neq j)$. Kuna $\sigma_2 \in \mathbb{S}_n$, siis

$$\sigma_2(i) \neq \sigma_2(j) \ \wedge\ \sigma_2(i), \sigma_2(j) \in [n].$$

Et $\sigma_1 \in \mathbb{S}_n$, siis eelneva põhjal

$$\sigma_1(\sigma_2(i)) \neq \sigma_1(\sigma_2(j)) \ \wedge\ \sigma_1(\sigma_2(i)), \sigma_1(\sigma_2(j)) \in [n].$$

Seega on $\sigma_1 \circ \sigma_2$ injektiivne funktsioon hulgast $[n]$ hullka $[n]$ ja järelikult bijektsioon, mistõttu $\sigma_1 \circ \sigma_2 \in \mathbb{S}_n$.
G1) Olgu $\sigma_1, \sigma_2, \sigma_3 \in \mathbb{S}_n$, $i \in [n]$. Näitame, et

$$[(\sigma_1 \circ \sigma_2) \circ \sigma_3](i) = [\sigma_1 \circ (\sigma_2 \circ \sigma_3)](i).$$

Tähistame $\tau_1 := \sigma_1 \circ \sigma_2$ ja $\tau_2 := \sigma_2 \circ \sigma_3$. Siis

$$[(\sigma_1 \circ \sigma_2) \circ \sigma_3](i) = (\tau_1 \circ \sigma_3)(i) = \tau_1(\sigma_3(i)) = \sigma_1(\sigma_2(\sigma_3(i)))$$

ja

$$[\sigma_1 \circ (\sigma_2 \circ \sigma_3)](i) = (\sigma_1 \circ \tau_2)(i) = \sigma_1(\tau_2(i)) = \sigma_1(\sigma_2(\sigma_3(i))).$$

G2) Olgu $\sigma \in \mathbb{S}_n$, $i \in [n]$. Näitame, et rühma ühikelemendiks sobib permutatsioon $e \in \mathbb{S}_n$, mis on antud seosega $e(j) = j \quad \forall j \in [n]$. Tõepoolest,

$$(\sigma \circ e)(i) = \sigma(e(i)) = \sigma(i)$$

ning

$$(e \circ \sigma)(i) = e(\sigma(i)) = \sigma(i).$$

G3) Olgu $\sigma \in \mathbb{S}_n$, $i \in [n]$. Teame, et igal bijektiivsel funktsioonil leidub üheselt määratud pöördfunktsioon. Seega leidub permutatsioonil σ pöördpermutatsioon $\sigma^{-1} \in \mathbb{S}_n$, mille korral

$$\sigma(k) = l \Leftrightarrow \sigma^{-1}(k) = l \quad (k, l \in [n]),$$

kusjuures

$$(\sigma \circ \sigma^{-1})(i) = (\sigma^{-1} \circ \sigma)(i) = i = e(i).$$

□

Definitsioon 2.24. Olgu $p, q \in \mathbb{N}$, $p \leq q$ ning olgu $\sigma \in \mathbb{S}_p, \tau \in \mathbb{S}_q$. Öeldakse, et permutatsioonis τ on *muster* σ, kui leiduvad naturaalarvud

$$1 \leq i_1 < \ldots < i_p \leq q$$

selliselt, et

$$\tau(i_{\sigma^{-1}(1)}) \ < \ \ldots \ < \ \tau(i_{\sigma^{-1}(p)}).$$

Näide 2.25. Olgu $\sigma = (1, 3, 2)$ ja $\tau = (4, 2, 5, 1, 3)$. Võttes $i_1 = 2$, $i_2 = 3$ ja $i_3 = 5$ saame, et

$$\tau(i_{\sigma^{-1}(1)}) = \tau(i_1) = \tau(2) = 2,$$
$$\tau(i_{\sigma^{-1}(2)}) = \tau(i_3) = \tau(5) = 3,$$
$$\tau(i_{\sigma^{-1}(3)}) = \tau(i_2) = \tau(3) = 5.$$

Seega on permutatsioonis $(4, 2, 5, 1, 3)$ muster $(1, 3, 2)$. Mustri tekitab siinkohal arvude 2, 5 ja 3 paigutus permutatsioonis τ. Arvud $1, 3, 2$ selles järjekorras paiknevad reaalteljel samas järjekorras kui vastavalt arvud $2, 5, 3$. Näeme seda, sest $2 < 5$, $2 < 3$, $5 > 3$ ja $1 < 3$, $1 < 2$, $3 > 2$. Samuti paiknevad nad vastavates permutatsioonides vasakult paremale.

Olgu meil nüüd fikseeritud mingi permutatsioon $\sigma \in \mathbb{S}_p$ ning olgu $\sigma_* \in \mathbb{S}_{p+1}$ permutatsioon, mis on saadud naturaalarvu $p+1$ lisamisel positsioonile l $(l = 1, \ldots, p+1)$ permutatsioonis σ, teisisõnu

$$\sigma_*(i) = \begin{cases} \sigma(i) & , 1 \le i < l \\ p+1 & , i = l \\ \sigma(i-1) & , l < i \le p+1 \end{cases} \qquad (i = 1, \ldots, p+1).$$

Lause 2.26. *Olgu $\sigma \in \mathbb{S}_p$ ning olgu $\sigma_* \in \mathbb{S}_{p+1}$ permutatsioon, mis on saadud naturaalarvu $p+1$ lisamisel permutatsiooni σ. Kui permutatsioonis $\tau \in \mathbb{S}_q$ $(p < q)$ on muster σ_*, siis permutatsioonis τ on ka muster σ.*

Tõestus. Eelduse kohaselt leiduvad naturaalarvud $1 \le i'_1 < \ldots < i'_{p+1} \le q$, et

$$\tau(i'_{\sigma_*^{-1}(1)}) \; < \; \ldots \; < \; \tau(i'_{\sigma_*^{-1}(p+1)}).$$

Olgu $\sigma_*^{-1}(p+1) = l$. Paneme tähele, et $\sigma_*^{-1}(j) \ge \sigma^{-1}(j) \quad \forall j = 1, \ldots, p$. Seega võttes

$$i_m = \begin{cases} i'_m & , m < l \\ i'_{m+1} & , m > l \end{cases} \quad (m = 1, \ldots, p),$$

saame

$$\tau(i_{\sigma^{-1}(1)}) \; < \; \ldots \; < \; \tau(i_{\sigma^{-1}(p)}).$$

□

Olgu $c : S \to \mathbb{T}_k$ permutatsioonikood. Tähistame

$$C := \{\sigma \in \mathbb{T}_k \mid \exists s \in S : c(s) = \sigma\}.$$

Definitsioon 2.27. Permutatsioonikood $c : S \to \mathbb{T}_k$ on *mustrivaba*, kui ühegi kasutusel olevas koodsõnas ei ole mingi teise koodsõna mustrit. See tähendab, et ühegi naturaalarvupaari $p, q \in [k], p \le q$ ning ühegi $\sigma \in \mathbb{S}_p \cap C$, $\tau \in \mathbb{S}_q \cap C$ $(\sigma \ne \tau, \sigma \ne \epsilon, \tau \ne \epsilon)$ korral ei leidu naturaalarve $1 \le i_1 < \ldots < i_p \le q$, et

$$\tau(i_{\sigma^{-1}(1)}) \; < \; \ldots \; < \; \tau(i_{\sigma^{-1}(p)}).$$

Näide 2.28. Olgu $S = \{x, y, z\}$ ning $c : S \to \mathbb{T}_4$ permutatsioonikood, kus $c(x) = 123$, $c(y) = 3412$ ja $c(z) = 4312$. Siis c on mustrivaba.

Definitsioon 2.29. Olgu $c : S \to \mathbb{T}_k$ permutatsioonikood parameetritega $a_1, \ldots, a_k$. Permutatsioonikoodile c vastavaks *mustrikonstandiks* nimetakse mittenegatiivset reaalarvu

$$Q_c := \frac{a_1}{1!} + \cdots + \frac{a_k}{k!} = \sum_{l=1}^{k} \frac{a_l}{l!}.$$

Järeldus 2.30. *Olgu $c : S \to \mathbb{T}_k$ permutatsioonikood parameetritega $a_1, \ldots, a_k$ ning P_c tema parameetritele vastav permutatsioonikonstant. Siis $P_c \leq Q_c$.*

Lause 2.31. *Olgu $c : S \to \mathbb{T}_k$ mustrivaba permutatsioonikood parameetritega $a_1, \ldots, a_k$ ning olgu $\{1 \leq l < k \mid a_l > 0\} \neq \varnothing$. Olgu kood c^* saadud koodist c pikkusega $i = min\{1 \leq l < k \mid a_l > 0\}$ koodsõnade asendamisel koodsõnadega, mis on saadud sümboli $i+1$ lisamisel kõikvõimalikesse positsioonidesse vastavates permutatsioonides. Siis on kood c^* mustrivaba ning $Q_c = Q_{c^*}$.*

Tõestus. Koodi mustrivabadusest järeldub, et koodis c^* ei teki korduvaid pikkusega i+1 koodsõnu. Seega on koodi c^* definitsioon korrektne.
1) Näitame, et c^* on mustrivaba. Oletame väite vastaselt, et üks uutest koodsõnadest σ' rikub koodi mustrivabaduse. Seega leidub $\tau \in \mathbb{T}_k$ nii, et permutatsioonis τ on muster σ'. Olgu σ koodsõnale σ' vastav algne koodsõna. Lause 2.26 kohaselt peab siis permutatsioonis τ olema ka muster σ. See on aga vastuolus eeldusega.
2)Näitame, et $Q_c = Q_{c_*}$. Olgu $a_1, \ldots, a_k$ ja $a_1^*, \ldots, a_k^*$ vastavalt koodide c ja c^* parameetrid. Ilmselt $a_j = a_j^*$, kui $j > i+1$ ning $a_j = a_j^* = 0$, kui $0 < j < i$. Ilmselt $a_i^* = 0$. Et

$$a_{i+1}^* = a_{i+1} + a_i \cdot (i+1),$$

siis

$$\frac{a_{i+1}^*}{(i+1)!} + \frac{a_i^*}{i!} = \frac{a_{i+1}}{(i+1)!} + \frac{a_i}{i!}.$$

Seega kuna ülejäänud liikmed mustrikonstandi arvutamise valemis on samuti võrdsed, on võrdsed ka vastavad mustrikonstandid. □

Teoreem 2.32. *Olgu $c : S \to \mathbb{T}_k$ mustrivaba permutatsioonikood. Siis $Q_c \leq 1$*

Tõestus. Oletame vastuväiteliselt, et $Q_c > 1$. Olgu $a_1, \ldots, a_k$ koodi c parameetrid ning $i = \min\{1 \leq l \leq k \mid a_l > 0\}$. Kui $i < k$, siis muudame koodi c koodiks c_i^*, asendades pikkusega i koodsõnad koodsõnadega, mis on saadud sümboli $i+1$ lisamisel kõikvõimalikesse positsioonidesse vastavates permutatsioonides. Lause 2.31 põhjal on kood c_i^* mustrivaba ning $Q_{c_i^*} = Q_c > 1$. Kui

nüüd $i+1 < k$, muudame eelpool kirjeldatud viisil koodi c_i^* koodiks c_{i+1}^*. Analoogiliselt jätkates jõuame lõpuks koodini c^*, kus iga koodsõna on pikkusega k. Eelneva põhjal on c^* mustrivaba ning $Q_{c^*} > 1$, mistõttu

$$\frac{a_k}{k!} > 1.$$

Samas aga $a_k \leq k!$, kuna pikkusega k koodsõnasid leidub ülimalt $k!$ tükki. Seega

$$\frac{a_k}{k!} \leq 1.$$

Oleme jõudnud vastuoluni. □

Järeldus 2.33. *Olgu $c : S \to \mathbb{T}_k$ mustrivaba permutatsioonikood. Siis koodi c parameetritele vastav permutatsioonikonstant on ülimalt võrdne ühega.*

Kui c on mustrivaba permutatsioonikood parameetritega $a_1, \ldots, a_k$, leidub järelduse 2.33 ja teoreemi 2.16 põhjal prefiksivaba üheselt dekodeeritav permutatsioonikood nende parameetritega. Osutub, et iga mustrivaba permutatsioonikood on prefiksivaba, mistõttu üheks selliseks koodiks sobib kood c ise.

Lause 2.34. *Olgu $l, p \in \mathbb{N}$ sellised, et $l < p$ ning olgu $\sigma = (n_1, \ldots, n_l) \in \mathbb{S}_l$, $\tau = (n_1', \ldots, n_p') \in \mathbb{S}_p$. Kui sõne $n_1 \ldots n_l$ on prefiksiks sõnele $n_1' \ldots n_p'$, siis permutatsioonis τ on muster σ.*

Tõestus. Eelduse kohaselt $n_j = n_j'$ kui $1 \leq j \leq l$, seega

$$\tau(\sigma^{-1}(j)) = j.$$

Võttes nüüd $i_j = j$, saame

$$\tau(i_{\sigma^{-1}(1)}) < \ldots < \tau(i_{\sigma^{-1}(l)}).$$

□

Järeldus 2.35. *Olgu $c : S \to \mathbb{T}_k$ mustrivaba permutatsioonikood. Siis on c prefiksivaba ja üheselt dekodeeritav.*

Näide 2.36. Olgu $k \in \mathbb{N}$, $k > 1$ .Vaatleme koodi $c : [k] \to \mathbb{T}_k$, $m \mapsto (1, \ldots, m)$. Ilmselt on koodi c parameetrid $a_1 = \ldots = a_k = 1$. Seega

$$Q_c = \sum_{l=1}^{k} \frac{1}{l!} > 1,$$

mistõttu pole c mustrivaba. Kuna aga

$$P_c = \sum_{l=1}^{k} \frac{a_l \cdot (k-l)!}{k!} = \sum_{l=1}^{k} \frac{1 \cdot (k-l)!}{k!} = \frac{1}{k} \cdot \sum_{l=1}^{k} \frac{(k-l)!}{(k-1)!} \leq \frac{1}{k} \cdot \sum_{l=1}^{k} 1 = 1,$$

siis leidub prefiksivaba üheselt dekodeeritav permutatsioonikood parameetritega $a_1, \ldots, a_k$.

Näide 2.37. Olgu $c : [4] \to \mathbb{T}_4$ antud võrdustega

$$\begin{cases} c(1) = 312 \\ c(2) = 213 \\ c(3) = 231 \\ c(4) = 1234. \end{cases}$$

Ilmselt on c mustrivaba. Seega on c ka prefiksivaba ja üheselt dekodeeritav.

Kirjandus

[NB] N. L. Biggs, *Codes: An Introduction to Information Communication and Cryptography*, Springer Verlag (2008), 1-25.

[JMSB] A. Jiang, R. Mateescu, M. Schwartz ja J. Bruck, *Rank Modulation for Flash Memories*, IEEE Transactions on Information Theory, **55** (2009), 2659-2660.

[MBZ] A. Mazumdar, A. Barg, ja G. Zémor, *Constructions of Rank Modulation Codes*, IEEE Transactions on Information Theory, **59** (2012), 1018-1019.

Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks

Mina, Kristo Visk (sünnikuupäev: 14.03.1993)

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) enda loodud teose

   "Permutatsioonikoodid ja allikakodeerimine",

   mille juhendaja on Ago-Erik Riet,

   1.1. reprodutseerimiseks säilitamise ja üldsusele kättesaadavaks tegemise ees-märgil, sealhulgas digitaalarhiivi DSpace'i lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;

   2.2. üldsusele kättesaadavaks tegemiseks Tartu Ülikooli veebikeskkonna kaudu, sealhulgas digitaalarhiivi DSpace'i kaudu kuni autoriõiguse kehtivuse tähtaja lõppemiseni.

2. olen teadlik, et punktis 1 nimetatud õigused jäävad alles ka autorile,

3. kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest tulenevaid õigusi.

Tartus, **05.06.2015**