Šifrování, kódování a jejich aplikace - ak. rok 2016/17

(zkratka předmětu: KAP/SKA, počet kreditů: 6)

Předmět je zakončen zkouškou, které musí předcházet získání zápočtu. Podmínky pro získání zápočtu a zkoušky jsou uvedený níže.

Vzhledem k počtu studentů kombinované formy, kteří mají v ak. roce 2016/17 předmět SKA zapsán, bude výuka probíhat formou konzultací. Jejich počet, termíny i obsah si studenti domlouvají s přednášejícím individuálně, tj. dle jejich potřeby.

Zápočet

- Student vypracuje zápočtový test – zadání viz níže. Řešení úloh musí být správné a dostatečně podrobně komentované. Formální stránka zpracování musí odpovídat VŠ úrovni.
- Vyřešené úlohy je třeba zaslat elektronickou poštou na univerzitní adresu přednášejícího ve formě souboru (v některém z formátů - doc, docx, rtf, pdf), a to nejpozději do konce zkouškového období ZS ak. roku 2016/17.
- O výsledku zápočtu bude student informován mailem do tří pracovních dnů od doručení řešení zápočtových úloh.

Zkouška

- Student s platným zápočtem se hlásí na zkoušku prostřednictvím systému stag. Zkušební termíny budou uveřejněny ve stagu nejpozději v zápočtovém týdnu zimního semestru ak. roku 2016/17.
- Zkouška má písemnou a ústní část. K ústní části postoupí pouze student, který uspěl v písemné části (tj. získá alespoň 70 % bodů z celkového možného počtu).

20. října 2016

doc. RNDr. Miroslav Koucký, CSc.
přednášející

Sylabus předmětu SKA v ak. roce 2016/17

Základy bezeztrátové komprese

- nejkratší kód, Huffmanova konstrukce nejkratšího kódu (binární + *n*-ární případ);
- konstrukce standardizovaného Huffmanova kódu (jednoznačnost kódu);

Aritmetické kódy - metoda DFWLD (kódování, dekódování); dyadické zlomky (jejich konstrukce);
Metody vyšších řádů (substituční schémata 1. řádu – Huffman, standardizovaná konstrukce)
Adaptivní metody (Huffman – kódování/dekódování);
Slovníkové metody - LZ77;

Detekční/opravné kódy

- základní pojmy, detekce chyb, chybové slovo, kód detekuje/opravuje *t*-násobné chyby, BSC (Binary Symmetric Channel); dekódování - opravování chyb (strategie MLD (CMLD/IMLD))

Lineární binární kódy

- definice (n, k); (n, k, d)- kódu, základní vlastnosti a pojmy (informační/kontrolní znaky, systematický kód, ekvivalence kódů, počet slov), min. vzdálenost $d = \min\{w(\boldsymbol{x}) | \boldsymbol{x} \in K - \{\mathbf{0}\}\}$, kód generovaný množinou slov *S* (lineární obal: $K = \langle S \rangle$), duální kód $K^{\perp}$;
- generující matice G, vlastnosti; kontrolní matice H, vlastnosti, vztah mezi G, H;
- rozklad ${}^{Z_2}/_{K}$, základní vlastnosti tříd rozkladů $\boldsymbol{e_i} + K$ (všechna slova z jedné třídy stejný syndrom) počet tříd rozkladů, jejich mohutnost; standardní dekódování; zásady volby chybových slov $\boldsymbol{e_i}$; nevýhody standardního dekódování;
- Základní nerovnosti (Hammingův, Singletonův, Gilbert-Varshamovův odhad), využití;
- Rozšířený kód, definice, vlastnosti, tvar G^*, H^*;

Perfektní kódy, definice, existenční věta, vlastnosti (perfektní pro opravy *t*-násobných chyb).

- Hammingův kód, dekódování.
- Golayův kód, vlastnosti, dekódování; rozšířený Golayův kód, vlastnosti, dekódování
- Reed-Mullerovy kódy R(r,m), rekurentní definice kódu i generující matice G(r,m), vlastnosti, dekódování RM(1,m)

Cyklické kódy

- základní pojmy: cyklický posun slova, polynomiální reprezentace slov, věta o dělení polynomů se zbytkem, $f(x) \equiv g(x)\ mod\ q(x)$; $f(x) = \big(g(x)\ mod\ q(x)\big)$
- definice cyklického kódu, generující polynom; informační polynom, kontrolní polynom, syndrom;

Úvod do šifrování

- základní pojmy (kryptologie = kryptografie + kryptoanalýza; steganografie), šifrovací systém/schéma;
- třídění šifrovacích metod (symetrické × asymetrické; transpozice × substituce; monoalfabetické × homofonní × polyalfabetické);
- Metoda RSA.
- blokové šifrování - Vernam, Feistel;
- hash funkce (vlastnosti, typy), jednosměrná (se zadními vrátky) funkce;
- digitální podpis, certifikát veřejného klíče, certifikační autorita; problém výměny šifrovacích klíčů (Diffie-Hellman, odolnost proti aktivnímu/pasivnímu protivníkovi);
- zmínka o jednoduché transpozii, jednoduchá substituce;

Doporučená literatura (dostupná na vyžádání u přednášejícího)

· Menezes, Oorschot, Vanstone: Handbook of applied cryptography. CRC Press
· Hankerson, Hoffman: Coding theory and cryptography: the essentials. Marcel Dekker
· Hankerson, Harris, Johnson: Introduction to Information Theory and Data Compression. CRC Press

Úspěšné absolvování předmětu SKA předpokládá základní znalosti následujících pojmů (získané v rámci předchozího Bc. studia)

- abeceda, slovo, délka slova, zřetězení, prefix, jazyk nad abecedou;
- grupy $(Z_m, +), (Z_m^*, \cdot), (S_n, \cdot)$; cyklické grupy; Lagrangeova věta;
- konečná tělesa $F_q = GF(p^k), q = p^k$; polynomy nad tělesy Z_p;
- vektorový prostor F_q^n nad tělesem F_q, lineární kombinace, lineární obal, lineární (ne)závislost, hodnost matice, REF, RREF, dimenze, báze; Hammingova vzdálenost/váha;
- Šifrování (kryptologie = kryptografie + kryptoanalýza; steganografie), základní pojmy, symetrické × asymetrické (s veřejným klíčem); transpozice × substituce;
- Kódování - komprese (ztrátová, bezeztrátová), detekční/opravné
    - základní pojmy, Kraftová nerovnost (důkaz), McMillanova věta, důsledek (prefixové kódy stejně obecné jako všechny jednoznačně dekódovatelné kódy);

Šifrování, kódování a jejich aplikace - zápočtové příklady pro ak. rok 2016/17

1. Určete, které z následujících kódů K1, K2, K3, K4 jsou jednoznačně dekódovatelné a které prefixové.
   $K_1 = \{0{,}01{,}011{,}0111{,}01111{,}011111\}$; $K_2 = \{xxx, xxy, xyx, yxx, xyy, yyx\}$; $K_3 = \{0{,}001{,}111{,}110{,}101{,}011\}$;
   $K_4 = \{0{,}1{,}20{,}21{,}220{,}221\}$

2. Kolik znaků musí mít kódová abeceda pro zakódování všech znaků anglické abecedy (26 znaků), jestliže požadujeme, aby kódová slova samohlásek (5) měla délku 1, ostatní délku 2.

3. Uvažujte zdrojovou abecedu $\begin{matrix} Znak \\ Pst. \end{matrix}\left|\begin{matrix} x & y & ! & ? \\ 0{,}4 & 0{,}2 & 0{,}2 & 0{,}2 \end{matrix}\right.$ (druhý řádek obsahuje četnost výskytu daného znaku). Pro následující kódování K1 a K2 odhadněte délku zakódované zprávy obsahující 150.

Znak	x	y	!	?
K1	0	10	110	111
K2	0	1	20	21

4. Pomocí Huffmanovy konstrukce určete nejkratší ternární kódování následujících abeced. Dále určete střední délku kódového slova.
   i) $\begin{matrix} a & b & c & d & e & f & g & h & i & j \\ 19/60 & 2/15 & 7/60 & 7/60 & 1/10 & 1/15 & 1/20 & 1/20 & 1/30 & 1/60 \end{matrix}$
   ii) $\begin{matrix} a & b & c & d & e & f & g & h & i \\ 17/50 & 7/50 & 7/50 & 2/25 & 2/25 & 2/25 & 3/50 & 3/50 & 1/50 \end{matrix}$

5. Nalezněte dyadické zlomky čísel: i) $\frac{25}{512}$ ii) $\frac{4}{7}$ iii) 5,40625

6. Uvažujte zdrojovou abecedu $\begin{matrix} a & b & c & d & e \\ 0{,}3 & 0{,}25 & 0{,}2 & 0{,}15 & 0{,}1 \end{matrix}$. Pomocí arit. kódování, metoda DFWLD, zakódujte slovo "cdea".

7. Uvažujte zdrojovou abecedu $\begin{matrix} a & b & c & d \\ 0{,}35 & 0{,}3 & 0{,}25 & 0{,}1 \end{matrix}$. Dekódujte následující slova, které vznikla aritmetickým kódováním metodou DFWLD: i) 11, ii) 010001. V obou případech byla délka zdrojového slova 4.

8. Uvažujte následující abecedy $\begin{matrix} Znak & a & b & c & d & e & f \\ Pst. & 0{,}25 & 0{,}2 & 0{,}2 & 0{,}15 & 0{,}15 & 0{,}05 \\ Pst. & 0{,}38 & 0{,}24 & 0{,}095 & 0{,}095 & 0{,}095 & 0{,}095 \\ Pst. & 5/13 & 2/13 & 2/13 & 2/13 & 2/13 & \end{matrix}$. Zkonstruujte standardizovaný Huffmanův kód.

9. Uvažujte adaptivní kódování (standardizovanou Huffmanovu konstrukci) abecedy $S = \{s_1, s_2, s_3, s_4, s_5, s_6\}$. Dekódujte 01100011001011111100

10. Uvažujte $BSC(0{,}98)$. Bylo přijato slovo $\boldsymbol{w} = 00110$. Které z následujících kódových slov 01101, 01001, 10100, 10101 bylo nejpravděpodobněji vysláno? (BSC … binární symetrický kanál)

11. Nechť $\boldsymbol{u} = 11010, \boldsymbol{v} = 01100$. Spočtěte $w(\boldsymbol{u}+\boldsymbol{v}), w(\boldsymbol{u}) + w(\boldsymbol{v}), d(\boldsymbol{u},\boldsymbol{v})$, kde w označuje Hammingovu váhu a d Hammingovu vzdálenost.

12. Rozhodněte, zda kód $K = \{001{,}101{,}110\}$ detekuje a) chybové slovo $\boldsymbol{e_1} = 010$, b) chybové slovo $\boldsymbol{e_2} = 100$.

13. Které z následujících kódů jsou lineární:
a) $\{101{,}111{,}011\}$ b) $\{000{,}001{,}010{,}011\}$ c) $\{0000{,}0001{,}1110\}$ d) $\{0000{,}1001{,}0110{,}1111\}$

14. Určete kód $\langle S \rangle$, tj. lineární kód generovaný množinou S, kde: a) $S = \{0100{,}0011{,}1100\}$ b) $S = \{010{,}011{,}111\}$

15. Uvažujte binární kód celkové kontroly parity délky *n*. a) Rozhodněte, zda je lineární (v kladném případě sestavte generující a kontrolní matici), dále rozhodněte, zda je systematický; b) určete minimální vzdálenost; c) určete kolikanásobné chyby objevuje/opravuje.

16. Uvažujte binární lineární kód s generující maticí $\begin{pmatrix} 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 1 \end{pmatrix}$. a) Sestavte kontrolní matici. b) Určete počet kódových slov a počet chybových slov, která je kód schopen opravit. c) Uvažujte chybová slova $e_1 = (000001), e_2 = (000010), e_3 = (100000)$ a dekódujte slova $y_1 = (010101), y_2 = (110011), y_3 = (000111), y_4 = (111000)$.

17. Uvažujte binární Hammingův kód řádu 3. Zapište jeho kontrolní matici a dekódujte slova (1100000), (1011101).

18. Dešifrujte text „CSUCDD“, který vzniknul zašifrování pomocí Hillovy šifry s šifrovací maticí $\begin{pmatrix} 8 & 23 & 11 \\ 15 & 0 & 13 \\ 10 & 20 & 17 \end{pmatrix}$.

19. Uvažujte dvoustupňové Feistel šifrování, které pro oba cykly používá šifrovací funkci
$$f(x_1, x_2, x_3, x_4) = (\overline{x_1}, x_2 \oplus x_4, x_3 + x_1, x_2 \cdot x_4).$$
Zašifrujte text „ON“. Pro převod otevřeného textu na binární řetězec použijte ASCII kód.
(označení logických operací: $\overline{\ }$ … negace; $\oplus$ … vylučující nebo (xor); $+$ … nebo; $\cdot$ … a)