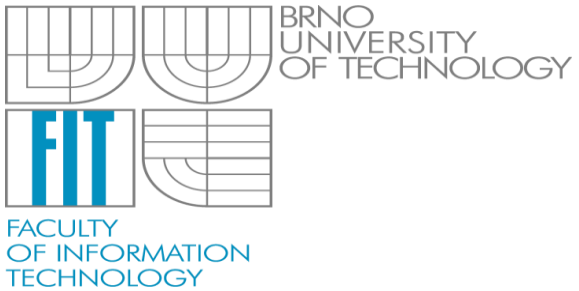
Substituční šifry a frekvenční analýza

Mgr. Radim Janča
ijanca@fit.vutbr.cz



Harmonogram

- Celkově 4 cvičení v P256
- Prezentace z cvičení budou zveřejňovány na http://buslab.fit.vutbr.cz/kib/
- 3 samostatné projekty
- Projekty jsou povinné bodované 0-6 body
- Z každého projektu musíte získat min. 3 body
- Bonusové body z projektů

Základní pojmy

- Kryptografie
- Kryptoanalýza
- Prostý text /plaintext - `mala pismena`
- Zašifrovaný text / ciphertext - `VELKA PISMENA`
- Šifra / cipher
- Šifrování / encryption
- Dešifrování /decryption

Motivace

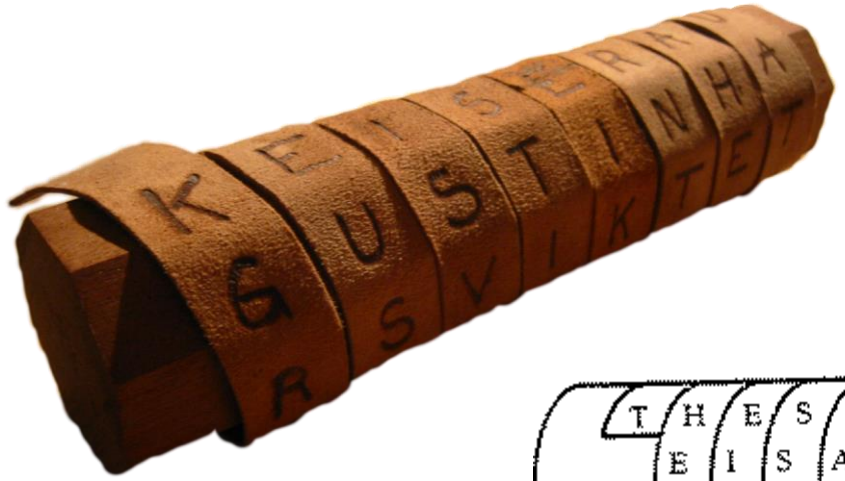
- Sdílení tajných informací bez strachu z prozrazení

- Zasílání válečných rozkazů
- Znalost nepřátelských plánů může změnit výsledek války
- Bezdrátová komunikace bez šifrování?

- Dnes setkáváme s kryptografií prakticky na každém kroku -> internet
- Elektronické platby, autentizace webových stránek, HTTPS…

Transpoziční šifra

- Změna pořadí znaků v textu
- Použití geometrického obrazce
- Pro dešifrování je použit reverzní postup
- Příklad – šifra Scytale



Substituční šifra

- Nahrazení znaků prostého textu podle klíče

- Caesarova šifra – Caesar šifroval tak, že nahradil každé písmeno třetím následujícím v abecedě
- `caesar= FDHVDU`
  - Nevýhoda – každý kdo zná algoritmus, může zprávu dešifrovat

- Vylepšení 1
  - Odesílatel si zvolí klíč od 1 do *n*, kde *n* je počet znaků abecedy
  - Šifra je stále slabá, příliš malý prostor klíčů

- Vylepšení 2
  - Použití permutace abecedy jako klíče

```
abcdefghijklmnopqrstuvwxyz
VWXYZEFGHIJKLMABCDNOPQRSTU
```

  - `medvedik cistotny = LZYQZYHJ XHNOAOMT`
  - Takto vytvořených klíčů je 26! = 4x10^26
- Šifra nerozluštitelná celé první tisíciletí našeho letopočtu

- Každý jazyk má svá specifika

- Frekvence jednotlivých písmen v textu
- Frekvence digramů, trigramů
- Frekvence slov

A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**	**M**
8.2	1.5	2.8	4.3	12.7	2.2	2.0	6.1	7.0	0.2	0.8	4.0	2.4
N	**O**	**P**	**Q**	**R**	**S**	**T**	**U**	**V**	**W**	**X**	**Y**	**Z**
6.7	7.5	1.9	0.1	6.0	6.3	9.1	2.8	1.0	2.4	0.2	2.0	0.1

- Vlastnosti pro angličtinu
  - Nejčastější písmena: etaoinshrdlu
  - Digramy: th, he, in, er, an, re, ed, on, es, st, en, at, to, nt, ha, nh, ou, ea, ng, as, or, ti, is, et, it, ar, te, se, hi, of
  - Trigramy: the, ing, and, her, ere, ent, tha, nth, was, eth, for, dth
  - Stejné znaky tvořící dvojice: ss, ee, tt, ff, ll, mm, oo
  - Dále může napomoci frekvence slov, frekvence slov za čárkou atd.

  - Pro ztížení dešifrování je vhodné prostý text zbavit mezer, interpunkce a diakritiky.

Frekvenční analýza

- Poprvé popsal Al-Kindi v 9 století n.l.
- Určení jazyka pomocí indexu koincidence
- Určení početnosti písmen
- Určení početnosti digramů a trigramů
- Postupné odkrývání textu

- Pokud je znám účel textu je vhodné vytvořit si vlastní frekvenční tabulky
- Vojenská korespondence bude mít jiné frekvence než běžná komunikace

Index koincidence

- Způsob jak zjistit jazyk zprávy
- Jazyky mají různé indexy koincidence
- *IK* – pravděpodobnost, že dvě náhodná písmena z textu budou stejná
- f_i – počet znaků *i* v textu
- *N* – celkový počet znaků textu

$$IK = \frac{\sum_{i="A"}^{"Z"} f_i(f_i - 1)}{N(N-1)}$$

čeština	0.06027
angličtina	0.06689
dánština	0.07073
finština	0.07380
francouzština	0.07460
holandština	0.07981
němčina	0.07667

Příklad

YAQZOSXYGJEXMJXNHMEIESIKIZEIZUASYHDXYGMEIESXLISJESH
HMAYMAYJEHJEAOMIYZIYZWXSMJYXGEJQEHYMEHEIZHYZHZJEHJI
BHAWNIISOWMEHSAMHIYZDXMMHZJEHGSAOYZUHWASHEXNMILXYGG
SHIJDXYGWASIJEAOMIYZIYZAYHYXGEJMXEIFHUHHYSHPAOYJXYG
JALAOJEHWIUBHMAWRIMJIGHMIYZJEHBHGHYZMAWIYPXHYJDXYGM
NILXNIDHMAUABZIMJAPSIFHIWIFASAWLAOSNITHMJL

- Index koincidence - 0.06563 – jedná se pravděpodobně o angličtinu
- Celkový počet znaků 337

Frekvence jednotlivých znaků

H	11.1111%	I	10.4377%	Y	9.0909%
A	7.7441%	M	7.7441%	E	7.4074%
J	7.0707%	S	5.7239%	X	5.7239%
Z	5.3872%	G	4.0404%	W	3.3670%
O	2.6936%	N	2.0202%	L	2.0202%
D	1.6835%	U	1.6835%	B	1.3468%
P	1.0101%	F	1.0101%	Q	0.6734%
T	0.3367%	K	0.3367%	R	0.3367%

Příklad

- Nejčastější znaky: H I Y
- Pravděpodobně odpovídají e t a – to nemusí platit vždy
- H I tvoří digram s většinou znaků, jedná se pravděpodobně o samohlásky
- Y v textu nesousedí s 15 znaky, jedná se pravděpodobně o souhlásku

Příklad

- Nejčastější trigram v angličtině the
- t a e mají vysokou četnost, h střední
- Častý výskyt digramu he, ale nízký u eh
- Nejčastější trigramy: XYG, IYZ, JEH
- JEH -> the
- IYZ -> and
- XYG ->ing

Příklad

```
n..d..ingthi.ti.e.hah.a.adhad...ne.ing.hah.i.a.th.e
e..n..ntheth...andand.i..tnight.hen.hehadendedtheta
.e...aa....he...eand.i..edtheg...nd.e...ehi..a.ingg
.eat.ing...ath...andand.nenight.iha.e.een.e...nting
t....the.a..e....a.tage.andthe.egend...an.ient.ing.
.a.i.a.e.....da.t...a.ea.a..........a.e.t.
```

- Vidíme slova `th.ee,  .een,  .egend`
- Stanovíme S -> r, U -> b, B -> l

Příklad

```
n..d.ringthi.ti.e.hahra.adhadb.rne.ing.hahri.arthre
e..n..ntheth...andand.ir.tnight.hen.hehadendedtheta
le...aar...her..eand.i..edthegr..ndbe..rehi..a.ingg
reat.ing..rath...andand.nenight.iha.ebeenre...nting
t....the.able....a.tage.andthelegend...an.ient.ing.
.a.i.a.e..b.lda.t..ra.ea.a..r.....r.a.e.t.
```

- Vidíme slova `ha.e, .he, d.ring`
- Stanovíme F -> v, M -> s, O -> u

```
nowduringthistimeshahrazadhadbornekingshahriyarthre
esonsonthethousandandfirstnightwhenshehadendedtheta
leofmaarufsheroseandkissedthegroundbeforehimsayingg
reatkingforathousandandonenightsihavebeenrecounting
toyouthefablesofpastagesandthelegendsofancientkings
mayimakesoboldastocraveafavorofyourmajesty
```

- Nakonec najdeme celý klíč

```
ABCDEFGHIJKLMNOPQRSTUVWXYZ
ol_khvgeatzysmucwprjb_find
```

Now during this time Shahrazad had borne King Shahriyar three sons. On the thousand and first night, when she had ended the tale of Maaruf, she rose and kissed the ground before him, saying: "Great King, for a thousand and one nights I have been recounting to you the fables of past ages and the legends of ancient kings. May I make so bold as to crave a favor of your majesty?

Epilogue, Tales from the Thousand and One Nights

- Dešifrování textu
- Monoalfabetická substituční šifra
- Zveřejnění zadání 24.9.2015 na http://www.securityfit.cz/kib/
- Odevzdání projektu do 8.10.2015 na malinka@fit.vutbr.cz
- Předmět emailu “KIBProjekt 1”
- Odevzdaná zpráva ve formátu pdf bude obsahovat dešifrovaný text, klíč postup dešifrování.
- Možný bodový zisk: 10 bodů

Děkuji za pozornost