Frekvenční analýza a substituční šifry

Marek Kumpošt

Laboratoř bezpečnosti
a aplikované kryptografie

Fakulta informatiky
Masarykova univerzita
Brno

Obsah

Rozdělení šifer podle způsobu šifrování

- Transpoziční šifry
  - Písmena zprávy se uspořádají jiným způsobem – přesmyčka
  - Přeuspořádání písmen otevřeného textu
  - Problém při dešifrování – je potřeba použít stejný postup
  - Vhodnější použít jednodušší „předpis“
  - Např. zpráva navinutá na tyči konkrétního průměru
  - Např. rozdělení zprávy do dvou řádků a napojení za sebe
- Substituční šifry
  - Nahrazení písmen otevřeného textu podle nějakého pravidla
  - A→V, D→X, H→B, I→G, K→J, M→C, O→Q, R→L, S→N, U→E, W→F, Y→P, Z→T
  - schuzka o pulnoci → NMBETJV Q YERSQMG
  - Caesarova (posuvná) šifra – písmeno zprávy nahrazeno písmenem o tři pozice dále v abecedě (celkem 25 odlišných šifer)
  - Nemusíme nutně posouvat, lze i přeskládat abecedu ($4 * 10^{26}$ možností)
  - a b c d e f g h i j k l m n o p q r s t u v w x y z
  - V E S L O A B C D F G H I J K M N P Q R T U W X Y Z

Substituční šifry – algoritmus a klíč

- Algoritmus je šifrovací metoda použita k šifrování
- Klíč specifikuje detaily použitého šifrování – např. posun o n písmen
- Bez klíče by nemělo být možné dešifrovat zprávu
- Kerckhoff (1883) – bezpečnost šifrovacího systému nesmí záviset na utajení algoritmu, ale pouze na utajení klíče
- Prostor klíčů musí být značný (Caesarova šifra vs. substituční algoritmus)
- Snadné použití a vysoký stupeň bezpečnosti
  - Odesílatel definuje nějaké přeuspořádání písmen abecedy
  - Útočník musí hledat toto přeuspořádání
  - Klíč musí sdílet obě komunikující strany – jednoduchost klíče $\longrightarrow$ snížení rizika nedorozumění
  - Jako klíč můžeme použít *klíčové slovo* nebo *klíčovou frázi*
  - Např. klíč JULIUS CAESAR $\longrightarrow$ JULISCAER
  - a b c d e f g h i j k l m n o p q r s t u v w x y z
  - J U L I S C A E R T V W X Y Z B D F G H K M N O P Q

Celkem tedy...

- Díky této jednoduchosti dominovala substituční šifra v tajné komunikaci velmi dlouhou dobu
- Natolik bezpečný systém, že neexistovala motivace pro jeho vylepšování
- Věřilo se, že obrovské množství klíčů zajišťuje nerozluštitelnost
- Nakonec nalezen způsob, jak výrazně luštění urychlit...
  - Unikátní kombinace lingvistiky a statistiky
  - *Kryptoanalýza* – dešifrování zprávy bez znalosti klíče (poprvé Arabové, 9. století)
- Monoalfabetická šifra
  - Nahrazuje písmeno textu vždy stejným znakem v šifrované podobě
  - a je např. vždy šifrováno jako `T`
  - Četnosti písmen v otevřeném i zašifrovaném textu se **zachovávají**
- Polyalfabetická šifra
  - Nahrazuje písmeno textu vždy různým znakem v šifrované podobě
  - a je šifrováno jako `T`, jindy jako `S`
  - V šifrovaném textu se **nezachovávají** četnosti písmen otevřeného textu

Kryptoanalýza – snaha dešifr. text bez znalosti klíče

- Zaměříme se na frekvenční analýzu
- Poprvé Arabové, založeno na znalostech lingvistiky a statistiky
- Známe-li jazyk zprávy, nalezneme odlišný otevřený text v tomtéž jazyce a spočteme výskyty jednotlivých písmen. Písmena setřídíme podle četnosti. Totéž provedeme se šifrovaným textem. Písmena šifrovaného textu nahradíme odpovídajícím způsobem podle četností písmen v otevřeném textu.
- V současnosti známe rozložení písmen nejběžnějších jazyků
- *Angličtina* – **a:8,2**; b:1,5; c:2,8; d:4,3; **e:12,7**; f:2,2; g:2,0; h:6,1; **i:7,0**; j:0,2; k:0,8; l:4,0; m:2,4; n:6,7; **o:7,5**; p:1,9; q:0,1; r:6,0; s:6,3; **t:9,1**; u:2,8; v:1,0; w:2,4; x:0,2; y:2,0; z:0,1
  - Vzorek 100 365 znaků anglické abecedy (noviny a beletrie)
- Problém u krátkých zpráv (From Zanzibar to Zambia and Zaire, ozone zones make zebras run zany zigzags)
- Alespoň sto písmen pro naději na úspěch

Frekvenční analýza – postup při hledání klíče

- Zjištění, v jakém jazyce může být zpráva napsána
- Určení četností písmen
- Určení četností digramů a trigramů
- Identifikace nejčastěji se vyskytujících písmen
- Postupné odkrývání původního textu a určování dalších písmen
- Dešifrování zprávy a zjištění použitého klíče

Frekvenční analýza – příklad

- Text je v angličtině
- Zašifrován monoalfabetickou substituční šifrou
- Neznáme klíč

```
PCQVM JYPDL BYKLY SOKBX BJXWX VBXVZ CJPOE YPDKB XBJYU
XJLBJ OOKCP KCPLB OLBCM KXPVX PVIYJ KLPYD BLQBO PKBOB
XVOPV OVLBO LXROC ISXXJ MIKBO JCKOX PVEYK KOVLB ODJCM
PVZOI CJOBY SKXUY PDDJO XLEYP DICJX LBCMK XPVXP VCPOP
YDBLK YBXNO ZOOPJ OACMP LYPDL CUCML BOIXZ ROKCI FXKLX
DOKXP VLBOR ODOPV KCIXP AYOPL EYPDK SXUYS XEOKC ZCRVX
KLCAJ XNOXI XNCJC IUCMJ SXGOK LUOFY RCDMO LXROK IJCSL
BOLBC MKXPV XPVCP OPYDB LK
```

- 337 znaků – není to mnoho na mechanickou frekvenční analýzu
- Zjistíme četnosti jednotlivých písmen šifrované zprávy

Frekvenční analýza – příklad

- Zjištění četností jednotlivých písmen šifrované zprávy

```
O    38   ######################################
X    34   ##################################
P    31   ###############################
C    27   ###########################
K    26   ##########################
B    25   #########################
L    25   #########################
Y    19   ###################
J    18   ##################
V    18   ##################
D    14   ##############
I    11   ###########
M    10   ##########
```

Frekvenční analýza – příklad

- Nejčastější znaky jsou `O P X`
- Nejspíše budou odpovídat nejčetnějším písmenům anglické abecedy `e t a`
- Nemůžeme si být jisti pořadím
- Podíváme se, jaká písmena nejčastěji sousedí s `O P X`
- `O`, `X` se páruje s většinou písmen (kromě 7, resp. 8 z nich)
- `P` se nepáruje s 15ti písmeny – pravděpodobně souhláska
- `OO` nalezneme dvakrát, `XX` ani jednou – častý výskyt `ee` v angl. textech
- Takže pravděpodobně `O=e`, `X=a`
- Následně se snažíme odhalit písmeno `h` – velmi častý výskyt před `e` ale vzácně po něm (`the`, `then`, `they`)
- `B` se před `O` vyskytuje devětkrát, takže pravděpodobně `B=h`
- Nejčastější trigramy `XPV YPD LBO` – zřejmě `XPV=and LBO=the`

Frekvenční analýza – příklad

- Dešifrujeme text s dosavadními znalostmi
- `O=e X=a B=h L=t P=n V=d`

```
n--d- --n-t h--t- -e-ha h-a-a dhad- --ne- -n--h ah---
a-th- ee--n --nth eth-- -anda nd--- -tn-- ht-he n-heh
adend edthe ta-e- --aa- ---he ---ea nd--- -edth e----
nd-e- --eh- --a-- n---e at--n ----a th--- andan d-nen
--ht- -ha-e -een- e---n t-n-t ----t he-a- -e--- -a-ta
-e-an dthe- e-end ---an --ent --n-- -a--- a-e-- ---da
-t--- a-ea- a---- ----- -a-e- t-e-- ----e ta-e- ----t
heth- --and and-n en--h t-
```

- Vidíme následující možná slova: `th- ee; -heh adend edthe; th--- andan d-nen --ht; -ha-e -een;`
- Proto stanovíme `J=r K=s C=o M=u N=v Z=b`

Frekvenční analýza – příklad

- Dešifrujeme text s dosavadními znalostmi
- `O=e X=a B=h L=t P=n V=d J=r K=s C=o M=u N=v Z=b`

```
no-du r-n-t h-st- -esha hra-a dhadb orne- -n-sh ahr--
arthr eeson sonth ethou sanda nd--r stn-- ht-he nsheh
adend edthe ta-eo --aar u-she rosea nd--s sedth e-rou
ndbe- oreh- -sa-- n--re at--n --ora thous andan donen
--hts -have beenr e-oun t-n-t o-out he-ab -eso- -asta
-esan dthe- e-end so-an --ent --n-s -a--- a-eso bo-da
sto-r avea- avoro --our -a-es t-e-- -o-ue ta-es -ro-t
hetho usand andon en--h ts
```

- Vidíme následující možná slova: `no-du r-n-; ta-e; -rou ndbe-oreh-; thous andan donen --hts -have been`
- Proto stanovíme `Q=w Y=i D=g R=l I=f`

Frekvenční analýza – příklad

- Dešifrujeme text s dosavadními znalostmi
- `O=e X=a B=h L=t P=n V=d J=r K=s C=o M=u N=v Z=b Q=w Y=i D=g R=l I=f`

```
nowdu ringt histi -esha hra-a dhadb orne- ingsh ahri-
arthr eeson sonth ethou sanda ndfir stnig htwhe nsheh
adend edthe taleo f-aar ufshe rosea nd-is sedth egrou
ndbef orehi -sa-i nggre at-in gfora thous andan donen
ights ihave beenr e-oun tingt o-out hefab lesof -asta
gesan dthel egend sofan -ient -ings -a-i- a-eso bolda
sto-r aveaf avoro f-our -a-es t-e-i logue tales fro-t
hetho usand andon enigh ts
```

- Určíme zbytek klíče `A=c G=j E=k S=m F=p H=q T=x U=y V=z`

Frekvenční analýza – příklad

- Dešifrujeme text
- O=e X=a B=h L=t P=n V=d J=r K=s C=o M=u N=v Z=b Q=w Y=i
  D=g R=l I=f A=c G=j E=k S=m F=p H=q T=x U=y V=z

  nowdu ringt histi mesha hraza dhadb ornek ingsh ahriy
  arthr eeson sonth ethou sanda ndfir stnig htwhe nsheh
  adend edthe taleo fmaar ufshe rosea ndkis sedth egrou
  ndbef orehi msayi nggre atkin gfora thous andan donen
  ights ihave beenr ecoun tingt oyout hefab lesof pasta
  gesan dthel egend sofan cient kings mayim akeso bolda
  stocr aveaf avoro fyour majes tyepi logue tales fromt
  hetho usand andon enigh ts
- a vložíme správně mezery.

Frekvenční analýza – příklad

- Výsledný text

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

- a b c d e f g h i j k l m n o p q r s t u v w x y z
- X Z A V O I D B Y G E R S P C F H J K L M N Q T U W

Dotazy?

Díky za pozornost a přeji hodně úspěchu při luštění!

xkumpost@fi.muni.cz