Projekt č. 1 - Vigenèrova šifra
KRY

Bc. Marián Kapišinský (xkapis00)

4.4.2021

Obsah

1 Index koincidencie

V implementácií sa využíva výpočet indexu koincidencie $I_c(x)$ (1.1), ktorý je definovaný ako pravdepodobnosť, že dva náhodné prvky x sú identické, kde $x = x_1, x_2, ..., x_n$ je reťazec znakov dĺžky n, a f_i je frekvencia i-teho znaku daná celým číslom ako počet výskytov i-teho znaku v skúmanom zašifrovanom texte (uvažujeme 26 znakovú anglickú abecedu) [1].

$$I_c(x) = \frac{\sum_{i=0}^{25} f_i(f_i - 1)}{n(n-1)} \tag{1.1}$$

2 Friedmanov test

Friedmanov test vychádza z indexu koincidencie (1.1). Uvažujeme 26 znakovú anglickú abecedu, pre ktorú určím hodnotu indexu koincidencie reťazca v anglickom jazyku (2.1), a hodnotu indexu koincidencie zcela náhodného reťazca, t.j. všetky znaky majú rovnakú frekvenciu (2.2) [1]. Dĺžku klúča r potom určím podľa vzorca 2.3, kde n je dĺžka a $I_c(x)$ je index koincidencie zašifrovaného textu [2].

$$I_c \approx \sum_{i=0}^{25} p_i^2 = 0.065 \tag{2.1}$$

$$I_c \approx 26\left(\frac{1}{26}\right)^2 = 0.0385 \tag{2.2}$$

$$r \approx \frac{(0.065 - 0.0385)n}{(n-1)I_c(x) - 0.0385n + 0.065} \tag{2.3}$$

Pozn.: Hodnota p_i značí frekvenciu výskytu i-teho znaku anglickej abecedy podľa tabulky 2.1.

Znak	Frekvencia	Znak	Frekvencia
a	0.08167	n	0.06749
b	0.01492	o	0.07507
c	0.02782	p	0.01929
d	0.04253	q	0.00095
e	0.12702	r	0.05987
f	0.02228	s	0.06327
g	0.02015	t	0.09056
h	0.06094	u	0.02758
i	0.06966	v	0.00978
j	0.00153	w	0.02360
k	0.00772	x	0.00150
l	0.04025	y	0.01974
m	0.02406	z	0.00074

Tabuľka 2.1: Frekvencie znakov anglického jazyka [2]

3 Kasiského test

Pre Kasiského test [2] som zvolil dĺžku n-gramov 3 až 8 vrátane. Pre každý n-gram v zašifrovanom texte vyhľadám všetky pozície, na ktorých sa vyskytuje. N-gramy, ktoré sa vyskytnú len raz a n-gramy, ktoré sa už preskúmali, sa preskočia. Následne sa z pozíc vypočítajú vzdialenosti medzi všetkými n-grammi ako rozdiel aktuálnej a predchádzajúcej pozície. Zo vzdialeností sa potom nájde ich najväčší spoločný deliteľ. Náhodné výskyty n-gramu sú ošetrené tak, že ak nejaká vzdialenosť spôsobí, že najväčší spoločný deliteľ je rovný 1, tak sa táto hodnota zahodí. Výsledné hodnoty (potenciálne dĺžky klúča) menšie ako 3 sa zahodia taktiež. Následne sa zo všetkých hodnôt vyberie najčastejšia hodnota a tá sa považuje za správnu dĺžku klúča.

4 Určenie dĺžky klúča

Pre určenie dĺžky hesla opäť využívam index koincidencie. Vieme, že pre anglický jazyk je jeho hodnota približne 0.065. Táto hodnota je približne rovnaká pre zašifrovaný text vytvorený monoalfabetickou šifrou. Takže, ak od prvého znaku textu zoberieme každý r-tý znak (tj. znaky na indexoch 1, 1+r, 1+2r,...), kde r je dĺžka klúča, hodnota indexu koincidencie tohto podreťazca by mala byť podobná hodnote v anglickom jazyku [1]. Za predpokladu, že minimálna dĺžka klúča je 4, pre každú hodnotu, počínajúc minimálnou dĺžkou až po hodnotu 200 vrátane[1], vytvorím podreťazec $c = c_1 c_{1+r} c_{1+2r} c_{1+3r}...$, a ak bol tento podreťazec vytvorený správnou hodnotou r, tak hodnota indexu koincidencie bude približne 0.065. V implementácií teda hľadám prvé r, pre ktoré bude hodnota indexu koincidencie vačšia, rovná 0.060. Prvé r beriem z dôvodu, že všetky násobky tejto hodnoty mali podobného hodnoty, a ak som vyberal maximum zo všetkých hodnôt, veľmi často bola vybraná nespráva hodnota r.

[1] vlastné zvolené obmedzenie dĺžky klúča

5 Určenie klúča

V tomto kroku predpokladáme už známu dĺžku hesla r. Z pôvodného zašifrovaného textu vytvoríme podreťazce podľa vzorca 5.1. Následne na každý z týchto podreťazcov aplikujem vzorec 5.2 pre každé g, kde g je posuv abecedy, p_i je frekvencia i-tého znaku podľa tabulky frekvencií 2.1, f_i je frekvencia i-tého znaku v podreťazci a n je dĺžka podreťazca [1]. Pre každý podreťazec dostanem pole hodnôt M_g a za výsledný znak klúča vyberiem to g, pre ktoré je hodnota najvyššia, t.j. ak má najvyššiu hodnotu napr. M_3, znamená to, že znak klúča bude c.

$$\begin{aligned} c_1 &= c_1 c_{1+r} c_{1+2r} c_{1+3r} \cdots \\ c_2 &= c_2 c_{2+r} c_{2+2r} c_{2+3r} \cdots \\ &\vdots \\ c_r &= c_r c_r c_{2r} c_{3r} \cdots \end{aligned} \tag{5.1}$$

$$M_g = \sum_{i=0}^{25} \frac{p_i f_{i+g}}{n} \tag{5.2}$$

Literatúra

[1] Hanáček, P. *Kryptografie, Část 1, Klasická kryptografie: Souhrnné materiály.* 2021.

[2] Lewand, R. E. *Cryptological Mathematics.* 1. vyd. The Mathematical Association of America, 2000. ISBN 0-88385-719-7.