Kryptografie (KRY)

Dokumentace k 1. projektu: *Lámání Vigenerovy šifry*

Vladimír Dušek, xdusek27

1 Úvod

Cílem projektu je prolomení Vigenerovy šifry. Jedná se o historickou symetrickou polyalfabetickou substituční šifru. Vigenerova šifra šifruje text pomocí série různých Caesarových šifer v závislosti na písmenech klíče. Princip lámání spočívá v odhadnutí délky hesla pomocí Kasiského testu a jeho upřesnění pomocí Friedmanova testu. Pokud je zjištěna délka hesla, lze problém rozdělit na n textů zašifrovaných Caesarovou šifrou (kde n je délka klíče). Caesarovu šifru lze pak prolomit hrubou silou.

2 Kasiského test

Kasiského test spočívá v hledání shodných posloupností písmen v šifrovaném textu. Využívá toho, že je velmi pravděpodobné, že tyto posloupnosti byly zašifrovány stejnou částí klíče a tedy i původní význam je shodný. Tyto posloupnosti mohou být různě dlouhé. Čím delší vstupní text, tím delší posloupnosti je možné hledat. U delších posloupností je nižší pravděpodobnost, že se jedná pouze o náhodu, a že je text skutečně zašifrován shodnou částí klíče.[1]

Po nalezení daného počtu takovýchto posloupností, jsou spočítány jejich jednotlivé vzájemné vzdálenosti. Ze vzdáleností je spočítán největší společný dělitel, který je pravděpodná délka klíče. U každé vzdálenosti jsou zaznamenány její četnosti. Pokud je četnost pod určitou mezí, tak vzdálenost není zahrnuta do počítání největšího společného dělitele. Tím lze částečně eliminovat chybné vzdálenosti.

Vzhledem k časovému omezení běhu programu je hledáno maximálně 2000 shodných posloupností. Délka posloupnosti je určována dynamicky v závislosti na délce zašifrovaného textu. Do 5000 znaků jsou hledány trojice, do 25000 znaků čtveřice a cokoliv výše hledá pětice. Vzdálenosti, které nemají aspoň 3 výskyty jsou eliminovány.

3 Friedmanův test

Friedmanův test využívá tzv. index koincidence, který měří rovnoměrnost výskytů jednotlivých písmen v textu. Využívá toho, že v jednotlivých jazycích jsou frekvence různých písmen odlišné. Pokud je známo, že zašifrovaný text je v Angličtině, lze této skutečnosti využít.[2]

Pokud známe pravděpodobnost, že dva náhodně vybrané znaky z textu jsou shodné, tak můžeme délku klíče odhadnout vzorcem 1. Pro Angličtinu je to asi $\kappa_p = 0,067$ a pro čistě náhodný text $\kappa_r = 0,0385$.

$$friedman\_test = \frac{\kappa_p - \kappa_r}{\kappa_o - \kappa_r} \tag{1}$$

Parametr κ_o je spočítán podle rovnice 2. N je délka textu, c je počet znaků abecedy a n_i je počet výskytů daného znaku v zašifrovaném textu.

$$\kappa_o = \frac{\sum_{i=1}^{c} n_i(n_i - 1)}{N(N-1)} \tag{2}$$

[1] `https://en.wikipedia.org/wiki/Kasiski_examination`

[2] `https://en.wikipedia.org/wiki/Vigenere_cipher`

Pokud se výsledek Friedmanova testu výrazně liší od výsledku Kasiského testu, je proveden další výpočet. Hranice pro výskyt vzdálenosti je inkrementována, vzdálenosti co novou hranici nesplňují jsou eliminovány a je znovu spočítán největší společný dělitel (pokud ještě nějaké vzdálenosti zbyly). Tímto lze odhadnutou délku hesla upřesnit, pokud Kasiského test vzal v potaz chybnou vzdálenost.

4 Prolomení klíče

Pokud je známá délka klíče n, je možné text rozdělit na n částí, kde každá část je šifrovaná Caesarovou šifrou. Vzniká n podtextů, každý šifrovaný jedním znakem. Tyto podtexty lze prolomit útokem hrubou silou. Jsou vyzkoušeny každý ze 26 možných posuvů a jazykovou analýzou je zjištěn ten nejpravděpodobnější, tedy ten co nejvíce odpovídá Angličtině. Po aplikaci stejného principu na každý z podtextů, jsme získali pravděpodobný klíč k původnímu textu zašifrovaným Vigenerovou šifrou.

Nejpravděpodobnější posuv je spočítán pomocí metody podobné indexu koincidence. Tentokrát jako pravděpodobnost výskytu každého jednotlivého znaku v daném jazyce.

$$M_g = \sum_{i=0}^{25} \frac{p_i f_{i+g}}{n} \tag{3}$$

Přesný výpočet je zachycen na rovnici 3. Kde n je délka podřetězce, g je jeden z možných posuvů v rámci Caesarovy šifry, $f_i + g$ značí frekvence písmen v podřetězci a p_i frekvence písmen v Angličtině. Pokud g je správny posuv, pak bude hodnota M_g odpovídat zhruba hodnotě $0,065$.[3]

5 Testování

Pro účely testování byl napsán bashový skript `test.sh`, který program testuje na náhodných anglických textech o délce 3500 – 50 000 znaků. Texty jsou šifrovány náhodnými klíči v délce 1 – 200.

Skript kontroluje návratové kódy, výsledky testů, odhadnuté heslo a délku běhu programu, která nesmí přesáhnout 30 sekund. Na konci je vypsán report.

```
$ ./test.sh
```

[3] `https://wis.fit.vutbr.cz/FIT/st/cfs.php.cs?file=\%2Fcourse\%2FKRY-IT\%2Flectures\%2FKRY01_Klas_MNG.pdf`