1. projekt pro předmět KRY 2012
Vigenèrova šifra

Mikulka Jiří
<xmikul39@stud.fit.vutbr.cz>

30. března 2012

1 Rozbor šifry

Vigenèrova šifra[1] patří mezi polyalfabetické substituční šifry pracující s tzv. *Vigenèrovým čtvercem*. Tento čtverec (viz tabulka 1) obsahuje 26 sloupců a 26 řádků (na každém řádku je zapsána celá uspořádaná abeceda a abeceda na následujícím řádku je vždy zrotována o 1 písmeno vlevo). Díky použití více abeced je tato šifra odolná proti frekvenční analýze, protože stejný znak prostého textu může být zašifrován na různá písmena.

	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
A	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
B	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	A
⋮													⋮													
Y	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
Z	Z	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

Tabulka 1: Vigenèrův čtverec.

Prostý text je znak po znaku zašifrován za použití klíčového slova. Nad prostý text je zapsáno klíčové slovo cyklicky se opakující nad celou délkou prostého textu. Zašifrování znaku spočívá v nalezení písmene ležícího na průsečíku sloupce, který je dán znakem prostého textu, a řádku daného příslušícím písmenem klíčového slova.

Šifrování (resp. dešifrování) lze matematicky zapsat následovně (uvažujeme anglickou abecedou se 26 alfabeticky seřazenými písmeny, M_i jako $i-$tý znak prostého textu, K_j jako $[i \ \mathbf{mod} \ length(K)]-$tý znak klíče a C_i jako $i-$tý znak zašifrovaného textu):

- *šifrování*:
$$C_i = (M_i + K_j) \ \mathbf{mod} \ length(alphabet)$$
- *dešifrování*:
$$M_i = (C_i - K_j) \ \mathbf{mod} \ length(alphabet)$$

2 Kryptoanalýza šifry

Jelikož se jedná o polyalfabetickou substituční šifru, není možné využít frekvenční analýzu k její prolomení. K odhalení délky hesla, resp. počtu monoalfabetických substitučních šifer, lze využít Friedmanův a/nebo Kasiského test.

2.1 Friedmanův test

Tento test je také někdy označován jako *kappa test*. Využívá tzv. *index koincidence*, který určuje pravděpodobnost, že 2 náhodně vybraná písmena z textu jsou totožná. Tento index je pro každý přirozený jazyk specifický a závisí na frekvenci jednotlivých písmen v textech v daném jazyce. Dále budeme uvažovat pouze anglické texty, pro které je index koincidence $\kappa_p = \sum_{i=A}^{Z} {p_i}^2 = 0.0654966995$. Pro zcela náhodný text je index koincidence

[1] Vigenèrova šifra na Wikipedia.org – `http://en.wikipedia.org/wiki/Vigenere_cipher`.

$\kappa_r = 26(\frac{1}{26})^2 = \frac{1}{26} = 0.0384615385$. Pomocí těchto (pro konkrétní jazyk zašifrovaného textu) konstat můžeme určit Friedmanovým testem teoretickou délku klíče:

$$length_{Friedman}(key) = \frac{\kappa_p - \kappa_r}{\kappa_o - \kappa_r},$$

kde

$$\kappa_o = \frac{\sum_{i=A}^{Z} f_i(f_i - 1)}{n(n-1)}.$$

2.2 Kasiského test

Tato metoda je starší než Friedmanův test, přesto dokáže poměrně přesně určit délku klíče. Princip této metody spočívá ve vyhledání všech trigramů (a delších podřetězců pro zpřesnění výsledku). Pro každý trigram jsou vypočteny rozdíly vzdáleností mezi všemi výskyty a prvním výskytem. Trigramy mohly s největší pravděpodobností vzniknout při šifrování stejných 3 písmen prostého textu za použití stejných 3 písmen klíče, proto délka klíče bude určena *největším společným dělitelem* vzdáleností všech výskytů od prvního výsktu trigramu.

2.3 Odhalení klíče

Jakmile máme vypočtenu délku klíče pomocí Friedmanova testu a Kasiského testu, můžeme odhalit podobu klíče. Pro každý (zatím neznámý) znak klíče použijeme metodu podobnou indexu koincidence – vytvoříme podřetězec každého $i-$tého znaku zašifrovaného textu a tento podřetězec budeme šifrovat $length(alphabet)$ monoalfabetickými šiframi s posunem $g = 0, 1, \ldots, length(alphabet)$. Pro takto vzniklé podřetězce vypočteme

$$M_g = \sum_{i=A}^{Z} \frac{p_i f_{i+g}}{n},$$

kde n je délka podřetězce. Pokud je hodnota posunu g správná, pak M_g bude přibližně rovno $\kappa_p = 0.0654966995$ (zde je nutné uvažovat určitou toleranci). Posuv g tedy určuje pozici znaku v anglické abecedě a tento znak je hledaným $i-$tým znakem klíče. Takto postupujeme, dokud neodhalíme všechny znaky klíče.

Ověření správnosti délky a podoby klíče je možné provedením dešifrováním a zašifrováním dešifrované zprávy a porovnáním těchto zpráv. Pokud je klíč (délka, podoba) určen správně, texty budou totožné:

$$ciphertext = encipher(decipher(ciphertext, key), key)$$

3 Implementace

Testy zmíněné výše spolu s odhalením klíče jsem implementoval v jazyce `C++` s využitím možností `STL`. Průběh kryptoanalytického útoku provádím podle následujícího scénáře:

1. načtení a upravení zašifrovaného textu (tzn. odstranění všech ne-alfabetických znaků, převod na kapitálky)
2. provedení Friedmanova a Kasiského testu
3. určení skutečné délky klíče na základě předchozích testů, určení jednotlivých znaků klíče
4. (ověření správnosti klíče na zašifrování dešifrovaného vstupního textu za použití získaného klíče)

Implementace zahrnuje tyto soubory:

- main.cpp – hlavní program, který provádí jednotlivé kroky (viz výše)
- friedman.cpp, friedman.h – implementace Friedmanova testu
- kasiski.cpp, kasiski.h – implementace Kasiského testu
- key.cpp, key.h – určení skutečné délky a podoby hesla (s využitím výsledků obou testů)
- vigenere.cpp, vigenere.h – implementace Vigenèrovy šifry (pouze pro testovací účely)
- Makefile – pravidla pro program make
- doc.pdf – tato dokumentace k projektu