Luštění Vigenérovy šifry (Kasiského test)

Následující šifrový text vznikl z anglického textu, ve kterém byly mezery nahrazeny písmenem Z, pomocí Vigenerovy šifry, tj. šifry s periodickým klíčem. Zjistěte délku klíče, klíč a původní otevřený text.

```
HQEOT FNMKP ELTEL UEZSI KTFYG STNME GNDGL PUJCH QWFEX FEEPR
PGKZY EHHQV PSRGN YGYSL EDBRX LWKPE ZMYPU EWLFG LESVR PGJLY
QJGNY GYSLE XVWYP SRGFY KECVF XGFMV ZEGKT LQOZE LUIKS FYLXK
HQWGI LF
```

Řešení. Prohlédneme šifrový text a zjistíme, že šest bigramů se vyskytuje alespoň třikrát:

```
EL na místech 11, 14 a 140,
FY na místech 23, 119 a 146,
GN na místech 31, 64 a 103,
HQ na místech 1, 40, 58 a 151,
LE na místech 70, 91 a 109,
YG na místech 24, 66 a 105.
```

Další zkoumání ukazuje, že bigram GN na místech 64 a 103 je v těchto dvou případech začátkem opakovaného oktogramu GNYGYSLE. Je velmi nepravděpodobné, že by opakované oktogramy vznikly náhodně (pravděp., že by k něčemu takovému došlo náhodně, byla menší než $1 : 2^{10}$), a tak budeme předpokládat, že toto opakování je významné. Vzdálenost mezi těmito oktogramy je 103 - 64 = 39 = 3 * 13, proto můžeme celkem spolehlivě předpokládat, že délka klíče je buď 3 nebo 13 nebo 39. Podíváme se proto ještě na vzdálenosti dalších opakujících se bigramů:

výskyty EL mají vzdálenosti 3 a 126 = 3 * 42,

výskyty HQ mají vzdálenosti 39, 18 a 93, všechny jsou násobkem 3.

To naznačuje, že 3 je zdaleka nejpravděpodobnější délkou klíče. Další krok spočívá v nalezení samotného klíče. Hypotéza: při šifrování byla použita tři různá posunutí. První posunutí bylo použito na písmena na místech 1, 4, 7, 10, atd., druhé posunutí na místa 2, 5, 8, 11, atd., a třetí posunutí na místa 3, 6, 9, 12, atd. Šifrový text si proto napíšeme do třech sloupců podle následujícího vzoru.

1	2	3
H	Q	E
O	T	F
N	M	K
P	E	L
T	E	L
U	E	Z
S	I	K
T	F	Y
G	S	T
N	M	E
...	...	...

První sloupec má 53 písmen, druhý a třetí mají po 52 písmenech. Nyní spočítáme počty výskytů jednotlivých písmen v každém z těchto tří sloupců a zapíšeme je do tabulky:

sloupec	A	B	C	D	E	F	G	H	I	J	K	L	M	…	Z
1	0	1	0	0	0	3	13	4	0	0	1	7	1	…	1
2	0	0	0	0	13	6	0	0	3	2	2	1	2	…	1
3	0	0	2	2	4	1	1	1	0	1	5	5	1	…	3

Pokud by frekvence jednotlivých písmen byly náhodné, mohli bychom očekávat, že počty výskytů jednotlivých písmen v jednotlivých řádcích jsou přibližně 2. Protože ale čísla v jednotlivých řádcích odpovídají frekvencím písmen v přirozeném jazyce, můžeme očekávat počty od 0 do 10, přičemž nejvyšší výskyt s největší pravděpodobností odpovídá tomu písmenu v šifrovém textu, který šifruje písmeno Z nahrazující mezeru v otevřeném textu. To je proto, že každý sloupec je tvořený písmeny otevřeného textu posunutými o stejný počet písmen. V prvním sloupci je nejčastější písmeno G, o kterém tedy předpokládáme, že je šifrovým ekvivalentem Z. První posunutí je tak pravděpodobně o 7 písmen. V tom případě by nejfrekventovanějšímu písmenu E v otevřeném textu mělo odpovídat písmeno L v šifrovém textu, které je skutečně druhým nejčastějším písmenem v prvním sloupci. To dále podporuje naši hypotézu, že první posunutí je o 7 písmen. Ve druhém sloupci je nejčastějším písmenem E, které je tak vhodným kandidátem pro šifrovou obdobu písmena Z, tj. mezery, v otevřeném textu. To znamená, že druhé posunutí je pravděpodobně o 5 míst. Druhá dvě nejčastější šifrová písmena v druhém sloupci jsou F a Q, ze kterých se při posunutí o 5 míst zpět stanou písmena A a L, která jsou skuteční hojně frekventovaná v otevřených textech. Naproti tomu šifrová verze jiného hojně frekventovaného písmene v otevřených textech, písmene E, by po posunutí o pět míst vpřed byla J, které se ve druhém sloupci vyskytuje pouze dvakrát. Naše hypotéza, že druhé posunutí je o 5 míst dopředu, tak není příliš přesvědčivě podepřená. Ve třetím sloupci není žádné šifrové písmeno příliš časté a tak nemáme žádný rozumný odhad, jaké je třetí posunutí. Nejčastějšími písmeny ve třetím sloupci jsou Y, K a L, jedno z nich bude patrní nahrazovat otevřenou mezeru Z, nevíme ale které. Zkusíme si proto napsat počátek šifrového textu a k němu odpovídající otevřená písmena, která dostaneme našimi odhady velikosti prvního a druhého posunutí.

```
HQEOTFNMKPELTELUEZSIKTFYGSTNMEGNDGLPUJCHQWFEXFEEPRPGKZY
AL.HO.GH. I.M  .
```

První slovo vypadá jako ALTHOUGH, a pokud tomu tak je, potom otevřené T je při třetím posunutí nahrazeno šifrovým E, což znamená posunutí o 11 míst dopředu. V tom případě by otevřené mezeře, tj. písmenu Z, odpovídalo šifrové písmeno K, které je skuteční jedním z nejpravděpodobnějších kandidátů pro šifrovou obdobu mezery při třetím posunutí. Můžeme tedy shrnout, že šifrovací klíč je 7,5,11 a dešifrovací klíč pak 19,21,15. Pokud tento dešifrovací klíč použijeme, dostaneme otevřený text

```
ALTHOUGH I AM AN OLD MAN NIGHT IS GENERALLY MY TIME FOR WALKING
IN THE SUMMER I OFTEN LEAVE HOME EARLY IN THE MORNING AND ROAM
ABOUT FIELDS AND LANES ALL DAYS
```

což je počátek jednoho z románů Charlese Dickense.