1 SUBSTITUČNÉ A TRANSPOZIČNÉ ŠIFRY

1.1 ÚVOD

V rámci cvičenia budeme pracovať so správami písanými štandardnou anglickou abecedou, ktorá má $N = 26$ písmen a v ďalšej časti ju budeme označovať zápisom $\mathbb{Z}_N$ resp. $\mathbb{Z}_{26}$. Väčšina šifrovacích algoritmov má matematický charakter, alebo môže byť opísaná matematickými prostriedkami, preto je výhodné pracovať v dvoch rôznych abecedách – **alfabetickej** a **numerickej**, ktorých súvislosť vyjadruje nasledujúca tabuľka

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
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25

Toto vyjadrenie nám umožní priamu aplikáciu **aritmetických operácií** na alfabetické znaky.

V rámci cvičenia budú využívané dve základné šifrovacie techniky označované ako substitúcia a transpozícia. V **substitúcii** nahradzujeme písmena inými písmenami a v **transpozícii** meníme usporiadanie znakov.

1.2 MONOALFABETICKÉ (SUBSTITUČNÉ) ŠIFRY

Medzi najjednoduchšie šifry tejto kategórie patria c**ézarovské šifry**[1]

$$c_i = E(p_i) = C_k(p_i) = p_{(i+k) \bmod N} \tag{1.1}$$

pričom písmeno otvoreného textu p_i je zašifrované písmenom c_i.

Príklad
Akú šifru dostaneme postupným aplikovaním dvoch Cézarových šifier?

Tzv. **Affinná šifra** je určená transformáciou

$$c_i = p_{(a*i+k) \bmod N} = A_{a,k}(p_i) \tag{1.2}$$

pričom a, k sú celé čísla, pričom a je nesúdeliteľné s N (t.j. $GCD(a, N) = 1$). V prípade, že $k = 0$ hovoríme o **multiplikatívnej** šifre.

[1] Pre k=3 dostávame šifru, ktorú použil už Július Cézar.

Príklad
*Aké sú prípustné hodnoty pre parameter a v prípade affinnej šifry s N=26? Pre vybraný kľúč (hodnoty a, k) zašifrujte a **dešifrujte** text ANALYZA, pričom pre spätnú transformáciu využite vzťah*

$$p_i = c_{\left(a^{-1}*i-a^{-1}*k\right)\bmod N} = A_{a^{-1},-a^{-1}*k}\left(c_i\right) \tag{1.3}$$

pričom

$$a * a^{-1} \equiv 1 \bmod N \tag{1.4}$$

Predchádzajúce monoalfabetické šifry pracovali nad abecedou Z_{26}, t.j. využívali prosté zobrazenie $\mathbb{Z}_{26} \leftrightarrow \mathbb{Z}_{26}$. Podobné princípy je možné využiť aj v prípade **polygrafických šifier**, t.j. šifier, v ktorých sa šifrovanie realizuje po skupinách n znakov. Typickým príkladom je **Hillovská šifra**, ktorá využíva regulárne matice **H** typu $n \times n$

$$\mathbf{H} = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n,1} & a_{n,2} & \cdots & a_{n,n} \end{pmatrix}, \text{ kde } a_{i,j} \in \mathbb{Z}_N \text{ pre } 1 \le i, j \le n \tag{1.5}$$

pričom danej n-tici $\left(x_1, x_2, \ldots, x_n\right)$ znakov $x_i \in \mathbb{Z}_N$ je priradená tá n-tica $\left(y_1, y_2, \ldots, y_n\right)$ znakov $y_i \in \mathbb{Z}_N$ zašifrovaného textu, ktorá je určená vzťahom

$$y_i = \sum_{j=1}^{n} a_{i,j} * x_j \bmod N, \quad \text{pre} 1 \le i \le n \tag{1.6}$$

Hilovská šifra vyžaduje, aby matica **H** vo vzťahu (1.5) bola regulárnou maticou nad $\mathbb{Z}_N$. Hovoríme, že matica **H** typu $n \times n$ je **regulárna nad** $\mathbb{Z}_N$**,** ak existuje matica **G** typu $n \times n$ nad $\mathbb{Z}_N$ taká, že pre ich maticový súčin platí

$$\mathbf{H} * \mathbf{G} = \mathbf{I}_N \tag{1.7}$$

pričom $\mathbf{I}_N$ je rovný jednotkovej matici. Maticu **G** nazývame inveznou maticou k matici **H** a označujeme $\mathbf{H}^{-1}$. Je možné ukázať, že regulárna matica je taká matica, ktorej determinant[2] má v $\mathbb{Z}_N$ vzhľadom na násobenie inverzný prvok.

Príklad
Pre Hillovskú šifru s kľúčom

$$\mathbf{H} = \begin{pmatrix} 3 & 2 \\ 9 & 23 \end{pmatrix}$$

[2] Determinant nad $\mathbb{Z}_N$ sa formálne počíta tak, ako nad množinou reálnych čísel, operácie násobenia, sčítania a odčítania sa vykonávajú modulo N. Aby bola definovaná operácia delenia (inverzie), N musí byť prvočíslo.

*zašifrujte a **dešifrujte** text ANALYZA.*

1.3 POLYALFABETICKÉ ŠIFRY

Monoalfabetické šifry sú málo bezpečné preto, že distribúcia frekvencií výskytu zašifrovaného textu je odrazom distribúcie otvoreného textu. Tento nedostatok odstraňujú polyalfabetické šifry.

Systém polyalfabetických šifier
nad abecedou $\mathbb{Z}_N$ tvorí konečná alebo nekonečna postupnosť monoalfabetických transformácií $(T_1, T_2, \ldots, T_n, \ldots)$ nad abecedou $\mathbb{Z}_N$. Tieto postupnosti tvoria priestor kľúčov tohto systému

$$K = \{T_1, T_2, \ldots, T_n, \ldots\} \tag{1.8}$$

Systém vignerovských šifier
Je špeciálny prípad polyalfabetických šifier, ktorý tvoria konečné postupnosti cézarovských kryptografických transformácií. Napr. pri základnej abecede $\mathbb{Z}_N$ môžu byť kľúčovacie postupnosti zapísané ako

$$K = \{k_1, k_2, \ldots, k_n\} \tag{1.9}$$

kde $k_i \in \mathbb{Z}_N$ pre $1 \le i \le n$. Kryptografická transformácia textu $(x_1, x_2, \ldots, x_t)$ odvodená od tohto kľúča je daná vzťahmi

$$y_1 \equiv (x_1 + k_1) \bmod N \tag{1.10}$$

$$y_2 \equiv (x_2 + k_2) \bmod N$$

$$\ldots$$

$$y_n \equiv (x_n + k_n) \bmod N$$

$$y_{n+1} \equiv (x_{n+1} + k_1) \bmod N$$

$$\ldots$$

$$y_t \equiv x_t + k_{t_1} \bmod N$$

pričom $t_1 \equiv t \bmod n$ a $1 \le t \le n$. Číslo n sa nazýva **periódou** vignerovskej šifry, alebo aj **dĺžkou kľúča**.

1.4 ZÁKLADY KRYPTOANALÝZY

Jednou zo základných úloh **kryptoanalytika** je po obdržaní zašifrovaného textu rozhodnúť, aký typ šifry bol použitý. Napr. jedným z dôležitých úvodných rozhodnutí je posúdenie, či bola použitá monoalfabetická, alebo polyalfabetická kryptografická transformácia. V monoalfabetickej šifre sa **zachovávajú rozdiely** medzi frekvenciami jednotlivých písmen, mení sa len ich **poloha**. V textoch šifrovaných polyalfabetickými šiframi je charakteristické **zmenšenie rozdielov** medzi špičkami a údoliami vrcholov v grafe frekvencií znakov v zašifrovanom texte oproti frekvenciám v priamom texte nad tou istou abecedou.

1.4.1 INDEX KOINCIDENCIE

Jedným zo základných prostriedkov na meranie rozdielov medzi relatívnymi frekvenciami jednotlivých znakov zašifrovaného textu je **index koincidencie**, ktorý je mierou rozptylu výskytu jednotlivých znakov.

Ak by boli všetky znaky v $\mathbb{Z}_{26}$ rovnako pravdepodobné, každý znak by sa vyskytoval s pravdepodobnosťou

$$Prav_i = 1/26 \cong 0,0384\,, \quad i = 1,2,\ldots,26 \tag{1.11}$$

a rozptyl distribúcie (výskytu jednotlivých znakov) by bol nulový. Ak znaky nebudú rovnako pravdepodobné, pre rozptyl distribúcie platí:

$$rozptyl = \sum_{i=1}^{26}\left(Prav_i - \frac{1}{26}\right)^2 = \sum_{i=1}^{26}\left(Prav_i^2 - \frac{2}{26}Prav_i + \left(\frac{1}{26}\right)^2\right) = \tag{1.12}$$

$$= \sum_{i=1}^{26} Prav_i^2 - \frac{2}{26}\sum_{i=1}^{26} Prav_i + \sum_{i=1}^{26}\left(\frac{1}{26}\right)^2 = \sum_{i=1}^{26} Prav_i^2 - \frac{2}{26} + 26\left(\frac{1}{26}\right)^2 =$$

$$= \sum_{i=1}^{26} Prav_i^2 - 0,0384$$

Hodnota 0,0384 je nezávislá na aktuálnej distribúcii a preto je ju možné z ďalších úvah vynechať a pre n-prvkovú postupnosť je možné rozptyl aproximovať vzťahom

$$IC = \sum_{i=1}^{26} \frac{Freq_i\left(Freq_i - 1\right)}{n\left(n-1\right)} \tag{1.13}$$

Ak množstvo zašifrovaného textu bude dostatočne veľké a otvorený text bude mať štandardnú distribúciu písmen, potom je možné index koincidencie využiť na predikciu počtu použitých abecied, čo je znázornené v nasledujúcej tabuľke[3]

[3] Údaje platia pre angličtinu.

počet použitých substitúcií	1	2	3	4	5	10	veľa
IC	0,068	0,052	0,047	0,044	0,044	0,041	0,038

1.4.2 KASISKIHO METÓDA

Táto metóda využíva systematičnosť národných jazykov, v ktorých sa vyskytujú často sa opakujúce skupiny písmen alebo dokonca slov. Angličtina napr. veľmi často používa koncovky **-th**, **-ing**, **-ed**, **-ion**, **-tion**, **-ation**, predpony **im-, in-**, **un-**, **re-**, a špecifické štruktúry **-eek**, **-oot**, **-our**. Často sa vyskytujú aj krátke slová ako napr. **of**, **and**, **to**, **with**, **are**, **is**, **that** ...

Kasiskiho metóda využíva fakt, že ak je správa šifrovaná n abecedami s cyklickou rotáciou a ak sa určité slovo alebo skupina písmen vyskytuje v otvorenom texte k-krát, potom toto slovo alebo skupina písmen by mali byť šifrované rovnako približne k/n-krát. Ak je tento pomer väčší ako jedna (čo je pre krátke kľúčové slovo a dostatočne dlhý text splnené) budú sa aj v šifrovanom texte vyskytovať opakujúce sa štruktúry. Tieto je možné využiť na určenie periódy polyalfabetickej šifry.

Podrobnejšie metódu popíšeme pri riešení príkladu na zistenie dĺžky kľúča v prípade zašifrovaného textu.

Príklad
*Overte funkčnosť Kasiskiho metódy pre nasledujúci text šifrovaný pomocou kľúča dickens (na automatizované spracovanie je možné použiť využiť program, ktorý je v súbore **kasiski.zip**, ktorý bol vytvorený študentmi v predchádzajúcich rokoch).*

dicke nsdic kensd icken sdick ensdi ckens dicke nsdic kensd icken sdick
itwas thebe stoft imesi twast hewor stoft imesi twast heage ofwis domit

ensdi ckens dicke nsdic kensd icken sdick ensdi ckens dicke nsdic kensd
wasth eageo ffool ishne ssitw asthe epoch ofbel iefit wasth eepoc hofin

LITERATÚRA

[1] Přibil, J. – Kodl, J.: Ochrana dat v iformatice. Vydavatelství ČVUT, Praha 1996, ISBN.

[2] Grošek, O. – Porubský, Š.: Šifrovanie – algoritmy, metódy, prax. Grada, 1992, ISBN 80-85424-62-2.