Classical Cryptography

Stanislav Palúch

University of Žlina/Department of Mathematical Methods

19. októbra 2017

A group $(G, \otimes)$ is a set G with a binary operation „$\otimes$“ assigning to every two elements $a \in G$, $b \in G$ an element $a \otimes b$ (shortly only ab) such that it holds:

1. $\forall a, b \in G \;\; a \otimes b \in G$
2. $\forall a, b, c \in G \;\; (a \otimes b) \otimes c = a \otimes (b \otimes c)$ – associative law
3. $\exists\, 1 \in G$ such that $\forall a \in G \quad 1 \otimes a = a \otimes 1 = a$
   – existence of a neutral element
4. $\forall a \in G \; \exists a^{-1} \in G \quad a \otimes a^{-1} = a^{-1} \otimes a = 1$ – existence of an inverse element

Abelian Groups

The group G is **commutative** if it holds $\forall a, b \in G \ a \otimes b = b \otimes a$. Commutative groups are also called **Abel groups**. In this case

- An additive notation of group binary operation is used, i. e., we write $a \oplus b$ instead of $a \otimes b$.
- The neutral element is denoted by 0 and called **null element** or **zero element** or zero.
- The inverse element of a will be denoted by $(-a)$ or simply $-a$ instead of a^{-1} and will be called **opposite element**.

Axioms for a commutative group can be rewritten as follows:

1. $\forall a, b \in G \ a \oplus b \in G$
2. $\forall a, b \in G \ a \oplus b = b \oplus a$ – commutative law
3. $\forall a, b, c \in G \ (a \oplus b) \oplus c = a \oplus (b \oplus c)$ – associative law
4. $\exists \ 0 \in G$ such that $\forall a \in G \quad 0 \oplus a = a \oplus 0 = a$ – existence of a neutral element
5. $\forall a \in G \ \exists (-a) \in G \quad a \oplus (-a) = 0$ – existence of an opposite element

Fields

A field $(F, \oplus, \otimes)$ is a set F containing at least two elements 0 and 1 together with two binary operations $\oplus$ and $\otimes$ such that it holds:

1. The set F with binary operation $\oplus$ is a commutative group with null element 0.
2. The set $F - \{0\}$ with binary operation $\otimes$ is a commutative group with neutral element 1.
3. $\forall a, b, c \in G \quad a \otimes (b \oplus c) = a \otimes b \oplus a \otimes c$ – distributive law

Examples

The set $\mathbb{R}$ of all real numbers with ordinary addition $+$ and multipication . is a field.

The set of all rational numbers with ordinary addition $+$ and multipication . is a field.

The set of all complex numbers with addition of complex numbers $+$ and multiplication of complex numbers . is a field.

Fields

Maybe the properties of fields are better visible if we rewrite conditions 1., 2., 3. of the definition of the field into single conditions:

Field is a set F containing at least two elements 0 and 1 together with two binary operations $\oplus$ and $\otimes$ such that it holds:

F1 $\forall a, b \in F \ a \oplus b \in F, \ a \otimes b \in F.$

F2 $\forall a, b, c \in F \ a \oplus (b \oplus c) = (a \oplus b) \oplus c,$
$a \otimes (b \otimes c) = (a \otimes b) \otimes c$ – associative laws

F3 $\forall a, b \in F \ a \oplus b = b \oplus a, \quad a \otimes b = b \otimes a$ – commutative laws

F4 $\forall a, b, c \in F \ a \otimes (b \oplus c) = a \otimes b \oplus a \otimes c$ – distributive law

F5 $\forall a \in F \ a \oplus 0 = a, \quad a \otimes 1 = a$

F6 $\forall a \in F \ \exists (-a) \in F \ a \oplus (-a) = 0$

F7 $\forall a \in F, \ a \neq 0 \ \exists a^{-1} \in F \ a \otimes a^{-1} = 1$

A commutative ring with 1 is a set R containing at least two elements $0 \in R$ and $1 \in R$ together with two operations $\oplus$ and $\otimes$, in which **F1** till **F6** hold.

Examples.
The set $\mathbb{Z}$ of all integers with operations $+$ and $.$ is commutative ring with 1.
However, the structure $(\mathbb{Z}, +, .)$ is not a field since **F7** does not hold.

The set $\mathbb{N} = \{1, 2, 3, \dots\}$ of all natural numbers with common addition and multiplication is not even a ring, since it has no zero element.

Factor ring mod p.

Let us have the set $\mathbb{Z}_p = \{0, 1, 2, \ldots, p-1\}$. Define two binary operations $\oplus$, $\otimes$ on the set $\mathbb{Z}_p$:

$$a \oplus b = (a+b) \mod p \qquad a \otimes b = (ab) \mod p,$$

where $n \mod p$ is the remainder after integer division of the number n by p. Structure $(\mathbb{Z}_p, \oplus, \otimes)$ is called a **factor ring modulo p.**
It can be easily shown that for an arbitrary natural number $p > 1$ the structure $(\mathbb{Z}_p, \oplus, \otimes)$ is a commutative ring with 1, i. e., it fulfills conditions (F1) till (F6).

Example $(\mathbb{Z}_8, \oplus, \otimes)$

$\oplus$	0	1	2	3	4	5	6	7
0	0	1	2	3	4	5	6	7
1	1	2	3	4	5	6	7	0
2	2	3	4	5	6	7	0	1
3	3	4	5	6	7	0	1	2
4	4	5	6	7	0	1	2	3
5	5	6	7	0	1	2	3	4
6	6	7	0	1	2	3	4	5
7	7	0	1	2	3	4	5	6

$\otimes$	0	1	2	3	4	5	6	7
0	0	0	0	0	0	0	0	0
1	0	1	2	3	4	5	6	7
2	0	2	4	6	0	2	4	6
3	0	3	6	1	4	7	2	5
4	0	4	0	4	0	4	0	4
5	0	5	2	7	4	1	6	3
6	0	6	4	2	0	6	4	2
7	0	7	6	5	4	3	2	1

Factor ring mod p.

$\oplus$	0	1	2	3	4	5	6	7
0	0	1	2	3	4	5	6	7
1	1	2	3	4	5	6	7	0
2	2	3	4	5	6	7	0	1
3	3	4	5	6	7	0	1	2
4	4	5	6	7	0	1	2	3
5	5	6	7	0	1	2	3	4
6	6	7	0	1	2	3	4	5
7	7	0	1	2	3	4	5	6

$\otimes$	0	1	2	3	4	5	6	7
0	0	0	0	0	0	0	0	0
1	0	1	2	3	4	5	6	7
2	0	2	4	6	0	2	4	6
3	0	3	6	1	4	7	2	5
4	0	4	0	4	0	4	0	4
5	0	5	2	7	4	1	6	3
6	0	6	4	2	0	6	4	2
7	0	7	6	5	4	3	2	1

Opposite element of 2 is 6, since $2 \oplus 4 = 0$.

Inverse elment of 5 is 5, since $5 \otimes 5 = 1$. Elements $2,\ 4,\ 6$ have no inverse element.

Condition 3. resp. F7 does not hold therefore $(\mathbb{Z}_8, \oplus, \otimes)$ is not a field.

Structure $(\mathbb{Z}_8 - \{0\}, \otimes)$ is not a group since it contains elements without corresponding inverse element.

The following theorem holds:

Theorem A factor ring $(\mathbb{Z}_p, \oplus, \otimes)$ is a field if and only if p is a prime number. The only finite fields are factor rings $\mathbb{Z}_p$ where p is a prime number and Galios fields $GF(p^n)$ having p^n elements. Two finite fields with the same number of elements are isomorphic.

Ceazar cipher

100 – 44 b.c. Ceasar used this table to encipher his messages shifting every character three positions rearwords

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

This way of enciphering is not a cryptography system since it uses no key. Generalization shift by k digits

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

We will use this representation (= coding) of alphabet characters $\{\mathrm{A}, \mathrm{B}, \dots, \mathrm{Z}\}$

$$\mathrm{A} \equiv 0,\ \ \mathrm{B} \equiv 1,\ \ \mathrm{C} \equiv 2,\ \ \mathrm{D} \equiv 3,\ \ \dots,\ \ \mathrm{Y} \equiv 24,\ \ \mathrm{Z} \equiv 25$$

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	15	18	19	20	21	22	23	24	25

Ceasar cipher

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	15	18	19	20	21	22	23	24	25

We can then consider that alphabet is the ring $\mathbb{Z}_{26}$ – the set $\{0, 1, \dots, 25\}$ with operations $\oplus$, $\otimes$ defined as follows

$$\forall a \in \mathbb{Z}_{26}, \quad b \in \mathbb{Z}_{26}$$
$$a \oplus b = (a + b) \mod 26 \qquad a \otimes b = (a.b) \mod 26 \tag{1}$$

Original Ceasar's enciphering algorithm:

enciphering: $y = E(x) = x \oplus D$ deciphering: $x = D(y) = y \ominus D$

Generalised cipher – called **Ceasar cipher** with key $k \in \mathbb{Z}_{26}$:

enciphering: $y = E_k(x) = x \oplus k$ deciphering: $x = D_k(y) = y \ominus k$

Attack against Ceasar Cipher

$\boxed{\text{Plaintext}} \to \boxed{\text{Ciphertext}}$

A cryptographic system is an ordered quadruple $(\mathcal{K}, \mathcal{M}, \mathcal{C}, \mathcal{T})$ where

- $\mathcal{K}$ is a key set
- $\mathcal{M}$ is a set o plintextss
- $\mathcal{C}$ is a set of ciphertexts
- $\mathcal{T}$ is a mapping $\mathcal{T} : \mathcal{K} \times \mathcal{M} \to \mathcal{C}$ which assignes an enciphered message $C \in \mathcal{C}$ to every couple $K \in \mathcal{K}$, $M \in \mathcal{M}$ and such that if $\mathcal{T}(K, M) = \mathcal{T}(K, M')$ then $M = M'$.

The set $\mathcal{M}$ of plaintexts in Ceasar's cryptosystem is the set of all possible sequences of characters – words or sentences of a real language.
Enciphering function enciphers these sequences character by character – Ceasar cipher is an instance of so called **monoalphabetic cipher**
The set of keys is $\mathcal{K} = \{\mathsf{A}, \mathsf{B}, \dots, \mathsf{Z}\}$. characters of both plaintext and key set $\mathcal{K}$ can be considered as elements of $\mathbb{Z}_{26}$.
Key $k = \mathsf{A} \equiv 0$ is unusable since enciphered text is equal to plaintext.
Brute force attack – trying at most 24 keys until understandable deciphered text belonging to $\mathcal{M}$ is obtained. „ciphertext only attack“.

Affine cipher

Affine cipher is monoalphabetic cipher.
Key – a couple of elements $k_1,\ k_2$ of $\mathbb{Z}_{26}$ such that there exists inverse element $k_1^{-1} \in \mathbb{Z}$ to k_1 (i.e. $k_1 \otimes k_1^{-1} = 1 \equiv \mathrm{B}$).

$$
\begin{aligned}
\text{enciphering:} \quad y &= E_{k_1,k_2}(x) = (x \otimes k_1) \oplus k_2 \\
\text{deciphering:} \quad x &= D_{k_1,k_2}(y) = (y \ominus k_2) \otimes k_1^{-1}
\end{aligned}
$$

The set of keys $\mathcal{K}$ – is the set of all ordered couples $(k_1,\ k_2)$ such that there exists $k_1^{-1} \in \mathbb{Z}$.

$k_1 \in \{1, 3, 5, 7, 9, 11, 15, 17, 19, 21, 23, 25\}$ – 12 possibilities

$k_2 \in \{0, 1, 2, \dots, 24, 25\}$ – 26 possibilities

The weak key is $(k_1,\ k_2) = (1, 0)$ since it does not change the plaintext.

Known Plaintext Attack against Affine Cipher

Brute force attack – Ciphertext only attack requires to try at most 311 keys.

We received message:

N	I	N	M	T	Y	M	D	V	J	M	Z	G	N	I	S	H	M	T	E	M	S

This message originated by enciphering of a plaintext by affine cipher using key $E_{k_1,k_2}(x) = (k_1 \otimes x) \oplus k_2$, wher $k_1 = 9$ a $k_2 = 12$.
Enciphering process in in the following table:

D	O	D	A	V	K	A	Z	B	R	A	N	I	D	O	S	L	A	V	C	A	S
3	14	3	0	21	10	0	25	1	17	0	13	8	3	14	18	11	0	21	2	0	18
13	8	13	12	19	24	12	3	21	9	12	25	6	13	8	18	7	12	19	4	12	18
N	I	N	M	T	Y	M	D	V	J	M	Z	G	N	I	S	H	M	T	E	M	S

First row – characters of plaintext,
Second row – their codes – representation in $\mathbb{Z}_{26}$ (A=0, B=1,..., Z=25),
Third row – ciphertest in $\mathbb{Z}_{26}$
Last row – ciphertext in text form.

Known Plaintext Attack against Affine Cipher

Cryptanalyst does not know numbers k_1, k_2.
Suppose he succeeds to guess that the character K was enciphered to Y and the character R was enciphered to J.

$$E_{k_1,k_2}(K) = Y, \quad E_{k_1,k_2}(R) = J,$$

i.e.

$$E_{k_1,k_2}(10) = 24, \quad E_{k_1,k_2}(17) = 9$$

Two last equations can be rewritten as a system of linear equations in $\mathbb{Z}_{26}$

$$k_1 \otimes 10 \oplus k_2 = 24 \tag{2}$$

$$k_1 \otimes 17 \oplus k_2 = 9 \tag{3}$$

Substraction of (2) from(3) gives

$$k_1 \otimes 7 = (-15) \bmod 26 = 11 \tag{4}$$

The inverse of 7 is 15, since $7 \otimes 15 = (7 * 15) \bmod 26 = 95 \bmod 26 = 1$.
Multiplication of equation (4) by number 15 gives:

$$k_1 = (11 * 15) \bmod 26 = (165) \bmod 26 = 9 \tag{5}$$

Known Plaintext Attack against Affine Cipher

We are solving tis system of linear equations in $\mathbb{Z}_{26}$

$$
\begin{aligned}
k_1 \otimes 10 \ \oplus k_2 &= 24 \\
k_1 \otimes 17 \ \oplus k_2 &= 9
\end{aligned}
$$

Till now we have calculated that $k_1 = 9$.
Substitution of 9 for k_1 into (3) gives

$$
\begin{aligned}
(9 \otimes 17) \ \oplus k_2 &= 9 \qquad (6) \\
23 \ \oplus k_2 &= 9 \qquad (7) \\
k_2 &= 9 \ominus 23 = 12 \qquad (8)
\end{aligned}
$$

Provided that we have correctly guessed that $E_{k_1,k_2}(K) = Y$, $E_{k_1,k_2}(R) = J$, the the searched key is the couple $(9, 12)$, what can be acknowlidged by deciphering received ciphertext.
We have solved one system of linear equation instead of trying 311 keys.
The difference between brute force and known plaintext attack is more visible if our alphabet would be the 256 character set of all 8-bit bytes, where brute force attack requires at most 256*128-1 while known plaintext attack means to solve one system of two linear equations in $\mathbb{Z}_{256}$.

General Monoalphabetic Cipher

Ceasar cipher uses for substitution equation $y = E_k(x) = x \oplus k$, affine cipher enciphers as $y = E_{k_1 k_2} = x \otimes k_1 \oplus k_2$.

General monoalphabetic cipher enciphers using formula $E_\pi = \pi(x)$ where π is arbitrary permutation of alphabet $\mathbb{Z}_{26}$.

Every permutation is a bijection, therefore there exists an inverse permutation π^{-1} to every permutation π.

Therefore corresponding deciphering function to enciphering function $y = E_\pi(x) = \pi(x)$ is the function $x = D_\pi(y) = \pi^{-1}(y)$.

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

Plaintext containing a sequence of characters is enciphered character by character using formula

$$y = E_\pi(x) = \pi(x).$$

Deciphering is done also character by character using formula

$$x = D_\pi(y) = \pi^{-1}(y).$$

The key space $\mathcal{K}$ is enormous $|\mathcal{K}| = 26! \approx 10^{27}$. In spite of fact that it contains large part of weak keys a brute attack against it is not possible.

Sources of Information

Cryptanalysis of general mohoalphabetical cipher makes use the fact, that the set $\mathcal{M}$ of plaintexts is the set of outcomes of certain source of information.

A source of information is defined by its alphabet X and by a collection of probabilities $P(x_1, x_2, \dots, x_n)$ for $n = 1, 2, \dots$ and all $x_i \in X$.

The number $P(x_1, x_2, \dots, x_n)$ expresses the probability of the event that the source from its start up generates the character x_1 in time moment 1, the character x_2 in time moment 2 etc., and the character x_n in time moment n. In other words, $P(x_1, x_2, \dots, x_n)$ is the probability of transmitting the word $x_1, x_2, \dots, x_n$ in n time moments starting with the moment of source start up. Number $P(x_1, x_2, \dots, x_n)$ have to fulfill followint conditions:

$$P() = 1 \tag{9}$$

$$\sum_{x_1}\sum_{x_2}\cdots\sum_{x_n} P(x_1, x_2, \dots, x_n) = 1 \tag{10}$$

$$P(x_1, x_2, \dots, x_n) = \sum_{y_1}\sum_{y_2}\cdots\sum_{y_m} P(x_1, x_2, \dots, x_n, y_1, y_2, \dots, y_m) \tag{11}$$

Sources of Information

Probability $P_n(x_1, x_2, \dots x_m)$ of transmitting the word $(x_1, x_2, \dots x_m)$ from time moment n – more exactly in time moments $n, n+1, \dots, n+m-1$ can be calculated as follows:

$$P_n(x_1, x_2, \dots x_m) = \sum_{y_1} \sum_{y_2} \cdots \sum_{y_{n-1}} P(y_1, y_2, \dots, y_{n-1}, x_1, x_2, \dots x_m) \quad (12)$$

Stationary source – $P_n(x_1, x_2, \dots x_m)$ does not depend on n

Independent source – transmitting arbitrary two words in two nonoverlapping two time intervaL are two independent events.

Cryptanalysis of general mohoalphabetics cipher makes use of mainly three probabilities $P(x_1)$, $P(x_1, x_2)$, $P(x_1, x_2, x_3)$ – probabilities of single characters, probabilities of digrams and probabilities of trigrams.

Entropy of a Source of Information

One character x_i of source alphabet with probability $P(x_i)$ carries with it information determined by SHANNON-HARTLEY formula

$$I(x_i) = -\log P(x_i) \tag{13}$$

Mean value of information per one character is

$$H_1 = \sum_{x_1} -P(x_1)\log P(x_1) \tag{14}$$

Mean value of information per ordered couple of charactesr is

$$H_2 = \sum_{x_1}\sum_{x_2} -P(x_1, x_2)\log P(x_1, x_2) \tag{15}$$

Mean value of information per one sequence containing n characters is

$$H_n = \sum_{x_1}\sum_{x_2}\cdots\sum_{x_n} -P(x_1, x_2, \dots, x_n)\log P(x_1, x_2, \dots, x_n) \tag{16}$$

Mean information per one character in words of length n is $H = \frac{1}{n}H_n$.
Limit of this value for $n \to \infty$ is an entropy of source.

Entropy of source is defined as $$\mathcal{H} = \lim_{n\to\infty} \frac{1}{n}H_n \tag{17}$$

Our assessment: $\mathcal{H}(\text{slov. lng}) = 1,57[\text{bit/char}]$, $\kappa = 0,0553$.

Znak

Písmeno	Pravdepodobnosť		Písmeno	Pravdepodobnosť	
	slovenčina	čeština		slovenčina	čeština
A	0,07340	0,054	Ň	0,00139	0,015
Á	0,01545	0,021	O	0,08308	0,068
Ä	0,00060	—	Ó	0,00075	0,000
B	0,01124	0,014	Ô	0,00128	—
C	0,02295	0,019	P	0,02538	0,027
Č	0,01077	0,008	Q	0,00000	0,000
D	0,02919	0,026	R	0,03783	0,029
Ď	0,00141	0,005	Ŕ	0,00006	—
E	0,06927	0,073	Ř	—	0,009
É	0,00669	0,010	S	0,04051	0,040
Ě	—	0,007	Š	0,00918	0,008
F	0,00266	0,002	T	0,04294	0,039
G	0,00222	0,002	Ť	0,00771	0,007
H	0,02050	0,020	U	0,02327	0,030
I	0,05594	0,034	Ú,Ů	0,00875	0,005
Í	0,00996	0,025	V	0,04057	0,039
J	0,01920	0,022	W	0,00011	0,000
K	0,03172	0,033	X	0,00047	0,001
L	0,02976	0,034	Y	0,01341	0,016
Ĺ	0,00006	—	Ý	0,00981	0,008
Ľ	0,00307	—	Z	0,01811	0,019
M	0,02539	0,029	Ž	0,00817	0,009
N	0,05185	0,040	␣	0,13489	0,163

Tabuľka 3.2.1. Relatívna frekvencia výskytu znakov pre zjednodušenú slovenskú a českú abecedu s medzerou

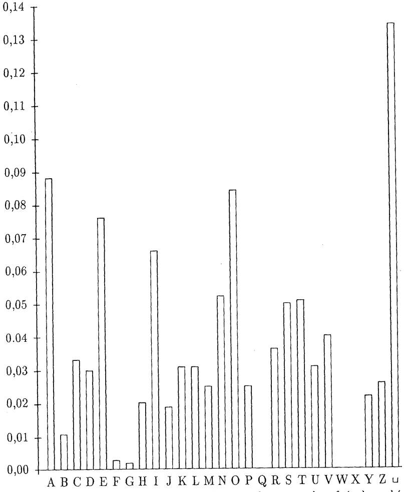
Relatíve Frequency of Characters of Slovak Alphabet with Space

Písmeno	Pravdepodobnosť slovenčina	Pravdepodobnosť čeština	Písmeno	Pravdepodobnosť slovenčina	Pravdepodobnosť čeština
A	0,08945	0,065	O	0,08511	0,067
B	0,01124	0,012	P	0,02538	0,016
C	0,03372	0,024	Q	0,00000	0,001
D	0,01124	0,031	R	0,03789	0,052
E	0,07596	0,107	S	0,04969	0,050
F	0,00266	0,023	T	0,03265	0,086
G	0,00222	0,013	U	0,03202	0,021
H	0,02050	0,043	V	0,04057	0,008
I	0,06590	0,056	W	0,00011	0,016
J	0,01920	0,001	X	0,00047	0,001
K	0,03172	0,003	Y	0,02322	0,016
L	0,03189	0,028	Z	0,02628	0,001
M	0,02539	0,020	␣	0,13489	0,182
N	0,05324	0,058			

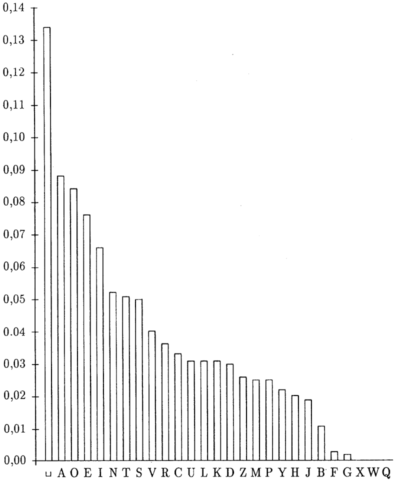
Tabuľka 3.2.2. Relatívna frekvencia výskytu znakov pre telegrafnú slovenskú a anglickú abecedu s medzerou

Zdroj nasledujúcich tabuliek a grafov: Grošek, Porubský : Šifrovanie. Grada 1992, ISBN 80-85424-62-2

Graph – Frequencies of Characters of Slovak Alphabet



Obrázok 3.2.1. Histogram frekvencie výskytu znakov pre telegrafnú slovenskú abecedu



Obrázok 3.2.2. Histogram usporiadaných frekvencií výskytov znakov pre telegrafnú slovenskú abecedu

Frequencies of Characters of Slovak Alphabet

Písmeno	Pravdepodobnosť		Písmeno	Pravdepodobnosť	
	slovenčina	angličtina		slovenčina	angličtina
A	0,11160	0,0856	N	0,05949	0,0707
B	0,01778	0,0139	O	0,09540	0,0797
C	0,02463	0,0279	P	0,03007	0,0199
D	0,03760	0,0378	Q	0,00000	0,0012
E	0,09316	0,1304	R	0,04706	0,0977
F	0,00165	0,0289	S	0,06121	0,0607
G	0,00175	0,0199	T	0,05722	0,1045
H	0,02482	0,0526	U	0,03308	0,0249
I	0,05745	0,0627	V	0,04604	0,0092
J	0,02158	0,0019	W	0,00001	0,0149
K	0,03961	0,0042	X	0,00028	0,0017
L	0,04375	0,0339	Y	0,02674	0,0199
M	0,03578	0,0249	Z	0,03064	0,0008

Tabuľka 3.2.3. Relatívna frekvencia výskytu znakov pre zjednodušenú slovenskú a anglickú abecedu bez medzery

Počty výskytov dvojíc písmen

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
A	0	50	245	238	0	3	16	77	4	222	221	439	160	298
B	56	0	5	6	62	0	0	0	50	13	3	38	5	20
C	99	1	0	0	170	0	0	527	428	0	159	28	1	134
D	160	12	21	2	237	0	0	4	160	0	25	22	18	174
E	16	95	139	408	0	12	14	128	1	317	102	194	132	400
F	9	0	0	0	26	0	0	0	77	0	0	3	0	1
G	26	0	0	0	19	0	0	0	20	0	0	1	2	4
H	81	0	6	0	27	0	0	0	19	2	3	69	3	33
I	408	16	345	38	472	8	2	41	20	19	95	153	101	191
J	63	4	3	7	260	0	0	4	46	0	2	4	18	11
K	181	0	4	13	204	0	0	0	4	0	0	73	5	52
L	340	11	1	4	268	0	1	1	314	0	31	0	7	87
M	174	3	1	0	220	1	0	0	198	0	3	17	0	43
N	613	0	30	7	598	6	6	0	577	0	26	0	1	29
O	2	192	265	329	3	36	32	91	2	116	143	242	338	110
P	68	0	5	0	90	0	0	0	39	0	3	72	0	18
Q	0	0	0	0	0	0	0	0	0	0	0	0	0	0
R	441	4	11	15	413	1	14	7	356	0	5	0	15	50
S	286	0	21	0	154	4	0	15	283	0	240	101	33	41
T	391	0	6	5	251	0	1	2	374	0	60	21	12	125
U	11	18	147	99	0	0	6	27	1	118	38	51	25	25
V	380	11	16	11	351	0	0	10	144	0	15	41	1	103
W	4	0	0	0	1	0	0	0	0	0	0	0	0	0
X	0	0	0	0	3	0	0	0	14	0	0	0	0	0
Y	0	20	242	2	0	0	0	19	0	2	43	8	109	17
Z	284	16	0	75	149	0	0	17	173	7	31	20	67	148
␣	650	143	275	364	50	70	26	117	190	202	433	94	293	710

Tabuľka 3.2.4. Relatívna frekvencia výskytu dvojíc znakov pre telegrafnú slovenskú abecedu (časť 1)

	O	P	Q	R	S	T	U	V	W	X	Y	Z	␣
A	4	42	0	152	229	408	22	258	3	5	0	174	1473
B	147	0	0	29	18	1	44	0	0	0	92	2	5
C	111	0	0	4	15	16	36	0	0	0	13	0	46
D	288	28	0	52	47	1	79	28	0	0	85	60	120
E	38	41	0	174	178	200	12	92	0	13	0	80	1242
F	14	0	0	5	0	0	3	0	0	0	1	1	1
G	23	0	0	14	0	0	5	0	0	0	3	0	1
H	297	1	0	30	0	14	41	3	0	0	52	0	406
I	43	31	0	18	174	273	38	125	0	0	0	109	774
J	31	4	0	4	52	9	155	7	0	0	0	0	334
K	380	0	0	72	8	182	131	20	0	0	194	0	159
L	306	0	0	0	60	8	99	4	0	0	47	1	101
M	156	15	0	6	0	6	135	0	0	0	29	0	339
N	385	0	0	1	53	66	105	2	0	0	234	6	79
O	3	54	0	318	350	155	157	577	0	0	0	253	745
P	467	0	0	534	13	3	16	0	0	0	4	0	12
Q	0	0	0	0	0	0	0	0	0	0	0	0	0
R	391	6	0	0	34	16	86	24	0	5	66	11	38
S	151	153	0	7	10	804	110	57	0	0	27	0	138
T	528	0	0	230	16	2	122	96	2	0	88	1	353
U	0	60	0	43	134	106	0	36	0	0	0	66	686
V	277	7	0	17	93	2	24	0	0	0	291	63	294
W	0	0	0	0	0	0	0	0	0	0	0	0	1
X	1	3	0	0	0	0	0	2	0	0	0	0	2
Y	0	16	0	19	85	29	16	34	0	0	0	21	549
Z	115	19	0	32	17	17	28	63	0	0	5	0	110
␣	357	864	0	248	1049	368	234	723	1	2	0	545	0

Tabuľka 3.2.4. Relatívna frekvencia výskytu dvojíc znakov pre telegrafnú slovenskú abecedu (časť 2)

Number of Trigrams

␣PR	455	OVA	166	ICK	131
␣NA	391	STA	166	A␣N	127
CH␣	377	␣JE	166	JE␣	127
␣A␣	362	HO␣	162	NOS	125
␣PO	302	␣ST	162	ENI	124
OST	251	A␣P	160	O␣S	122
EJ␣	248	PRI	157	A␣Z	118
YCH	233	E␣S	156	CIA	115
NE␣	231	TOR	155	OVE	115
NA␣	215	TI␣	150	E␣V	114
IE␣	210	ALI	149	LA␣	114
␣SA	210	␣DO	147	␣VE	114
␣ZA	197	␣V␣	143	EHO	113
A␣S	194	OU␣	142	␣SP	113
SA␣	186	TO␣	141	STR	112
␣VY	186	NIE	140	E␣N	111
PRE	180	␣RO	139	LI␣	110
OM␣	178	VED	137	NY␣	109
STI	176	E␣P	134	E␣A	108
IA␣	172	KTO	133	JU␣	108
␣NE	167	A␣V	132	␣KT	107

Tabuľka 4.3.1. Najčastejšie trojice v abecede s medzerou

YCH	270	IST	113	VAT	85
OST	236	ACI	111	TAT	84
OVA	197	AST	107	ENE	83
STI	181	NAS	107	EPR	82
PRE	180	EJS	105	NIC	82
STA	173	NOV	105	EDN	79
TOR	159	ICH	104	CKE	78
PRI	157	ALE	99	ENA	78
ALI	156	EST	98	ITA	78
ANI	148	SPO	98	NIA	78
NIE	141	NEJ	97	POD	78
ENI	140	LAD	95	RAV	78
VED	140	NYC	94	RED	78
KTO	138	CIT	92	AKO	77
ICK	131	IAL	91	LOV	77
NOS	128	INA	91	SKO	77
PRA	127	APR	90	TIC	77
OVE	126	OCI	90	AJU	76
EHO	122	EDO	87	STO	75
STR	118	VAN	87	VOJ	75
CIA	117	ANA	85	CHO	73

Tabuľka 4.3.2. Najčastejšie trojice v abecede bez medzery

Most frequent characters of Slovak alphabet are space and

A, O, E, I, N, T, S

Procedure of cryptanalysis of general monoalphbetical cipher (Grošek, Porubský):

- If encryption permutation enciphers space to space (or if we can guess which characters of ciphertext are encrypted spaces) then it is necessary to analyze shorter words which offer less space for combinations.
- It is convenient to search for characteristic combinations of characters (triplets, quadruplets). Such combinations often apper on biginnins or ends of words.
- To guess using „side information“, which words could appear in text.
- To assess which characters are vowels and which ones are consonants.

Cryptanalysis of General Monoalphabetic Cipher

Several hints how to guess vowels:

- vowel are often fenced by consonants
- consonants are often fenced by vowels
- characters with small number of different neighbours are often consonants and those neighbours are vowels
- If a couple XY occurs often also in reverse order YX one of them is probably a vowel
- almost in every normal word occurs a vowel.

Cryptanalysis of General Monoalphabetic Cipher

- p_{ij} probability of bigram $a_i a_j$ in languae
- r_{pq} relative frequency of bigram $a_p a_q$ in ciphertext
- $x_{ip} = \begin{cases} 1 & \text{if } a_i \text{ was enciphered to } a_p \\ 0 & \text{otherwise} \end{cases}$

Minimalize

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{p=1}^{n} \sum_{=1}^{n} x_{ip} x_{jq} (p_{ij} - r_{pq})^2$$

subject to

$$\begin{aligned} \sum_{i=1}^{n} x_{ip} &= 1 \quad \text{pre } p = 1, 2, \ldots, n \\ \sum_{p=1}^{n} x_{ip} &= 1 \quad \text{pre } i = 1, 2, \ldots, n \\ x_{ip} &\in \{0, 1\} \end{aligned}$$

Polyalphabetic Ciphers.

A great disadvantage of monoaphabetical cipher is, that relative count of enciphered characters depends on probabilities of corresponding inverse images in used languae.

New idea originated – to continue to encipher character by character but to encipher every character of plaintext with another key.

Polyalphabetic cipher divides plaintext

$$x_1, x_2, x_3, \dots$$

into substring of the length n

$$x_1, x_2, x_3, \dots = \underbrace{x_1, x_2, \dots, x_n}_{\text{1.th substring}}, \underbrace{x_{n+1}, x_{n+2}, \dots, x_{2n}}_{\text{2.-nd substring}}, \underbrace{, x_{2n+1}, x_{2n+2}, \dots, x_{3n}}_{\text{3.-d substring}}, \dots$$

Ciphertext $y_1 y_1 \dots y_n$ is obtained from plaintext $x_1 x_1 \dots x_n$ as follows:

$$\begin{aligned} y_1 &= E_{K_1}(x_1) \\ y_2 &= E_{K_2}(x_2) \end{aligned}$$

Vigenère Cipher

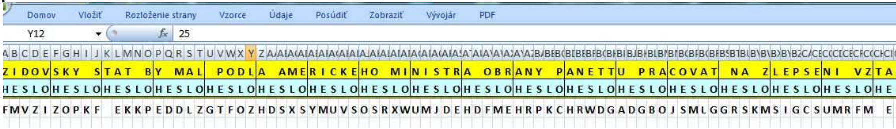
The simplest way is to choose a secret key – e.g. „HESLO“ and then to calculate:

$$
\begin{aligned}
y_1 &= x_1 \oplus H \\
y_2 &= x_2 \oplus E \\
y_3 &= x_3 \oplus S \\
y_4 &= x_4 \oplus L \\
y_5 &= x_1 \oplus O \\
y_6 &= x_6 \oplus H \\
y_7 &= x_7 \oplus E \\
y_8 &= x_8 \oplus S \\
y_9 &= x_9 \oplus L \\
&\ldots
\end{aligned}
$$

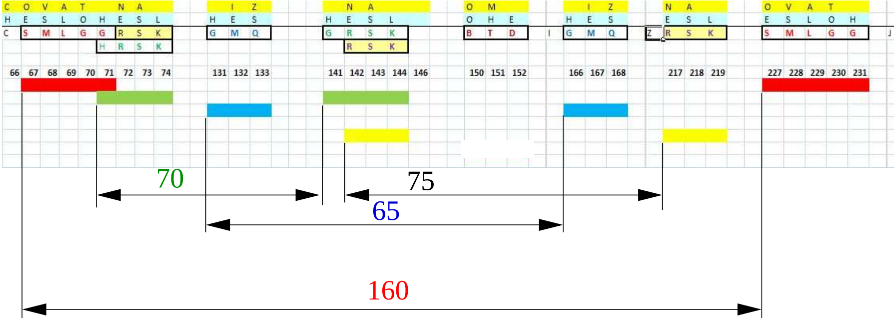
This cipher is called **Vigenère Cipher** although its real inventor was Giovan Battista Bellaso who had invented the cipher earlier (around 1467). Vigenère developed similar (stronger ?) autokey cipher (pubished in 1586).
Vigenère Cipher cipher was consider to be unbreakable for the long time.

Kasiski Key Length Test

This method was first published by Friedrich Kasiski in 1863.



This method searches for appearances of the same substrings in plaintext. If two occurences of the same substring are ciphertexts of the same substrings of plaintext then the distance of these occurences has to be an integer multiply of key lenght.



Kasiski Key Length Test (2)

Prvvý výskyt	Druhý výskyt	Offset	Trojica
67	227	160	S M L
68	228	160	M L G
69	229	160	L G G
71	141	70	G R S
72	142	70	R S K
72	217	145	R S K
131	166	35	G M Q
142	217	75	R S K
192	244	52	W B L

The key length is probably the greatest common divisor of distancess of the same appearances.

Index of Coincidence

Our problem:

We are searching a way how to numerically express inequalities of probabilities of characters.

If all characters of an alphabet $A = \{a_1, a_2, \ldots, a_q\}$ with q elements have the same probablity, then $p(a_i) = \dfrac{1}{q}$.

How to characterise the measure of chaos in probabilities?

$$\sum_{i=1}^{q}(p(a_i) - \frac{1}{q})^2$$

$$\sum_{i=1}^{q}(p(a_i) - \frac{1}{q})^2 = \sum_{i=1}^{q} p(a_i)^2 - \underbrace{2 \cdot \sum_{i=1}^{q} p(a_i)\frac{1}{q}}_{=2\frac{1}{q}} + \underbrace{\sum_{i=1}^{q}(\frac{1}{q})^2}_{=\frac{1}{q}} = \sum_{i=1}^{q} p(a_i)^2 - \frac{1}{q}$$

For $q = 26$

$$\sum_{i=1}^{26} p(a_i)^2 - 0,03846$$

Index of Coincidence (2)

Definition:
The number $\sum_{i=1}^{q} p(a_i)^2$ is called **index of coincidence**.

The greater is the index coincidence than $\frac{1}{q}$, the more the probability distribution differs from uniform distribution.

Index of coincidence of Slovak capital alphabet without space is approsimately equal to $0,06027$, while $\frac{1}{q} = 0,03846$.
Index of coincidence for Slovak alphabet with with diacritic, numeral characters, and punctuation marks was estimated to $0,0553$.

Index of Coincidence (3)

Another meaning of index of coincidence:

Let us compute probability of the eventthat two characters chosen from a source at random will be the same
Probability of the event that two random characters both will be equal to a_i is $p^2(a_i)$.
The event that two random characters will be equal is union of following disjoint events:

- both characters will be equal to a_1 – probability $p(a_1)^2$
- both characters will be equal to a_2 – probability $p(a_2)^2$
- $\cdots\cdots\cdots$
- both characters will be equal to a_q – probability $p(a_q)^2$

The probability of th event that two random characters will be equal is the sum of just listed events $\sum_{i=1}^{q} p(a_i)^2$.

Assessment of Index of Coincidence

Let us have a text (no matter if plaintext of ciphertext) containing n characters – n_1 characters a_1, n_2 characters a_2, e.t.c. till n_q characters a_q.
The number of non ordered couples with both charaters equal to a_i in this text is $\dfrac{n_i(n_i-1)}{2}$, the number of non ordered couples of arbitrary characters in this text is $\dfrac{n(n-1)}{2}$.
The probablity that both characters will be equal to a_i is

$$p(a_i)^2 \approx \frac{n_i(n_i-1)/2}{n(n-1)/2} = \frac{n_i(n_i-1)}{n(n-1)}$$

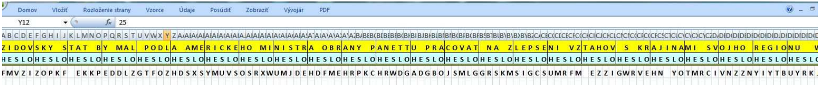
The probability of the event that both characters will be equal we can asses by

$$\kappa = \frac{\sum_{i=1}^{q} n_i(n_i-1)}{n(n-1)} \tag{18}$$

In the case of monoalphabetic cipher, index of coincidence of plaintext is equal to index of coicidence of corresponding ciphertext, since number of every character in plaintext is equal to the number of its images in ciphertext.

If the index of coincidence of ciphertext is close to the one of used language then probably a monoalphabetic cipher was used.

If the index of coincidence is close to $1/q$ then a polyalphabetic or block cipher was used.



Index of coincidence of Slovak language – alphabet with space is $\kappa = 0,062$.

Index o coincidence of our ciphertext is $\kappa = 0,04116$, while $1/27 = 0,03704$.

Therefore we can conclude that a polyaplphabetic cipher was used.

Estimation of key length by method of coincidence

Let us have two plaintexts:

$$\mathbf{r} = r_1 r_2 \dots r_n,$$
$$\mathbf{s} = s_1 s_2 \dots s_n$$

Probablility of the event that $r_i = s_i$ je equal to the index of coincidence κ of used language.

Let those texts are enciphered character by character both with te same key as follows

$$\bar{\mathbf{r}} = E_{K_1}(r_1)E_{K_2}(r_2)\dots E_{K_n}(r_n),$$
$$\bar{\mathbf{s}} = E_{K_1}(s_1)E_{K_2}(s_2)\dots E_{K_n}(s_n).$$

Probablity of the event that $E_i(r_i) = E_i(s_i)$ si he sam as the probablility of the event that $r_i = s_i$, becaus $E_i(r_i) = E_i(s_i)$ hold if and only if $r_i = s_i$. Hence

$$P(T_i(r_i) = T_i(s_i)) = P(r_i = s_i) = \kappa$$

Assume that we have ciphertext $\bar{\mathbf{r}}$ enciphered by a Vigeneére cipher.

Let $\overline{\mathbf{s}_d}$ be a ciphertext $\bar{\mathbf{r}}$ shifted by d characters to the right.

If we observe the number of the same characters on the same positions of ciphertext $\bar{\mathbf{r}}$ and shifted ciphertext $\overline{\mathbf{s}_d}$ then the number of equalities should considerable rise if d eauals to the length of key since compared characters are enciphered by the same key.

Friedman's test

Friedman's test is based on the similar principle as the method of coincidence.
Let us have a ciphertext

$$\mathbf{s} = s_1 s_2 \dots s_n.$$

Arrange characters of **s** into table with k columns.

1	2	...	...	k
s_1	s_2	...	...	s_k
s_{k+1}	s_{k+2}	...	...	s_{2k}
s_{2k+1}	s_{2k+2}	...	...	s_{3k}
s_{3k+1}	s_{3k+2}	...	...	s_{3k}

If k is equal to the length of key then every column is enciphered by a monoalphabetic cipher.
In this case indices of coincidence of all columns should significantly rise.

Determining Characters of Key

Let us have the characters of ciphertext $\mathbf{s} = s_1 s_2 \ldots s_n$ arraged into the following table, where k is the key length:

1	2	...	...	k
s_1	s_2	...	...	s_k
s_{k+1}	s_{k+2}	...	...	s_{2k}
s_{2k+1}	s_{2k+2}	...	...	s_{3k}
s_{3k+1}	s_{3k+2}	...	...	s_{3k}

Let $Z_1,\ Z_2, \ldots, Z_t$ are most frequent characters in the first column. There is lare probability, that the sequence $Z_1,\ Z_2, \ldots, Z_t$ contains at least one characeter which is enciphered one of most frequent characters of used language – for Slovak language

$$\text{A, O, E, I.}$$

Therefore the first character of key could be found probably among characters

$$Z_i - \text{A},\ Z_i - \text{O},\ Z_i - \text{E},\ Z_i - \text{I}$$

where $i = 1, 2 \ldots, t$.

Hill Cipher.

The Hill cipher is a block cipher based on linear algebra. It was invented by Lester S. Hill in 1929. Let us have a plaintext in q-characters alphabet

$A = \{a_0, a_1 \dots, a_{q-1}\}$.
We identify the characters of alphabet A with element of the ring $\mathbb{Z}_q$.
There are operations $\oplus$ a $\otimes$ defined on the alphabet A.

If moreover q is a prime number, then $\mathbb{Z}_q$ is a field and for every $a \in A$ $a \neq 0$ there exists inverse element $a^{-1} \in A$ such that $a \otimes a^{-1} = 1$.

If q is a composite number, then inverse elements exists only for such elements of $\mathbb{Z}_q$ which are coprime with q.
Therefore, if it is possible we prefere q prime number.

Besides finete fiels with prime number of elements there exit also finite field wits $q = p^n$ elements where p is a prime number, namely Galois fields denoted as $GF(p^n)$.

There is no way how to define operations $\oplus$ a $\otimes$ on alphabets whose number of characters is not equal to p or p^n where p is prime such that sgtructure $(A, \oplus, \otimes)$ is a field.

Hill Cipher

Hill cipher is a block cipher enciphering the whole n-character block of a plaintext at once.
The plaintext to encipher is divided into blocks with n charqacters as follows:

$$\underbrace{x_{11}x_{12}\dots x_{1n}}_{\mathbf{x}_1}\underbrace{x_{21}x_{22}\dots x_{2n}}_{\mathbf{x}_2}\dots\dots\dots\underbrace{x_{m1}x_{m2}\dots x_{mn}}_{\mathbf{x}_m} \tag{19}$$

Kye is a square matrix $\mathbf{K}$ of the type $n \times n$ such that there exists for it an inverse matrich $\mathbf{K}^{-1}$.

$$\mathbf{K} = \begin{pmatrix} k_{11} & k_{12} & \dots & k_{1n} \\ k_{21} & k_{22} & \dots & k_{2n} \\ \dots & \dots & \dots & \dots \\ k_{n1} & k_{n2} & \dots & k_{nn} \end{pmatrix} \tag{20}$$

Hill Cipher – Enciphering and deciphering

Enciphering function is as follows:

$$\mathbf{y} = \mathbf{K}\mathbf{x} \quad \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix} = \begin{pmatrix} k_{11} & k_{12} & \dots & k_{1n} \\ k_{21} & k_{22} & \dots & k_{2n} \\ \dots & \dots & \dots & \dots \\ k_{n1} & k_{n2} & \dots & k_{nn} \end{pmatrix} . \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{pmatrix} \tag{21}$$

$$\begin{aligned} y_1 &= k_{11}x_1 + k_{12}x_2 + \cdots + k_{1n}x_n \\ y_2 &= k_{21}x_1 + k_{22}x_2 + \cdots + k_{2n}x_n \\ \dots \\ y_n &= k_{n1}x_1 + k_{n2}x_2 + \cdots + k_{nn}x_n \end{aligned}$$

Deciphering:

$$\mathbf{x} = \mathbf{K}^{-1}\mathbf{y}$$

Deciphering is correctly defined, since

$$\mathbf{K}^{-1}\mathbf{y} = \mathbf{K}^{-1}.(\mathbf{K}.\mathbf{x}) = (\mathbf{K}^{-1}.\mathbf{K}).\mathbf{x} = \mathbf{I}.\mathbf{x} = \mathbf{x} \tag{22}$$

Hill Cipher – Example

Alphabet:

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}$\equiv \mathbb{Z}_{26}$.

Key matrix:

$$\mathbf{K} = \begin{pmatrix} 17 & 4 & 3 & 9 \\ 1 & 13 & 21 & 16 \\ 10 & 12 & 5 & 9 \\ 13 & 6 & 3 & 12 \end{pmatrix}$$

Regularity of matrix **K** can be ascertained by the following way:
Calculate the determinant of K (e.g. in a spreadshed).
For our **K** is $\det \mathbf{K} = -11305$.
$-11305 \mod (26) = 5$ is a number which is coprime with 26 and therefore it has an inverse in $\mathbb{Z}_{26}$ – namely 21.
Therefore **K** is a regular matrix in $\mathbb{Z}_{26}$.

Calculation of an inverse matrix.

Most of spreadsheets can not compute an inverse matrix in $\mathbb{Z}_q$ in one step.

This is a procedure how to calculate an inverse matrix manualy.

All operations are operation in $\mathbb{Z}_{26}$

We start with the matrix $(\mathbf{K}|\mathbf{I})$:

$$\left(\begin{array}{cccc|cccc} 17 & 4 & 3 & 9 & 1 & 0 & 0 & 0 \\ 1 & 13 & 21 & 16 & 0 & 1 & 0 & 0 \\ 10 & 12 & 5 & 9 & 0 & 0 & 1 & 0 \\ 13 & 6 & 3 & 12 & 0 & 0 & 0 & 0 \end{array}\right)$$

We apply Gauss-Jordan elimination matrix $(\mathbf{K}|\mathbf{I})$. This elimination uses elementary row operations in ored to obtain an matrix of the form $(\mathbf{I}|\mathbf{L})$ equivalent with matrix $(\mathbf{K}|\mathbf{I})$. Then $\mathbf{L} = \mathbf{K}^{-1}$.

Hill Cipher – Example

An elementary row operation is any one of the following moves:

1. Swap: Swap two rows of a matrix.
2. Scale: Multiply a row of a matrix by a nonzero constant.
3. Pivot: Add a multiple of one row of a matrix to another row.

$$\left(\begin{array}{cccc|cccc} 17 & 4 & 3 & 9 & 1 & 0 & 0 & 0 \\ 0 & 25 & 4 & 17 & 3 & 1 & 0 & 0 \\ 0 & 2 & 17 & 19 & 4 & 0 & 1 & 0 \\ 0 & 6 & 16 & 25 & 13 & 0 & 0 & 1 \end{array}\right)$$

$$\left(\begin{array}{cccc|cccc} 17 & 4 & 3 & 9 & 1 & 0 & 0 & 0 \\ 0 & 25 & 4 & 17 & 3 & 1 & 0 & 0 \\ 0 & 0 & 25 & 1 & 10 & 2 & 1 & 0 \\ 0 & 0 & 14 & 23 & 5 & 6 & 0 & 1 \end{array}\right)$$

$$\left(\begin{array}{cccc|cccc} 17 & 4 & 3 & 9 & 1 & 0 & 0 & 0 \\ 0 & 25 & 4 & 17 & 3 & 1 & 0 & 0 \\ 0 & 0 & 25 & 1 & 10 & 2 & 1 & 0 \\ 0 & 0 & 0 & 11 & 15 & 8 & 14 & 1 \end{array}\right)$$

Hill Cipher – Example

Now we have calculated an upper triangular matrix wich is equvialent with original matrix $(\mathbf{K}|\mathbf{I})$. All the entries below the main diagonal of our last matrix are zero.
Now it is necessary to achieve that all the entries above the main diagonal are zero.

$$\left(\begin{array}{cccc|cccc}
17 & 4 & 3 & 0 & 10 & 10 & 24 & 11 \\
0 & 25 & 4 & 0 & 20 & 17 & 2 & 15 \\
0 & 0 & 25 & 0 & 11 & 6 & 21 & 7 \\
0 & 0 & 0 & 11 & 15 & 8 & 14 & 1
\end{array}\right)$$

$$\left(\begin{array}{cccc|cccc}
17 & 4 & 0 & 0 & 17 & 2 & 9 & 6 \\
0 & 25 & 0 & 0 & 12 & 15 & 8 & 17 \\
0 & 0 & 25 & 0 & 11 & 6 & 21 & 7 \\
0 & 0 & 0 & 11 & 15 & 8 & 14 & 1
\end{array}\right)$$

$$\left(\begin{array}{cccc|cccc}
17 & 0 & 0 & 0 & 13 & 10 & 15 & 22 \\
0 & 25 & 0 & 0 & 12 & 15 & 8 & 17 \\
0 & 0 & 25 & 0 & 11 & 6 & 21 & 7 \\
0 & 0 & 0 & 11 & 15 & 8 & 14 & 1
\end{array}\right)$$

Hill Cipher – Example

Last matrix from previos page:

$$\begin{pmatrix} 17 & 0 & 0 & 0 & | & 13 & 10 & 15 & 22 \\ 0 & 25 & 0 & 0 & | & 12 & 15 & 8 & 17 \\ 0 & 0 & 25 & 0 & | & 11 & 6 & 21 & 7 \\ 0 & 0 & 0 & 11 & | & 15 & 8 & 14 & 1 \end{pmatrix}$$

It holds in $\mathbb{Z}_{26}$ $17^{-1} = 23$, $25^{-1} = 25$, $11^{-1} = 19$. Multipying of the firs, second, third and forth row of last matrix in sequence by 23, 25, 19 and 25 (all in $\mathbb{Z}_{26}$ gives:

$$\begin{pmatrix} 1 & 0 & 0 & 0 & | & 13 & 22 & 7 & 12 \\ 0 & 1 & 0 & 0 & | & 14 & 11 & 18 & 9 \\ 0 & 0 & 1 & 0 & | & 15 & 20 & 5 & 19 \\ 0 & 0 & 0 & 1 & | & 25 & 22 & 6 & 19 \end{pmatrix}$$

We have:

$$\mathbf{K}^{-1} = \begin{pmatrix} 13 & 22 & 7 & 12 \\ 14 & 11 & 18 & 9 \\ 15 & 20 & 5 & 19 \\ 25 & 22 & 6 & 19 \end{pmatrix}$$

Hill Cipher – Axample

$$\mathbf{K}^{-1} = \begin{pmatrix} 13 & 22 & 7 & 12 \\ 14 & 11 & 18 & 9 \\ 15 & 20 & 5 & 19 \\ 25 & 22 & 6 & 19 \end{pmatrix}$$

$$\mathbf{K}.\mathbf{x} = \begin{pmatrix} 17 & 4 & 3 & 9 \\ 1 & 13 & 21 & 16 \\ 10 & 12 & 5 & 9 \\ 13 & 6 & 3 & 12 \end{pmatrix} \begin{pmatrix} A \equiv 0 \\ B \equiv 1 \\ C \equiv 2 \\ D \equiv 3 \end{pmatrix} = \begin{pmatrix} L \equiv 11 \\ Z \equiv 25 \\ X \equiv 23 \\ W \equiv 22 \end{pmatrix}$$

$$\mathbf{K}^{-1}.\mathbf{y} = \begin{pmatrix} 13 & 22 & 7 & 12 \\ 14 & 11 & 18 & 9 \\ 15 & 20 & 5 & 19 \\ 25 & 22 & 6 & 19 \end{pmatrix} \begin{pmatrix} L \equiv 11 \\ Z \equiv 25 \\ X \equiv 23 \\ W \equiv 22 \end{pmatrix} = \begin{pmatrix} A \equiv 0 \\ B \equiv 1 \\ C \equiv 2 \\ D \equiv 3 \end{pmatrix}$$

Demonstration of changing one character in a block:

$$\mathbf{K}.\mathbf{x}' = \begin{pmatrix} 17 & 4 & 3 & 9 \\ 1 & 13 & 21 & 16 \\ 10 & 12 & 5 & 9 \\ 13 & 6 & 3 & 12 \end{pmatrix} \begin{pmatrix} P \equiv 15 \\ B \equiv 1 \\ C \equiv 2 \\ D \equiv 3 \end{pmatrix} = \begin{pmatrix} G \equiv 6 \\ O \equiv 14 \\ R \equiv 17 \\ J \equiv 9 \end{pmatrix}$$

Known plaintext attack against Hill Cipher

Suppose we know couples of n blocks of plaintexts and corresponding ciphertexts.

$$\mathbf{y}_1 = \mathbf{K}\mathbf{x}_1,\ \mathbf{y}_2 = \mathbf{K}\mathbf{x}_2,\ \dots,\ \mathbf{y}_n = \mathbf{K}\mathbf{x}_n \tag{23}$$

Create squere matrices $\mathbf{X}$, $\mathbf{Y}$, both of the typeu $n \times n$ columne of those matrices will be created by vectors $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$, resp. $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n$, i.e.:

$$\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n), \qquad \mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n).$$

Relations (23) can be stated in matrix form as follows:

$$\mathbf{Y} = \mathbf{K}.\mathbf{X} \tag{24}$$

The equation (24) multipied with matrix $\mathbf{X}^{-1}$ from the right (provided that $\mathbf{X}^{-1}$ does exist) yields:

$$\mathbf{Y}.\mathbf{X}^{-1} = (\mathbf{K}.\mathbf{X}).\mathbf{X}^{-1} = \mathbf{K}.(\mathbf{X}.\mathbf{X}^{-1}) = \mathbf{K}.\mathbf{I} = \mathbf{K}$$

Transposition Cipher

Transposition Cipher is a method of encryption by which the positions held by characterss of plaintext are shifted according to a regular system, so that the ciphertext constitutes a permutation of the plaintext.
That is, the order of the characters is changed (the plaintext is reordered).
Mathematically a permutation is used on the characters' positions to encrypt and an inverse permutation to decrypt.

Transposition cipher is a special case of Hill Cipher. If transposed position of i-th character of plaintext in ciphertext is j then j-th entry in i-th column of key matrix $\mathbf{K}$ is 1, i.e. $k_{ji} = 1$. All other entries of matrix $\mathbf{K}$ are zeros.

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array} \begin{array}{c} \begin{array}{cccccc} 1 & 2 & 3 & 7 & 5 & 6 \end{array} \\ \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} \end{array} \otimes \begin{pmatrix} 10 \\ 20 \\ 30 \\ 40 \\ 50 \\ 60 \end{pmatrix} = \begin{pmatrix} 30 \\ 10 \\ 20 \\ 60 \\ 40 \\ 50 \end{pmatrix}$$