Chapter 1

1.1 Terminology

1.1.1 Cryptology, Cryptography and Cryptanalysis

Cryptology is the science of secure communications. It can be divided to

- **cryptography**, a discipline of encryption methods and
- **cryptanalysis** , a discipline of breaking encrypted messages

Cryptographer and **Cryptanalyst** are experts of these disciplines

Finland has also some significant industries in the field of cryptopgrahy. SSH is a software package that enables secure system administration and file transfers over insecure networks.

Developer of SSH is Finnish Tatu Ylönen. SSH software is used by EU and NASA.

1.1.2 Information security services

Internet Engineering Task Force (IETF) has defined a list of **information security services.**

The implementation of these services needs cryptographic methods.

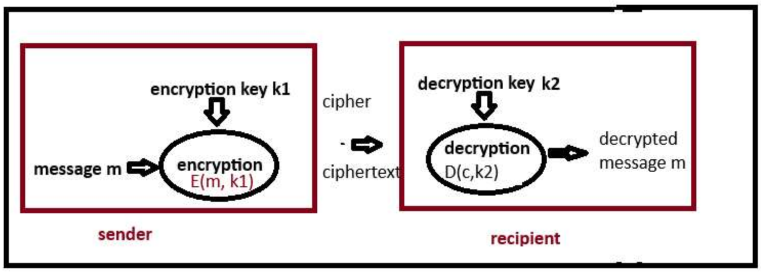
Service	Cryptographic method
Confidentiality	Data encryption
Integrity	Hash functions
Authentication	User authentication of TLS
Non repudiation	Digital Signature
Access control	Password Tokens, Cryptographic authentication
Availability	No cryptoalgorithm for this

Terminology:

Confidentiality: Only those who are entitled to information can access to it
Integrity: Information is unchanged during transfer. Sender is authenticated
Authentication: The sender is provably identified
Non-repudiation: A communicating party cannot deny the contents of his/her messages or deny later having participated in the interaction.

Availability: System and Network are working reliably.

1.1.3 Basic terminology of encryption: Diagram presentation



Message m (called also plaintext) is encrypted using encryption function E(m, k1), where the second argument is encryption key k1. Output of the function is called cipher or ciphertext c .

Decryption is done with decryption function D(c, k2) with arguments ciphertext c and decryption key k2. Output is the original plaintext m

1.1.4 Symmetric and asymmetric encryption

SYMMETRIC ENCRYPTION: Encryption key k1 and decryption key k2 are the same.

- The parties have to agree on the session key before they start communicating (Usually agreeing on symmetric key is done using ”key exchange protocol” (f.e DH, RSA or ECDHE)

ASYMMETRIC ENCRYPTION: Encryption key k1 and decryption key k2 are different.

- Most common form of asymmetric encryption is public key encryption (PKI), in which **every user has two keys** : **Public key** , which is used to encrypt messages to the user, and its pair the **private key**, which only the user knows and uses for decryption of received messages.

1.1.5 Kerckhoff’s principle

"A cryptosystem should be secure even if all its details are public and only the key is secret".

(Auguste Kerckhoff 1835 - 1903 was a Dutch linguistic and cryptographer)

In recent decades DES encryption algorithm and GSM mobile network algorithm were kept at first secret. It did not succeed. Both algorithms leaked to the public.
Todays algorithms are completely transparent and open for everyone who wants to to examine them

1.1.6 Key space

A good, well planned cryptosystem should not have vulnerabilities or backdoors.

The best option for a cryptoanalyst to break a secure cryptosystem is to use **Brute Force attack**: trying to find the key among all possible keys by trial and error or systematic key search.

The most important security factor against brute force is **the number of all possible keys**, which is called the **key space**. Key space is usually measured in **bits**.

Theoretical minimum against brute force is 80 bits, which means that there are $2^{80} = 1.2*10^{24}$ possible keys.
Practical **recommended minimum** for key space, which has some safety margin, is 128 bits

1.1.7 Effective key space

For most cryptosystems it is possible to find methods of breaking the key with less searches than the **theoretical key space** would indicate.

Effective key space means the average number of steps needed to break the key using best known methods of cryptanalysis. Effective key space is smaller than the theoretical key space. In good cryptosystems the difference of theoretical and effective key spaces is small.

Most common symmetric cipher is AES.

AES128	key space 128 bits, effective 126.1
AES256	key space 256 bits, effective 254.4

Some older algorithms have vulnerabilities and difference is bigger

DES	key space 64 bits, effective 54
3DES	key space 168 bits, effective 112

1.1.8 Key space can be applied also to passwors lengths

The 80 bit security minimum $2^{80} = 1.2 \cdot 10^{24}$ can be applied also to security of passwords against Brute Force – attack. Following examples show how to calculate secure password lengths.

Example 1. Password length = 8 english characters. No distinction between uppercase and lower case letters. (=26 characters). Password space = $26^{8} = 2 \cdot 10^{11}$, which is not secure

Example 2. Password length = 11 english characters, distinction between lower case and upper case letter + numbers 0…9: (= 62 characters)
Password space = $62^{11} = 5 \cdot 10^{19}$, which is also not secure

Example 3. Password length = 13 english characters (lower case+upper case) + numbers 0…9 + ten special characters: (=72 characters). Password space $72^{13} = 1.4 \cdot 10^{24}$ provides adequate security.

1.1.9 Mathematical concepts related to cryptography

Hard problem

= **mathematical problem which is hard and time consuming to solve due to its complexity**. The security of cryptoalgorithms is often based on some known ”hard problem” .
RSA algorithm is based on the difficulty of factoring large integers. Diffie-Hellman key exchange is based on Discrete Logarithm Problem (DLP).

One-way function

= function y = f(x), that is fast and easy to calculate if argument x is known, but the i**nverse function** which calculates x for known y **is difficult** or impossible **to calculate without extra knowledge**.

Backdoor

= **Additional knowledge which makes the calculation of inverse function of an one way function possible.**
In RSA encryption calculation of the decryption key from the public key of the user is practically impossible. The public key is a large integer. However **knowledge of the factors of public key is a backdoor to RSA**: calculation of decryption key becomes trivial.

In some known cases a cryptoalgorithm has had an intentional backdoor, which security authorities have used to break encryption of communication.

1.2 Classical ciphers

Contents

1.2.1 Classifications of classical ciphers

Mono- and polyalphabetic ciphers

A) In a **monoalphabetic cipher** message is encrypted character at the time. Every character has always the same image character. Examples of monoalphabetic ciphers are Caesar -cipher and simple substitution.

Frequency analysis is an effective method in breaking monoalphabetic ciphers.

B) **Polyaphabetic cipher** is a system in which the same character can have different image characters at different points of the message. Example of a polyalphabetic cipher is Vigenère cipher. Frequency analysis is less effective against this cipher type.

Division of ciphers based on the block size

In **Monographic ciphers** message is encrypted one character at the time.
In **Polygraphic ciphers** message is divided to blocks of several characters and encryption is done one block at the time.

1.2.2 Caesar cipher

Encryption is based on rotation of alphabet. Key is the amount of rotation.

The Romans probably used a disc like in the figure

The ciphertext of message “AAMU” is “NNZH” in the picture.

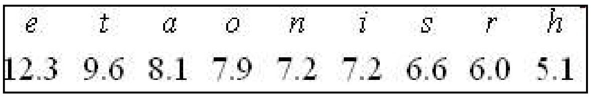
Key space of Caesar cipher = 25 (the number of possible rotations of english alphabet)

Cipher is easily broken with *Brute Force attack* (trying all 25 possible keys)
With *frequency analysis* breaking is even faster.

1.2.3 Frequency analysis of monoalphabetic ciphers

Frequency analysis can be used to break many classical ciphers.
It is most effective in breaking monoalphabetic ciphers.
It is also a basic tool of cryptanalysis of polyalphabetic ciphers in all cases, when the characters of cipher text are not evenly distributed.

Breaking monoalphabetic ciphers (of messages in English) uses the table of relative frequencies of most common characters in english texts



e	*t*	*a*	*o*	*n*	*i*	*s*	*r*	*h*
12.3	9.6	8.1	7.9	7.2	7.2	6.6	6.0	5.1

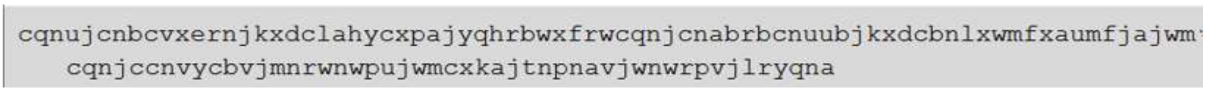
Letter **e** is clearly most common character of english lanquage.

In monoalphabetic cipher the character with highest frequency is very probably the image of **e** (can also be t or a)

For other languages similar tables are found in Wikipedia.
In Finnish language the 6 most common characters in orded are a, i , t , n , e , s

Example of frequency analysis of Caesar

Task is to break following ciphertext.



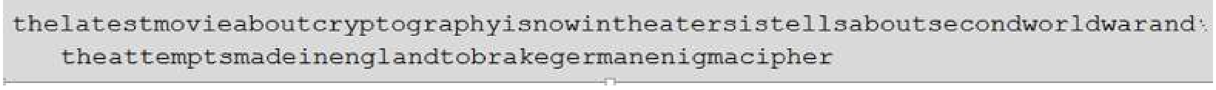
```
cqnujcnbcvxernjkxdclahycxpajyqhrbwxfrwcqnjcnabrbcnuubjkxdcbnlxwmfxaumfjajwm
   cqnjccnvycbvjmnrwnwpujwmcxkajtnpnavjwnwrpvjlryqna
```

Cryptanalysis starts with calculation of frequencies of characters of the ciphertext

n : 15 , c : 14, j : 13 , w : 9 , a : 8 , x : 8 , ….

Character n has biggest frequency. Hypothesis**:** n is the image of e**.** The amount of shift from e to n is 9, which is the most probable candidate for encryption key.

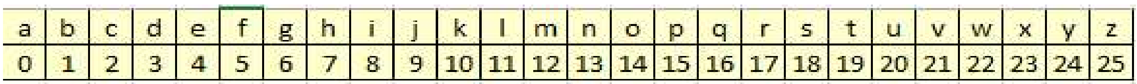
Test the hypothesis by shifting the cipher characters 9 steps in opposite direction. Result is the following meaningful english message:



```
thelatestmovieaboutcryptographyisnowintheatersistellsaboutsecondworldwarand
   theattemptsmadeinenglandtobrakegermanenigmacipher
```

1.2.4 Affine cipher

1. English characters are coded to integers 0 … 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	17	18	19	20	21	22	23	24	25

Encryption formula: **$c = a*m + b \pmod{26}$** , where
m = message character, c = image character in ciphertext, pair (a, b) is the encryption key; a,b є Z_{26}

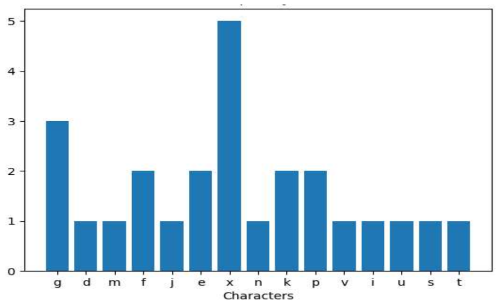
Decryption formula: **$m = a^{-1} c – a^{-1} b \pmod{26}$,** where a^{-1} = inverse of a (mod 26) *)
The inverse of a can be calculated a) manually with ExtendedGCD algorithm as presented in part1 of the course.
With WolframAlpha.com calculator calculation is straightforward: a^-1 mod 26.
Key space = 11*26 = 286. *(b can be any number of Z_{26}, but only φ(26) = 11 numbers of Z_{26} have inverse elements)*

Example of affine encryption

Encrypt message ”kemi” using key (a = 11 , b = 3)
1. Coded to integers message m = (10, 4 , 12 , 8)
2. Encryption formula gives a cipher
c = (11*10 + 3 , 11*4 + 3, 11*12 + 3, 11*8 + 3) mod 26 = (9 , 21 , 5 , 13) = ”jvfn”

Decryption using formula: m = a^{-1} c – a^{-1} b (mod 26)
1. First we calculate the inverse a^{-1} = 11^{-1} mod 26 = 19 *)WolframAlpha
2. Decrypted cipher is
(19*9 – 19*3 , 19*21 – 19*3 , 19*5 – 19*3 , 19*13 – 19*3) mod 26 = (10 , 4 , 12 , 8) = ”kemi”

Example of cryptanalysis of affine encryption



Break affine cipher ”gdmfjgexnxfgexkpvikxxupst” .

Frequency diagram on the right show,

that **x** and **g** have biggest frequencies:

The best guess we can make is that x is the image of e and g is the image of t .

Using codes we get a pair of equations:
a*4 + b ≡ 23 and a*19 + b ≡ 6 (mod 26)

Subtracting (1) from (2) gives 15*a ≡ -17 ≡ 9 => **a ≡** 15^{-1}*9 ≡ 7*9 ≡ **11** (mod 26)
Substituting a = 11 gives **b ≡** 23 - 11*4 **≡ 5** (mod 26)

Now we need to test, if our guess (hypothesis) about images was right.
First we calculate the inverse of a, which is 11^{-1} (mod 26) = 19.

As an example we decrypt the first two letters ”g” (code 6) and ”d” (code 3) of cipher:
(19*6 – 19*5) mod 26 = 19 = ”t” and (19*3 – 19*5) mod 26 = 14 = ”o” .
Continuing in the same way we get the whole message, which is ”todaytheweatherisfreezing”

In solving Moodle exercises instead of manual decryption use Excel template ”classical ciphers”, which is found in Moodle.

1.2.5 Hill cipher (Matrix cipher)

A natural extension of affine cipher is Hill cipher (1929) which encrypts message blocks instead of single letters. Hill cipher is polyalphabetic: a character of message has several image characters. **Hill cipher uses matrix multiplication for encryption**.

Example: Encrypt block ”act” using matrix key ((G,Y,B),(N,Q,K),(U,R,P)) as multiplier

$$\begin{pmatrix} 6 & 24 & 1 \\ 13 & 16 & 10 \\ 20 & 17 & 15 \end{pmatrix}\begin{pmatrix} 0 \\ 2 \\ 19 \end{pmatrix} = \begin{pmatrix} 67 \\ 222 \\ 319 \end{pmatrix} \equiv \begin{pmatrix} 15 \\ 14 \\ 7 \end{pmatrix} \pmod{26} \quad = \text{”poh”}$$

Decryption of ”poh” uses inverse matrix of the key matrix

$$\begin{pmatrix} 6 & 24 & 1 \\ 13 & 16 & 10 \\ 20 & 17 & 15 \end{pmatrix}^{-1} \pmod{26} \equiv \begin{pmatrix} 8 & 5 & 10 \\ 21 & 8 & 21 \\ 21 & 12 & 8 \end{pmatrix}$$

Decryption of ciphertext is done with matrix multiplication

$$\begin{pmatrix} 8 & 5 & 10 \\ 21 & 8 & 21 \\ 21 & 12 & 8 \end{pmatrix}\begin{pmatrix} 15 \\ 14 \\ 7 \end{pmatrix} = \begin{pmatrix} 260 \\ 574 \\ 539 \end{pmatrix} \equiv \begin{pmatrix} 0 \\ 2 \\ 19 \end{pmatrix} \pmod{26} \quad = \text{”act”}$$

Hill cipher fulfills **Shannon’s diffusion principle:**

”Each character of the ciphertext should depend on several characters of the message, obscuring the connections between the two.”

For block size of 4 letters, Hill cipher uses 4x4 matrices, e.t.c

Key space: If the block size is 3 letters, the key space is 26^{9} which is approximately 10^{12}

First versions of German Enigma cipher machine used in WW2 were based of Hill Cipher’s ideas.

1.2.6 Simple substitution cipher

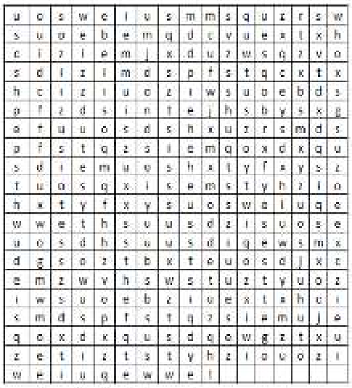
Encryption key is a permutation of alphabet (e.e. a->k, b->z, c→q,…).

Key space = number of permutations = 26! = $4*10^{26}$, which is secure against Brute Force attack.

With **frequency analysis** breaking simple substition cipher is relatively easy and can be compared to solving a difficult crossword puzzle.

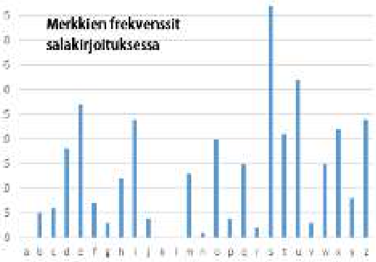
Below on the left we have ciphertext, in the middle its frequency table, on the right part of the message is already found. The method is based on hypothesis that E(e) = s and E(t) = u because e and t are the most common English characters and in cipher s and u have highest frequencies. Also ”the” is the most common trigram of English language (appears in ”the”, ”these”, ”they”,…). So we can fill the missing character of ”t_e” combinations with ”h”.
Netx phase could be finding trigrams corresponding ”and” in the ciphertext. Then we would know already the images of 6 quite common characters : ”t”,”e”,”h”,”a”,”n”,”d”.

Cipher:

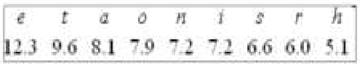


u	o	s	w	e	l	u	s	m	m	s	q	u	z	r	s	w
s	u	o	e	b	e	m	q	d	c	v	u	e	x	l	x	h
c	i	z	l	e	m	j	x	d	u	z	w	s	q	z	v	o
s	d	l	z	l	m	d	s	p	f	s	t	q	c	k	t	x
h	c	i	z	i	u	o	z	i	w	s	u	o	e	b	d	s
p	f	z	d	s	i	n	f	e	j	h	s	b	y	s	x	g
a	f	u	u	o	s	d	s	h	x	u	z	r	s	m	d	s
p	f	s	t	q	z	s	i	e	m	q	o	x	d	k	q	u
s	d	i	e	m	u	o	s	h	x	l	y	f	x	y	s	z
f	u	o	s	q	x	l	s	e	m	s	t	y	h	z	l	o
h	x	t	y	f	x	y	s	u	o	s	w	e	l	u	q	e
w	w	e	t	h	s	v	u	s	d	z	i	s	u	o	s	e
u	o	s	d	h	s	u	u	s	d	i	q	e	w	s	m	x
d	g	s	o	z	t	b	x	t	e	u	o	s	d	j	x	c
e	m	z	w	v	h	s	w	s	t	u	z	t	y	u	o	z
i	w	s	u	o	e	b	z	i	u	e	x	t	a	h	c	i
s	m	d	s	p	f	s	f	q	z	s	i	e	m	u	j	e
q	o	x	d	k	q	u	s	d	q	o	w	g	z	t	x	u
z	e	t	i	z	t	s	t	y	h	z	i	o	u	o	z	i
w	e	i	u	q	e	w	w	e	l							

Frequency analysis

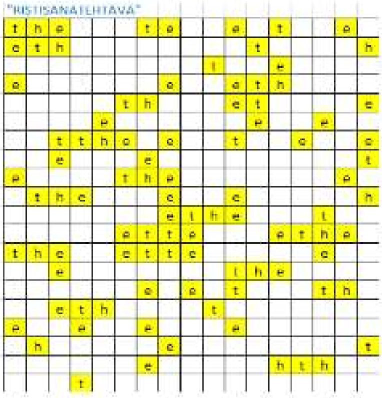


Alla englanninkielen tekstin merkkien suht. frekvenssit



e	*t*	*a*	*o*	*n*	*i*	*s*	*r*	*h*
12.3	9.6	8.1	7.9	7.2	7.2	6.6	6.0	5.1

Initial phase of cryptanalysis



"RISTISANATEHTAVA"

t	h	e				t	e			e		t			e	
e	t	h									t					h
									t			e				
e							e			e	t	h				
					t	h				e	t					e
				e							e			e		
		t	t	h	e		e			t			e			e
		e				e										t
e					t	h	e								e	
	t	h	e				e			e						h
							e	t	h	e				t		
					e	t	t	e				e	t	h	e	
t	h	e			e	t	t	e						e		
		e								t	h	e				
						e		e		t				t	h	
		e	t	h					t							
e			e			e				e						
	h						e									t
						e						h	t	h		
			t													

1.2.7 Vigenere cipher

Blaise de Vigenère 1523 -1596 was a french linguistic and cryptographer

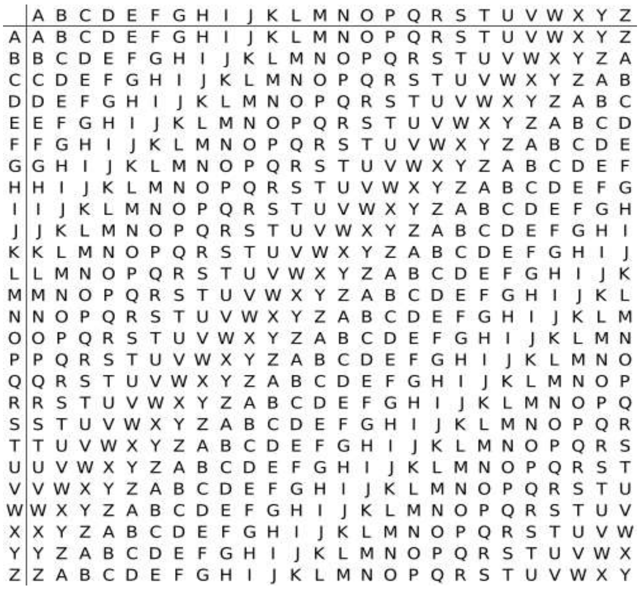
Encryption key is a password or pass phrase, which is extended by copying to the length of the whole message.

Encryption

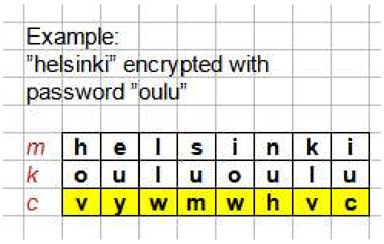
The characters of message and key are added using the addition table below.
Another way is to code alphabet to numbers 0...25 and use formula **$c_i = m_i + k_i$ (mod 26)**.

Decryption

Cipher is decrypted using the addition table in reverse order or subtracting ***$m_i = c_i - k_i$ (mod 26)***



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



Example:
"helsinki" encrypted with
password "oulu"

m	h	e	l	s	i	n	k	i
k	o	u	l	u	o	u	l	u
c	v	y	w	m	w	h	v	c

Key space of Vigenere depends on password length. Key space is 25^{n} ,where n = password length.

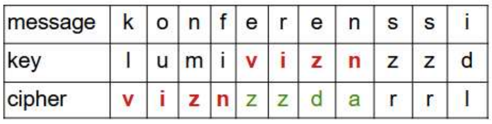
If f.e the password length n = 20, the key space = $9*10^{27}$ (secure against Brute Force).

If the cipher is arranged to columns so that the number of columns bilir the password length, each column is encrypted using the same key character k_i corresponding Caesar cipher.

In 1800's Prussian officer Kasiski found a method of breaking Vigenere cipher. Kasiski's cryptanalysis is based on frequency analysis of 2- 3 character long "substrings" in ciphertext, which helped to find the period of the key.

The vulnerabilty caused by the periodicity of extended key can be avoided by changing the way of extending the key. In **autokey cipher** next block of the extended key is the cipher of previous block.

Encrypt ”konferenssi” with key ”lumi”



message	k	o	n	f	e	r	e	n	s	s	i
key	l	u	m	i	v	i	z	n	z	z	d
cipher	v	i	z	n	z	z	d	a	r	r	l

1.2.8 One Time Pad

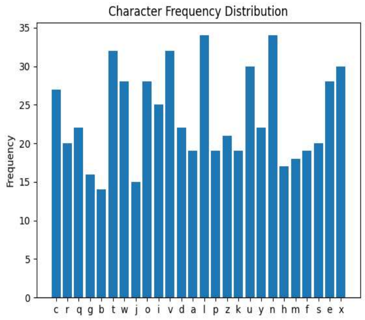
”The Unbreakable Cipher”

If Vigenère encryption is used with **a random, one-time key** of same length as the message**,** encryption is impossible to break. (even in future and unlimited resources).

This is obvious, because for each cipher text c and each possible message m there is a key k which encrypts m to c. It is impossible to distinguish the original message from other possible messages of the same length.

The cipher produced by One Time Pad has ”**maximal entropy”**. (**Entropy** is in physics and also in information theory a measure of disorder and randomness).

Frequency table of characters of ciphertext created with One Time Pad shows that frequencies are evenly distributed – except random fluctuation.



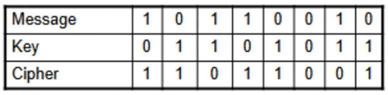
Also in modern cryptoalgorithms, f.e in hash functions and block ciphers, a randomness of output is one of the basic criterions for goodness of the algorithm.

Vernam cipher (1919) = a binary version of One Time Pad.

Message m and a random key k of same lentgh are binary sequences. In encryption m and k are added using **XOR addition** (rules: 0 + 0 = 1 + 1 = 0, 0 + 1 = 1 + 0 = 1) . Decryption is done adding the same key k to the cipher. *During the cold war Moscow and Washington had a direct telex connection which used Vernam cipher.*

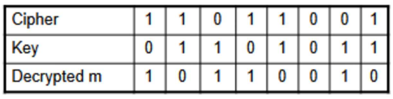
The practical problem of Vernam is that both communicating parties must have identical ”codebooks” containing very long one-time passwords. Using the same key twice will break the security.

Encryption:



Message	1	0	1	1	0	0	1	0
Key	0	1	1	0	1	0	1	1
Cipher	1	1	0	1	1	0	0	1

Decryption:



Cipher	1	1	0	1	1	0	0	1
Key	0	1	1	0	1	0	1	1
Decrypted m	1	0	1	1	0	0	1	0

Time of classical ciphers ended at the same time as WW2. Computers were invented in late 1940’s. At the beginning of 1960’s computer networks started to build up. Modern cryptography was needed to secure the connections.