Chapter 1

Concepts and principles

Classical ciphers

Cryptology, Cryptography, Cryptanalysis

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

- cryptography, which is a discipline of encryption methods
- 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 cryptography. 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.

IETF's list of Information Security Services

Service	Cryptographic method
Confidentiality	Data encryption
Integrity	Hash functions
Authentication	User authentication of TLS
Non repudiation	Digital Signature
Access control	Cryptographic authentication
Availability	

Internet Engineering Task Force (IETF) has defined a list of **information security services.**
The implementation of these services needs cryptographic methods.

The names of services are **confidentiality, integrity, authentication, non-repudiation, access control and availability.**

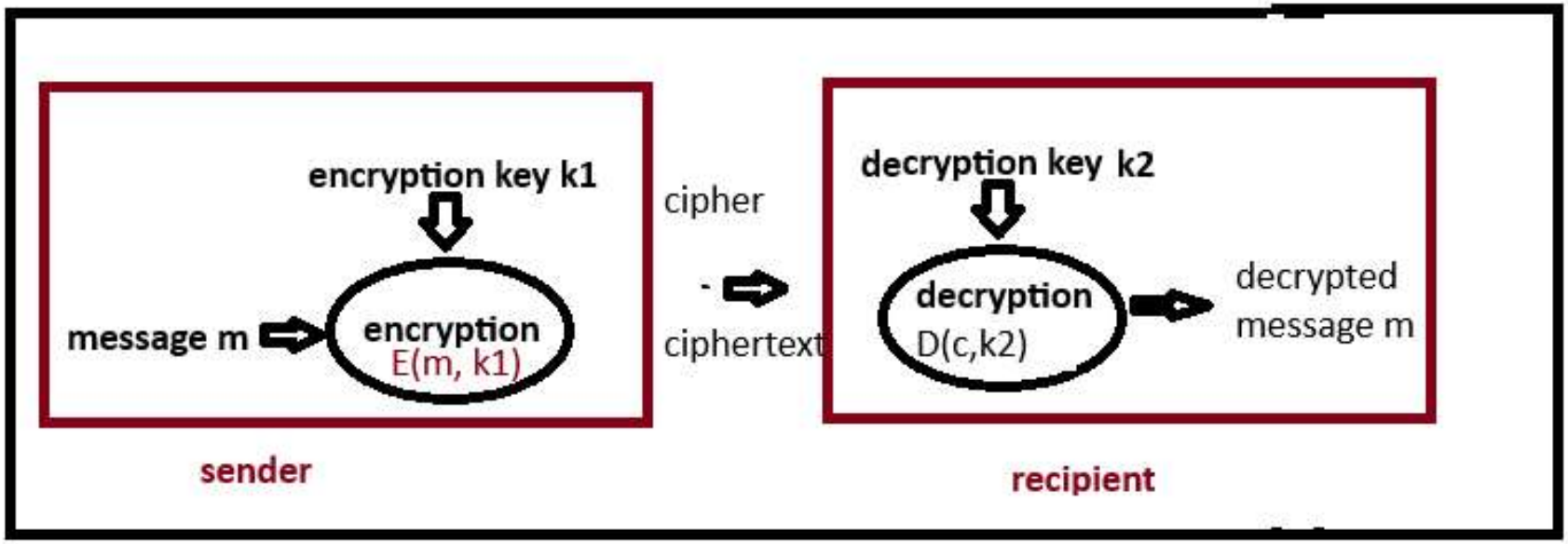
Cryptography offers solutions for these services. **Confidentiality** means that only those who are entitled to information can access it. Encryption of files and messages provides confidentiality. **Integrity** means that information remains unchanged in data transfer. Cryptographic hash functions provide integrity. **Authentication** of users also uses cryptographic methods, like RSA. **Non-repudiation** means that a party cannot deny later the contents of his messages or deny later having participated in the interaction. Digital signature is the cryptographic method which provides non-repudiation.
Access control can use for example one-time passwords created with cryptoalgorithms.

Basic concepts and principles

Topics:

1. Encryption of messages as a diagram
2. Symmetric and asymmetric encryption
3. Kerckhoff principle
4. Key space, effective key space
5. Mathematical concepts

1. Encryption : diagram presentation



Basic terminology of encryption

Message m , which is called also plaintext, is encrypted using encryption function E(m,k1), which has two arguments: message m and encryption key k1. Output of the function is called cipher or ciphertext, which is here denoted as c .

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

2. SYMMETRIC AND ASYMMETRIC ENCRYPTION

In SYMMETRIC ENCRYPTION the 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)

In ASYMMETRIC ENCRYPTION the 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.

3. Kerckhoff 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 examine them

4. Key space

A secure cryptosystem should not have vulnerabilities or backdoors.

The best option for a cryptanalyst to break a secure cryptosystem is 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} or in decimal representation $1.2*10^{24}$ possible keys.
Key space includes all keys from 00...0 to 11...1 (80 bit binary numbers)

Practical **recommended minimum** for key space, which has some safety margin, is 128 bits

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	theoretical key space 128 bits, effective 126.1 bits
AES256	theoretical key space 256 bits, effective 254.4 bits
In some older algorithms with vulnerabilities difference is bigger	
DES	theoretical key space 64 bits, effective 54 bits
3DES	theoretical key space 168 bits, effective 112 bits

Key space can be applied also to password 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 is 8 english characters. There is no distinction between uppercase and lower case letters.
Password space is $26^8 = 2*10^{11}$,which is less than the security minimum.

Example 2. Password length is 11 english characters + numbers from 0 to 9. There is distinction between lower case and upper case letter. The size of character set is 62.
Password space is $62^{11} = 5*10^{19}$. This is also too low.

Example 3. Password length is 13, where the character set is the previous 62 + ten special characters. Password space is $72^{13} = 1.4*10^{24}$, which exceeds security minimum and provides adequate security.

5. Mathematical concepts

HARD PROBLEM

Is a 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” . F.e RSA algorithm is based on the difficulty of factoring large integers.
Diffie-Hellman key exchange is based on Discrete Logarithm Problem (DLP).

ONE WAY FUNCTIONS

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

BACKDOOR

is some additional knowledge which makes the calculation of inverse function of an one way function possible

In RSA 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 changes the game and is a backdoor to RSA**: calculation of decryption key becomes trivial.

In some cases a cryptoalgorithm has had an **intentional backdoor**, which security authorities can use to break encryption of communication.

Classical ciphers

cipher = encryption algorithm, encryption method

Classifications of classical ciphers
Some classical ciphers:
1. Caesar cipher
 - Frequency analysis
2. Affine encryption
3. Hill cipher (Matrix cipher)
4. Enigma
5. Simple substitution
6. Vigenere cipher with key phrase
7. Autokey cipher
8. One Time Pad: ”the only unbreakable cipher”
Legacy of classical ciphers

Classifications of classical 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.

Another division of ciphers based on the block size

In **Monographic ciphers** message is encrypted one character at a time.
In **Polygraphic ciphers** message is divided to blocks of several characters and encryption is applied on one block at the time. Modern ciphers are polygraphic.

1. Caesar cipher

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

The ciphertext of message "AAMU" is "NNZH" which can be read from the disc of the picture

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

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

Frequency analysis

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 cases, when the characters of cipher text are not evenly distributed.

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

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: frequency analysis of Caesar cipher

Following ciphertext is obtained with Caesar cipher.

```
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 , number **9** is a 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
```

2. Affine cipher

1. Characters are coded to numbers 0 – 25 using the coding below

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

2. Let m = character of the message, c = image character of m. Key is the pair (a,b) of Z_{26}

Encryption formula : $c = a\,m + b \pmod{26}$

3. Ciphertext is a sequence of numbers c, which are decoded using the table above

Decryption

4. Multiplication encryption formula by a^{-1} (= inverse of a mod 26) gives $a^{-1}c = (a^{-1}a)m + a^{-1}b$. Rearranging gives:

Decryption formula: $m = a^{-1}\,c - a^{-1}\,b \pmod{26}$

5

- There are several ways of calculating the inverse of a.
- 1. Extended Euclid's algorithm
- 2. Formula a^{-1} = a^(ϕ(26)-1) mod 26 = a^11 mod 26 (based on Euler's theorem)
- 3. WolframAlpha calculator accepts: a^-1 mod 26

The key space size is 26*ϕ(26) = 26*12 = 312
(only those 12 values of a can be keys, which are coprime with 26; gcd(a,26) = 1)

Example of affine cipher encryption and decryption

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

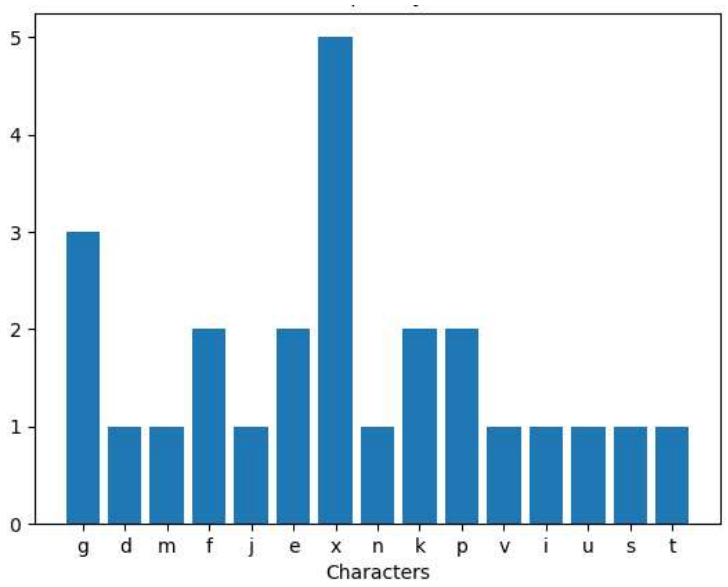
Decryption: $m = a^{-1} c - a^{-1} b \pmod{26}$
1. Calculate inverse $a^{-1} = 11^{-1} \bmod 26 = 19$ *)WolframAlpha
2. Decrypted cipher is
(19*9 – 19*3 , 19*21 – 19*3 , 19*5 – 19*3 , 19*13 – 19*3) mod 26 =
(10 , 4 , 12 , 8) = ”kemi”

***)WolframAlpha.com calculator calculates inverse easily: 11^-1 mod 26.**

Example of cryptanalysis

Break affine cipher ”gdmfjgexnxfgexkpvikxxupst”

Frequency diagram on the right show that **x** and **g** have biggest frequencies:



Hypothesis E(e)=x and E(t)=g gives
a*4 + b ≡ 23 and a*19 + b ≡ 6 (mod 26)
Subtract (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)

Test hypothesis (a,b) = (11,5)
Inverse $a^{-1} \equiv 11^{-1} \equiv 19$.

In Moodle there is Excel document ”classical ciphers” , which has a page Affine ciphers, which can be used to decrypt the cipher. Message is

”todaytheweatherisfreezing”

3. 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**.

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

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

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

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

Taking the previous example ciphertext of 'POH', we get:

$$\begin{pmatrix} 8 & 5 & 10 \\ 21 & 8 & 21 \\ 21 & 12 & 8 \end{pmatrix} \begin{pmatrix} 15 \\ 14 \\ 7 \end{pmatrix} = \begin{pmatrix} 260 \\ 574 \\ 539 \end{pmatrix} \equiv \begin{pmatrix} 0 \\ 2 \\ 19 \end{pmatrix} \pmod{26}$$ = ”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
If the block size is 3 letters, the key space is 26^9 around 10^{12}

First versions of Enigma cipher machine use by Germans in WW2 were based of Hill Cipher’s ideas.

4. Enigma Encryption machine

In World War II the Germans used Enigma encryption machine in communication with submarines.

The first computers were made in England for breaking the encrypted messages of Enigma. (”Tummy – machine”)

Videos: 15 min

https://www.youtube.com/watch?v=GBsfWSQVtYA Lorentz machine

https://www.youtube.com/watch?v=b4WBINgRMTY Tummy machine

The era of classical ciphers ends to the invention of computers and communication networks.

Modern encryption algorithms appear in business and administration in 1970’s

5. Simple substitution cipher

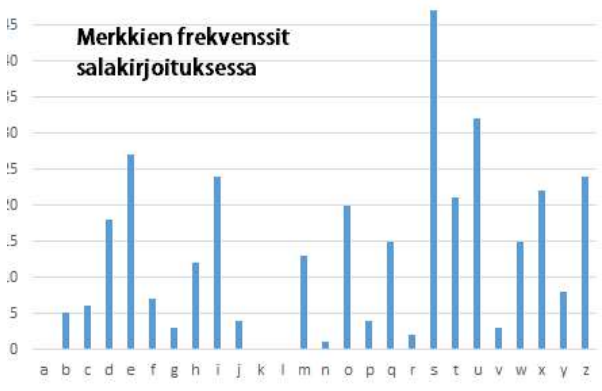
Encryption key is a permutation of alphabet (e.e. a->k, b->z, c->q,…).

Key space = number of permutations = 26! = $4*10^{26}$. Brute force is powerless.
With frequency analysis breaking of simple substition cipher is easier and can be compared to solving a difficult cross words puzzle.

CIpher:

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

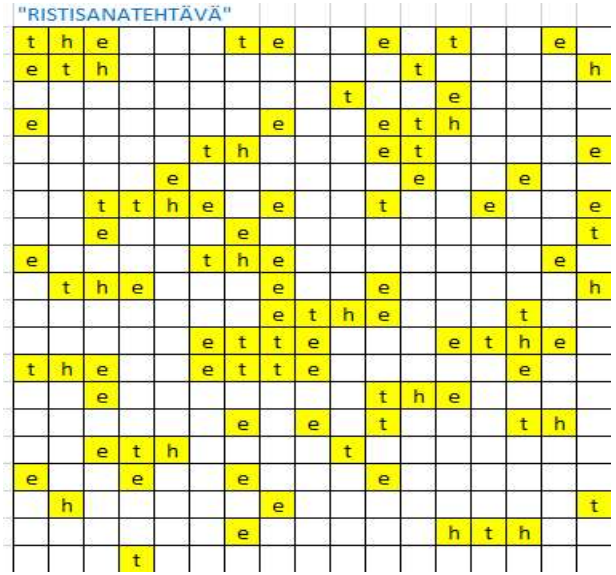
Frequency analysis



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



*) The table on the right is based on hypothesis, that the character s with the highest frequency s is the image of e, and u of cipher is image of t. In addition the middle character of s_u of chiper (t_e in original) is h, because cobination ”the” is most common in english language (appears in ”the”, ”these”, ”they”, ”them”, ….). This helps a lot in solving the ”cross words puzzle”.

5. Vigenère cipher using key word or phrase

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

- Encryption key is a password
- The whole key is obtained by copying password to the length of the message
- In encryption the characters of message and key are added using the table shown in the next slide

- Vigenere encryption is vulnerable if the password is moderately short. In 1800's Prussian officer Kasiski broke it. Kasiski's cryptanalysis is based on frequency analysis of 2- 3 character long "substrings" in ciphertext

Traditional Vigenere encryption with addition table

	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

Encrypt message “helsinki” with password “oulu”

H E L S I N K I

O U L U O U L U

==========

V Y W M W H V C

Key space is 25^n ,where n = password length. If f.e the password length n = 20, the key space = $9*10^{27}$ (large enough against Brute Force)

6. Autokey cipher: improvement of Vigenère

- Encryption primary key is a password
- Message is divided into blocks of length of the password
- First cipher block is calculated adding first message block with the message
- After that the new key used for encryption of next block is always the previous cipher block
- This procedure can be called CBC : ”cipher block chaining”.

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

Decrypt ”viznzzdaffl” with key ”lumi”

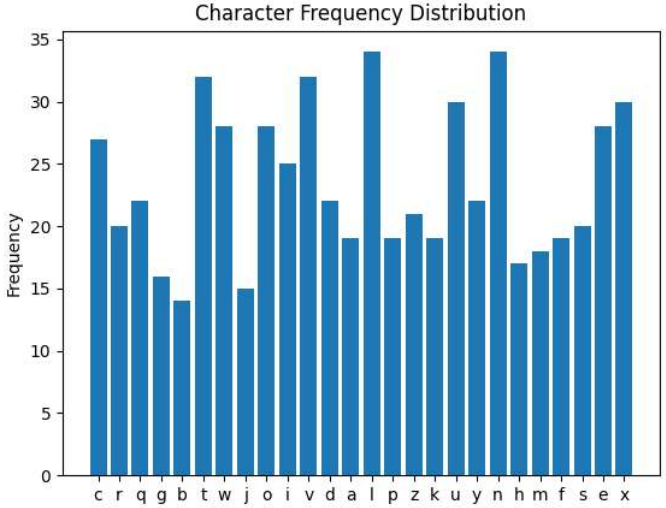
”vizn” – ”lumi” = ”konf”
”zzdz”-”vizn” = ”eren”
”rrl”-”zzs” = ”ssi”

Why Autokey is safer than Vigenère? Key string has no periodicity, which makes cryptanalysis with Kasiski method impossible.

7. One time pad – the only unbreakable cipher

- If Vigenère encryption is used with a random, one-time password of same lenght as the message, encryption is impossible to break. (true even with quantum computers)

- 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 right message from all other possible messages.



Frequency analysis of ciphers produced by One Time Pad encryption show nearly even distribution of characters. The longer is the message, the smaller are the differences of frequencies.

A basic property of secure encryption algorithm is that no information about the message or key can be retrieved from frequencies.

7. Binary One Time Pad (Vernam 1919)

A binary version of One Time Pad is Vernam cipher (1919)

Message m and a random key k of same lentgh are binary sequences.
In encryption m and k are added using XOR addition.

Decryption is done adding the same key k to the cipher.

XOR addition : 0 + 0 = 0, 1 + 1 = 0 , 1 + 0 = 1, 0 + 1 = 1

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

Vernam cipher was used in encryption of TELEX line between Moscow and Washington during the cold war.
The problem which makes both versions of One Time Pad unpractical is that both communicating parties must have similar ”codebooks” containing very long one-time passwords.

The legacy of classical ciphers

Attempts to break encryption of Enigma cipher machine led to **invention of first computers** during the end of WW2

Development of modern cryptoalgorithms started gradually after WW2 when administration and business were computerized.

Many ideas and principles of modern cryptography are copied from reseach of classical ciphers:
- key space calculations
- Kerckhoff's priniciple
- frequency analysis

American mathematician **Claude Shannon**, "The Father of Information Theory" laid the foundations for modern cryptography in his publications "Mathematica Theory of Information" and "Communication Theory of Secrecy System" (both in 1948).