

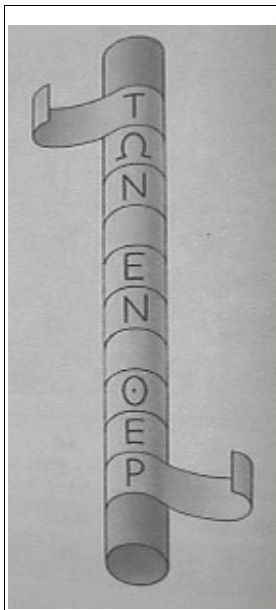
## Lecture 1 - Steganography, codes and ciphers.

We will start our course with a short introduction to the history of cryptology. It was given this name only in the 20<sup>th</sup> century, but already in the ancient times it occupied an important place in the human culture. Knowledge of these facts has many times turned out to be useful in the work of cryptologists, therefore they should be also mentioned to the cryptology students.

The age we live in is usually referred to as the century of knowledge or information, although information is not always identical to knowledge. The key role of knowledge in civilisation was obvious already to its earliest representatives. In order to make it useful for the society, it had to be gathered and handed over, but in a way that it does not fall into the wrong hands. Therefore the methods of hiding knowledge appeared almost at the same moment as the knowledge itself and the society that gathered it.

### STEGANOGRAPHY

Ancient Spartans hid their messages by winding spiral, narrow strips of leather around a wooden stick called "skytale", and then writing the message down along the stick; after unwinding of the leather strip, the characters appeared to be random. The addressee read the message by winding it around the stick of the same diameter - you can say that the "skytale's" diameter constituted the key to the code. A certain Greek, Histiayos, serving at the court of the great king of Persia, decided to send information about a convenient moment to stir uprising against the Persian rulers to his cousin in Miletus. He summoned one of his slaves, shaved his head and tattooed the text of warning on it. When the hair of the messenger grew back, he sent him to his country. A bit later, Julius Ceasar, running a campaign in Gaul, coordinated operations of several military units by sending couriers with orders to their commanders. Taking into consideration the risk of the courier falling into the hands of enemy, he encrypted letters by replacing each character of the Latin alphabet with a character situated three positions further.



A scytale

Three basic elements of hiding messages - steganography, codes and ciphers - appeared quite early. The first was probably steganography, which may also be referred to as hidden text. If we simplify things a little, we might say that this anonymous Greek used it by sending messages tattooed on the head of the messenger. The essence of steganography is not only hiding the contents of the message but also the fact that the message is being sent. In time, its methods became more sophisticated, as the users came to the conclusion that the fact of sending a courier or a parcel arouses suspicions of the enemy if no innocent and credible reason for sending them is provided. They were sending apparently innocent messages, hiding the proper message in their text or on their carrier. A typical method was the use of the so-called invisible ink. On a sheet of papyrus, parchment or paper, the real contents of the message was written with

milk or lemon juice; then, when the invisible ink dried, an ordinary letter, e.g. greetings for a cousin, was written with a normal ink. Even during a detailed inspection the messenger could claim that he was transporting only a family correspondence. The addressee heated the message over a burning candle or chafed the sheet with a cloth saturated with an appropriate reagent thus deciphering the hidden message.

With appearance of newspapers and other magazines, another method of transmitting hidden messages became popular. It was enough to take a copy of a newspaper and, in a place previously agreed with the addressee, puncture holes, invisible to the naked eye, for example below subsequent letters of the text. Another method was elaborated just before the World War II by ingenious Germans. They benefited from the advancement of German microphotography, applying it in their spying techniques. A photo of any document was subsequently reduced in size so that several pages of print or handwriting could be contained on a microphotograph 1 millimetre wide. The microphotograph was cut out of the plate with medical needle and stuck in a place agreed with the addressee - this technique was called the microdots technique. When Americans, warned by a double agent, discovered such message for the first time in August 1941, the microphotograph pretended to be a dot on a postal stamp.



If the sender and the addressee do not have such technical means at their disposal, they can try to exchange letters of completely neutral contents, in which only determined characters, e.g. on the first position in a given line or word, or written with a different font, are significant. As a rule, the choice of words in given positions is not easy and the letters hiding secret text strike with unnatural style or vocabulary. The contemporary technology led to a renaissance of steganographic techniques.

## CODES

The news of the method used by Julius Ceasar can indicate that ciphers appeared as the second, after steganography, method of hiding information, but we will concentrate first on codes. Our everyday language also constitutes a code. In our early childhood, the uncalled notions that function in our mind are given names in the language used by our environment. The analogy with foreign languages shows the code nature of our language: people speaking one language use one code, allowing them to communicate. However, it is enough to use this same code when speaking to a person of a different language circle to encounter informational barrier. Besides, there are particular variations of the language the goal of which is to hide information from third persons; prison or smugglers slang, as well as the language of the majority of sciences.

It is easy therefore to conclude that it is enough to elaborate a secret glossary to hide the contents of the discussion or correspondence. And such glossary does not have to include all the

words of the natural language, but only a chosen group of notions playing the key role in the dialogue.

In its extreme and yet the most natural form, the code should encompass possibly largest group of language words. The codes elaborated in the 19<sup>th</sup> and at the beginning of the 20<sup>th</sup> century were of monstrous dimensions; one of telegraphic codes contained equivalents

Ezdig	After what has taken place.
Ezdom	After you (they) have.
Ezdro	After you have seen them.
Ezdus	Agree to the proposal.
Fabmi	Agree with them if possible.
Fabok	Aid them all you can.
Fabup	All goes well.
Facax	Always the same.
Facby	Am able to .....
Facda	Am acting under legal advice.
Faceb	Am alone. Where can I meet you?
Fache	Am coming over.
Facif	Am not coming over.
Facil	Am not coming over at present.
Facol	Am not quite .....

Code book

of over 300 000 notions. Codes in such forms had at least two defects. They were published as large books, difficult to hide from the enemy, especially if they were destined for use on his territory. Therefore the use of codes became popular mainly where the sender and the addressee of the information operated in stationary and closely-guarded premises, e.g. embassy of a given country. Another

disadvantage of codes was the long time and considerable effort necessary to elaborate and safely copy them. The process of code production lasted several months and its use was extended for years, if not for dozens of years. Of course, in such a long period of time the opponent had the possibility to intercept many secret telegrams, coded with the same code, what gave him a perfect attack point.

In the period of the Renaissance, the practice of replacing only the most often appearing notions with their code equivalents gave birth to a popular tool for hiding information - the so-called nomenclator. In its most typical form, it contained code equivalents of notions the most often used in correspondence and an accompanying simple cipher, with which were secured those fragments of the text for which there were no equivalents in the code. Here is an example of a simple nomenclator:

<b>COMPANY</b>	<b>8381</b>
<b>BATTALION</b>	<b>9211</b>
<b>REGIMENT</b>	<b>1153</b>
<b>DIVISION</b>	<b>0455</b>
<b>ATTACKS</b>	<b>9802</b>
<b>DEFENDS</b>	<b>6653</b>
<b>WITHDRAWS</b>	<b>1388</b>
<b>PROVIDE AMMUNITION</b>	<b>9077</b>
<b>SEND REINFORCEMENTS</b>	<b>6615</b>

**ABCDEFGHIJKLMNPOQRSTUVWXYZ**  
**QWERTYUIOPASDFGHJKLZXCVBNM**

A report: "Third company attacking. Deliver ammunition", encrypted with the use of the nomenclator above will look as follows:

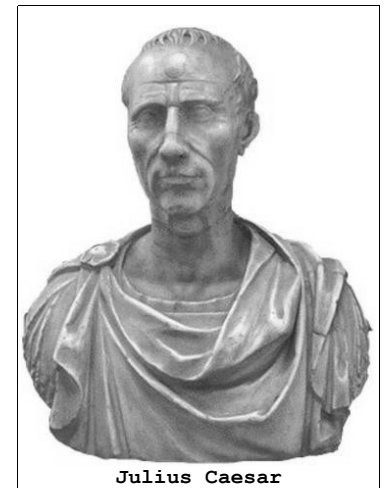
**ZITZIOKR838198029077**

Codes and ciphers do not contain equivalents for the interval between words or for punctuation marks (full stops, commas etc.); they are omitted before the text is encoded or replaced with a letter or a couple of characters rarely appearing in a given language.

A successful attack on the code required gathering a large number of coded messages. The nature of correspondence caused some notions to appear more often than others. Moreover, the process of coding replaces the words of the natural language but does not deform the natural structure of statements in this language. A cryptologist could assume that at the beginning of the message there will be a designation of its addressee ("To the command of the division", "To the Ministry of Foreign Affairs" etc.), afterwards a social formula or a formula beginning the message ("Dear Minister", "Dear General" or "The commander of the regiment reports"), then a series of sentences of the structure consistent with the nature of the language and the character of the message. In subsequent intercepted messages the same elements begin to repeat themselves on the same or similar positions, giving the attacker the code of increasingly better hypotheses concerning their meaning. When he manages to identify the meaning of several code groups, he reviews earlier messages searching for them, trying to identify the groups of codes often accompanying the analysed ones and determine their meaning. The authors of code perfectly knew their weaknesses and tried to prevent them, for example by introducing many equivalents of the most often used words in the code books. E.g. in the code for the diplomatic use, the words MINISTER and AMBASSADOR usually had several code equivalents and the task of the coder was to use them interchangeably, without any regularity.

## CIPHERS

The third method of hiding information are ciphers. Let's get back to Julius Ceasar, who, in his messages to his legates, replaced each letter with a character located three positions further. His own name, IULIUS CEASAR, thus became LXNLXV FDHVDU. By doing so he used encryption, the essence of which is to replace each letter of the clear text with a letter of the ciphertext. By sending encrypted messages, Ceasar could feel confident as the literacy was rare among the Celtic tribes fighting against him. If, however, a Roman deserter had found his way to the opponents, additionally aware of the Ceasar's tricks, he would not have had many problems with breaking his cipher.



Julius Caesar

Two 26-character alphabets can be moved one with regard to the other in 25 combinations, so that the encrypted character is not identical to the clear text character; the Ceasar's cipher had only 25 different keys. But let's recall the simple cipher constituting part of the nomenclator described above.

**ABCDEFGHIJKLMN OPQRSTUVWXYZ**  
**QWERTYUIOPASDFGHJKLZXCVBNM**

Its upper line is called clear alphabet, the lower - secret alphabet. For the cipher to work, the secret alphabet characters do not have take the alphabetical order, as it was in case of Ceasar. What is more, the secret alphabet characters do not have to be identical to the clear alphabet characters; the cipher below will be just as good:

**ABCDEFGHIJKLMNOPQRSTUVWXYZ**  
**¢£¥§©®µ¶;€∂∫≈∆▲⊙♠♦◆◊♀♣□△**

For two parties to communicate safely using a cipher, they have to dispose of a pair of identically assigned alphabets: the clear and the secret one. Transferring them in the form presented above is inconvenient and memorising it - impossible. Therefore, to define the mutual assignment of the clear and secret alphabet, a key word is usually used. Let's assume that the key word for our cipher is CRYPTOLOGY; on this basis we create a pair of alphabets in the following manner: in the upper line we write the clear alphabet in alphabetical order. In the lower line we write the key word, omitting the characters that repeat themselves (in our case it will be O). Having written the key word, we write those signs that do not appear in the key word in alphabetical order. We obtain the following encrypted substitution.

**ABCDEFGHIJKLMNOPQRSTUVWXYZ**  
**CRYPTOLGABDEFHIJKMNQSUVWXZ**

Codes are based on words of the clear text and ciphers - on single characters. It is significant for the code-cipher breaking methods. As we recall, the work on breaking codes was started with identification of the most often appearing elements, then, based on their place in the text, attempts were made to determine their meaning. The code hides the notions of the clear text, but it does not hide the structure of the entire text, characteristic for the language and the subject matter. Therefore, people chosen to break codes knew well the structure of the clear text language. Their experience was less useful when analysing a cipher, however. As each letter is transformed individually, a large part of purely linguistic properties of the clear text becomes obliterated in ciphers.

**In order to better assimilate and memorise the most important issues concerning steganography, codes and ciphers, we prepared three riddles for you. The success in resolving them will be your pass to the next part of the course, as every student of cryptology should show some talent in this domain. The riddles will be presented exactly one month after the publication of this lecture, on the 10<sup>th</sup> of December 2011 at 19.29.**