INFO-F-405 : Security
Encryption and statistical analysis

The goal of this session is to learn about encryption and decryption as well as about first cryptographic tools and basic techniques of cryptanalysis. We are going to use shift ciphers during this exercises.

1 Mono-alphabetic encryption

One of the first ciphers was the Caesar's cipher, it is a mono-alphabetic shift cipher. Long time ago this cipher was used with shift of 3 letters in the alphabet and it was enough to secure messages, since most of people did not know how to read at all. This method became completely obsolete with the discovery of an attack based on frequency analysis.

There exist other mono-alphabetic ciphers like substitution cipher, but we are going to use a special case which is the shift cipher.

1.1 Encryption and decryption

1.1.1 Encryption

The encryption is relatively strait forward. Let us represent letters of the latin alphabet by numbers between 0 and 25, in other words:

$$A \to 0\ ,\ B \to 1,\ ..., Z \to 25$$

Now we can write an encryption function $E_k(x)$ for the shift cipher in $\mathbb{Z}_{26}$ with a secret key k, where $k \in [0; 25]$, this value k represents the amplitude of the shift. Here is the definition of the encryption function:

$$E_k(x) = (x + k) \bmod 26$$

1.1.2 Decryption

The decryption function is the inverse of the encryption function $E_k(x)$, we will note it as $D_k(x)$. We can write the decryption function:

$$D_k(x) = (x - k) \bmod 26$$

Exercise 1

A cipher with $\mathcal{M} = \mathcal{C}$ for certain keys $k \in \mathcal{K}$ is called *involutory* if its encryption and decryption procedures become identical, i.e. for all $m \in \mathcal{M} : E_k(m) = D_k(m)$. Under which keys $k \in [0, 25]$ the above Shift Cipher becomes involutory?

Solution: For $k = 13$ and $k = 26$. For instance, for $k = 13$ plaintext letter `A` will be encrypted to the ciphertext letter `N`, which is equivalent to computing $14 = 1 + 13 \pmod{26}$ whereas the decryption will compute $1 = 14 + 13 \pmod{26}$. For $k = 13$ the Shift Cipher is sometimes called ROT13. Note that if $k = 26$ the encryption procedure becomes an identity function, i.e. $\#c_i = \#m_i$. That is, using $k = 26$ is not interesting since the ciphertext is identical to the plaintext.

1.2 Attacks

1.2.1 Brute-force attack

The exhaustive key search a.k.a. brute-force is very easy to implement and in this case it is also very easy (and fast) to execute since we only have to test 26 different values for the key. Afterwards a human can easily detect which decryption was successful. Sometimes one can immediately spot the key while looking at the ciphertext by finding small frequently used words like "the", "of" or "to".

Exercise 2

Write a program that uses brute-force attack in order to find the secret key that decrypts the following message:

```
Fbzr crbcyr jrne Fhcrezna cnwnznf. Fhcrezna jrnef Puhpx Abeevf cnwnznf.
```

Here is another ciphertext:

```
'Yvccf, nficu!' - zj fev fw kyv wzijk kyzexj gvfgcv kip
kf gizek nyve kyvp cvrie r evn gifxirddzex crexlrxv.
```

1.2.2 Known-Plaintext Attacks

In one of the previous exercices we performed cryptanalysis by knowing only the ciphertexts. Here we consider known-plaintext attacks where the analyst may know plaintext/ciphertext pairs.

Exercise 3

(a) How many pairs of plaintext/ciphertext characters must be known in order to determine the key of the Shift Cipher?
Solution: One pair (m_i, c_i) is enough since the same key $k = \#m_i - \#c_i \pmod{26}$ is used to encrypt all characters.

(b) Give an answer to question (a) for the Vigenère Cipher.
Solution: Since in Vigenère Cipher the key k consists of d characters that are periodically used to encrypt the plaintext the analyst must known first d characters of the plaintext/ciphertext pair (m, c). Then each of these d characters can be found as in Exercise 1.2(a).

A	B	C	D	E	F	G	H	I	J	K	L	M
8.17	1.49	2.78	4.25	12.70	2.23	2.02	6.09	6.97	0.15	0.77	4.03	2.41
N	O	P	Q	R	S	T	U	V	W	X	Y	Z
6.75	7.51	1.93	0.09	5.99	6.33	9.06	2.76	0.98	2.36	0.15	1.97	0.07

Table 1: Relative frequencies of letter in a text in English (in %).

1.2.3 Statistical analysis

One of the first statistical attacks on this cipher is an attack based on frequency analysis. The idea is to calculate the frequency of each letter of the ciphertext and to compare them to the reference table that contains frequencies of letter use in a given language. Using this technique we can find the correct shift i.e. the secret key that was used. Table 1 gives frequencies of letters for a text written in English.

Exercise 4

Write a program that uses frequency analysis in order to find the key. Test it on the previous exercise and compare the result with the previous one (that used the brute-force attack).

1.3 Kasiski examination/Classical ciphers

Kasiski Method. This test allows to get a good approximation of the size of the key. The idea is to search patterns of $2, 3$ or 4 letters (N-gramms) that appear regularly. Once we find some patterns, we can calculate the distance between them. Once it is done, we can calculate the *gcd* between these distances and approximate the size of the key.

Vigenère cipher. The Vigenère cipher is a polyalphabetic cipher, its development was a very important step in the history of cryptography. The Vigenère cipher was not broken for the period of 200–300 years after its development.

Exercise 5

The following ciphertext was generated using the Vigenère cipher from the plaintext in English language alphabet containing only capital letters (ABC...XYZ). That is, $\#c_i = \#m_i + \#k_j \pmod{26}$ where $k = k_1 \dots k_d$ is the key, $m = m_1 \dots m_n$ is the plaintext, and $c = c_1 \dots c_n$ is the ciphertext. Vigenère cipher assumes the following numbering of alphabet letters: $\#\mathtt{A} = 0, \#\mathtt{B} = 1, \dots, \#\mathtt{Y} = 24, \#\mathtt{Z} = 25$.

```
LNRKE RRWLZ HCUEG ZAQEO PQLGY EBDMY QWFRS YLLCG TMVEY DNBKE
RRLBX SMIHG VLGIE GQTFH LYQDM ISSEM YUILH SQRWC VAGOE BXPRR
TFHSS QTGOL UHFMX NBWHC VEYRF EUECQ ALGWC OITHD ZHNCD TFWHC
                ZATHS GQOSU YCOLM ZSSEM YUILH
```

(1) The most frequent digramms in the ciphertext are `HC`, `HS` and `RR`, each occurs three times. Find the length d of the key using the Kasiski Method.
Solution: The positions of `HC` are 11, 119, 149 and the differences $\delta_{\mathtt{HC},1} = 119 - 11 = 108$,

$\delta_{\texttt{HC},2} = 149 - 119 = 30$, $\delta_{\texttt{HC},3} = 149 - 11 = 138$. The positions of HS are 85, 103, 154 and the differences $\delta_{\texttt{HS},1} = 103 - 85 = 18$, $\delta_{\texttt{HS},2} = 154 - 103 = 51$, $\delta_{\texttt{HS},3} = 154 - 85 = 69$. The positions of RR are 6, 51, 99 and the differences $\delta_{\texttt{RR},1} = 51 - 6 = 45$, $\delta_{\texttt{RR},2} = 99 - 51 = 48$, $\delta_{\texttt{RR},3} = 99 - 6 = 93$. The greatest common devisor is $\gcd(108, 30, 138, 18, 51, 69, 45, 48, 93) = 3$. The period is $d = 3$.

(2) Assume that digramm HC decrypts to HE and HS decrypts to ES. Determine the letters of the key k.
Solution: The digramm HC is located on position 11. Since $11 \bmod 3 = 2$ the digramm HC was encrypted using letters k_2 and k_3. Since ciphertext letter H decrypts to plaintext letter H we get $k_2 = 0$, which corresponds to letter A. Since ciphertext letter C (with $\#\texttt{C} = 2$) decrypts to plaintext letter E (with $\#\texttt{E} = 4$) from $2 = 4 + k_1 \bmod 26$ we get $k_1 = 24$, which corresponds to key letter Y. The digramm HS is located on position 85. Since $85 \bmod 3 = 1$ the digramm HS was encrypted using k_1 and k_2. Since ciphertext letter H ($\#\texttt{H} = 7$) decrypts to plaintext letter E ($\#\texttt{E} = 4$) from $7 = 4 + k_1 \bmod 26$ we get $k_1 = 3$, which corresponds to key letter D. The resulting keyword k is DAY.

Playfair cipher.

- was used in WW I by British army
- substitutes digrams (more efficient than substituting single characters)
- key k is a 5x5 matrix filled with alphabet letters (without J)

Let $m = m_0, ..., m_n$. $Enc(k, m) = c_0, ..., c_n$ according to the rules:

1. m_i, m_{i+1} in one row: c_i right to m_i, c_{i+1} right to m_{i+1}
2. m_i, m_{i+1} in one column: c_i below m_i, c_{i+1} below m_{i+1}
3. m_i, m_{i+1} span a rectangle: c_i opposite to m_i, c_{i+1} opposite to m_{i+1} (within the same row)
4. $m_i = m_{i+1}$ or $|m|$ is odd: inject or append some pre-specified dummy letter.

Exercise 6

Compute the plaintext from the following Playfair ciphertext using the 5×5-key matrix below.

FK SM VS WV ST QC

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

Solution: The plaintext CR YP TO GR AP HY can be obtained by performing the encryption process in the reverse order. For instance, FK spans a rectangle where within the same rows C is in the opposite corner to F and R in the opposite corner to K.

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 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 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 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 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 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 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 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 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 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 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 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 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 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 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 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 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 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 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 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 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 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 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 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 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 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

Figure 1: Vigenère square.