INFO-F-405 : Security
Encryption and statistical analysis

created by: Naïm Qachri
updates: Nikita Veshchikov

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

On 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

Write a program that decrypts the following message, the value of `k = 'i'`:

```
Kzgxbwozixpg qa ijwcb kwuucvqkibqwv qv bpm xzmamvkm wn iv ildmzaizg.
```

1.2 La cryptanalysis

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 Cryptanalysis

One of the fist 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 3

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).

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 %).

2 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.

2.1 Encryption and decryption

The encryption uses what is called the Vigenère square or Vigenère table, see Figure 1. This square contains all 26 possible shifts (as in shift cipher). In order to encrypt a message one must choose a key or the secret word. Then one must copy the secret word many times in order to obtain a text of the same size as the message that he wants to encrypt. Final step is to apply the shift cipher on each letter of the message using the letter of the key that is associated to it.

Here is an example:

Message	T	O	U	T	A	U	N	E	M	O	R	A	L	E	.....
Key	L	E	W	I	S	L	E	W	I	S	L	E	W	I	.....
Ciphertext	E	S	Q	B	S	F	R	A	U	....					

Decryption uses the same principle, but the shift cipher is applied in the other direction.

Exercise 4

Decrypt the following cipher text using "edgar" as the key:

```
prtdfrlytyifgpzxdrowkukakfuotrmqotjtrrikmfglvgrtodmf
gnugxrtlvdrcvrwxezxvunvsizhvpdxgvwwiikmhyiexkkwfvoji
kwsuplpdzifrlymfvhzhrrhremiqsicplungirvlvprtdfrlyszx
xgtvhrttyiuovvvwnadivzhvglzyzwykrpsojaehekalxllucmwc
ajjranuigsoiiwnaexzutysxyaehbkaiwdmo
```

```
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.

2.2 Index of coincidence

This test is used in cryptanalysis and allows to determine the language that was used in order to write a message thus allowing to use appropriate statistical analysis afterwards. The idea is based on the fact that regardless of the shift the global frequency of letters does not change. For the letter i, knowing the p_i – the probability of its appearance, we are going to calculate:

$$I_c(x) = \sum_{i=0}^{25} {p_i}^2$$

Exercise 5

Use Table 1 in order to calculate the index of coincidence of English language. How might it be helpful for the cryptanalysis?

2.3 Kasiski examination

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 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.

Exercise 6

Use Kasiski examination in order to find possible size of the secret key.

2.4 Automatic test using index of coincidence

This test works in the following way, the ciphertext should be divided into m parts $z_1, z_2, ..., z_m$ writing them column by column. In other words we are going to build a $m \times \frac{n}{m}$ matrix, where n is the length of the message. Each value at position i of a line k is a letter of the ciphertext y. Lines of this matrix give us parts z_i, $1 \leq i \leq m$, z_i that represents a line i of the matrix is given by: $z_i = y_i \ y_{m+i} \ y_{2m+i} \ ....$

A mono alphabetic cipher does not alter the value of the index of coincidence, which is not the case of Vigenère cipher. In Vigenère, if the length of each part z_i (in the matrix) is equal to the size of the key, then its index of coincidence would be close to an index of some language. Index of coincidence of random text is $I_c = 0,038$, thus we can distinguish it from a non-random text e.g. in English.

Exercise 7

Get the size of the key given by Kasiski examination and to confirm it using the test described here above. If tests agree with each other, than index of coincidence $m =$*size of the key* would be close to its value for the English language.

2.5 Combination of attacks and cryptanalysis of Vigenère

We saw how to determine the language of a message as well as how to calculate the size of the secret key. Thus, now we can decrypt the cihphertext.

You can easily run a frequency analysis on each z_i in order to obtain shifts. A test based on the index of coincidence can speed up this computation. For all g such as $0 \leq g \leq 25$, the quantity $\sum_{i=0}^{25} p_i \cdot \frac{f_{i+g}}{n/m}$, where p_i gives us the frequency of appearance of a letter, f_{i+g} is the number of times when the letter $(i+g)$ appears in z_i and n/m is the length of z_i.

If g is the correct shift for the letter k_i of the key for z_i then the quantity above would be approximatively equal to the index of coincidence of a given language.

Exercise 8

Find the shift value of each part of the message (in form of a matrix) and find the secret key. Try this attack on the following message:

```
oybfhvzeihisftflnchusflpahrfsiugyavrlbfxonmexoyofivaqiylfqstmaxaasmdvvbvlbfmskevr
vmguvvbbwwpfzmvbrvkjfdsvkbrvqjfhxcfxkoqfoeqaieyuqrjxrvkydanubbraaifmfagpqwmauugys
ombdolvqjazuqeelrkjyfecvvgmylbpeewexzybetnqahkusauelsmnynqtvzftwltqcfvqkmmueugbbi
aaifiebuiwpfzieogseisdonubwlfzolfcqczchtszvglwgjptnmaxaytzoibuismuiielfwwpfztfbrr
eyuqrjxrvkydanubbqglsaweqtllgpetcgppgoekiebuiwpfzieoglwhdxerzvryfpisrzbyfxuiecdrr
gluteragaaheetyzriliguvvurxwltbeiargghe
```