C Y B E R S E C U R I T Y

L A B 2 / Solution

1. Introduction

The laboratory covers some of the basic cryptographic algorithms that were used for providing a secure way of communicating the messages between entities in the olden days. In these cryptographic algorithms, we assign numbers (or) algebraic elements to the given input message to be communicated between two entities. If the assigned numbers (or) algebraic values are in intelligent form, then it is considered as plaintext which is also called plaintext. This intelligible plaintext is converted into an unintelligible form called ciphertext. To convert the intelligible plaintext into the unintelligible ciphertext, an encryption function is used on the sender's side. Similarly, a decryption function is used in the receiver side to find intelligible plaintext from the unintelligible ciphertext. The process of converting the intelligible plaintext into unintelligible ciphertext and back into intelligible plaintext is called cryptography. This laboratory is discussing cryptanalysis of classical encryption algorithms such as *substitution, transposition,* and mixed algorithms.

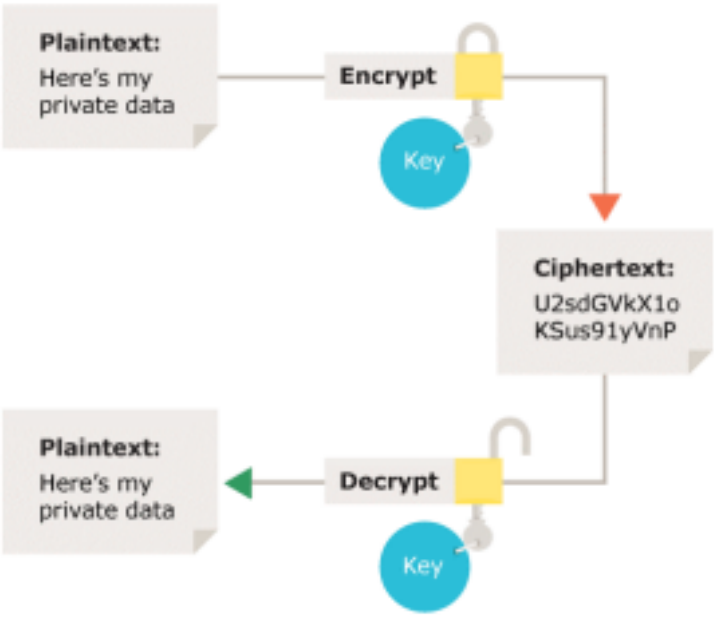


Figure 1 Encryption and decryption process. Source https://ico.org.uk/

Entropy is the amount of information a given message or variable contains. Information entropy is sometimes described as "the number of bits required to communicate information." From a mathematical point of view, it is the sum of the probabilities of a given sign and the number of occurrences within the data. It turns out that the level of entropy of different types of data is different. Thus, a text in natural language will have a different level of entropy than machine code, and this in turn will have a different level of entropy than compressed data. The entropy of an encrypted string of data, depending on the encryption algorithm used, may not change at all or increase significantly.

Histogram shows relative frequency of each of the characters in the document. It allows us to compare the prevalence of different letters and can help derive the nature of the information and income cases, even revealing some of the character's representation.

Autocorrelation is used to compare the text with its displaced copy and overlapping

characters are counted. It turns out that if you move by the length of the key (or a multiple of it), there are many more overlapping characters. Also known as the coincidence index, it is an indicator of how likely it is that, when any two texts are compared letter by letter, the two characters being compared will be the same. Autocorrelation charts show the number of overlapping characters according to the offset of the text. Based on the autocorrelation graphs, the length of the used encryption key can be derived.

Periodicity searches for repeated fixed patterns in a document that start from any point and continue through to the end of the document. The pattern has two notable features, its offset and period. The offset is the start position of the cycle, counting from the first character of the document, and the period is the length of the set of char acters that are repeated.

N-gram divides the message into blocks of N size and checks the frequency of phrases. It allows for evaluating the keys.

Floating Frequency breaks the message down into blocks of 64 characters and then displays the range of different letters for the block surrounding each letter. These results are then displayed to the user in the form of a line graph to allow them to see what parts of a message have high levels of variance and which have comparatively low variance. The purpose of this excursive is often to derive the nature of the data that has been encrypted. Frequency analysis is the basic tool for breaking most classical ciphers. In natural languages, certain letters of the alphabet appear more frequently than others. By examining those frequencies, you can derive some information about the key that was used. This method is very effective against classic ciphers such as Caesar, Vigenère, and so on. It is far less effective against modern methods, however. In fact, with modern methods, the most likely result is that you will get some basic information about the key, but you will not get the key.

2. Run and get familiar with the tool used in the laboratories - Cryptool.

a. The CrypTool program is a free tool developed by the universities of Siegen and Darmstadt available at www.cryptool.org. The program is the basic platform used for several consecutive laboratory exercises. https://www.cryptool.org/en/
b. The aim of the CrypTool creators was and still is to provide good quality teaching aids for illustrating and demonstrating the operation of selected mechanisms in cryptology.
c. The program package includes several encryption algorithms, high-quality documentation on cryptography, and a demonstration of selected algorithms.
d. Try to install and look at the capabilities of the program even before the classes.
e. The first classes will mainly use the tools available in the Encrypt/Decrypt -> Symmetric (Classic) and Analysis -> Tool for analysis tabs.

Attention!

For exercises on historical algorithms, it is best to use input data in the form of a string of characters limited to the Latin alphabet, without special characters or punctuation.

For example:

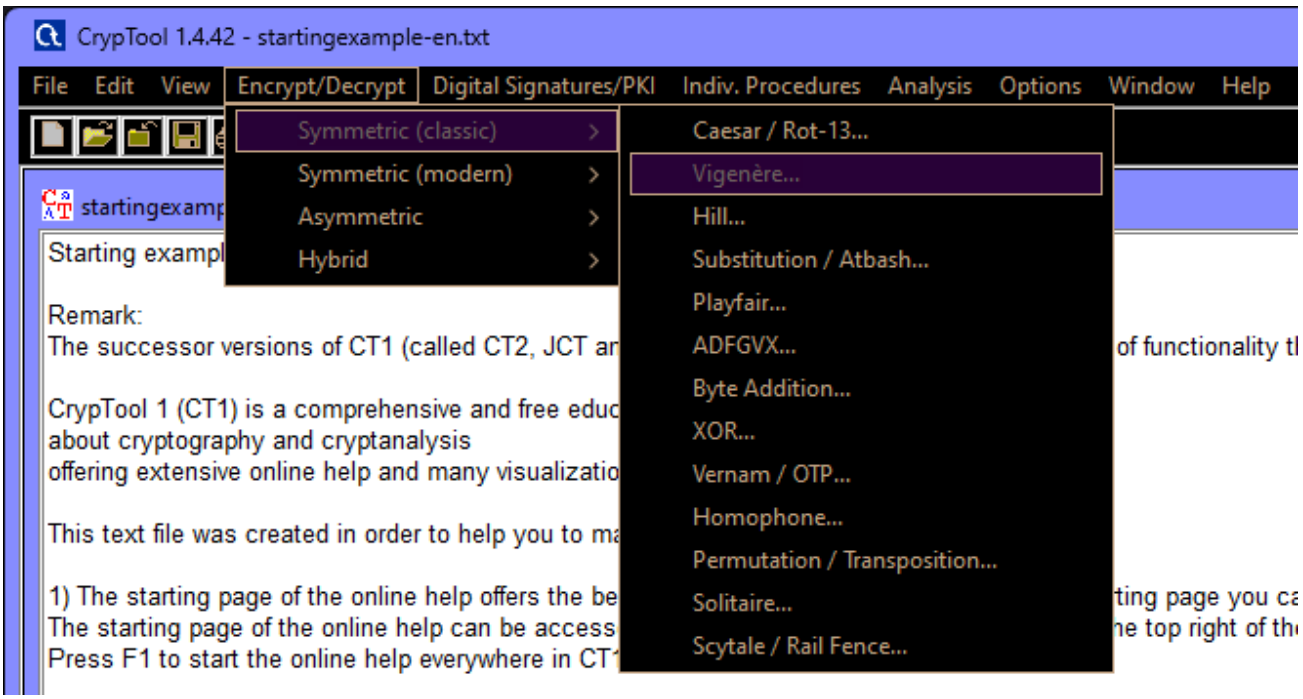
DLACWICZENDOTYCZACYCHALGORYTMOWHISTORYCZNYCHNAJLEPIEJKORZYSTACZDAN
YCHWEJSCIOWYCHWPOSTACICIAGUZNAKOWOGRANICZONYCHDOALFABETULACINSKIEG
OBEZ ZNAKOWSPECJALNYCHIINTERPUNKCYJNYCH

3. Historical algorithms:

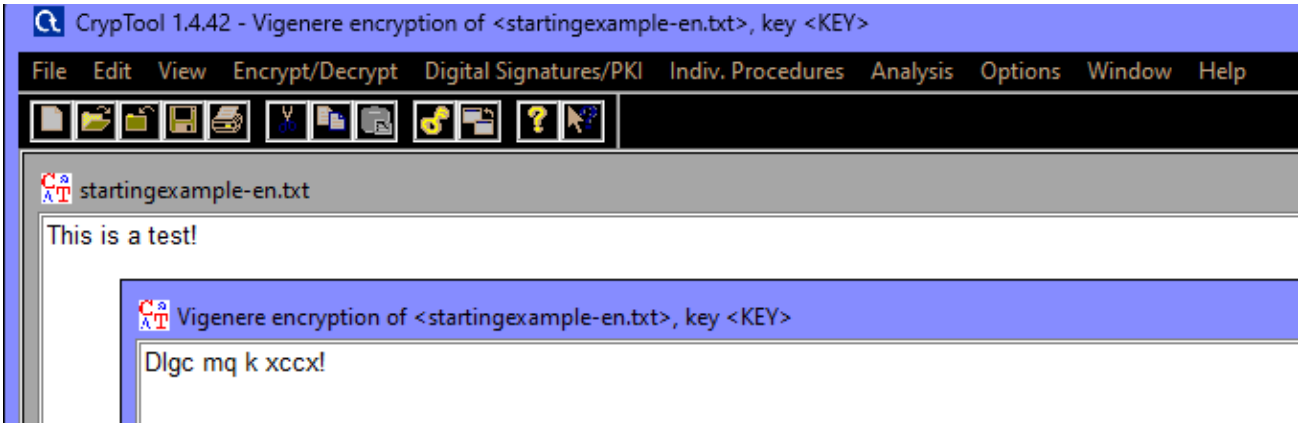
This section aims to learn about the operation of historical algorithms.

Tasks:

1. Check the available historical algorithms in tab *Encrypt/Decrypt -> Symmetric (Classic)*.



2. Choose a few of them and check their properties.



>> ex: a test using vigenere

Question:

1. What can we say about multiple encryption in the context of historical algorithms (consider their different classes)? How does this affect the ability to decrypt a ciphertext? The answer to this question is illustrated by the result of an experiment conducted in Cryptool.

- **Dealing with Historical algorithms, multiple encryption is needed as they take longer to decrypt yet not strong enough to protect the data.**

2. How many different keys are there in the classical substitution algorithm?
   - **There are 26!**

3. What is the space (number) of keys in polyalphabetic algorithms?
   - **K! * kd - 1 , where k is the size of the alphabet**

4. How many different keys are there in the Playfair and Hill algorithm?
   - **For Playfair it is 25! On 5*5 matrix &**
     Pk 2(1-p-1)(1-p-2)...(1-p-k) for Hill where p is the size of the alphabet and k is the size of the matrix.

5. What does the number of keys in transposition algorithms depend on?
   - **It is based on a two-phase hill climbing algorithm, a two dimensional fitness score and a special transformation on key segments.**

6. Which of the selected algorithms do you consider the strongest, why?
   - **Hill's algorithm is based on linear algebra where Playfair is more complex. The answer is Playfair.**

Conclusion

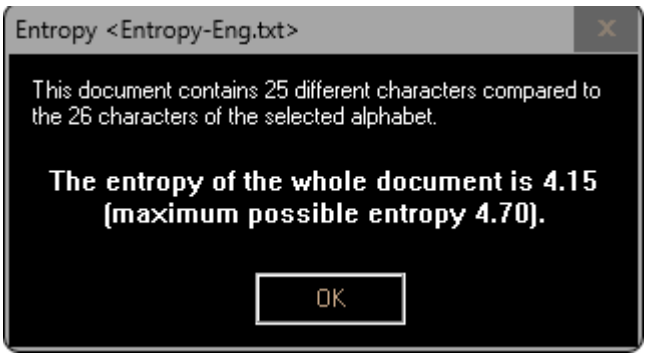
In this part I learned about the intro to Cryptool and some of the algorithms used.

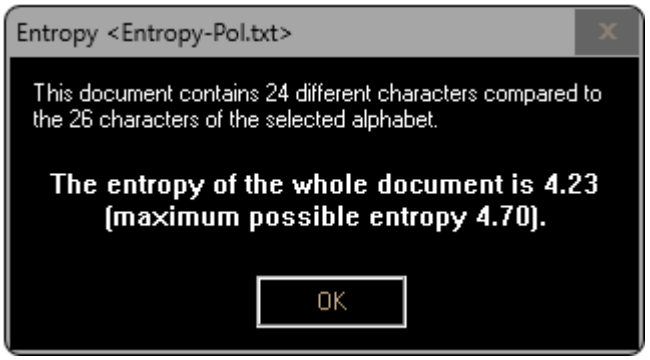
4. Property analysis of available algorithms.

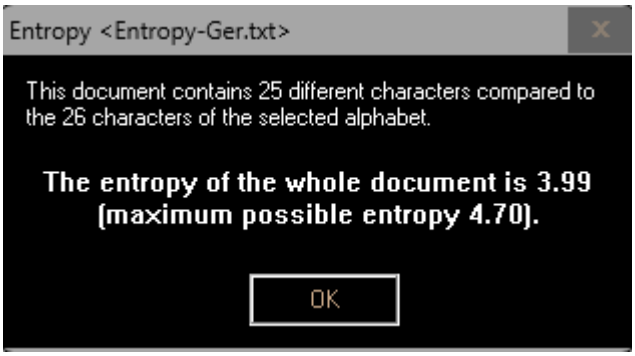
I. Tasks:

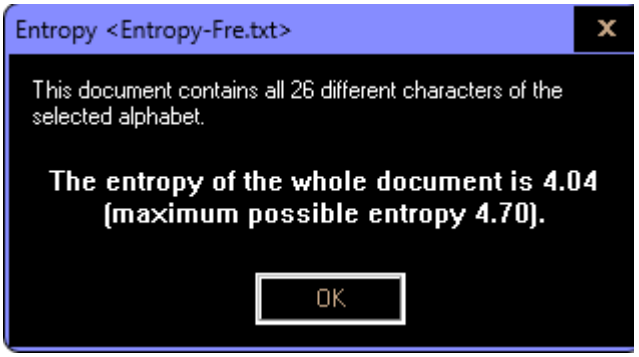
1. Compare the entropy values of plaintexts for different languages (English, Polish, German, French, Italian, Spanish, ...)

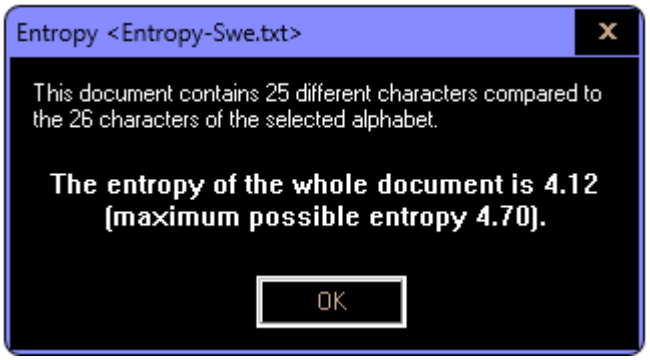
   - **I used William Shakespeare, Henry V as an example and translated to the given languages to make the entropy value.**











2. Compare the entropy values of plaintext and ciphertext depending on the algorithm. The entropy of the ciphertext for fixed plaintext should be calculated and compared when encrypting with algorithms:

- **To compare the entropy values I used only one text from the previous task which was English.**

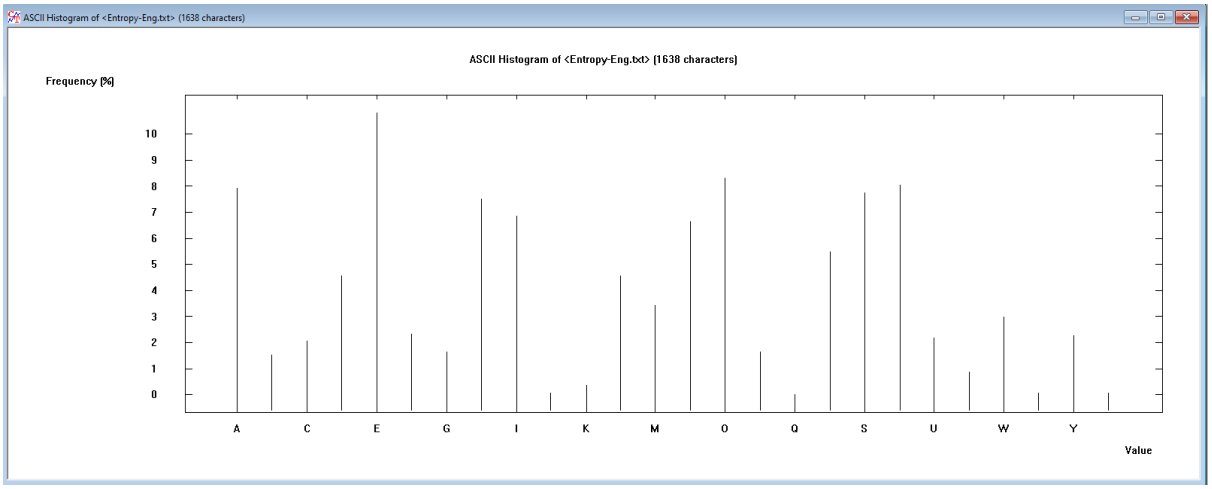
  >> Caesar , 4.15
  >> playfair, (key is “KEY”), 4.47
  >> adfgvx, (random matrix, transposition key, “Key”), 2.26
  >> homophones, 7.77 (max 8.0)

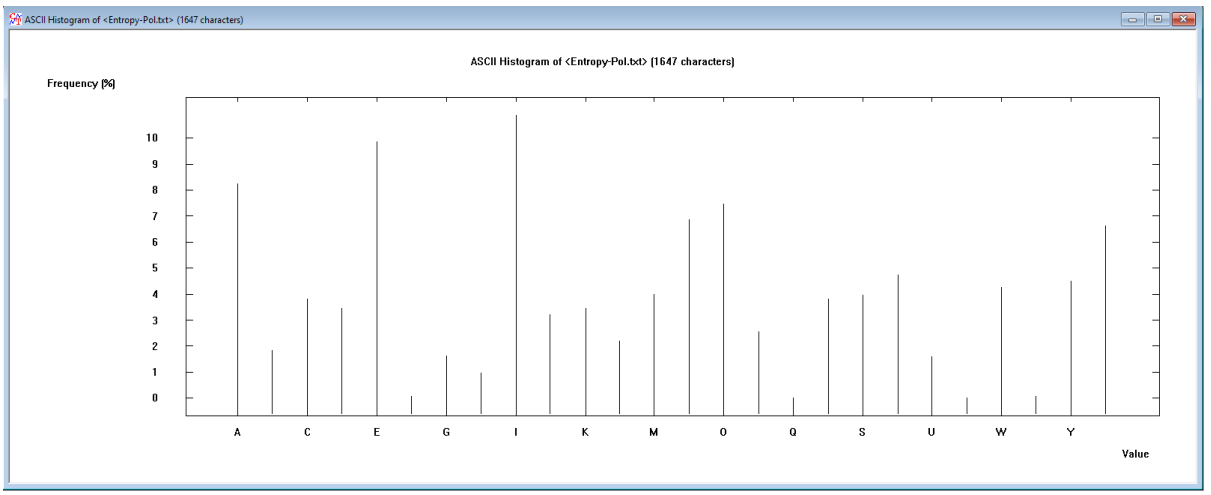
>> vigenere (for different key lengths), 4.54 with “KEY”
4.62 with “VIGENERE”

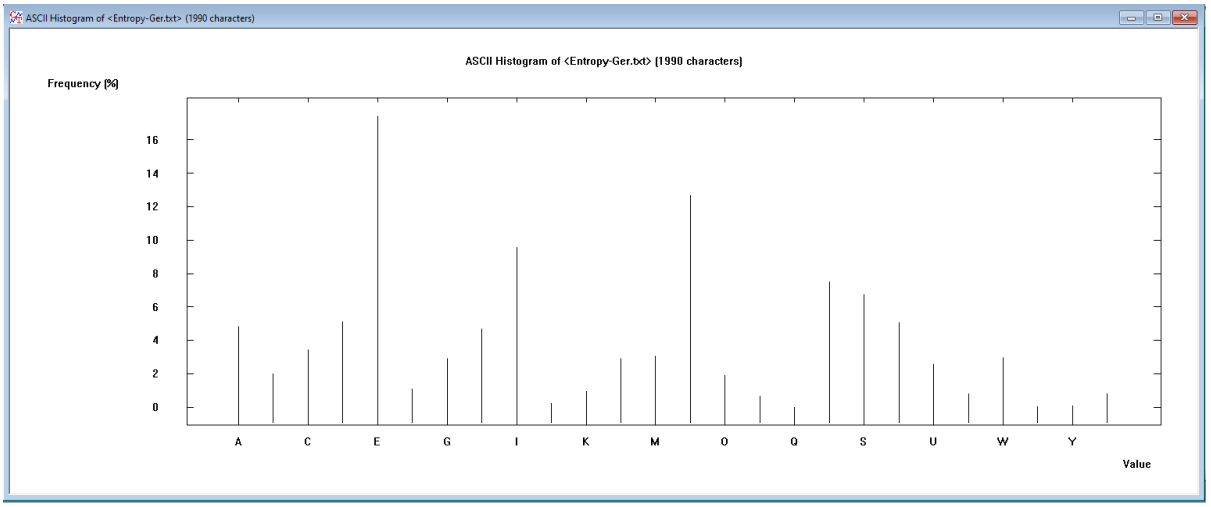
>> hilla (for different matrix sizes),
2*2 matrix with random key is 4.63
5*5 matrix with random key is 4.68

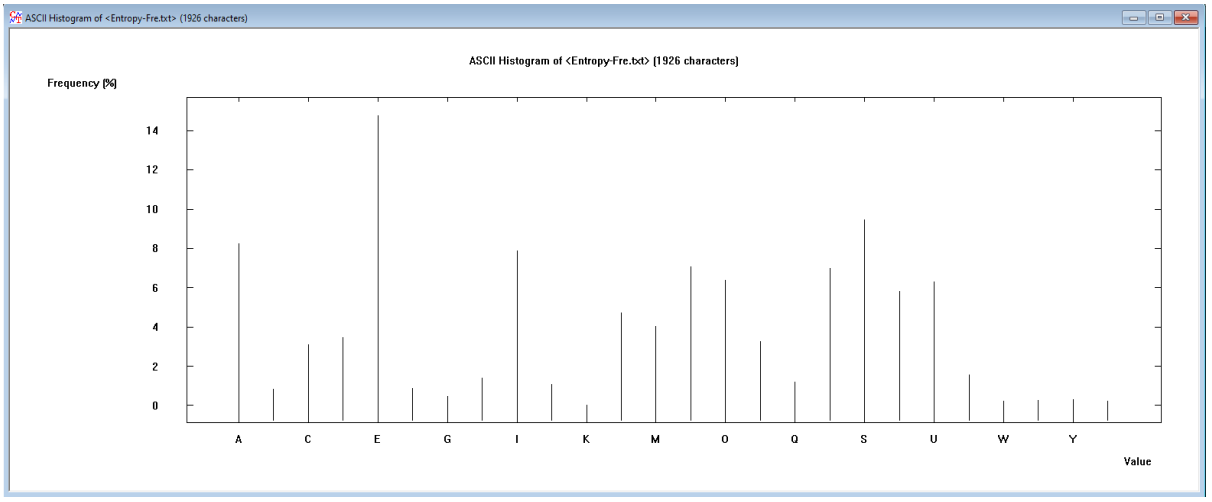
3. Compare histograms of plaintexts for selected 3 different languages (English, Polish, German, French, Italian, Spanish, ...) .

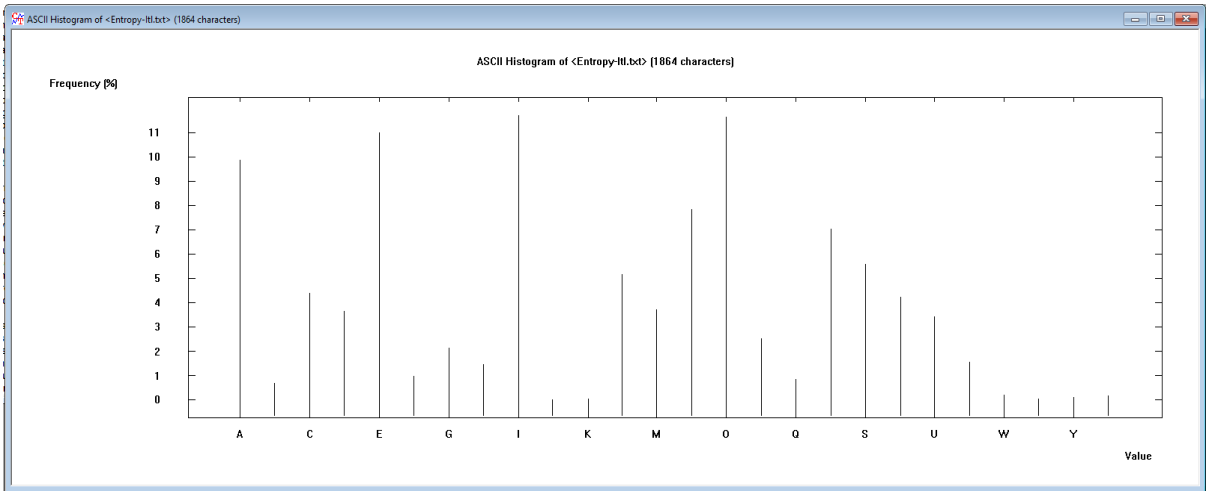
- **The histogram are in given order,**

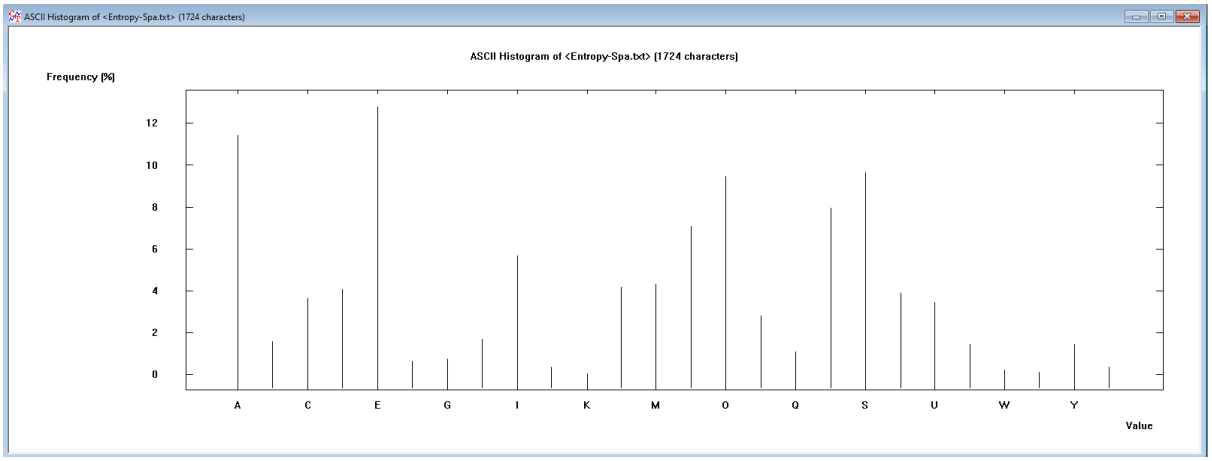




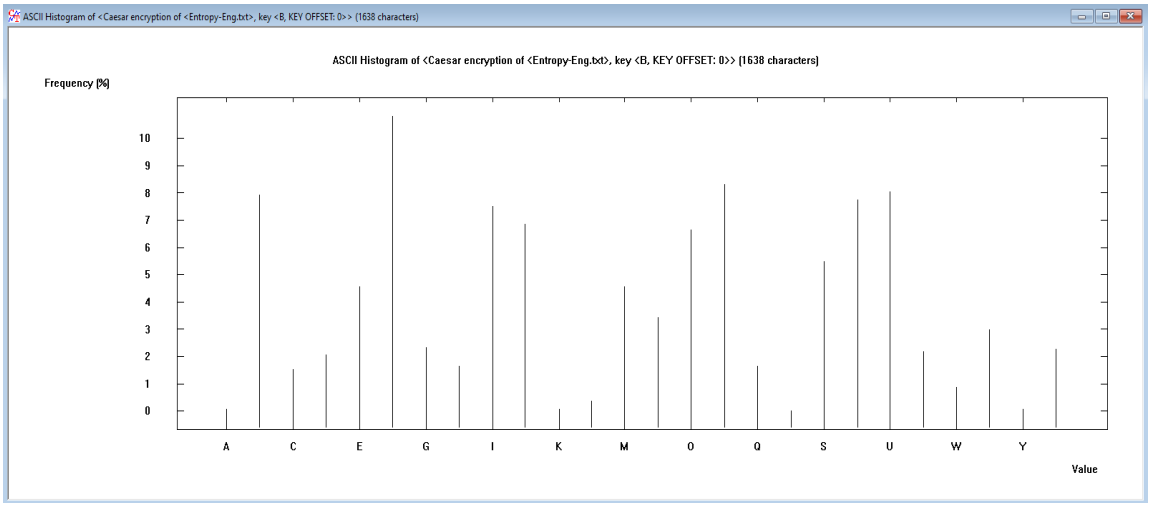


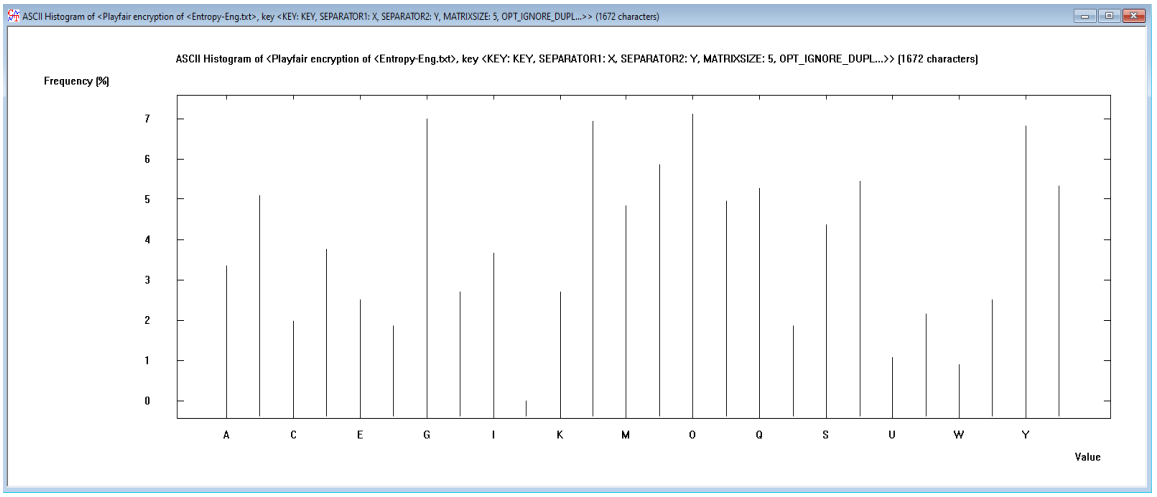


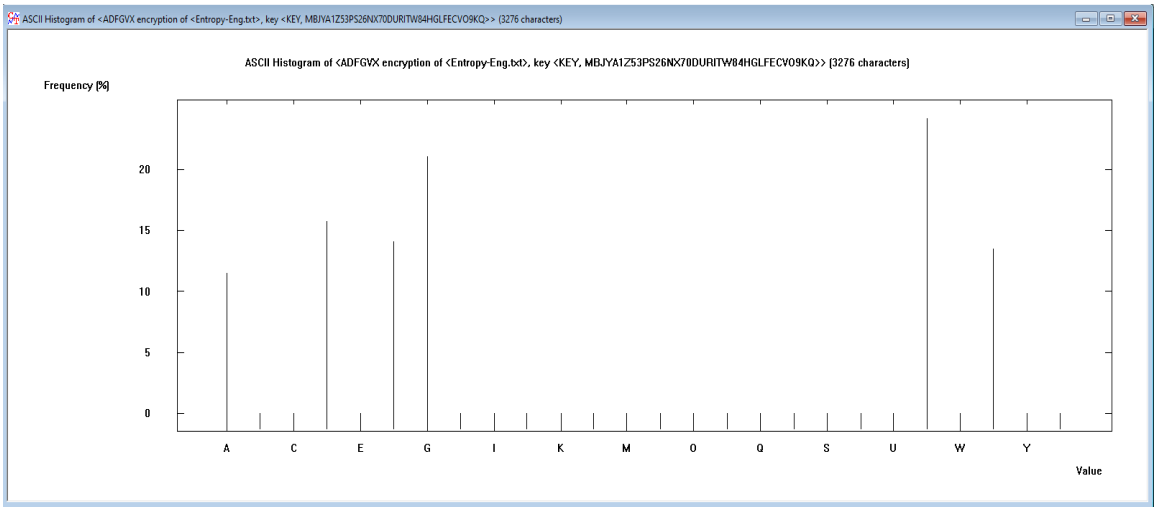




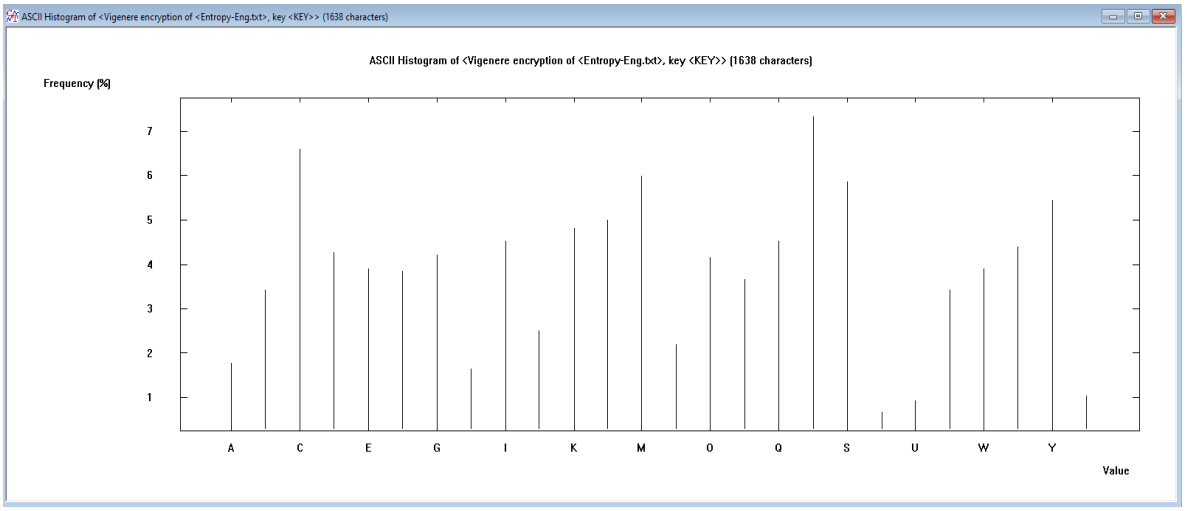
4. Compare histograms of plaintext and ciphertext depending on the algorithm. For algorithms such as in point 2.
   - **For this i am choosing only english text to compare the algorithms mentioned in point 2 one by one;**







— For Homophones if I click to find histogram after encryption the cryptool app crashes !!!



5. Compare n-grams (bi/tri/n) of plaintexts for selected 3 different languages (English, Polish, German, French, Italian, Spanish, ...)

- **From the previous tasks translated text, bi-grams of English, Polish, German;**

No.	Character seq...	Frequency in %	Frequency
1	TH	4.2010	51
2	HE	3.0478	37
3	AN	2.8007	34
4	IS	2.6359	32
5	HA	2.3064	28
6	HI	2.0593	25
7	ND	1.9769	24
8	OU	1.8946	23
9	IN	1.8122	22
10	SH	1.6474	20
11	LL	1.5651	19
12	RE	1.5651	19
13	NO	1.4827	18
14	OR	1.4827	18
15	AR	1.4003	17
16	AT	1.4003	17
17	ME	1.4003	17
18	AL	1.2356	15
19	EN	1.2356	15
20	ES	1.2356	15
21	ER	1.1532	14
22	ON	1.1532	14
23	OT	1.1532	14
24	TO	1.1532	14
25	WI	1.1532	14

No.	Character seq...	Frequency in %	Frequency
1	IE	5.3068	64
2	NI	3.8143	46
3	WI	2.4876	30
4	CZ	2.2388	27
5	TE	1.9900	24
6	NA	1.9071	23
7	JE	1.8242	22
8	ST	1.8242	22
9	AN	1.7413	21
10	TA	1.7413	21
11	DZ	1.5755	19
12	OW	1.4096	17
13	ZI	1.4096	17
14	ZY	1.4096	17
15	ZE	1.3267	16
16	GO	1.2438	15
17	SZ	1.2438	15
18	LI	1.1609	14
19	PO	1.1609	14
20	RZ	1.1609	14
21	ZA	1.1609	14
22	CH	1.0779	13
23	EG	0.9950	12
24	IA	0.9950	12

No.	Character seq...	Frequency in %	Frequency
1	EN	5.3756	83
2	ER	4.4689	69
3	CH	3.4974	54
4	IN	3.4326	53
5	EI	2.9145	45
6	ND	2.2021	34
7	IE	2.1373	33
8	DE	2.0725	32
9	ES	2.0725	32
10	SE	2.0725	32
11	NE	2.0078	31
12	GE	1.9430	30
13	IC	1.9430	30
14	TE	1.6839	26
15	BE	1.6192	25
16	AN	1.5544	24
17	IR	1.4249	22
18	WI	1.4249	22
19	NN	1.3601	21
20	UN	1.3601	21
21	HE	1.2306	19
22	WE	1.2306	19
23	ME	1.1658	18
24	RD	1.1658	18
25	DI	1.1010	17

tri-grams of English, Polish, German;

No.	Character seq...	Frequency in %	Frequency
1	THE	2.8360	23
2	AND	2.4661	20
3	HIS	2.0962	17
4	ALL	1.4797	12
5	HAT	1.4797	12
6	SHA	1.4797	12
7	DAY	1.2330	10
8	HAL	1.2330	10
9	THI	1.2330	10
10	NOT	1.1097	9
11	THA	1.1097	9
12	CRI	0.8631	7
13	FOR	0.8631	7
14	ISP	0.8631	7
15	RIS	0.8631	7
16	SPI	0.8631	7
17	MAN	0.7398	6
18	MOR	0.7398	6
19	OUR	0.7398	6
20	ARE	0.6165	5
21	ING	0.6165	5
22	ITH	0.6165	5
23	MEN	0.6165	5
24	OLD	0.6165	5
25	ORE	0.6165	5

No.	Character seq...	Frequency in %	Frequency
1	NIE	3.4279	29
2	DZI	2.0095	17
3	ZIE	1.6548	14
4	EGO	1.4184	12
5	OWI	1.4184	12
6	WIE	1.4184	12
7	CZY	1.1820	10
8	RZE	1.1820	10
9	EST	1.0638	9
10	STA	0.9456	8
11	ANI	0.8274	7
12	SPI	0.8274	7
13	TEN	0.8274	7
14	AMI	0.7092	6
15	IEK	0.7092	6
16	JEG	0.7092	6
17	JES	0.7092	6
18	PRZ	0.7092	6
19	TAN	0.7092	6
20	IAN	0.5910	5
21	TEM	0.5910	5
22	ACH	0.4728	4
23	AKI	0.4728	4
24	ALE	0.4728	4

No.	Character seq...	Frequency in %	Frequency
1	EIN	2.6572	30
2	ICH	2.5686	29
3	INE	1.6829	19
4	WIR	1.5943	18
5	UND	1.4172	16
6	CHT	1.3286	15
7	DIE	1.2400	14
8	SEI	1.2400	14
9	DER	1.0629	12
10	GEN	1.0629	12
11	SCH	1.0629	12
12	DEN	0.9743	11
13	ESE	0.9743	11
14	IRD	0.9743	11
15	CHE	0.8857	10
16	IES	0.8857	10
17	MEI	0.8857	10
18	ANN	0.7972	9
19	NEN	0.7972	9
20	STE	0.7972	9
21	TAG	0.7972	9
22	AND	0.7086	8
23	EHR	0.7086	8
24	HEN	0.7086	8
25	NIC	0.7086	8

6. Compare n-games (bi/tri/n) of plaintext and ciphertext depending on the algorithm. For algorithms such as in point 2.

7. Analyze the autocorrelation values of the ciphertext obtained by encryption XOR and Vigenere algorithms for different password lengths.

II. Questions/questions.

1. How do the observed parameters change?

- **Entropy changes depending on the randomness and the language as for different languages and parameters for the given text - frequency of individual alphabets changes.**

2. How can the text analysis tools available in CrypTool be used to determine the encryption algorithm for a given ciphertext?

- **This tool can be used to find the frequency of characters and find patterns to find encryption of algorithms.**

3. How can the text analysis tools available in the CrypTool program be used to determine the password used for encryption?

- **Did not advance that far to answer this question.**

Conclusion

This section I learned to do analysis with Cryptool and find patterns to understand a given encryption better.

5. Analysis of delivered files.

I. Tasks.

1. Try to recognize which encryption algorithm has been used based on an analysis of the ciphertext program only. All cryptograms were created based on the same plaintext. Files 1_X.txt,..,

— Caesar , Playfair, ADFGVX, Homophones, Vignere, Hilla

2. Try to decrypt by hand (using premises such as type algorithm, histogram, and other tools) or by using the automatic cipher-breaking option (Analysis -> Symmetric Encryption (Classic) -> Ciphertext-only) encryption programs placed in files 2_x.txt, ...

— Vignere, Substitution, Playfair

3. Recognize what encryption algorithm was used, and then try to decrypt the provided ciphertext programs. Files: 3_x.txt, ...

— Did not do.

II. Questions.

1. What does the strength of the algorithm depend on?

- **I would say KEY.**

2. How can the encryption strength be increased for known ciphers?

- **Not using common `key` & giving a complex input for KEY**