Teaching and Promoting Cryptology at Faculty of Science University of Hradec Králové

Michal Musílek
Faculty of Science
University of Hradec Kralove
Rokitanskeho 62, Hradec Kralove
michal.musilek@uhk.cz

Štěpán Hubálovský
Faculty of Science
University of Hradec Kralove
Rokitanskeho 62, Hradec Kralove
stepan.hubalovsky@uhk.cz

Abstract

University of Hradec Králové is one of the smaller public higher education institutions in the Czech Republic. At present, there are about 7,000 students studying here. Even though at any of their faculties it is not possible to study a study program closely focused on cryptology, we pay attention to the historical and modern cryptology at the Faculty of Science in several subjects. Basic concepts, principles and methods of modern computer cryptology form an important part of the subject Computer and Data Protection. Principles of encrypting, decrypting and deciphering of basic substitution and transposition ciphers and the history of cryptology have become part of the subjects of Computer Science and Structured Programming in various degrees and with different approaches.

1 Introduction

Overview of the history of cryptology is included in the subject History of Computer Science as an expanding and enlightening of the curriculum. Special attention is paid to rotate encryption machines such as Enigma, the Lorenz SZ42 cryptographic telegraph and to machines used to break them as well - Turing's bomb, the first British computers, Colossus Mark 1 and 2 (Boone, 2005; Singh, 2000). Part of the exercises in the subject History of Computer Science is devoted to the encryption, decryption and deciphering of some basic types of hand ciphers, because this course also has practical lessons in which students learn, for example, to multiply numerical numbers using a counter or to perform calculations on a logarithmic ruler.

Different way to learning of cipher is used in the subject of Programming subject. The cipher system is used to enter interesting and motivating programming tasks during Structured Programming course. In addition to cryptology, attention is also paid to the problems associated with wireless transmission and plain coding of information (Musílek, 2012; Musílek, Hubálovský, & Hubálovská, 2017).

2 Theoretical Background

A deeper insight into the issue of cryptology is then made possible for students who are interested in this topic within realization of the bachelor as well as diploma thesis. The Department of Cybernetics of the Faculty of Science of the University of Hradec Králové, with the two authors of this article, prepares future lower and upper secondary schools teachers of Informatics in the Czech Republic. Teacher preparation is always carried out for two teaching subjects and significant space is devoted not only to these two areas, but also to basic pedagogical-psychological, general and subject didactics preparation. Interestingly, although our students are studying many different combinations of subjects, interest in historical ciphers has so far been shown only by students of two different combinations, namely Informatics - Mathematics and Informatics - History.

The individual bachelor theses of the students of the program Mathematics and Informatics with a focus on education had following topics:

Proceedings of the 1st Conference on Historical Cryptology, pages 137– 143,
Uppsala, Sweden, 18-20 June, 2018

"Deciphering of substitution ciphers with computer support" (Procházka, 2012), "Deciphering of ciphers with computer support" (Hanzalová, 2014; Hájková, 2015). The diploma theses in the follow-up master's program Teaching of Mathematics and Informatics for Secondary Schools was oriented more didactically with following topics: "Ciphers as motivation in the teaching of algorithms and programming" (Bukáček, 2013); "Board games, puzzles, anagrams and ciphers as motivation in the teaching of algorithms and programming" (Procházka, 2014); "Fundamentals of cryptology as a teaching topic in subject "Informatics" at lower secondary school" (Hájková, 2017). The topics of bachelor's thesis in the study program Informatics and History were "Computer Analysis of Encrypted Correspondence of House of Piccolomini" (Vlnas, 2017) and "History of ciphering of transposition ciphers with computer support" (Musílek, 2017).

Thesis focused on specific historical clues used by Marshal Ottavio Piccolomini during the Thirty Years War links the work of the historian - investigator in the archive - with the approach of informatics. The routine decryption algorithms were realized in macros in Visual Basic for Applications in a MS Excel spreadsheet. This thesis was evaluated by the Dean of the Faculty of Science of the University of Hradec Králové for the best bachelor thesis in the study program Informatics.

3 Samples of advancement of students

In the following text, we will discuss in detail two bachelor's theses: work for the automatic analysis of text encoded by monoalphabetic substitution based on the trigram frequency analysis (Hanzalová, 2014) and work analyzing real historical ciphers of the early 17th century (Vlnas, 2017).

3.1 Using spreadsheet in cryptanalysis of short cipher text

The authors of the paper suggest new method and algorithm for deciphering and automation analysis of the ciphered monoalphabetic text. The method generalized frequency analysis of the bigram saved in a two-dimensional array to frequency analysis of the trigrams saved in a three-dimensional array with 26 x 26 x 26 = 14576 elements.

The algorithm for automatic deciphering of short simple substitution cipher text is similar to algorithm for deciphering of long ciphered text.

The algorithm for deciphering of long ciphered text is based on method of evaluation of frequency of pairs of consecutive letters and compares it with the frequency of bigrams of reference text using the evaluation function:

$$f = \sum_{i=1}^{26} \sum_{j=1}^{26} \mid D_{ij} - E_{ij} \mid \qquad (1)$$

where E_{ij} is matrix of bigrams of reference text and D_{ij} is matrix of bigrams of ciphered text.

Similarly, the algorithm for deciphering of short ciphered text is based on automatic analysis of evaluation function *f* of trigrams:

$$f = \sum_{i=1}^{26} \sum_{j=1}^{26} \sum_{k=1}^{26} \mid D_{ijk} - E_{ijk} \mid \qquad (2)$$

where E_{ijk} is matrix of trigrams of reference text and D_{ijk} is matrix of trigrams of ciphered text.

The algorithm consists o three relatively independent sub-procedures (Hanzalová, Hubálovský, & Musílek, 2012).

The first sub-procedure *Frequency* creates three-dimensional reference matrix ***E*** that corresponds to the frequencies of letters in the reference text.

The second sub-procedure *Trigrams* creates three-dimensional matrix ***D*** of relative frequencies of the trigrams of the cipher text. The part of the matrix ***D*** is shown on the Figure 2. In next step (Hanzalová, Hubálovský, & Musílek, 2012) the procedure evaluates the compliance with the reference text by using an evaluation function:

$$f = \sum_{i=1}^{26} \sum_{j=1}^{26} \sum_{k=1}^{26} \mid D_{ijk} - E_{ijk} \mid \qquad (2)$$

The third sub-procedure *Exchange* provides exchange of two *x*-vectors, two *y*-vectors and two *z*-vectors of three dimensional matrix ***D*** and creates new matrix ***D'***. The vectors are exchanged in the order of the frequencies of the letters in the cipher text from the most frequent to the least frequent based on following rules - see Hanzalová, Hubálovský, & Musílek (2012):

- the *x*-vector corresponding to the order of the first exchanged character is replaced by the x-vector corresponding to the second exchanged character;
- then *y*-vector corresponding to the order of the first exchanged character is replaced by the y-vector corresponding to the second exchanged character;
- then *z*-vector corresponding to the order of the first exchanged character is replaced by the y-vector corresponding to the second exchanged character;

After each substitution a new matrix ***D'*** is obtained, and the evaluation of the compliance of the relative frequency of the trigrams in the cipher text and the reference text is obtained using the evaluation function:

$$f' = \sum_{i=1}^{26} \sum_{j=1}^{26} \sum_{k=1}^{26} \mid D'_{ijk} - E_{ijk} \mid \quad (3)$$

After each substitution the values f and f' are compared and if $f' < f$, the procedure immediately stops the process of the letter substitution and the exchange in the conversion table is proposed, which will improve the compliance (lowering the value of evaluation function *f*). Finally, the sub-procedure will creates a new matrix ***D*** of relative frequencies of the trigrams of the cipher text, and it will provide a new assessment of compliance with the reference text using the evaluation function (3).

The sub-procedure *Frequency* is run only once at the beginning of the program to set the appropriate initial conditions. The sub-procedures *Trigrams* and *Exchange* are run alternately.

Above mentioned sub-procedures were realized in Visual Basic for Application in MS Excel Spreadsheet. Deciphering based on trigrams' analysis has been studied in cipher text with the length in the range from 200 to 500 characters. It was proved that algorithm for deciphering of short simple substitution cipher text based on automatic analysis of the trigrams enables decryption almost without manually performed exchanges

3.2 Computer Analysis of Encrypted Correspondence of House of Piccolomini

Bachelor thesis titled Computer Analysis of Encrypted Correspondence of House of Piccolomini interconnects the two author's study areas (Vlnas, 2017). The author is a student of Informatics and History in education. Archives, especially archives of aristocratic families whose members held important state, military or diplomatic positions, contain, in addition to open documents, very often-encrypted documents.

The analysis of archive ciphertexts is a complex task. Standardly, the task is necessary to do in large part by hand, such as recognizing different forms of written fonts (different types of ancient shape handwritings or special cipher characters), counting frequencies of individual characters, etc. Some monotonous tasks can be performed a computer. The author of the bachelor thesis had appropriately used custom-made macros in Visual Basic for Applications to decrypt encrypted texts transcribed into Excel spreadsheets. After identifying a given cipher system and determining the transmission table, it is a purely mechanical matter. Macro combined both basic principles used to make mono-alphabetic substitution, i.e. simple swapping and supposed words, and was created in such a way that allow the gradual uncovering of the spreadsheets that effectively supports cipher shredders in the phase of stepwise reconstruction of the encryption table.

The first phase of the work was to obtain appropriate cipher texts. The State Regional Archive of Zámrsk offers to researchers a family archive of the genus Piccolomini on microfilms. It has advantages and disadvantages. The advantage is the possibility of fast document browsing, where the scrolling of the microfilm in the reader can be significantly faster than working with original archives. The disadvantage is the lower contrast and the overall loss of quality and the worse possibility of photographic documentation. The student focused on documents related to the person of Ottavio Piccolomini and his activities during the Thirty Years' War. He found two microfilms with a number of cryptic texts, partially decrypted, apparently immediately after receiving the addressee, but partly un-decrypted. The student took photographs of the corresponding microfilm images. He thought he had captured only a small

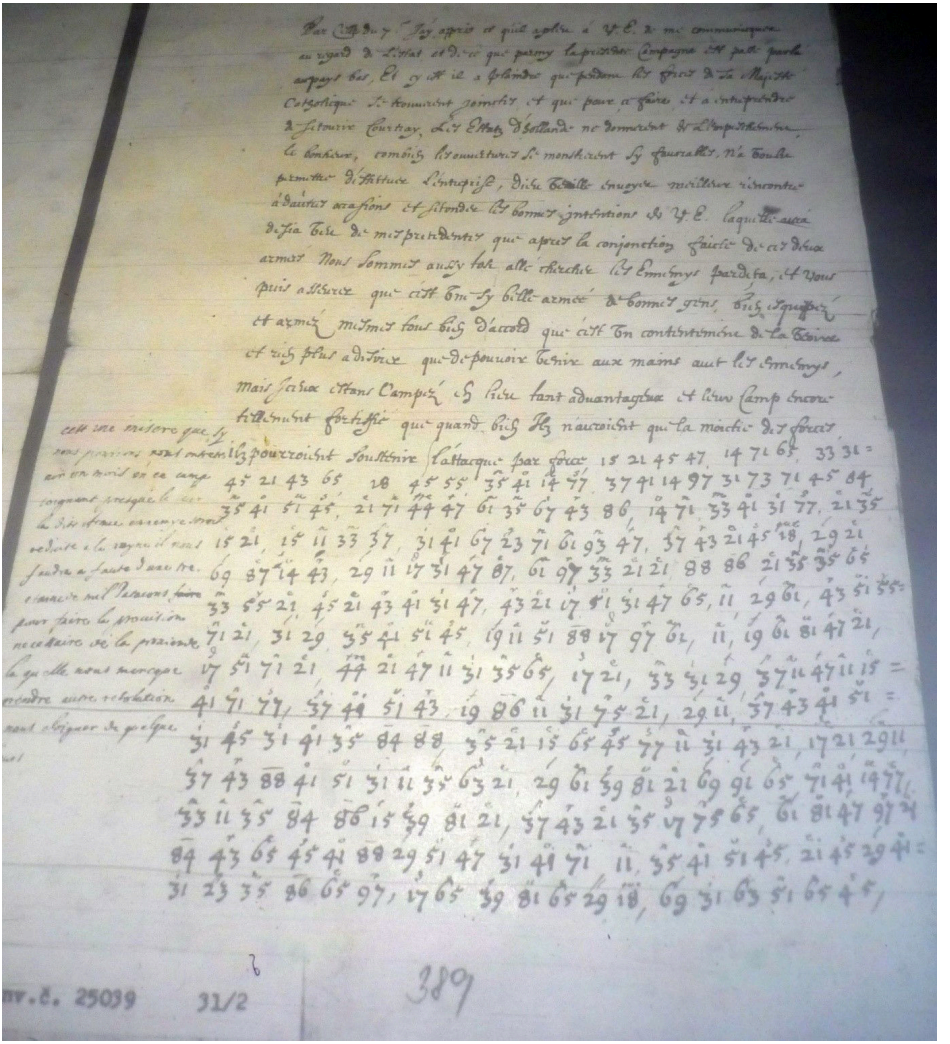

Figure 1: A homophone substitution cipher - pair of digits represents a letter or null character

part of Ottavio Piccolomini's cipher correspondence. The archive contains large number of the ciphertext than cannot be found within one business day.

Firstly, the document number 25039 was analyzed, see Figure 1. Part of this letter was written in plain text and part was written in numerical code. Some letters of the alphabet have been written above the two-digit codes. It allows reconstructed the decryption conversion table - see Table 1. The entire cipher text, composed only of digits, was read quite well. It was certainly easier to read digits then to read handwritten characters in shapes which had been written in individual manuscripts of many different scribes. This text was copied in the cell of Excel spreadsheet. Also deciphering table was placed in the same sheet. The deciphering table was step by step filled in and simultaneously was used for decrypting of prepared ciphertext. It was found that the open text in French was encrypted by simply replacing the alphabet letters with two-digit numbers, but with the use of vowel homophones and special codes for some selected short words, supplemented by several null

characters. It is a simple nomenclator. Partially decrypted text was a significant help, however, the use of macros in Visual Basic for Application significantly simplified the complete decryption of the encrypted part of the letter.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
1	a		b	u	C		d	que	f
2	e		g	et	L		a	e	l
3	i		m	sko	n		p		q
4	o		r	tre	s		t	m	o
5	u				y		s		
6	a	a	g		e		i		l
7	n		o		r		s	rc	t
8	u			*		*	e	*	
9	e		n				r		

Table 1: Encryption table of the document 25039
(* = null characters)

„C'est une misère que si nous puissions nous entretenir un mois dans ce camp Xabucd que fais-je là et lim skonpqor très tmouysaageilnorsrctu et en joignant presque le leur, ladite armée, ennemie seroit réduite à la ruine, il nous faudra, à faute d'une trentaine de mil patacons pour faire la provision nécessaire de la prouiange laquelle nous manque, que prendre autre résolution à nous? sloigcer de quelques ligues d'ici où nous puissions muuerles dit su iures et fourrage à fin de ne voir ce que dit une plais est réduit et cette belle armée en teees état misérable comme elle samuue du passé au camp skoprocle Bérenburg. En quoi consiste néanmoins la conservation ou perte du reste de l'empire gedlusrc?“

The similar system, based on a square table, with otherwise delimited letters, was used in several other cipher letters. From the cryptanalysis point of view, the letter with non-encrypted cipher sections consisting of letters and numbers was more interesting. This letter was included in document number 24873, see Figure 2. The cipher was taken by standard methods, i.e. by frequency analysis and predicted words. The plain cipher text was in Italian. Even this cipher contained the null characters represented by all even digits (2, 4, 6, 8). The cipher does not contain any homophones. If these null characters were omitted, the very simple monoalphabetic substitution cipher was reached - the consonants did not replace and the vowels, were successively replaced by the numbers 1, 3, 5, 7, 9. As soon as the eyes of the solver became used to the Italian handwriting of the first half of the 17th century, reading the text was relatively simple. Fig. 3.

b	c	d	f	g	h	j	k	l	m	n	p	q
b	c	d	f	g	h	j	k	l	m	n	p	q

r	s	t	v	1	2	3	4	5	6	7	8	9
r	s	t	v	a	*	e	*	i	*	o	*	u

Table 2: Encryption table of the document 24873
(* = null characters)

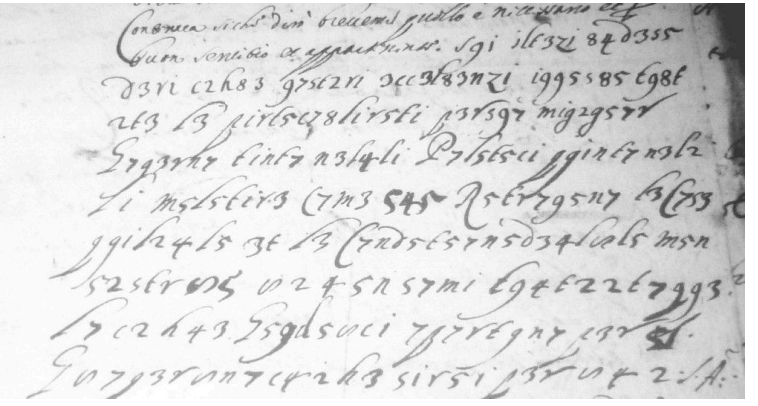

Figure 2: An example of a cipher text (cut-off from document 24873) that was not decrypted by the recipient

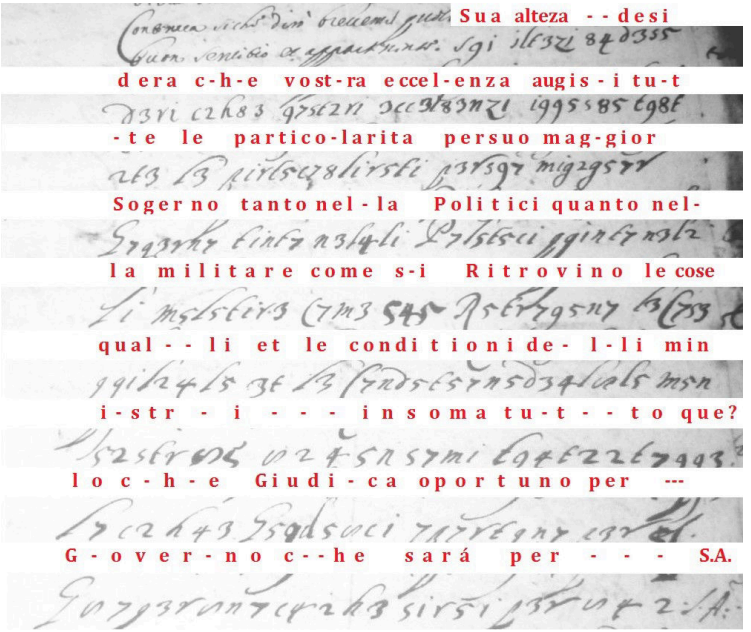

Figure 3: Sample of decryption of the text from the Figure 2 (cut-off from document 24873)

4 Conclusion

It is pleasing that some of the students of the Faculty of Science of the University of Hradec Králové, supervised by the authors of this paper, contributed to solution of tasks of historical cryptology. The important is the fact that not only students mentioned above, but also many other graduates of study program the teaching of informatics or also teaching mathematics or history will to motivate the secondary school students by examples of cipher systems, or by historical events that were affected by revealing of secret correspondence or by breaking of cipher systems.

Acknowledgments

The research was supported by Specific research project at Faculty of Science, University of Hradec Kralove, 2018.

References

Boone, J. V. 2005. *Brief History of Cryptology.* Annapolis: Naval Institute Press. 192 p. ISBN 1-59114-084-6.

Bukáček, D. 2013. *Ciphers as motivation in the teaching of algorithms and programming.* Diploma Thesis at Faculty of Science University of Hradec Králové. Supervisor . 118 p.

Hájková, S. 2015. *Deciphering of transposition ciphers with computer support.* Bachelor Thesis at Faculty of Science University of Hradec Králové. Supervisor . 41 p.

Hájková, S. 2017. *Fundaments of cryptology as teaching topic in subject "Informatics" at lower secondary school.* Diploma Thesis at Faculty of Science University of Hradec Králové. Supervisor . 74 p.

Hanzalová, P., Hubálovský, Š., & Musílek, M. 2012. Automatic cryptanalysis of the short monoalphabetic substituted cipher text. In *Proceedings of the 5th WSEAS International Conference on Visualization, Imaging and Simulation*. Sliema, Malta: Wseas Press. p. 199-204. ISBN: 978-1-61804-119-7.

Hanzalová, P. 2014. *Using of spreadsheet in cryptanalysis of short cipher text.* Bachelor Thesis at Faculty of Science University of Hradec Králové. Supervisor . 48 p.

Hubálovský, Š., & Musílek, M. 2010. Automatic cryptanalysis of the monoalphabetic substitution

as a method of the system approach in the algorithm development thinking. *International journal of applied mathematics and informatics*. 4 (4), 92-102. ISSN 2074-1278.

Hubálovský, S., & Musílek, M. 2014. Algorithm for Automatic Deciphering of Mono-Alphabetic Substituted Cipher Realized in MS Excel Spreadsheet. *Applied Mechanics and Materials*. p. 624-627

Musílek, M., Hubálovský, Š., & Hubálovská, M. 2017. Mathematical Modeling and Computer Simulation of Codes with Variable Bit-Length. *International Journal of Applied Mathematics and Statistics.* 56 (1), 1-12. ISSN 0973-7545.

Musílek, M. 2012. Morse telegraph alphabet and cryptology as a method of system approach in computer science education. In *Proceedings of 9th International Scientific Conference on Distance Learning in Applied Informatics (DIVAI)*. Štúrovo, Slovakia: Wolters Kluwer. p. 223-231. ISBN 978-80-558-0092-9.

Musílek, P. 2017. *History of ciphering of transposition ciphers with computer support.* Bachelor Thesis at Faculty of Science University of Hradec Králové. Supervisor . 58 p.

Procházka, L. 2014. *Board games, puzzles, anagrams and ciphers as motivation in the teaching of algorithms and programming.* Diploma Thesis at Faculty of Science University of Hradec Králové. Supervisor . 75 p.

Procházka, L. 2012. *Deciphering of substitution ciphers with computer support.* Bachelor Thesis at Faculty of Science University of Hradec Králové. Supervisor M. Musílek. 37 p.

Singh, S. 2000. *The code book: the science of secrecy from Ancient Egypt to quantum cryptography.* New York: Anchor Books. 411 p. ISBN 0-385-49532-3.

Vlnas, V. 2017. *Computer Analysis of Encrypted Correspondence of House of Piccolomini.* Bachelor Thesis at Faculty of Science, University of Hradec Králové. Supervisor M. Musílek. 58 p.