An Extended Version of the Polybius Cipher

Tabu S. Kondo
Computer Technologies and Applications
The University of Dodoma
Dodoma, Tanzania

Leonard J. Mselle
Computer Science
The University of Dodoma
Dodoma, Tanzania

ABSTRACT

This paper provides a security method which can be used for data that contains alphabets, numerals and some special symbols during their transmission. A discussion about cryptology and the existing Polybius cipher is made. The existing Polybius cipher is based on the use of a 5X5 matrix of letters constructed using numbers from 1 to 5. This square can allow the text that contains alphabets only. For this reason, we have proposed an improvement to the existing Polybius cipher, in which an 8X8 matrix can be constructed.

General Terms

Encryption, Decryption, Cryptography, Cryptology, Security

Keywords

Polybius square, Polybius cipher

1. INTRODUCTION

There are three basic approaches to securing information: prevention, restriction, and cryptography. Access to information can be prevented. If an attacker cannot access information, then the information is safe from the attacker. For network security, isolated networks and restrictive architectures are generally sufficient deterrents [1].

When networks cannot be isolated, access can still be restricted. Most remote login systems require a username and password. Different types of authentication exist for different security needs. Although authentication restricts access, it generally does not hinder eavesdropping-related attacks [1].

Cryptography is the most common approach for securing networks. Cryptography encodes data so that only the intended recipients can decode the message. Cryptographic systems include random number generators, hashes, ciphers, and encryption algorithms [1].

Information which is encrypted remains secure even when it is transmitted over a network that does not provide strong security even if the information is publicly available. In most versions of the UNIX operating system, for example, the file containing user passwords stores those passwords in encrypted form. Encryption protects these passwords effectively, to the point that if somebody does access the file, encryption would make it very difficult for an attacker who obtained the file to be able to decipher the passwords [2].

Whether you realize it or not, there are a lot of ways that you deal with some form of encryption every day. As businesses now rely heavily on the internet and other forms of networks to buy, sell, organize, inform, provide services, and form alliances, they also have to deal with the fact that sometimes these networks are transmitting very sensitive data. Some businesses decide on their own that protection of this data is a good thing, and others have either learn that through bad experiences or have to comply with new laws that deal with the protection of personal data [3].

Computers have become so insidious that many of us don’t even realize sometimes that we are interacting with them. Most of these systems are encrypting the data as it goes across the wires [3].

2. OVERVIEW OF CRYPTOLOGY

Cryptology is the science of secret communication. It has two main subfields (i.e. cryptography and cryptanalysis). Cryptography is the science of creating secret codes; Cryptanalysis is the science of breaking codes. These two aspects are closely related; when creating a secret code the analysis of its security plays an important role. There are five pillars of cryptology [4]:

- Confidentiality: keep communication private.
- Integrity: detect unauthorized alteration to communication.
- Authentication: confirm identity of sender.
- Authorization: establish level of access for trusted parties.
- Non-repudiation: prove that communication was received.

Cryptography itself splits into three main parts: Symmetric ciphers, Asymmetric ciphers and Cryptographic protocols.

Symmetric ciphers are what many people assume cryptography is about: two parties have an encryption and decryption method for which they share a secret key [5].

Symmetric ciphers also referred to as conventional encryption or single-key encryption was the only type of encryption in use prior to the development of public-key encryption in the 1970s. It remains by far the most widely used of the two types of encryption [6].

Asymmetric ciphers are the one in which encryption and decryption is performed using the different keys: a public key and a private key. It is also known as public-key encryption [6].

Cryptographic protocols are protocols which deal with the application of cryptographic algorithms. Symmetric and asymmetric algorithms can be viewed as building blocks with which applications such as secure internet communication can be realized. The Transport Layer Security (TLS) scheme, which is used in every web browser, is an example of a cryptographic protocol [5]. Figure 1 depicts the cryptology tree.

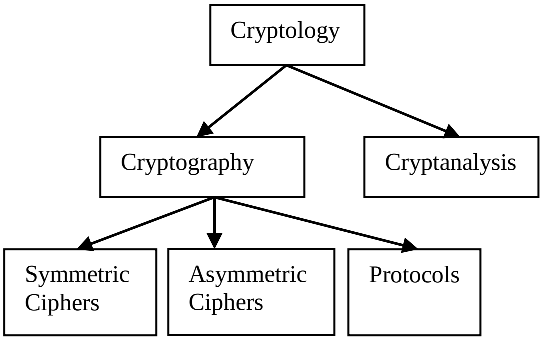

Fig 1: Cryptology Overview

3. IMPORTANT TERMS

3.1 Ciphers and Cryptosystems

The transformation represented by an encryption function and its corresponding decryption function is called a cipher [7]. Ciphers have been used in sharing information privately by government officials, military officers, spies, ambassadors, revolutionaries, business owners, religious leaders, and more.

A cryptosystem is the combination of three elements: the encryption engine, keying information, and operational procedures for their secure use. In other words, almost every encryption program can be considered a cryptosystem because it has everything together in one package [3].

3.2 Transposition and Substitution Ciphers

A transposition cipher rearranges the characters in the plaintext to form the ciphertext. The letters are not changed [8].

A substitution cipher changes characters in the plaintext to produce the ciphertext [8].

3.3 Stream and Block Ciphers

In a stream cipher, encryption and decryption are done one symbol (such as a character or a bit) at a time [9].

In a block cipher, a group of plaintext symbols of size m (where m > 1) are encrypted together, creating a group of ciphertext of the same size. Based on the definition, in a block cipher, a single key is used to encrypt the whole block even if the key is made of multiple values. In a block cipher, a ciphertext block depends on the whole plaintext block [9].

3.4 Confusion and Diffusion

An encrypting algorithm should take the information from the plaintext and transform it so that the interceptor cannot readily recognize the message. The interceptor should not be able to predict what will happen to the ciphertext by changing one character in the plaintext. We call this characteristic confusion. An algorithm providing good confusion has a complex functional relationship between the plaintext/key pair and the ciphertext. In this way, it will take an interceptor a long time to determine the relationship between plaintext, key, and ciphertext. It will therefore, take the interceptor a long time to break the code [10].

The cipher should also spread the information from the plaintext over the entire ciphertext so that changes in the plaintext affect many parts of the ciphertext. This principle is called diffusion, the characteristic of distributing the information from single plaintext letters over the entire output. Good diffusion means that the interceptor needs access to much of the ciphertext to be able to infer the algorithm [10].

3.5 Fractionating Ciphers

Fractionation is a method of splitting alphabets, numerals and special symbols so that each plaintext symbol is represented by two or more ciphertext symbols. For example, "a" could be represented by "01", "b" by "02", “c” by “03” and so on.

4. THE EXISTING POLYBIUS SQUARE

The Ancient Greek historian Polybius (203-120 BC), being responsible with the operation of a “telegraph” used to send at distance messages, invented a substitution cipher, known since that as Polybius square. The letters of the Latin alphabet (26) are arranged in a square of size 5X5 as shown in table 1. The letters I and J are combined in a unique character because the choice between them can be easily decided from the text meaning. The encryption consists of replacing each letter with the corresponding pair of numbers (the line and column crossing point) [11]. For instance, D is 14, K is 25 and Q is 41. Hence, the letters are plaintext and the numbers are ciphertext. To decrypt a message one has to find the letter that intersects the specified row and column. For instance, 14 is D, 25 is K and 41 is Q. Polybius’ substitution cipher has found great acceptance among cryptographers up to modern times, who have used it as the basis for numerous ciphers [12].

Table 1. The Polybius square

	1	**2**	**3**	**4**	**5**
1	A	B	C	D	E
2	F	G	H	I/J	K
3	L	M	N	O	P
4	Q	R	S	T	U
5	V	W	X	Y	Z

5. LIMITATIONS OF THE EXISTING POLYBIUS SQUARE

- The existing Polybius square has no key, and thus is easily broken.
- The Existing Polybius square works well on the plaintext message containing alphabets. This means that it cannot be used to encrypt a plaintext message containing numerals and special symbols. For example, the blank, or word separator, is not encoded. If you can't encode a blank, the phrase NOW IN can be decoded as NO WIN. This is a potential misread.
- The existing Polybius square is based on the use of a 5X5 matrix of letters. The 5X5 matrix can only allow 25 characters, hence the letters I/J count as one. If a plaintext having the letter I/J is encrypted, when the resultant ciphertext at the receiving end is decoded the receiver will fail to distinguish I from J. This can lead to the distortion of the original message. Some care therefore, needs to be exercised when decrypting the message to make sure the right letter is always used.

To overcome these limitations we are proposing an extension to the existing Polybius square in which we propose to:

- Introduce the idea of a key to make it more secure.
- Increase the size of the matrix in which we will include numerals and special symbols.

- Separate the letters I and J.

6. DESCRIPTION OF THE EXTENDED POLYBIUS SQUARE

The extended Polybius square proposed in this work is based on the use of an 8X8 matrix of alphabets, numerals and special symbols constructed using a keyword. The matrix is constructed by filling in the letters or numbers or special symbols of the keyword without repetition from left to right and from top to bottom; and the filling in of the remainder of the matrix with the rest of the letters of the alphabet in alphabetical order. Numerals are filled in ascending order form 0 to 9 and special symbols in the order in which they appear in the ASCII table as shown in Table 2.

If some or all of the numerals from 0 to 9 are not used as part of the keyword, they can be placed to the next cells of the last letter of the alphabet in an ascending order. Furthermore, if some or all of the special symbols are not used as part of the keyword, they can be placed from the next cells of the last numeral if and only if some of the numerals are not used as part of the keyword. In such case we have not counted I/J as one letter instead we are placing both I and J in two different cells in order to avoid the ambiguity to the user at the time of decipherment.

The extended Polybius square can accommodate the plaintext containing alphabets, numerals and special symbols. The user can easily encrypt alphabets, numerals and special symbols efficiently. Thus, the plaintext containing email address, date of birth, house numbers and other numerals and special symbols can be easily and efficiently encrypted using the extended Polybius square.

In the proposed extended Polybius square, we reorder the alphabets in the same way as we did for the traditional Polybius square before we put it in the grid. Moreover, we arrange the numerals in ascending order and special symbols in the order in which they appear in the ASCII table. That is, we use the letters of the alphabet or numerals or special symbols of the keyword first, ignoring any repeat. So using a keyword of POLY2013 we get the extended Polybius square as shown in Table 2. Note that, since the matrix size of the proposed extended Polybius square is 8X8, the minimum key leghth is proposed to be 8.

Table 2. The extended Polybius square

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1	P	O	L	Y	2	0	1	3
2	A	B	C	D	E	F	G	H
3	I	J	K	M	N	Q	R	S
4	T	U	V	W	X	Z	4	5
5	6	7	8	9		!	“	#
6	$	%	&	‘	(	)	*	+
7	,	-	.	/	:	;	<	=
8	>	?	@	[	\	]	^	_

For encryption, one has to look at the intersection of any row and column (with row number listed first and column number listed second) as the representation of the alphabet or numeral or special symbol in question in the odd positions (i.e. first, third, fifth, …) and with column number listed first and row number listed second as the representation of the alphabet or numeral or special symbol in question in the even positions (i.e. second, fourth, sixth, ...). By so doing, we will make frequency analysis attack more difficult. For example, if you encrypt the plaintext message SEE YOU the ciphertext will be 38522555142142 as shown in table 3.

Table 3. Encryption using the extended Polybius cipher

Plaintext	S	E	E		Y	O	U
Position	1	2	3	4	5	6	7
Ciphertext	38	52	25	55	14	21	42

As it can be seen from table 3, although the letter “E” appears twice in the plaintext, but the ciphertext is not the same simply because they appear in different positions (i.e. even and odd).

To decrypt a message you find the letter of the alphabet, numeral or special symbol that intersects the specified row and column in the odd and even positions. For example, if you decrypt the ciphertext message 38522555142142 the plaintext message will be SEE YOU as shown in table 4.

Table 4. Decryption using the extended Polybius cipher

Ciphertext	38	52	25	55	14	21	42
Position	1	2	3	4	5	6	7
Plaintext	S	E	E		Y	O	U

7. RESULTS

Although the proposed extended Polybius cipher is a substitution cipher, it is different from other substitution ciphers because each plaintext symbol is replaced by two ciphertext symbols. In this way, we can represent 26 letters of the alphabet, 10 numerals and 28 special symbols by only 8 ciphertext symbols. This process is called Fractionation. This is a very powerful tool in cryptography, and although the Polybius square itself is not that strong, using it in other ciphers will definitely improve the security.

Through this extended Polybius square we can encrypt the plaintext containing alphabets, numerals and special symbols with reasonable efficiency. Furthermore, the user does not face any ambiguity in the process of deciphering, because the characters i and j are placed in separate cells of the matrix. Moreover, the extended Polybius cipher proposed in this work can be used in extending other ciphers such as the Nihilist cipher.

8. CONCLUSIONS AND RECOMMENDATIONS

Although encryption is designed to provide authentication and privacy, it does not prevent attackers from intercepting a message in transit. With our proposed scheme, an attacker may not know the contents of a data transfer but can see that a message transfer occurred. Steganography addresses the visibility-related risks from a data transfer. Encryption protects data by preventing readability- the datum can be observed but not understood. Steganography prevents data from being seen. In steganographic encoding environments, the heuristics encode ciphertext, obscuring detection efforts [1]. Thus, it is advisable to combine encryption and steganography algorithms so as to prevent data from being seen and read.

By itself the Polybius square is not 100% secure, even if used with a keyword. The pairs of digits, taken together, just form a simple substitution in which the ciphertext symbols happen to be pairs of digits. However, the Polybius square offers the possibility of fractionating, leading toward Claude E. Shannon's confusion and diffusion. As such, it is a useful component in several ciphers such as the ADFGVX cipher, the Nihilist cipher and the bifid cipher. Therefore, it is suggested to employ the Polybius cipher at the early stage of teaching and learning cryptography so as to understand the ADFGVX cipher, the Nihilist cipher and the bifid cipher clearly.

Future work will focus on developing software that implements the proposed extended Polybius cipher.

9. ACKNOWLEDGMENTS

Our thanks to the management of the University of Dodoma for creating a good working environment that enabled us to prepare this paper.

10. REFERENCES

[1] Krawetz, N. 2007 Introduction to Network Security. Charles River Media.

[2] Lehtinen, R. 2006 Computer Security Basics. O'Reilly.

[3] Cobb, C. 2004 Cryptography for Dummies. John Wiley & Sons.

[4] Sedgewick, R. and Wayne, K. 2007 Introduction to Programming in Java: An Interdisciplinary Approach. Addison-Wesley.

[5] Paar, C. and Pelzl, J. 2010 Understanding Cryptography. springer.

[6] Stallings, W. 2011 Cryptography and Network Security Principles and Practices. Prentice Hall.

[7] Peterson, L. L. and Davie, B. S. 2007 Computer Networks: a systems approach. Morgan Kaufmann Publishers.

[8] Bishop, M. 2002 Computer Security: Art and Science. Addison Wesley.

[9] Forouzan, B. A. 2010 TCP/IP Protocol Suite. McGraw-Hill.

[10] Pfleeger, C.P. 2006 Security in Computing. Prentice Hall.

[11] Borda, M. 2011 Fundamentals in Information Theory and Coding. Springer.

[12] Mollin, R. A. 2005 Codes: The Guide to Secrecy from Ancient to Modern Times. Chapman & Hall/CRC.