ISSN: 2278 – 7798
International Journal of Science, Engineering and Technology Research (IJSETR)
Volume 2, Issue 4, April 2013

A cryptographic algorithm based on words database

Quist-Aphetsi Kester, MIEEE

***Abstract*— Secrecy and privacy are the key issues cryptography addresses. Through cryptography one can prevent a third party from understanding raw data during signal transmission. Classical ciphers have played a major role in ensuring safer communications up till today.**

One challenge in classical cryptography is the ability to encrypt a set of #n words and at the end of the process obtained #n-k words. That is to obtain a reduced total number of character counts of the ciphertext. A result of a reduced number of words strengthens the algorithm and makes the ciphertext more resistive to known cryptanalysis processes. This means a data compression technique have to be adopted within the encryption process to help reduce network traffic and also to increase the security strength of the ciphertext.

This paper sets out to contribute to the general body of knowledge in the area of cryptography application by developing a cryptographic cipher based on words database. This is done by mapping a plaintext word to many matching words within a database and substituting the numeric index position values back into the position of the plaintext words. The decimal numeric values are then operated on by a function to covert them back to an ASCII character code format. The output is then becomes the encrypted message. At the end, the ciphertext will not have the same text length compared to the plaintext. This will make the cipher more difficult to decipher using frequency attack.

***Index Terms*— cryptography, words, database, cipher, algorithm**

I. INTRODUCTION

Privacy is one of the key issues addressed by information Security. Through cryptographic encryption methods, one can prevent a third party from understanding transmitted raw data over unsecured channel during signal transmission. The cryptographic methods for enhancing the security of digital contents have gained high significance in the current era. Breach of security and misuse of confidential information that has been intercepted by unauthorized parties are key problems that information security tries to solve. [1]

Encryption of messages in this modern age of technology becomes necessary for ensuring that data sent via communications channels become protected and made difficult for deciphering. [2]Enormous number of transfer of data and information takes place through internet, which is considered to be most efficient though it's definitely a public access medium. Therefore to counterpart this weakness, many researchers have come up with efficient algorithms to encrypt this information from plain text into ciphers [3].

In information security, encryption is the process of transforming information using an algorithm to make it unreadable to anyone except those possessing special knowledge, usually referred to as a key. The result of the process is encrypted information. The reverse process is referred to as decryption [4]. There two main algorithmic approaches to encryption, these are symmetric and asymmetric. Symmetric-key algorithms [5] are a class of algorithms for cryptography that use the same cryptographic keys for both encryption of plaintext and decryption of cipher text. The keys may be identical or there may be a simple transformation to go between the two keys. The keys, in practice, represent a shared secret between two or more parties that can be used to maintain a private information link [6]. This requirement that both parties have access to the secret key is one of the main drawbacks of symmetric key encryption, in comparison to public-key encryption. Typical examples symmetric algorithms are Advanced Encryption Standard (AES), Blowfish, Triple Data Encryption Standard (3DES) and Serpent [7].

Asymmetric or Public key encryption on the other hand is an encryption method where a message encrypted with a recipient's public key cannot be decrypted by anyone except a possessor of the matching private key, presumably, this will be the owner of that key and the person associated with the public key used. This is used for confidentiality. [8]. Typical examples of asymmetric encryption algorithms are Rivest Shamir Adleman (RSA),Diffie–Hellman key exchange protocol and Digital Signature Standard(DSS), which incorporates the Digital Signature Algorithm (DSA)

Modern day cryptography entails complex and advance mathematical algorithm are applied to encryption of text and cryptographic techniques for image encryption based on the RGB pixel displacement where pixel of images are shuffled to obtained a cipher image [9][10][11].

This research paper is aimed at contributing to the general body of knowledge in the area of the application of cryptography by developing a new encryption algorithm based on words database. The encryption process is done by mapping a plaintext word to many matching words within a database and substituting the numeric index position values back into the position of the plaintext word. The decimal numeric values are then operated on by a function to covert them back to an ASCII character code format. The output is then becomes the encrypted message. At the end, the ciphertext will not have the same text length compared to the plaintext. This will make the cipher more difficult to decipher

***Quist-Aphetsi Kester, MIEEE**, Faculty of Informatics, Ghana Technology University College. Accra, Ghana, +233 209822141*

using frequency attack.

The paper has the following structure: section II consist of related works, section III of the methodology, section IV The algorithm section V Implementation, section VI Results and Analysis and section VII concluded the paper.

II. Related works

Caesar cipher, also known as the shift cipher, is one of the simplest and most widely known classical encryption techniques. It is a type of substitution cipher in which each letter in the plaintext is replaced by a letter some fixed number of positions down the alphabet. For example, with a shift of 3, A would be replaced by D, B would become E, and so on. The encryption step performed by a Caesar cipher is often incorporated as part of more complex schemes, such as the Vigenère cipher, and still has modern application in the ROT13 system. As with all single alphabet substitution ciphers, the Caesar cipher is easily broken and in modern practice offers essentially no communication security. [12]

The encryption can also be represented using modular arithmetic by first transforming the letters into numbers, according to the scheme, A = 0, B = 1... Z = 25. [13]

Encryption of a letter by a shift n can be described mathematically as, [14]

$$En(x) = (x + n) \bmod 26$$

Decryption is performed similarly,

$$Dn(x) = (x - n) \bmod 26$$

The Vigenère cipher is a method of encrypting alphabetic text by using a series of different Caesar ciphers based on the letters of a keyword. It is a simple form of polyalphabetic substitution [15][16]. The Cipher spoils the statistics of a simple Caesar cipher by using multiple Caesar ciphers. The technique is named for its inventor, Blaise de Vigenère from the court of Henry III of France in the sixteenth century, and was considered unbreakable for some 300 years [17].

Vigenère can also be viewed algebraically. If the letters A–Z are taken to be the numbers 0–25, and addition is performed modulo 26, then Vigenère encryption E using the key K can be written,[18]

$$Ci = EK(Mi) = (Mi + Ki) \bmod \{26\}$$

and decryption D using the key K,

$$Mi = DK(Ci) = (Ci - Ki) \bmod \{26\},$$

whereas M = M0 … Mn is the message, C = C0 …ts Cn is the ciphertext and K = K0 … Km is the used key.

Thus Given m, a positive integer, P = C = (Z26)n, and K = (k1, k2… km) a key, we define:

Encryption:

ek(p1, p2… pm) = (p1+k1, p2+k2…pm+km) (mod 26)

Decryption:

dk(c1, c2… cm) = (c1-k1, c2-k2 … cm- km) (mod 26)

Example:

Plaintext: C R Y P T O G R A P H Y

Key: L U C K L U C K L U C K

Ciphertext: N L A Z E I I B L J J I

A modified form of the Vigenère cipher, the alpha-qwerty cipher extended the original 26 character Vigenère cipher to a 92 characters case sensitive cipher including digits and some other symbols commonly used in the English language and can be written from a computer keyboard. The alpha-qwerty cipher also changes the mapping sequence used in the Vigenère cipher. The mapping takes from an extended alphabet sequence to extended qwerty keyboard sequence. To decrypt the code reverse mapping takes place (compliment of encryption) that is from extended QWERTY key-board to extended alphabet sequence. In short this proposed version extends and rearranges the original Vigenère table, therefore making it much more complex than the existing one. The greater character set allows more type of messages to be encrypted like passwords. It also increases the key domain and hence provides more security [19].

	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
A	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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. The Vigenère square

The algebraic description of the extended version is similar to that of the original cipher. It uses modulo 92 instead of modulo and cipher text Ci is derived using a sequence different from plain text sequence Pi.

$$Ci = EK(Pi) = (Pi + Ki) \bmod 92$$

and decryption D,

$$Pi = DK(Ci) = (Ci - Ki) \bmod 92$$

where, P = P0…Pn is the message,

C = C0….Cn is the ciphertext and K = K0.....Km is the used key.

Friedrich Kasiski was the first to publish a successful general attack on the Vigenère cipher. Earlier attacks relied on knowledge of the plaintext, or use of a recognizable word as a key. Kasiski's method had no such dependencies. He published an account of the attack, but it's clear that there

were others who were aware of it. Babbage was goaded into breaking the Vigenère cipher when John Hall Brock Thwaites submitted a "new" cipher to the Journal of the Society of the Arts. Thwaites challenged Babbage to break his cipher encoded twice, with keys of different length. Babbage succeeded in decrypting a sample, "The Vision of Sin", by Alfred Tennyson, encrypted according to the keyword "Emily", the first name of Tennyson's wife. Studies of Babbage's notes reveal that he had used the method later published by Kasiski [13] [20].

In cryptography, a transposition cipher is a process of encryption by which the positions held by units of plaintext are shifted according to a regular system or pattern, so that the ciphertext constitutes a permutation of the plaintext. That is, the order of the units is changed at the end of the shifting process. Mathematically, a bijective function is used on the characters' positions to encrypt and an inverse function to decrypt. The letters themselves are kept unchanged, which implies that the effect is only on their positions only, making their order within the message scrambled according to some well-defined scheme. Many transposition ciphers are done according to a geometric design [21][22].

In a columnar transposition, the message is written out in rows of a fixed length, and then read out again column by column, and the columns are chosen in some scrambled order. Both the width of the rows and the permutation of the columns are usually defined by a keyword [21]. Advanced forms of columnar encryption techniques are used for encryption in a matrix representation form [11].

Procedure for singe columnar transposition cipher:

1. Chose a key of a fixed length
2. Write the plain text row-by-row in rectangular form but with a fixed column which is equal to the chosen key.
3. Rearrange the column into alphabetical column using the key as the determinant.
4. Read the message column-by-column.
5. The message read becomes the ciphertext.

Example let the key be GERMAN and the plain text be "defend the east wall of the castle"

Then we obtain the following table

G	E	R	M	A	N
d	e	f	e	n	d
t	h	e	e	a	s
t	w	a	l	l	o
f	t	h	e	c	a
s	t	l	e	x	x

Rearranging the above we will obtain

A	E	G	M	N	R
n	e	d	e	d	f
a	h	t	e	s	e
l	w	t	l	o	a
c	t	f	e	a	h
x	t	s	e	x	l

The following ciphertext will be obtained: nalcxehwttdttfseeleedsoaxfseahl

This paper adopted columnar matrix approach where the words are stored in a tabular form in the database and are referenced during the encryption and the decryption process. The substituted characters are based on their row and column referenced values like the Vigenère approach.

III. METHODOLOGY

This paper uses an approach where set of words belonging to a table are viewed in a tabular form. The words are arranged in rows and columns and their index positions can easily be referenced using four numerical value. The first two values represent the table name, the third value represents the row value of the mapped word and the fourth value represents the column value of the mapped word. An algorithm will be built to accomplish the encryption and decryption process.

The encryption process will be done by mapping a plaintext word to a matching word within a database and substituting the word's numeric index position values back into the position of the plaintext word. The decimal numeric values are then operated on by a function to covert them back to an ASCII character code format. The output is then becomes the encrypted message. At the end, the ciphertext will not have the same text length compared to the plaintext. This will make the cipher more difficult to decipher using frequency attack.

IV. THE MATHEMATICAL ALGORITHM

For the plaintext, we have
Let P =plaintext
P = {Pi} = {P1, P2, P3, P4,……..Pn}
Where Pi ∈ P and i=1……n

For the tables, we have
Let T =table
T= {Ti} = {T1, T2, T3, T4,…….Tn}
Where Ti ∈T and i=1……n

If ∃pi ∈ Ti then ⇒Tirc=Pi
Where r =1……n and c =1……m
Where r=row number of Pi in Ti
and i=column number of Pi in Ti

Pi ∈ Tirc iff Pi=Tirc
Pi = [Tir.c] r = 1,...,m; c = 1,...,n and Ti = [Tr.c]m×n
Let Et Pi=CTi iff Pi=Tirc
Set CTi= KQA
K=#Ti
Q=#r^{th} of Tirc
and A= #c^{th} of Tirc

The Encryption algorithm
1. Start
2. Select Pi from P
3. Search for pi within Ti
4. If ∃pi ∈ Ti then ⇒Tirc=Pi

5. Set CT= KQA
6. Interchange the 1st and 3rd digit positions of CTi
7. Add the table number to each first two digits and the last two digits of CTi.
8. Find the ASCII character of the 1st and 2nd two decimal numbers of CTi
9. Repeat step 1 to 8 until Pi=Pn.
10. End

=

V. THE IMPLEMENTATION

Message = {Cryptography prior to the modern age was effectively synonymous with encryption, the conversion of information from a readable state to apparent nonsense. The originator of an encrypted message shared the decoding technique needed to recover the original information only with intended recipients, thereby precluding unwanted persons to do the same.}

Table 1, 2 and 3 be tables that consist of words and the words are displayed in their respective fields. Assuming that, the words in P can be found in the tables below in the fields displaying the respective words in P.

Let table 1= have a decimal value of 30
Let table 1= have a decimal value of 60
Let table 1= have a decimal value of 90

Table 1

					shared			
	Cryptography							
to				with			an	
	the	prior						
					of	effectively		
			modern					to
	age						a	
			was		synonymous			
	with			encryption				

Table 2

			information					
		the			encrypted			
	of							readable
from						to		
		precluding	do				originator	
	conversion				unwanted			
		apparent		of		state		
	the				message			nonsense
a				The				

Table 3

			needed					
				only				
	decoding					technique		
								recipients
persons	intended				the		same	
				thereby				
	to					recover		
		original		the				
								information

P= {Cryptography prior to the modern age was effectively synonymous}

To encrypt P we use the algorithm for the encryption process.

Pi =Tirc= {22 43 69 42 64 72 84 57 68 35 59 }

CTi= KQA = {3022 3043 3069 3042 3064 3072 3084 3057 3068 3035 3059}

Let G= Interchange the 1st and 3rd digit positions of CTi

G= {2032 4033 6039 4032 6034 7032 8034 5037 6038 3035 5039}

Let W=G+#Ti=G+30= {20 32 40 33 60 39 40 32 60 34 70 32 80 34 50 37 60 38 30 35 50 39} + 30

Let encrypted message = E (P)
E (P) = Chr(W) = Chr{50 62 70 63 90 69 70 62 60 64 100 62 110 64 80 67 90 68 60 65 80 69}

E (P)={2>F?ZEF><@d>n@PCZD<APE}

The encrypted message now becomes E (P) above. The inverse of the algorithm will yield the encrypted message back.

VI. RESULTS AND ANALYSIS

From the results obtained from the above, the length of the ciphertext is shorter than the length of the plaintext. Hence, this makes it difficult for one to apply frequency attack to the ciphertext to guess its corresponding values and then also the same word can have different values if it is coming from a different table. Frequency attack was avoided

.

VII. CONCLUSION

This proposed algorithm proved to be very difficult to break without the knowledge of the algorithm and the database full with tables containing words. This algorithm and technique can also be used to compress text data files in the future.

.

REFERENCES

[1] QA Kester (2013) A Hybrid Cryptosystem Based On Vigenère Cipher And Columnar Transposition Cipher. International Journal of Advanced Technology and Engineering Research(IJATER) 3

[2] Kester, Quist-Aphetsi. "A cryptosystem based on Vigenère cipher with varying key." International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) [Online], 1.10 (2012): pp:108-113. Web. 16 Jan. 2013

[3] Kester, Quist- Aphetsi., & Danquah, Paul. (2012). A novel cryptographic key technique. In Adaptive Sci-ence & Technology (ICAST), 2012 IEEE 4th Interna-tional Conference on (pp. 70-73).

[4] Abraham Sinkov, Elementary Cryptanalysis: A Mathematical Approach, Mathematical Association of America, 1966. ISBN 0-88385-622-0

[5] Nicolas Courtois, Josef Pieprzyk, "Cryptanalysis of Block Ciphers with Overdefined Systems of Equa-tions". pp267–287, ASIACRYPT 2002

[6] Delfs, Hans & Knebl, Helmut (2007). "Symmetric-key encryption". Introduction to cryptography: prin-ciples and applications. Springer, 2007

[7] Mullen, Gary & Mummert, Carl. Finite fields and applications. American Mathematical Society. p. 112. 2007

[8] IEEE 1363: Standard Specifications for Public-Key Cryptography

[9] Kester, Q. A., & Koumadi, K. M. (2012, October). Cryptographie technique for image encryption based on the RGB pixel displacement. In Adaptive Science & Technology (ICAST), 2012 IEEE 4th International Conference on (pp. 74-77). IEEE.

[10] Kester, Q. A. (2013). A cryptographic Image Encryption technique based on the RGB PIXEL shuffling. International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), 2(2), pp-848.

[11] Kester, Q. A. (2012, October). A public-key exchange cryptographic technique using matrix. In Adaptive Science & Technology (ICAST), 2012 IEEE 4th International Conference on (pp. 78-81). IEEE.

[12] Encryption. Wellesley college Computer Science De-partment lecture note retrieved from : http://cs110.wellesley.edu/lectures/L18-encryption/

[13] Caesar cipher. Retrieved from http://en.wikipedia.org/wiki/Caesar_cipher

[14] Luciano, Dennis; Gordon Prichett (January 1987). "Cryptology: From Caesar Ciphers to Public-Key Cryptosystems". The College Mathematics Journal 18 (1): 2–17. doi:10.2307/2686311. JSTOR 2686311.

[15] Bruen, Aiden A. & Forcinito, Mario A. (2011). Cryp-tography, Information Theory, and Error-Correction: A Handbook for the 21st Century. John Wiley & Sons. p. 21. ISBN 978-1-118-03138-4. http://books.google.com/books?id=fd2LtVgFzoMC&pg=PA21.

[16] [13] Martin, Keith M. (2012). Everyday Cryptography. Oxford University Press. p. 142. ISBN 978-0-19-162588-6. http://books.google.com/books?id=1NHli2uzt_EC&pg=PT142.

[17] Wobst, Reinhard (2001). Cryptology Unlocked. Wiley. pp. 19. ISBN 978-0-470-06064-3.

[18] [15] Vigenère cipher. Retrieved from http://en.wikipedia.org/wiki/Vigenère_cipher

[19] Rahmani, M. K. I., Wadhwa, N., & Malhotra, V. (2012). Advanced Computing: An International Jour-nal (ACIJ). Alpha-Qwerty Cipher: An Extended Vigenere Cipher, 3 (3), 107-118.

[20] Franksen, O. I. (1985) Mr. Babbage's Secret: The Tale of a Cipher—and APL. Prentice Hall..

[21] Classical cipher, Transposition ciphers, Retrieved from http://en.wikipedia.org/wiki/Classical_cipher

[22] Transposition ciphers, columnar transposition Re-trieved from http://en.wikipedia.org/wiki/Transposition_cipher

Quist-Aphetsi Kester, MIEEE: is a global award winner 2010 (First place Winner with Gold), in Canada Toronto, of the NSBE's Consulting Design Olympiad Awards and has been recognized as a Global Consulting Design Engineer. Currently the national chair for Policy and Research Internet Society (ISOC) Ghana Chapter, a world renowned body that provides international leadership in Internet related standards, education, and policy. He is the Chairman for the Centre of Research, Information Technology and Advanced computing-CRITAC. He is a law student at the University of London UK. He is a PhD student in Computer Science. The PhD program is in collaboration between the AWBC/ Canada and the Department of Computer Science and Information Technology (DCSIT), University of Cape Coast. He had a Master of Software Engineering degree from the OUM, Malaysia and BSC in Physics from the University of Cape Coast-UCC Ghana.

He has worked in various capacities as a peer reviewer for IEEE ICAST Conference, IET-Software Journal, lecturer, Head of Digital Forensic Laboratory Department at the Ghana Technology University and Head of Computer science department. He is currently a lecturer at the Ghana Technology University College.