7-2006

The Vigenére Cipher Expository Paper

Virginia L. Clark
University of Nebraska-Lincoln

The Vigenére Cipher
Expository Paper

Virginia L. Clark

In partial fulfillment of the requirements for the Master of Arts in Teaching with a
Specialization in the Teaching of Middle Level Mathematics
in the Department of Mathematics.
Jim Lewis, Advisor

July 2006

The Vigenére Cipher

French diplomat and cryptographer Blaise de Vigenére (1523-1596), developed the Vigenére Cipher in 16th century France in the mid-1580s. Vigenére was on the court of Henry III of France. Vigenére developed a polyalphabetic coding system in which one letter of plain text may be encrypted as different letters rather than one plain text letter represented as one cipher text letter throughout the encoded message.

Though the Vigenére Cipher was thought to be unbreakable for almost 300 years, it was Charles Babbage who first developed a method to successfully decrypt messages encrypted using the Vigenére Cipher. This fact wasn't known until the 1900's for two possible reasons: 1) Babbage had a habit of not completing paperwork, or 2) he wanted to protect the fact that he had broken the code so Britain could continue to decrypt messages encrypted using the Vigenére Cipher sent in the Crimean War (1853-1856). It was left to a Prussian officer, Friedrich Wilhelm Kasiski, to share a statistical method for breaking the Vigenére Cipher in 1863, which will be discussed later in this paper.

The Vigenére Cipher involves using a table that entails 26 shifts of the alphabet and a key word. The encoding process involves using a plain text message and matching letters with the key word. The intersection of the column using the plain text letter and the row of the key word letter determines the cipher text letter. The following example uses the table, the key word, a key phrase MATH ROCKS and the plain text PARALLEL LINES ignoring spaces and punctuation. Notice that the intersection of the 'P' column and the 'M' row gives a result of 'B' and the intersection of the 'A' column and the 'A' row gives the letter 'A' as the result. (It may

seem problematic that an A is encoded as an A, but looking at the completely encrypted message, you can see that the next 'A' is encoded as 'H'.)

Vigenére Square (Table)

Plain Text

Key Word	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

Plain text	P	A	R	A	L	L	E	L	L	I	N	E	S
Key Word	M	A	T	H	R	O	C	K	S	M	A	T	H
Cipher Text	**B**	**A**	**K**	**H**	**C**	**Z**	**G**	**V**	**E**	**U**	**N**	**X**	**Z**

The Vigenére Cipher is difficult to break because each letter of the plain text alphabet can be encoded as several different letters. In this case, each plain text letter could be encoded as four different letters. You can see that the four L's in PARALLEL LINES are encoded as a C, a Z, a V, and an E. The two A's are encoded as an A and an H. Another difficulty with decrypting this cipher is that the same letter can encrypt different letters. Notice in this example that an L

and the S are both encoded as a Z. This encryption method 'hides' the letter frequencies, so frequency analysis cannot be performed as easily to discover the secret message.

Hamilton & Yankosky (2004) discuss another method of encrypting messages using the Vigenére Cipher. This involves using A=0, B=1, C=2..., Y=24, Z=25 and arithmetic in modulo 26. This involves changing plain text and key word letters to corresponding numbers, then adding the numbers in modulo 26. The result is the number that corresponds with the cipher text letter. Using the same key word (MATH ROCKS) as before and the table below to encode the word GRADUATION, G=6 (from the plain text message) and M=12 (from the key word) we have $(6+12) \bmod 26 \equiv 18 \bmod 26$. In the table 18 represents the letter S, so plain text G is encrypted as an S. Then, R=17 (plain text) and A=0 (key word), so $(17+0) \bmod 26 \equiv 17 \bmod 26$. 17 in the table represents the letter R, so the plain text R is encoded as an R. Encrypting the A and D in the plain text message are found in the same way. Encoding the U requires more understanding of modular arithmetic and least residue. U=20 (plain text) and R=17 (key word), so $(20+17) \bmod 26 \equiv 37 \bmod 26$, but there is no 37 in our table. One must recall that $37 \bmod 26 \equiv (37-26) \bmod 26 \equiv 11 \bmod 26$. In the table, the number 11 corresponds to the letter L, so U in plain text is encrypted as an L. Any time the result is a number greater than 26, 26 (or a multiple of 26) should repeatedly be subtracted until the least residue, r is $0 \leq r < 26$. The completed encryption is shown below. This encryption can be verified using the Vigenére table used in the previous examples.

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
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

Plain text	G	R	A	D	U	A	T	I	O	N

Key Word	M	A	T	H	R	O	C	K	S	M
Cipher Text	**S**	**R**	**T**	**K**	**L**	**O**	**V**	**S**	**G**	**Z**

Decrypting a message using the Vigenére Cipher is quite easy, *if* you know the key word. Using the key word of MATH ROCKS again, consider the encrypted message MUZBJHVGWXFMO. Use the 'M' row (from the key word) in the Vigenére Square and continue across until you find the M (from the cipher text message), look at the top of this column to find the plain text letter of A. Continue with the 'A' row (from the key word) and look across this row until you find the U (from the cipher text), and go to the top of the column to find the plain text letter U. See the table below to find the completely decoded message in plain text. Again, notice that the plain text T has been encoded as an H, a V, and an M. If you didn't know the key word, this would have been very difficult to decrypt.

Plain text	**A**	**U**	**G**	**U**	**S**	**T**	**T**	**W**	**E**	**L**	**F**	**T**	**H**
Key Word	M	A	T	H	R	O	C	K	S	M	A	T	H
Cipher Text	M	U	Z	B	J	H	V	G	W	X	F	M	O

Decrypting using arithmetic in modulo 26 is similar to encrypting using this method, instead of finding the sum of the numerical values the difference is needed. Using the previous example of cipher text, M=12 (cipher text) and M=12 (key word). The difference in modulo 26 is $(12-12) \bmod 26 \equiv 0 \bmod 26$. A 0 represents an A in plain text, which is also verified in the table above. Decrypting the U and Z work in a similar manner with plain text results of U and G, respectively. Then, B=1 (cipher text), H=7 (key word) and $(1-7) \bmod 26 \equiv (-6) \bmod 26$. There is no -6 in our table and the least residue in modulo 26 should be $0 \leq r < 26$, so 26 (or a multiple of 26)

should be added until the result satisfies the inequality. So, (-6)mod26≡(-6+26)mod26 ≡20mod26. The number 20 represents the letter U, which is confirmed in the completed decryption. The remainder of the message can be decoded in a similar manner.

The Kasiski Method is a procedure used to decode a message that has been encoded using the Vigenére Cipher. As previously mentioned, the Kasiski Method was developed almost 300 years after the invention of the Vigenére Cipher (Cook, 1997). This method uses to repeated sequences of letters to determine possible key word lengths. The idea is that frequently used words will eventually be encrypted with the same key word letter if the message is long enough. The message is divided into smaller pieces of the message with the same length as the possible key words. Frequency analysis can then be used to find the key word and decrypt the message. The example that follows will illustrate how the Kasiski Method is used to decrypt a message encrypted using the Vigenére Cipher. It is taken from *Secret Codes: Real-World Mathematics Through Science* (Cook, 1997).

In the following cipher text, I need to look for repeated sequences (underlined):

FHXJMTBUFHXOMTTAQTALSRXLZEZNEAGKTAFIGTRL

DTELFHXAGRMSQOUZQROLPHHYFOGOMTVOFHXLSG

The number of spaces from the beginning of the first ‘FHX’ to the beginning of the next ‘FHX’ is 8, from the second to the third ‘FHX’ is 36 spaces and from the third to the fourth ‘FHX’ is 28 spaces. The key word length is likely to be a common factor of these three lengths. This is because the key word, of a fixed length, will be repeated a certain number of times until the repeated sequence appears again. Looking at the factors of 8, 36 and 28, there are common factors of two and four. If the key word length was two, we could not get repeated sequences of length three for the same plain text word, so the key word length must be four. We now group

the cipher text into sets of four (shown on the following page at the right) and find the most frequently occurring letter in each column (shown in the rectangle).

In the first column, the F occurs most often (five times). T occurs six times in the second column, making it the most frequently occurring letter in this column. X is the most frequently occurring letter in the third column, with a frequency of five. In the fourth column, L occurs most frequently by appearing six times. Now we can use what we know about letter frequency in the English language to help us determine what these letters may represent. Since the letter E is the most frequently used letter in the English language, we start by assuming the most frequently occurring letters in each column represent the letter E.

Using the letters FTXL we utilize the Vigenére Table on page two to try to decode the key word. Assuming that the F (cipher text) was generated from column E (plain text) row 'B' (key word) would generate F. So it's reasonable to say that the first letter of the key word *might be* B. Next, T in cipher text would be the result if we use column 'E' (plain text) and row 'P' (key word). The predicted second letter of the key word is then P. Using similar reasoning, the possible third and fourth letters of the key word are T and H respectively. The key word might look like BPTH, but that's not a word. If we try a different letter for the cipher text F and T, the letter T is the next most frequently used letter in the English language so we try that. Checking column 'T' (plain text) shows that the 'M' row would result in F (cipher text) from a plain text T, so now assume the first letter of the key word is M. Then, the 'A' row (key word) would generate a cipher text T from a plain text T. This would result in the word MATH for a key word. Using MATH to decrypt the message has the following outcome:

FHXJ
MTBU
FHXO
MTTA
QTAL
SRXL
ZEZN
EAGK
TAFI
GTRL
DTEL
FHXA
GRMS
QOUZ
QROL
PHHY
FOGO
MTVO
FHXL
SG
FTXL

THE CAT IN THE HAT ATE THE GREEN EGGS AND HAM BUT YERTLE THE TURTLE OBSERVED HORTON HATCH THE EGG.

So, the key word must have been MATH. Even though the first possible key word did not decode the message, the information was useful in finding the correct key word. Using the most frequently used letters made it possible to find a possible key word. If MATH had not decoded the message, knowing the key word length is enough information to continue the deciphering of a Vigenére Cipher. Knowledge of letter frequency will allow others to decode the message, as shown in this example.

It is possible to remove all traces of letter frequency by using a key 'word' that is made up of randomly generated letters and forming a very long key 'word' called a worm. This type of cipher is called a Vernam Cipher, named after American engineer Gilbert Vernam who developed the method. It is also called a one-time pad because the key is used only once. This makes the code almost impossible to break, even by cryptography experts.

All previous examples excluded spaces and any punctuation. It is possible to add these symbols (spaces, commas and periods) to the cipher by including them in the Vigenére Square, as shown below. Messages can be encoded and decoded in the same manner as previously discussed. When using modular arithmetic, modulo 29 would need to be used instead of modulo 26 because there are now 29 elements available to code and decode. The following has been encrypted using the key M SQUARED. We can use this key and the table on the following page to decrypt the message:

Y IXRIBBWTBP..DUPHK

XIXY VRGJFHNAUGXACEWNVEXMQZFCWT

	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
		,	.	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	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	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		,

Plain Text	**M**	**A**	**T**	**H**		**I**	**N**		**T**	**H**	**E**		**M**	**I**	**D**	**D**	**L**	**E**	**,**		**T**	**H**	**E**		**E**	**N**	**D**
Key	M		S	Q	U	A	R	E	D	M		S	Q	U	A	R	E	D	M		S	Q	U	A	R	E	D
Cipher Text	Y		I	X	R	I	B	B	W	T	B	P	.	.	D	U	P	H	K	X	I	X	Y		V	R	G

Plain Text		**I**	**S**		**J**	**U**	**S**	**T**		**T**	**H**	**E**		**B**	**E**	**G**	**I**	**N**	**N**	**I**	**N**	**G**	**.**
Key	M		S	Q	U	A	R	E	D	M		S	Q	U	A	R	E	D	M		S	Q	U
Cipher Text	J	F	H	N	A	U	G	X	A	C	E	W	N	V	E	X	M	Q	Z	F	C	W	T

Developed over 400 years ago, the Vigenére Cipher seemed ‘unbreakable’. During that time it was used to allow messages to be securely sent in wartime. As time passed and mathematical thinking became more sophisticated, that idea was proven wrong. Even so, this

cipher is still used in educational settings and would be ‘unbreakable’ for most without the key word. It is used as an introduction to more complex methods of encryption as opposed to simpler mono-alphabetic ciphers. Its use can also be associated with development of modular arithmetic skills. The Vigenére Cipher also led to the more complicated one-time pad where a ‘worm’ is used instead of a key word that leaves the message open for decryption using letter frequency analysis. In addition to being useful, cryptography can be an enjoyable pastime.

References

Beissinger, J. & Pless, V., (2005). *Cryptography: The Mathematics of Secret Codes* (draft). Wellesley, MA: AKPeters, Ltd.

Cook, N., (1997). *Secret Codes: Real-World Mathematics Through Science*. Palo Alto, CA: Dale Seymour Publications.

Hamilton, M. & Yankosky, B., (2004). The Vigenére Cipher with the TI-83. *Mathematics and Computer Education*. 38(1), 19-31.

Meyerscough, D., Ploger, D., McCarthy, L., Hopper, H., & Fegers, V., (1996). Cryptography: Cracking Codes. *Mathematics Teacher*. 89(9), 743-750.

http://www.cryptodox.com/History_of_Cryptography

http://www.cypher.com.au/crypto_history.htm

http://www.trincoll.edu/depts/cpsc/cryptography/vigenere.html