

Factorization and Cryptology

David Brown - dabrown@ithaca.edu

Department of Mathematics
Ithaca College

Math Day
April 6, 2009

Quick Introduction to Cryptology

Art and Science of Secret Writing

What is it?

- Cryptography - “making secret codes
- Cryptanalysis - “breaking secret codes
- Plaintext - message to be made secret
- Ciphertext - the enciphered message
- Key - the process that makes a message secret

Caesar Cipher

The idea of replacing one letter of the alphabet with another one dates back to Julius Caesar (and even earlier.) Caesar did this substitution by shifting the alphabet. In the table below, the upper alphabet is the normal English alphabet and the lower alphabet is the shifted alphabet, in this example, shifted by 3.

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
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

To make the message secret, you replace the letters with the corresponding shifted letters. For example, you replace T with W. To uncover a secret message, reverse the process.

For example, suppose that our message is **BUBBLES ARE FUN** and that we want to use a Caesar shift of 3. Using the table above, we see that B is replaced with E, U is replaced with X, and so on.

B \Rightarrow E
U \Rightarrow X
B \Rightarrow E
B \Rightarrow E
L \Rightarrow O
E \Rightarrow H
S \Rightarrow V
A \Rightarrow D
R \Rightarrow U
E \Rightarrow H
F \Rightarrow I
U \Rightarrow X
N \Rightarrow Q

So, the secret message to send is **EXEEOHVDUHIXQ**.

Now, if we have the secret message **G**DYHLVDOV**R**IX**Q** and we know that it was made secret by using a shift of 3, we can easily recover the original message. Look at the bottom row of the alphabet table. **G** comes from **D**. So the first letter of our message is **D**. **D** comes from **A**. So the second letter of our message is **A**. Continuing,

G	←←	D
D	←←	A
Y	←←	V
H	←←	E
L	←←	I
V	←←	S
D	←←	A
O	←←	L
V	←←	S
R	←←	O
I	←←	F
X	←←	U
Q	←←	N

So, the original message is **DAVE IS ALSO FUN**.

BREAKING CAESAR

MESSAGE:

(QIIX) (QIJS) (VGSJ) JIIN^N

(CIPHERTEXT)

FREQUENCY ANALYSIS:

Q - 2

I - 5

X - 1

J - 2

S - 2

V - 1

G - 1

N - 1

E - T - A - N - O

GUESS: E enciphered as I:

Plain →

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
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

CIPHER ALPHA

QIIX	QIJS	VGSJ	JIIN
M E E T	M E F O	R C O F	F E E J

Vigenère Cipher - shifting Caesar ciphers

- In a monoalphabetic cipher, each plaintext letter is always enciphered with the same ciphertext letter.
- This fact makes it vulnerable to frequency analysis.
- (1586) Blaise de Vigenère, French diplomat, put several monoalphabetic ciphers together simultaneously in order to encipher.
- Polyalphabetic ciphers! Became the standard of professional ciphers and used until WWII (Italy).

Question: How would you decipher a message using the Vigenère cipher? See if you can figure out the following message, which was enciphered with the Vigenère cipher using keyword SLEEP.

XLPPQ JPEOX KNSQX FRWSD F

- Looking at these examples, what do you think is the strength of the Vigenère cipher? Any weaknesses?
- Compare this cipher to the additive and multiplicative ciphers and to the keyword cipher you already know. Which one you would use and why?

Further Practice The following messages have been enciphered with a Vigenère cipher. Decipher them.

1. **Question:** What do you call a happy Lassie?

Answer: XAOJI PCMIC IC (keyword is xray)

2. **Question:** Knock, knock. *Who's there?* Cash. *Cash who?*

Answer: XSMDW NWTVE GMRNM TSHMD DNMTT (keyword is pizza)

3. **Question:** What is the noisiest dessert?

Answer: TMPTL LG (keyword is lunch)

Breaking the Vigenère Code

Prussian colonel, Friedrich Wilhelm Kasiski (1805-1881), published a method of attacking the Vigenère cipher in 1863. (Babbage may have figured it out 9 years earlier.)

Key Idea

- If a plaintext message repeats sequences of letters (eg, THE), then usually the ciphertext letters are all distinct.
- If, however, the first letter of a repeated sequence is enciphered by the same keyword letter, the next few subsequent letters in the repeated sequences will be enciphered the same!

Example

Suppose that part of a plaintext message is

“... on a plane. The plane is due...”

Encipher with keyword WATER:

W	A	T	E	R	W	A	T	E	R	W	A	T	E	R	W	A	T	E	R	W
O	N	A	P	L	A	N	E	T	H	E	P	L	A	N	E	I	S	D	U	E
K	N	T	T	C	W	N	X	X	Y	A	P	E	E	E	A	I	L	H	L	A

The two occurrences of PLANE are enciphered differently as TCWNX and PEEEA. No surprise here. Cipher is doing its job.

Example

Suppose however that the keyword is MILK.

M	I	L	K	M	I	L	K	M	I	L	K	M	I	L	K	M	I	L	K	M
O	N	A	P	L	A	N	E	T	H	E	P	L	A	N	E	I	S	D	U	E
A	V	L	Z	X	I	Y	O	F	P	P	Z	X	I	Y	O	U	A	O	E	Q

Both times, PLANE is enciphered as ZXIYO.

Why?

Notice that P in the second PLANE appears exactly 8 characters after the P in the first PLANE.

8 is a multiple of 4, the keyword length!!!!

Kasiski Method

If a string of ciphertext characters appears repeatedly in a polyalphabetic cipher, then it is possible (although not certain) that the distance between occurrences is a multiple of a keyword's length.

Let's try it ourselves.

Kasiski Method

- Examine the enciphered message and try to find repeated strings of letters.
- Record the distance between the first letters of consecutive repeated strings.
- Find the prime factorizations of each distance.
- Find common factors of all the distances.
- Conjecture the keyword length.

FYS:HISTORY OF SECRETS

MAGIC FROM NONSENSE

We suspect that the following message was encrypted using a Vigenère cipher.
How can we be sure and how can we decrypt it?

JAKXQ	SWECW	MMJBK	TQMCM	LWCXJ	BNEWS
XKRBO	IAOBI	NOMLJ	GUIMH	YTACF	ICVOE
BGOVC	WYRCV	KXJZV	SMRXY	VPOVB	UBIJH
OVCVK	RXBOE	ASZVR	AOXQS	WECVO	QJHSG
ROXWJ	MCXQF	OIRGZ	VRAOJ	RJOMB	DBMVS
CIESX	MBDBM	VSKRM	GYFHA	KXQSW	ECWME
UWXHD	QDMXB	KPUCN	HWIWF	NFCKA	SKXNF
DLJBY	RNOBI	YFSQN	HRIYV	IWRQS	WCGKC
BHRVN	SSWYF	SQNTS	ZNWCT	AWWIB	SFIWW
CTAWW	IWWXI	RGKRN	LZIAW	WIWHK	PNFBS
ASVIE	SXMBD	BMVSK	RMGYC	NGKPU	CNHWI
WFNFC	KASKX	NFDLJ	BYRNO	BIYFS	QNHRI
NBQMW	SOVBO	IWCVB	INWCT	AWWIO	WFIRG
ZVRAO	WNJOR	RGZVR	AORRB	OMBDB	MVSOP
NJORR	GZVRA	OXQWB	XNSXM	BDBMV	SPMOH
OIWWC	TAWWI				

We can figure out the length of the keyword by looking for patterns. Try to find repeated strings of letters, the longer the better.

RGZVRAO

MBDBMVS

WCTAWWI

FYS:HISTORY OF SECRETS
MAGIC FROM NONSENSE

We suspect that the following message was encrypted using a Vigenère cipher.
How can we be sure and how can we decrypt it?

12341 23412 34
 JAKXQ SWECW MMJBK TQMCM LWCXJ BNEWS
 XKRBO IAObI NOMLJ GUIMH YTACF ICVOE
 BGOVC WYRCV KXJZV SMRXY VPOVB UBIJH
 OVCVK RXBOE ASZVR AOXQS WECVO QJHSG
 ROXWJ MCXQF OIRGZ VRAOJ RJOMB DBMVS
 CIESX MBDBM VSKRM GYFHA KXQSW ECWME
 UWXHD QDMXB KPUCN HWIWF NFCKA SKXNF
 DLJBY RNOBI YFSQN HRIYV IWRQS WCGKC
 BHRVN SSWYF SQNTS ZNWCT AWWIB SFIWW
CTAWW IWWXI RGKRN LZIAW WIWHK PNFBS
 ASVIE SXMBD BMVSK RMGYC NGKPU CNHWI
 WFNFC KASKX NFDLJ BYRNO BIYFS QNHRI
 NBQMW SOVBO IWCVB INWCT AWWIO WFIRG
ZVRAO WNJOR RGZVR AORRB OMBDB MVSOP
NJORR GZVRA OXQWB XNSXM BDBMV SPMOH
OIWWC TAWWI

We can figure out the length of the keyword by looking for patterns. Try to find repeated strings of letters, the longer the better.

Repeated String	Distance	Prime factorization
-----------------	----------	---------------------

RGZVRAO

1-2: 256
 2-3: 12
 3-4: 24

2^8
 $2^2 \cdot 3$
 $2^3 \cdot 3$

WCTAWWI

12
 108
 76

$2^2 \cdot 3$
 $2^2 \cdot 3^3$
 $2^2 \cdot 19$

MBDBMVS

1-2 12
 2-3 152
 3-4 104
 4-5 28

$2^2 \cdot 3$
 $2^3 \cdot 19$
 $2^3 \cdot 13$
 $2^2 \cdot 7$

Common factors?

2, 4

1234

JAKX
QSWE
CWMM
JBKT
QMCM
LWCX
JBNE
WSXK
RBOI
AOBI
NOML
JGUI
MHYT
ACFI
CVOE
BGOV
CWYR
CVKX
JZVS
MRXY
VPOV
BUBI
JHOV
CVKR
XBOE
ASZV
RAOX
QSWE
CVOQ
JHSG
ROXW
JMCX
QFOI
RGZV
RAOJ
RJOM
BDBM
VSCI
ESXM
BDBM
VSKR
MGYF
HAKX
QSWE
CWME
UWXH

DQDM
XBKP
UCNH
WIWF
NFCK
ASKX
NFDL
JBYR
NOBI
YFSQ
NHRI
YVIW
RQSW
CGKC
BHRV
NSSW
YFSQ
NTSZ
NWCT
AWWI
BSFI
WWCT
AWWI
WWXI
RGKR
NLZI
AWWI
WHKP
NFBS
ASVI
ESXM
BDBM
VSKR
MGYC
NGKP
UCNH
WIWF
NFCK
ASKX
NFDL
JBYR
NOBI
YFSQ
NHRI
NBQM
WSOV

BOIW
CVBI
NWCT
AWWI
OWFI
RGZV
RA'OW
NJOR
RGZV
RAOR
RBOM
BDBM
VSOP
NJOR
RGZV
RAOX
QWBX
NSXM
BDBM
VSPM
OHOI
WWCT
AWWI

What about the keyword?

Now that we have a keyword length, how do we find the actual keyword?

Well, if we have the keyword length (4 in our case), we know that every 4th letter is enciphered using the same shift (Caesar) cipher within the Vigenère cipher.

So, use frequency analysis on the four sequences of every fourth letter.

What about the keyword?

Column 1	Column 2	Column 3	Column 4
J	A	K	X
Q	S	W	E
C	W	M	M
J	B	K	T
Q	M	C	M
L	W	C	X
J	B	N	E
W	S	X	K
R	B	O	I
⋮	⋮	⋮	⋮

Frequency analysis reveals

- N appears most often in column 1; so assume E is enciphered as N. Look across E row to find N, look up to see possible keyletter J.
- S appears most often in column 2; so assume E is enciphered as S. Look across E row to find S, look up to see possible keyletter O.
- O appears most often in column 3; so assume E is enciphered as O. Look across E row to find O, look up to see possible keyletter K.
- I appears most often in column 4; so assume E is enciphered as I. Look across E row to find I, look up to see possible keyletter E.

What about the keyword?

Key word is JOKE! Decipher the message.