Spring 2010
Chris Christensen
HNR 304

Caesar Ciphers

Suetonius, the gossip columnist of ancient Rome, says that [Julius] Caesar [100? – 44 B.C.] wrote to Cicero and other friends in a cipher in which the plaintext letters were replaced by letters standing three place further down the alphabet …
David Kahn, *The Codebreakers*

So, cryptology has existed for more than 2000 years. But, what is cryptology? The word ***cryptology*** is derived from two Greek words: *kryptos*, which means "hidden or secret," and *logos*, which means, "description." Cryptology means secret speech or communication.

Cryptology encompasses two competing skills – concealment and solution.

The concealment portion of cryptology is called ***cryptography***. The aim of cryptography is to render a message incomprehensible to the unauthorized reader. Cryptography is often called "code making."

The solution portion of cryptology is called ***cryptanalysis***. Cryptanalysis is often called "code breaking." The word cryptanalysis was coined (c. 1920) by the American cryptologist William Friedman.

William Friedman
Center for Cryptologic History photo

Friedman (1891 – 1969) is often called the dean of modern American cryptologists. He was a pioneer in the application of scientific principles to cryptology. During World War II, Friedman was the director of communications research for the Signal Intelligence Service (SIS). SIS later became the Army Security Agency (ASA). After World War II, Friedman served first as a consultant for ASA and then for the National Security Agency (NSA) after its birth in 1952. Friedman and his wife Elizebeth, who was also a cryptologist, jointly authored the book *The Shakespearean Ciphers Examined*.

Cryptography of Caesar Ciphers

Here is the key for a simple substitution cipher:

Plaintext letters:	`abcdefghijklmnopqrstuvwxyz`
Ciphertext letters:	`YNROTKMCPBDVXZALEWUSFQJHGI`

Could you remember the plaintext/ciphertext correspondences? Probably not; you would probably need a written copy of the key. But, having a written copy of the key could lead to problems with key security – the key might be lost or stolen. It is desirable to have a key that need not be written down. (Of course a person who has memorized the key might be coerced to give it up, but that it a different story.)

Caesar's cipher, to which reference was made in the David Kahn quote at the beginning of this section, was a simple substitution cipher, but it had a memorable key. For Caesar's cipher, "letters were replaced by letters standing three place further down the alphabet … ." Here is the key to Caesar's cipher:

Plaintext letters	`abcdefghijklmnopqrstuvwxyz`
Ciphertext letters	`DEFGHIJKLMNOPQRSTUVWXYZABC`

The key can be memorized because there is a pattern to it -- the ciphertext alphabet is just the plaintext alphabet shifted to the right three places. Sender and receiver just need to remember the shift.

Of course, other shifts could be used. All such shift, or translation, ciphers are now usually called Caesar ciphers. Here is the plaintext/ciphertext correspondence for a Caesar cipher with shift 8:

Plaintext letters	`abcdefghijklmnopqrstuvwxyz`
Ciphertext letters	`IJKLMNOPQRSTUVWXYZABCDEFGH`

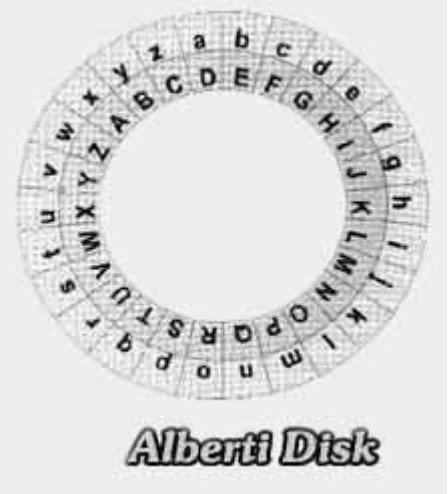
For each of these ciphers, the method of encryption is the Caesar cipher (which is a special case of the simple substitution cipher) and the key is the shift. Knowing the key, the sender and receiver can create the plaintext/ciphertext correspondence as needed. There is no need to keep a written copy of the plaintext/ciphertext correspondence; therefore, key security is less of an issue than it is for the more general simple substitution cipher.

Over the years, cryptographers have created disk or slide devices to show the plaintext/ciphertext correspondence for use when encrypting and decrypting.

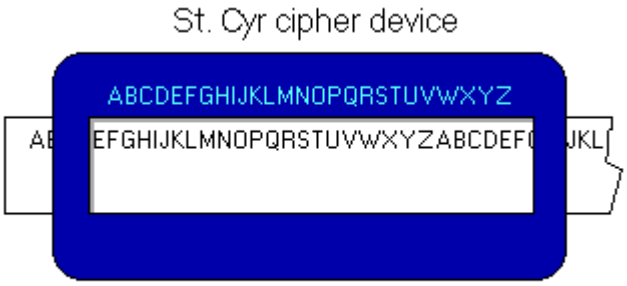
The Italian cryptologist Leon Battista Alberti (1404 – 1472), who is called the Father of Western Cryptology, developed a cipher disk.

> *"I make two circles out of copper plates. One, the larger, is called stationary, the smaller is called movable. … I divide the circumference of each circle into … equal parts. These parts are called cells. In the various cells of the larger circle I write the capital letters, one at a time …, in the usual order of the letters." … In each of the … cells of the movable circle [Alberti] inscribed "a small letter … [Alberti used a random ordering of the letters in the cells of the smaller circle] After completing these arrangements we place the smaller circle upon the larger so that a needle driven through the centers of both may serve as the axis of both and the movable plate may be revolved about it."* Leon Battista Alberti quoted in David Kahn's *The Codebreakers*.

CIPHER DISK



The disk that is shown has the letters in the cells in the usual order. Sender and receiver must agree which circle corresponds to plaintext and which circle corresponds to ciphertext. The disk that is pictured has ciphertext on the smaller circle and plaintext on the larger circle. The disk has been set to Caesar's original cipher – a shift of 3.



The Dutch cryptologist Auguste Kerkhoffs (1835 – 1903) named the cryptographic slide. In 1883, Kerkhoffs published *La Cryptographie militarie*, which became a major cryptological work.

> *[Kerkhoffs] called the slide the St.-Cyr system, after the French national military academy where it was taught. A St.-Cyr slide consists of a long piece of paper or cardboard, called the stator, with an evenly spaced alphabet printed on it and with two slits cut below and to the sides of the alphabet. Through these slits runs a long strip of paper – the slide paper – on which the alphabet is printed twice.* David Kahn, *The Codebreakers*.

A modern St.-Cyr slide is shown. Plaintext is on the stator, and ciphertext is on the slide.

> *[Kerkhoffs] pointed out that a cipher disk was merely a St.-Cyr slide turned round to bite its tail.* David Kahn, *The Codebreakers*.

Here is an example of a Caesar cipher with additive key 5.

Caesar cipher
Additive key = 5

Plaintext			Ciphertext
a	1	6	F
b	2	7	G
c	3	8	H
d	4	9	I
e	5	10	J
f	6	11	K
g	7	12	L
h	8	13	M
i	9	14	N
j	10	15	O
k	11	16	P
l	12	17	Q
m	13	18	R
n	14	19	S
o	15	20	T
p	16	21	U
q	17	22	V
r	18	23	W
s	19	24	X
t	20	25	Y
u	21	26	Z
v	22	1	A
w	23	2	B
x	24	3	C
y	25	4	D
z	26	5	E

Thinking of the ciphertext alphabet “turning round to bite its tail” Caesar ciphers are sometimes called rotation ciphers. When the additive key is 5, we can think of the letters of the alphabet as being rotated by 5 places. A Caesar cipher with an additive key of 5 is called a rot5 cipher. The original Caesar cipher is a rot3 cipher. Rot13 is often used on the internet to hide hints.

Encryption of a Message with a Caesar Cipher

Let us use the Caesar cipher with additive key 5 to encrypt the plaintext message:

> The book *Gadsby* by Ernest Vincent Wright does not contain the letter *e*.

Giving word length and punctuation gives the cryptanalyst too much information. We have already noted that although it is usually easy to solve simple substitution ciphers when word length and punctuation are given, it can be very difficult to solve simple substitution ciphers when word length and punctuation are not given.

Word length and punctuation provide patterns that permit us to quickly make sense of plaintext. Without word length and punctuation, even plaintext can be difficult to read. Here is an example of plaintext without word length and punctuation:

> CARDANOALSOACHIEVEDTHEDUBIOUSRENOWNOFBEING
> THEFIRSTCRYPTOLOGISTTOCITETHEENORMOUSNUMBERO
> FVARIATIONSINHERENTINACRYPTOGRAPHICSYSTEMASPR
> OOFOFTHEIMPOSSIBILITYOFACRYPTANALYSTSEVERREAC
> HINGASOLUTIONDURINGHISLIFETIME.

Usually cryptographers do not give word length and punctuation.

After the invention of the telegraph in the Nineteenth Century, nearly instantaneous communication over long distances became possible, but communication by telegraph involved handing messages to operators who transmitted them in Morse Code. Both the sending and receiving telegraph operators (and probably other telegraph employees) would have access to messages. Business communications and even personal communications were often encrypted. For the convenience of telegraph operators, messages were usually sent in blocks of letters which allowed momentary pauses for the operators' hands. Traditionally the blocks consisted of four or five letters. That practice became a tradition in cryptology. Often ciphertext messages are blocked in blocks of four or five letters. (We will use five-letter blocks.)

Because it hides word length, blocking makes the ciphertext much harder to cryptanalyze. It also makes the message harder for the authorized receiver to read after it is deciphered. Here is a plaintext message in five-letter blocks.

> THEMO STFAM OUSOF FICTI ONALD ETECT IVESS HERLO CKHOL MESEN COUNT EREDC IPHER SNOTO NCEBU TTHRE ETIME SINHI SDIST INGUI SHEDC AREER

Sometimes a letter, often an `x`, is inserted between words in the plaintext. This makes the decrypted ciphertext easier to read, but the encrypted `x` would be a frequent character in the ciphertext and would be easy to spot. Word length is not well concealed by this method.

Here is our plaintext message in five-letter blocks:

```
thebo  okgad  sbyby  ernes  tvinc  entwr  ightd
oesno  tcont  ainth  elett  ere
```

The partial block at the end may be left with only three letters or it may be padded with "nulls," meaningless letters, to complete the five-letter block. Adding nulls to the end of a message might make cryptanalysis more difficult because the cryptanalyst would expect the last letter of ciphertext to correspond to a "final letter" when, in fact, it is "junk." Of course, the nulls must be chosen in such a way that the authorized receiver who decrypts the message would recognize them as nulls.

Here is our message encrypted with a Caesar cipher with additive key 5:

```
thebo  okgad  sbyby  ernes  tvinc  entwr  ightd
YMJGT  TPLFI  XGDGD  JWSJX  YANSH  JSYBW  NLMYI

oesno  tcont  ainth  elett  ere
TJXST  YHTSY  FNSYM  JQJYY  JWJ
```

Cryptanalysis Using Brute Force

Unfortunately, Caesar ciphers have a small key space, and messages encrypted with Caesar ciphers can be easily broken by brute force if it is recognized that the message has been encrypted with a Caesar cipher.

How many distinct Caesar ciphers are possible? Well, a shift of 0 would not make any sense; we would still have plaintext. Shifts of 1, 2, 3, … 25 make sense. But, a shift of 26 would (because the alphabet returns to the beginning) be the same as a shift of 0. Similarly, a shift of 27 is the same as a shift of 1, a shift of 28 is the same as a shift of 2, etc. So, there are only 26 possible Caesar ciphers, and one of those is a shift of 0 which would provide no encryption at all.

Notice that with the exception of the Caesar cipher with additive key 26, when using a Caesar cipher, no letter substitutes for itself. Also, if we know one plaintext/ciphertext correspondence we know them all because the shift is the same for each letter.

Because of the small number of possible keys, a **brute force attack** is possible – we could try all possible keys and see which one yields plaintext.

Here is a brute force ciphertext attack on a Caesar cipher.

The following message is known to have been encrypted with a Caesar cipher:

```
VRRQS  HRSOH  EHJDQ  VOLGL  QJWKH  DOSKD  EHWEB
DPRXQ  WVGLI  IHUHQ  WWKDQ  WKUHH  WRGHW  HUPLQ
HFLSK  HUHTX  LYDOH  QWV
```

Begin with VRRQS, the first five-letter block of the ciphertext. Now beneath it write the five letters that would result by shifting each of the cipehrtext letters to the right by one. On the next line, write the result by shifting each of the ciphertext letters to the right by two. Do this for each of the 26 possible shifts. This attack on a Caesar cipher is sometimes called "running the alphabet."

VRRQS
WSSRT
XTTSU
YUUTV
ZVVUW
AWWVX
BXXWY
CYYXZ
DZZYA
EAAZB
FBBAC
GCCBD
HDDCE
IEEDF
JFFEG
KGGFH
LHHGI
MIIHJ
NJJIK
OKKJL
PLLKM
QMMLN
RNNMO
SOONP
TPPOQ
UQQPR

Now scan the column for something that makes sense. Notice near the bottom SOONP. This line corresponds to shifting the ciphertext alphabet to the right 23 places. The key inverse is 23. The additive key is 3.

Cryptanalysis Using a Known Plaintext Attack

Another possibility is to do a **known plaintext attack**. The name is a bit deceiving because sometimes we only "suspect" rather than "know" a piece of the plaintext message. Consider that in a message of reasonable length we should expect to find the word `the`. If it occurs in a message encrypted with a Caesar cipher, it was encrypted one of the following ways:

Trigraph	Shift
THE	0
UIF	1
VJG	2
WKH	3
XLI	4
YMJ	5
ZNK	6
AOL	7
BPM	8
CQN	9
DRO	10
ESP	11
FTQ	12
GUR	13
HVS	14
IWT	15
JXU	16
KYV	17
LZW	18
MAX	19
NBY	20
OCZ	21
PDA	22
QEB	23
RFC	24
SGD	25

Here is a message that is known to have been encrypted with a Caesar cipher:

```
FGWFM  FRXNS  PTAKN  WXYBT  WPJIF  XFHWD  UYTQT
LNXYB  NYMYM  JBFWI  JUFWY  RJSY
```

To determine the key, search through the ciphertext for a Caesar cipher ciphertext of `the`. Because the beginning and ending of words is hidden by the five-letter blocks, when searching for an encrypted `the`, we must check every three consecutive letters – every trigraph: `FGW GWF WFM FMF MFR FRX RXN XNS NSP SPT PTA TAK AKN KNW NWX WXY XYB YBT BTW TWP WPJ PJI JIF IFX FXF XFH FHW HWD WDU DUY UYT YTQ TQT QTL TLN LNX NXY XYB YBN BNY NYM MYM` **`YMJ`** `MJB JBF BFW FWI WIJ IJU JUF UFW FWY WYR YRJ RJS JSY`. The trigraph in bold is `the` encrypted with an additive key of 5. If we assume the message was encrypted with an additive key of 5, the message decrypts.

This technique of searching for an encrypted version of a word or phrase was used during World War II by the British codebreakers at Bletchley Park who broke the German Enigma messages. The Enigma machine had letters but no numbers on its keyboard; so, numbers were written out in plaintext messages. It was common that the word *Eins* (one) would appear in a message. With a lot of patience and having a catalog of the encrypted versions of *Eins*, the Enigma key might be determined.

The word `the` when used as we have in this process is called a *crib*. Gordon Welchman, one of the cryptologists at Bletchely Park writes:

> *Cryptologically speaking, however, one has a "crib" to a cipher text if one can guess the clear text from which some specific portion of the cipher text was obtained. As my analysis of the Enigma traffic began to reveal certain routine characteristics in the preambles of individual messages, I realized that, if we could somehow determine to whom they were addressed, or by whom they were sent, we might be able to guess a portion of the clear text either at the beginning or the end of each of the messages, and so have cribs*. Gordon Welchamn, *The Hut Six Story*

Stereotyped writing provides cribs. In cryptography, variety breeds security.

Recognition of a Caesar Cipher and Its Key by Frequency Analysis

A Caesar cipher is easy to break, but how do we recognize that a Caesar cipher was used? It is easy to spot a Caesar cipher from frequency analysis of the ciphertext.

Patterns occur in the letter frequencies of any language. Here are the patterns for English:

Frequencies for English

```
a    1111111
b    1
c    111
d    1111
e    1111111111111
f    111
g    11
h    1111
i    1111111
j
k
l    1111
m    111
n    11111111
o    1111111
p    111
q
r    11111111
s    111111
t    111111111
u    111
v    1
w    11
x
y    11
z
```

Abraham Sinkov (who was one of William Friedman's cryptanalysts during World War II) in his text *Elementary Cryptanalysis: A Mathematical Approach* points out the following patterns which are useful for elementary cryptanalysis:

1. a, e, and i are all high frequency letters (at the beginning of the plaintext alphabet), and they are equally spaced (four letters apart) with e the most frequent.
2. n and o form a high frequency pair (near the middle of the plaintext alphabet).
3. r, s, and t form a high frequency triple (about 2/3 of the way through the plaintext alphabet).
4. j and k form a low frequency pair (just before the middle of the plaintext alphabet).
5. u, v, w, x, y, and z form a low frequency six-letter string (at the end of the plaintext alphabet).

Because a Caesar cipher just translates the letters of the plaintext alphabet to the right, it translates to the right the frequency patterns we expect with plaintext.

Here are the expected frequencies for a Caesar cipher with additive key 5:

Frequencies Additive key = 5

```
A     1
B     11
C
D     11
E
F     1111111
G     1
H     111
I     1111
J     1111111111111
K     111
L     11
M     1111
N     1111111
O
P
Q     1111
R     111
S     11111111
T     1111111
U     111
V
W     11111111
X     111111
Y     111111111
Z     111
```

Notice that the usual frequencies have just shifted 5 places further in the alphabet.

Instead of having a, e, and i be all high frequency letters spaced four letters apart with e the most frequent, we now have that F, J, and N have that property with J being the most frequent letter.

Instead of n and o forming a high frequency pair (near the middle of the plaintext alphabet), we have that S and T form such a pair.

Instead of r, s, and t forming a high frequency triple, we have that W, X, and Y form such a triple.

Instead of j and k forming a low frequency pair, we now have that O and P form such a pair.

Instead of u, v, w, x, y, and z forming a low frequency six-letter string (at the end of the plaintext alphabet), we now have that Z, A, B, C, D, and E form such a string.

Such shifts of frequency patterns should be easy to spot. They identify a Caesar cipher, and they exhibit the shift – the key.

Here is a ciphertext message:

```
VRRQS  HRSOH  EHJDQ  VOLGL  QJWKH  DOSKD  EHWEB
DPRXQ  WVGLI  IHUHQ  WWKDQ  WKUHH  WRGHW  HUPLQ
HFLSK  HUHTX  LYDOH  QWV
```

Here is a frequency analysis of the ciphertext:

```
A
B    1
C
D    111111
E    111
F    1
G    111
H    111111111111111
I    11
J    11
K    11111
L    111111
M
N
O    1111
P    11
Q    11111111
R    11111
S    1111
T    1
U    1111
V    1111
W    111111111
X    11
Y    1
Z
```

Notice that the pattern of frequencies suggests that `H = e`. It is only necessary to determine one correspondence between a plaintext and ciphertext letter to determine the key – during encryption all plaintext letters are shifted by the same amount.

Exercises

1. Construct a plaintext-ciphertext correspondence for a Caesar cipher having additive key 9.

2. Encrypt the following message using a Caesar cipher with additive key 9. Use five-letter blocks.

```
The telegraph made cryptography what it is today.
```

3. Use frequency analysis to cryptanalyze the following ciphertext:

```
MAXGX   QMWTR   VKXPF   XFUXK   LHGMA   XFTWW
HQLBZ   AMXWY   BOXGH   KMAOB   XMGTF   XLXGT
ORTMM   TVDUH   TML
```

4. Use brute force to cryptanalyze the following ciphertext that is known to be encrypted with a Caesar cipher.

```
CQNQR   BCXAH   XOVXM   NAWLX   MNKAN   JTRWP
JWMCQ   NQRBC   XAHXO   LXVYD   CNABJ   ANLXC
NAVRW   XDB
```

5. The following was enciphered with a Caesar cipher. By running the alphabet on the first 5-letter block determine the shift and decipher the message.

```
dvysk dhyad vthyr zhjoh unlpu jyfwa vsvnf hsaov
bnoao lylhy lleht wslzv mthao lthap jphuz zabkf
punjv klzhu kjpwo lyzao yvbno vbaop zavyf dvysk
dhyad vthyr zaolw vpuah adopj ojpwo lyjby lhbzi
lnhua vyljy bpath aolth apjph uzmvy aolpy wyvis
ltzvs cpunh ipspa plz
```

6. The following was encrypted with a Caesar cipher. By frequency analysis determine the key and decrypt the message.

```
xojuu cqnya xkunv bnwlx dwcna nmrwk antjr wpcqn
uxanw ilxmn borwm rwpcq nfqnn uyjcc cnawb fjbcq
nvxbc mroor lduc
```

7. Search through the following ciphertext that is known to have been encrypted with a Caesar cipher and find an encrypted version of the word `the`. Determine the key and recover the plaintext.

```
IBYBC   KBSJS   BHCRO   MWGHV   SGDMK   OFHVO
HFOUS   ROHHV   SHCDC   THVSK   CFZR
```

8. Find a word (of more than one letter) and a Caesar cipher that translates that word into another word.

9. Here is an example of using a crib. The following is ciphertext of a message in German. We have reason to expect that the title *Generalleutnant der Waffen—SS* appears in the message. Try to use the pattern of letters in the crib to locate the ciphertext for this phrase in the message. Cryptanalyze as much of the message as you can. Exhibit as much of the key as you can. Remember, the plaintext is written in German; so, unless you know German and can figure out the remaining letter correspondences from context, you will only be able to get the letter correspondences that occur from the crib. This has **not** been encrypted with a Caesar cipher, but it has been encrypted with a simple substitution cipher.

```
ERKKG BCJJO RXCPB OBGOR OBEZZ OCTRE RTFOB UEXXO
RKKNO BGOBN OBZYR KKPEC JTEIT IYTFO BNYTT OCIEN
KJBEW POGOI EKKOU OYTOB GENO
```