Fall 2019
Chris Christensen
MAT/CSC 483
Section two

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*

Cryptography of Caesar Ciphers

Here is the key for a MASC:

Plaintext letters:	abcdefghijklmnopqrstuvwxyz
Ciphertext letters:	YNROTKMCPBDVXZALEWUSFQJHGI

The cipher alphabet has been randomly generated.

Could you remember the key? There are no patterns; so, it is hard to remember. You would probably need a written copy of the key. But, having a written copy of the key could lead to the key being lost or stolen. It is desirable to have a key that need not be written down.

Caesar's cipher, to which reference was made in the David Kahn quote above, was a MASC with 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

Notice that the cipher alphabet "wraps back on itself" – the letters that "fell off" the left-hand side returned to the right-hand side.

The key can be easily be recalled because it is only necessary to remember the shift.

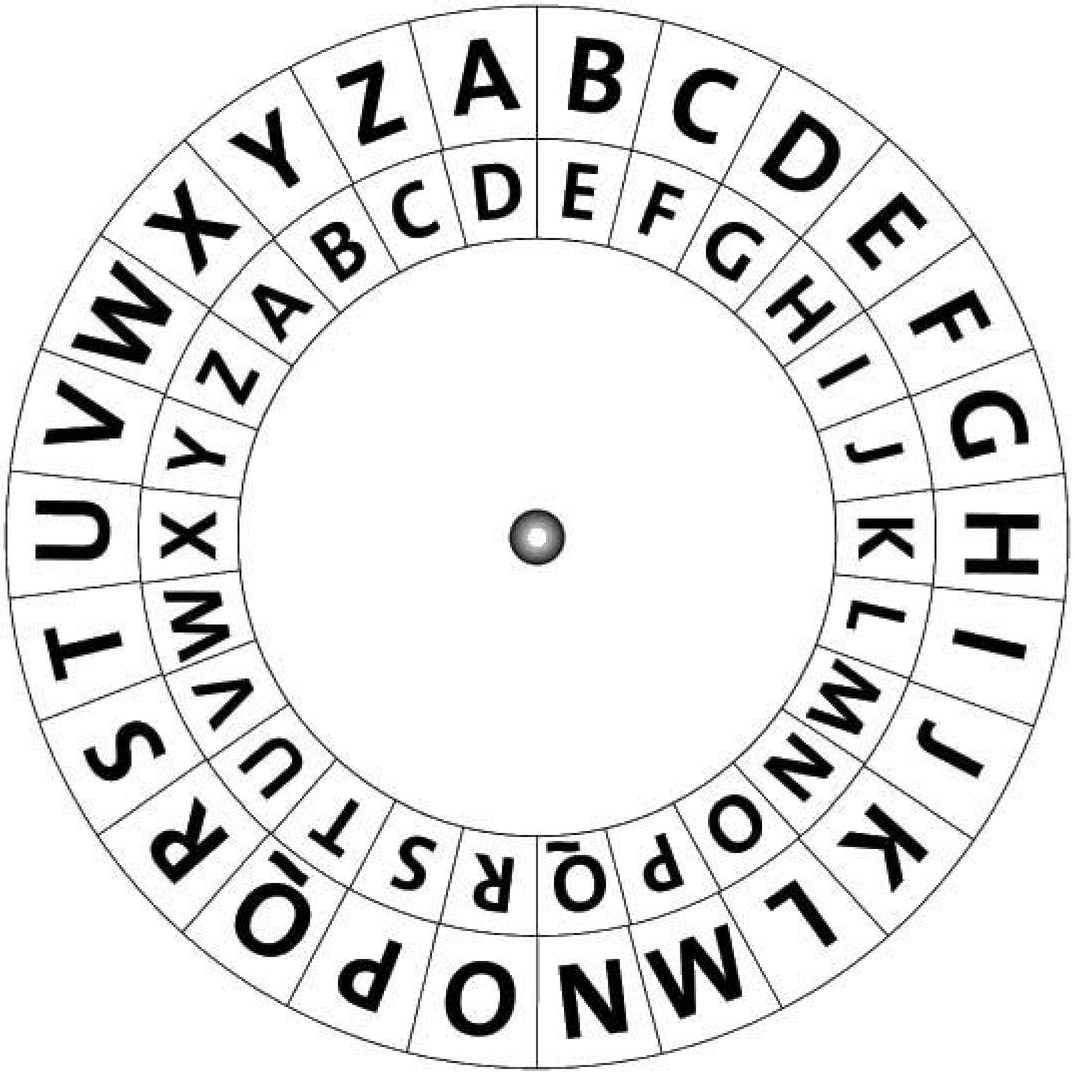
Of course, other shifts could be used. All such shift, or translation, ciphers are now usually called Caesar ciphers. Here is the key for a Caesar cipher with shift 8:

Plaintext letters	abcdefghijklmnopqrstuvwxyz
Ciphertext letters	IJKLMNOPQRSTUVWXYZABCDEFGH

For each of these ciphers, the method of encryption is the Caesar cipher, and the key is determined by the shift.

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

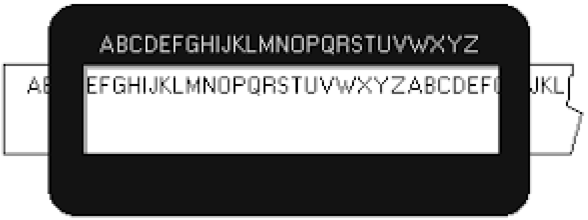
The idea of a cipher disk goes back to the Italian cryptologist Leon Battista Alberti (1404 – 1472), who is called the Father of Western Cryptology.


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

If on the disk above, we think of plaintext as being on the outside disk and ciphertext as being on the inside disk; then the disk has been set to Caesar's original cipher – a shift of 3.

Cipher disks exploit the fact that the ciphertext alphabet wraps back on itself.

The Dutch cryptologist Auguste Kerkhoffs (1835 – 1903) named the cryptographic slide the St. Cyr slide after the French national military academy where it was taught.



St. Cyr cipher device

If on the slide above, we think of plaintext as being on the top and ciphertext on the slide; then the slide has been set to a Caesar cipher with a shift of 5.

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

The Caesar Cipher is an Algebraic Cipher

The Caesar cipher can be thought of as an algebraic cipher. A shift to the right of three spaces, for example, can be symbolized as $C = p + 3$ where p represents a plaintext letter and C represents the corresponding ciphertext letter. More generally, a shift of *b* spaces to the right can be symbolized by $C = p + b$. The Caesar cipher can be described as C = p + key.

Of course, to make sense of this transformation, first, we must introduce numbers for the letters of the alphabet. Computer scientists would probably prefer a = 00, …, z = 25, but for the next several ciphers that we will discuss, there are reasons to prefer the numbering:

01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
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

Notice that we must make provision for sums larger than 26. For example, what happens to plaintext x (24) when we shift 3 places to the right?

When we write down the key, we do “the obvious” – we wrap back to the beginning of the alphabet.

Plaintext letters	abcdefghijklmnopqrstuvwxyz
Ciphertext letters	DEFGHIJKLMNOPQRSTUVWXYZABC

Mathematically, we accomplish the same thing by doing addition modulo 26; i.e., if a number is larger than 26, we divide 26 and take the remainder. For example, then 24 + 3 = 1.

Mathematicians usually write this as 24 + 3 = 1 mod 26. Often in programs this is written as 24 + 3 %26.

So `x` (24) shifts to `A`(1).

Caesar Cipher Example

Here is an example of a Caesar cipher with a shift of 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

Let us use the Caesar cipher with a shift of 5 to encrypt the plaintext message:

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

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
```

Decryption of a Message Encrypted with a Caesar Cipher

Of course, it is necessary to be able to decrypt any message that has been encrypted.

If a message were encrypted with a Caesar cipher with a shift of 5, then the enciphering key would be:

Plaintext letters	abcdefghijklmnopqrstuvwxyz
Ciphertext letters	FGHIJKLMNOPQRSTUVWXYZABCDE

The same key could be used to decrypt a message, but it might be more useful to place the ciphertext alphabet in alphabetical order. Here the corresponding "deciphering key:"

Ciphertext letters	ABCDEFGHIJKLMONPQRSTUVWXYZ
Plaintext letters	vwxyzabcdefghijklmnopqrstu

Let's consider what is happening mathematically.

What undoes addition modulo 26? Well, "subtraction mod 26," but remember that subtraction is just "adding the additive inverse." What undoes addition of 3 modulo 26 is addition of 23 modulo 26 because 3 + 23 = 26 = 0 mod 26. If we shift to the right by 3 and then by 23, we have shifted to the right be 26 and returned to plaintext.

$$\text{plaintext}\xrightarrow{+3 \bmod 26}\text{CIPHERTEXT}\xrightarrow{+23 \bmod 26}\text{plaintext}$$

So, every (enciphering) key has a corresponding "deciphering key," which is sometimes called the key inverse.

Key	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
Key inverse	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	09	08	07	06	05	04	03	02	01	26

Notice that for the deciphering key given above

Ciphertext letters	`ABCDEFGHIJKLMONPQRSTUVWXYZ`
Plaintext letters	`vwxyzabcdefghijklmnopqrstu`

Plaintext can be recovered from ciphertext by: $p = C + 23 \bmod 26$.

Cryptanalysis Using a Brute Force Attack

Unfortunately, Caesar ciphers have a small key space; there are only 26 possible keys (shifts), and one of those is plaintext (a shift of 0 or 26). So messages encrypted with Caesar ciphers can be easily broken by brute force – by trying all possible keys.

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
```

Here is a **brute force attack**.

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 ciphertext 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."

String	Shift forward by
VRRQS	0
WSSRT	1
XTTSU	2
YUUTV	3
ZVVUW	4
AWWVX	5
BXXWY	6
CYYXZ	7
DZZYA	8
EAAZB	9
FBBAC	10
GCCBD	11
HDDCE	12
IEEDF	13
JFFEG	14
KGGFH	15
LHHGI	16
MIIHJ	17
NJJIK	18
OKKJL	19
PLLKM	20
QMMLN	21
RNNMO	22
SOONP	**23**
TPPOQ	24
UQQPR	25

Now scan the column for something that makes sense. Notice near the bottom SOONP. This line corresponds to shifting the ciphertext alphabet to 23 places to recover plaintext; so, the key inverse is 23. Therefore, the 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:

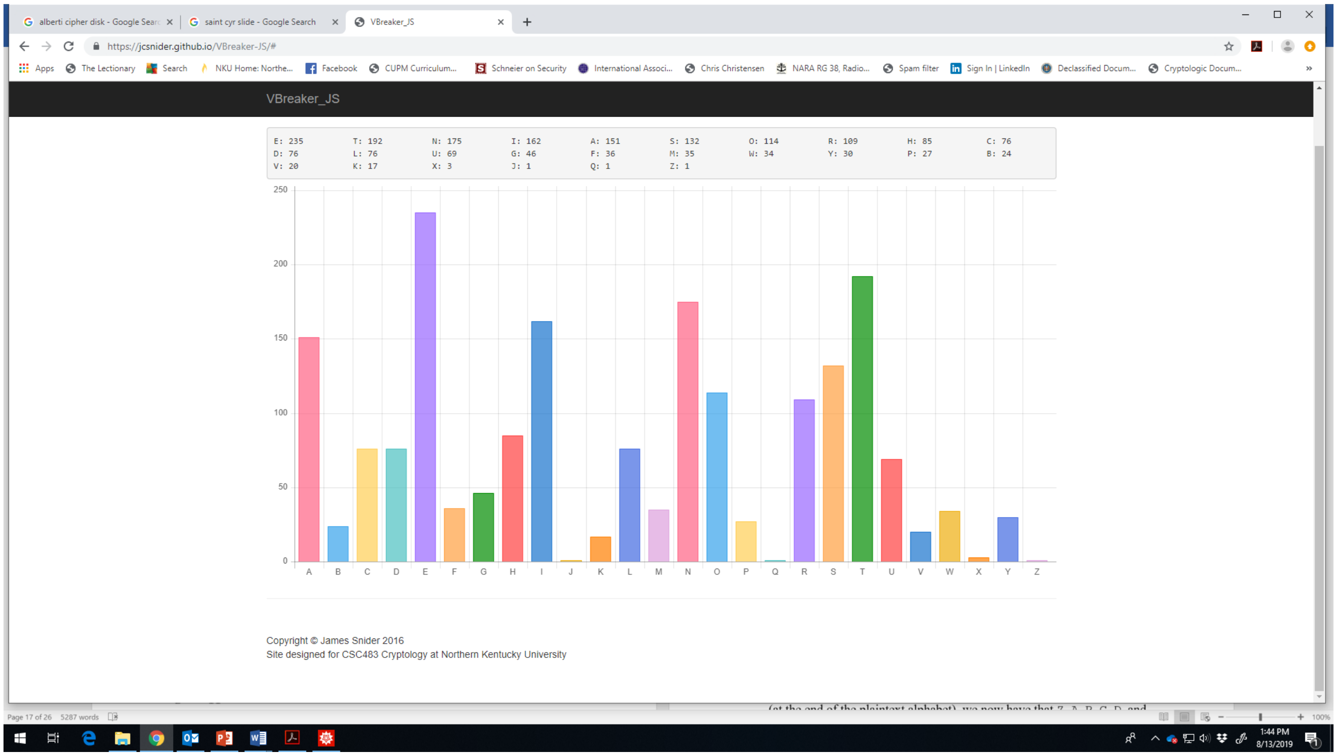
FGW GWF WFM FMF MFR FRX RXN XNS NSP SPT PTA TAK AKN
KNW NWX WXY XYB YBT BTW TWP WPJ PJI JIF IFX FXF XFH
FHW HWD WDU DUY UYT YTQ TQT QTL TLN LNX NXY XYB YBN
BNY NYM MYM **YMJ** MJB JBF BFW FWI WIJ IJU JUF UFW FWY
WYR YRJ RJS JSY.

The trigraph in bold is the encrypted with an additive key of 5. That suggests that the shift is 5. Assuming a shift of 5, the message decrypts.

Cryptanalysis using a Ciphertext Attack

Suppose that we are given ciphertext but do not know the method of encryption. Using **frequency analysis**, it is both easy to recognize that a message has been encrypted with a Caesar cipher and to determine the shift.

Here are the letter frequencies for typical plaintext message in English.



VBreaker_JS

```
E: 235    T: 192    N: 175    I: 162    A: 151    S: 132    O: 114    R: 109    H: 85     C: 76
D: 76     L: 76     U: 69     G: 46     F: 36     M: 35     W: 34     Y: 30     P: 27     B: 24
V: 20     K: 17     X: 3      J: 1      Q: 1      Z: 1
```

Copyright © James Snider 2016
Site designed for CSC483 Cryptology at Northern Kentucky University

Abraham Sinkov (who was a U. S. Army cryptanalysts during World War II) in his text *Elementary Cryptanalysis: A Mathematical Approach* points out the following patterns which are useful for cryptanalysis:

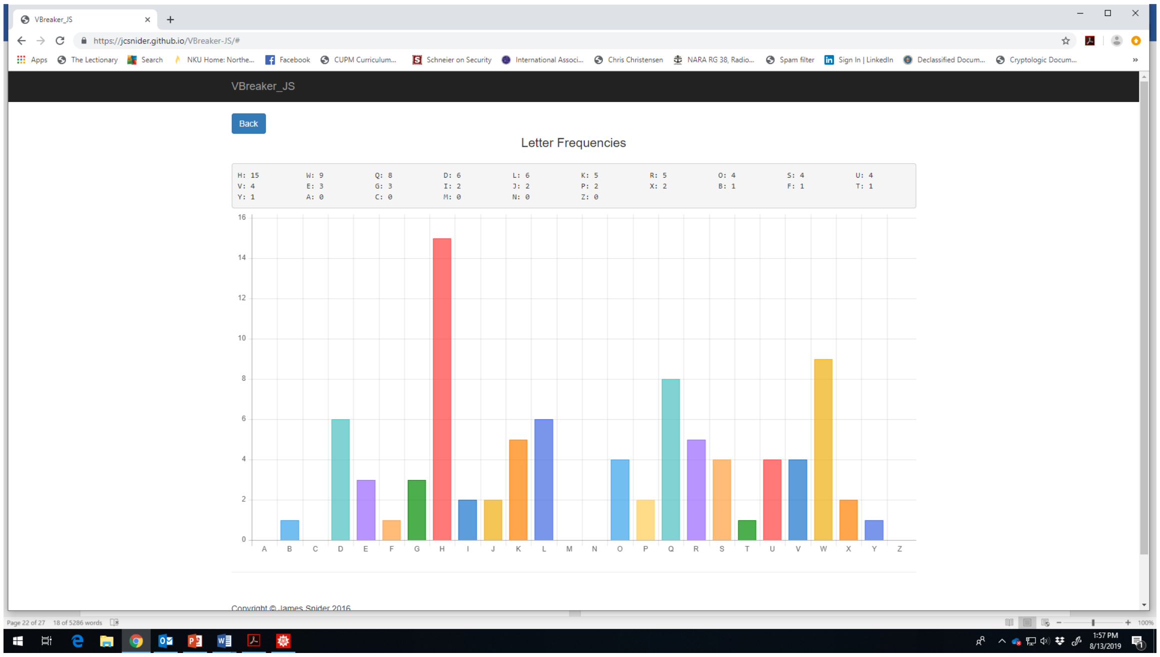
1. `a`, `e`, and `i` are all high frequency letters (at the beginning of the plaintext alphabet), and they are equally spaced (four letters apart) with `e` the most frequent.
2. `n` and `o` form a high frequency pair (near the middle of the plaintext alphabet).
3. `r`, `s`, and `t` form a high frequency 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, a ciphertext message that has been encrypted with a Caesar cipher should show the "usual" frequencies of the letters in English – but shifted to the right.

Here is a ciphertext message that has been encrypted with an unknown cipher:

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

Here are its ciphertext frequencies:



```
H: 15     W: 9     Q: 8     D: 6     L: 6     K: 5     R: 5     O: 4     S: 4     U: 4
V: 4      E: 3     G: 3     I: 2     J: 2     P: 2     X: 2     B: 1     F: 1     T: 1
Y: 1      A: 0     C: 0     M: 0     N: 0     Z: 0
```

Notice that the frequencies look like the “usual” English frequencies but shifted to the right. It appears that the method of encryption is Caesar cipher. The frequencies suggest that `H = e`.

It is only necessary to determine one correspondence between a plaintext and ciphertext letter to determine the key. The frequency patterns suggest that `H = e`; so, 8 = 5 + additive key. The additive key is 3, which is correct.