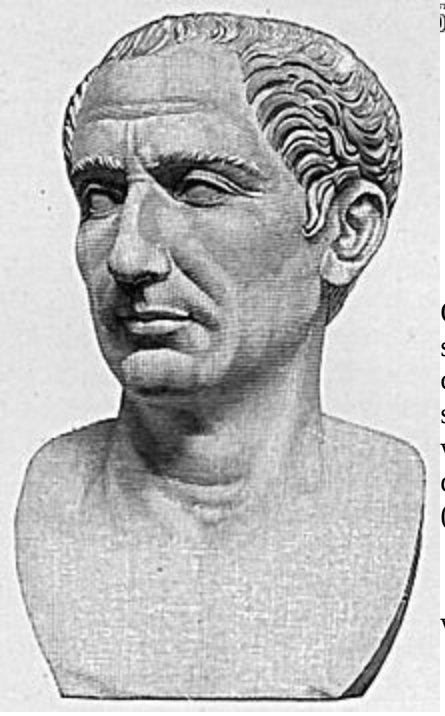


The Caesar Cipher

One of the simplest examples of a substitution cipher is the *Caesar cipher*, which is said to have been used by Julius Caesar to communicate with his army. Caesar is considered to be one of the first persons to have ever employed encryption for the sake of securing messages. Caesar decided that shifting each letter in the message would be his standard algorithm, and so he informed all of his generals of his decision, and was then able to send them secured messages. Using the Caesar Shift (3 to the right), the message,

"RETURN TO ROME"

would be encrypted as,

"UHWXUA WR URPH"

In this example, 'R' is shifted to 'U', 'E' is shifted to 'H', and so on. Now, even if the enemy did intercept the message, it would be useless, since only Caesar's generals could read it.

$$E_n(x) = (x + n) \mod 26.$$
$$D_n(x) = (x - n) \mod 26.$$

Thus, the Caesar cipher is a *shift cipher* since the ciphertext alphabet is derived from the plaintext alphabet by shifting each letter a certain number of spaces. For example, if we use a shift of 19, then we get the following pair of ciphertext and plaintext alphabets:

Plaintext: 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
Ciphertext: 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

To encipher a message, we perform a simple substitution by looking up each of the message's letters in the top row and writing down the corresponding letter from the bottom row. For example, the message

THE FAULT, DEAR BRUTUS, LIES NOT IN OUR STARS BUT IN OURSELVES.

would be enciphered as

MAX YTNEM, WXTK UKNMNL, EBXL GHM BG HNK LMTKL UNM BG HNKLXEOXL.

Essentially, each letter of the alphabet has been shifted nineteen places ahead in the alphabet, wrapping around the end if necessary. Notice that punctuation and blanks are not enciphered but are copied over as themselves.

Breaking a Caesar Cipher (Cryptanalysis)

Can a computer guess what shift was used in creating a Caesar cipher? The answer, of course, is yes. But how does it work?

The unknown shift is one of 26 possible shifts. One technique might be to try each of the 26 possible shifts and check which of these resulted in readable English text. But this approach has limitations. The main problem is that the computer would need a comprehensive dictionary in order to be able to recognize the words of any given cryptogram.

A better approach makes use of statistical data about English letter frequencies. It is known that in a text of 1000 letters of various English alphabet occur with about the following relative frequencies:

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
73	9	30	44	130	28	16	35	74	2	3	35	25	78	74	27	3	77	63	93	27	13	16	5	19	1

This information can be useful in deciding the most likely shift used on a given enciphered message.

Suppose the enciphered message is:

K DKVO DYVN LI KX SNSYD, PEVV YP CYEXN KXN PEBI, CSQXSPISXQ XYDRSXQ.

We can tally the frequencies of the letters in this enciphered message, thus

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
0	1	2	4	3	0	0	0	3	0	4	1	0	4	1	4	3	1	6	0	0	4	0	7	4	0

Now we can now shift the two tallies so that the large and small frequencies from each frequency distribution match up roughly. For example, if we try a shift of ten on the previous example, we get the following correspondence between English language frequencies and the letter frequencies in the message.

English Language Frequencies

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
73	9	30	44	130	28	16	35	74	2	3	35	25	78	74	27	3	77	63	93	27	13	16	5	19	1

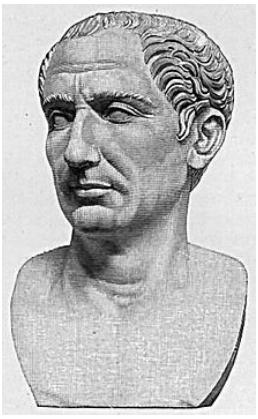
Enciphered Message Frequencies

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
4	1	0	4	1	4	3	1	6	0	0	4	0	7	4	0	0	1	2	4	3	0	0	0	3	0

Note that in this case the large frequencies for cipher X and Y correspond to large for English N and O, the bare spots for cipher T and U correspond to bare spots for English J and K. Also, an isolated large frequency for cipher S correpsonds to a similar one for English I. In view of this evidence we needn't even worry too much about the drastic mismatch for English E, which is usually the most frequent letter in a random sample of English text. If we now apply this substitution to the message we get:

A TALE TOLD BY AN IDIOT, FULL OF SOUND AND FURY, SIGNIFIYING NOTHING.

History and usage



The Caesar cipher is named for Julius Caesar, who used an alphabet with a left shift of three.

The Caesar cipher is named after Julius Caesar, who, according to Suetonius, used it with a shift of three to protect messages of military significance. While Caesar's was the first recorded use of this scheme, other substitution ciphers are known to have been used earlier.

If he had anything confidential to say, he wrote it in cipher, that is, by so changing the order of the letters of the alphabet, that not a word could be made out. If anyone wishes to decipher these, and get at their meaning, he must substitute the fourth letter of the alphabet, namely D, for A, and so with the others.
—Suetonius, *Life of Julius Caesar 56*

His nephew, Augustus, also used the cipher, but with a right shift of one, and it did not wrap around to the beginning of the alphabet:

Whenever he wrote in cipher, he wrote B for A, C for B, and the rest of the letters on the same principle, using AA for X.
—Suetonius, *Life of Augustus 88*

There is evidence that Julius Caesar used more complicated systems as well, and one writer, Aulus Gellius, refers to a (now lost) treatise on his ciphers:

There is even a rather ingeniously written treatise by the grammarian Probus concerning the secret meaning of letters in the composition of Caesar's epistles.
—Aulus Gellius, *Attic Nights 17.9.1–5*

It is unknown how effective the Caesar cipher was at the time, but it is likely to have been reasonably secure, not least because most of Caesar's enemies would have been illiterate and others would have assumed that the messages were written in an unknown foreign language. There is no record at that time of any techniques for the solution of simple substitution ciphers. The earliest surviving records date to the 9th century works of Al-Kindi in the Arab world with the discovery of frequency analysis.

A Caesar cipher with a shift of one is used on the back of the Mezuzah to encrypt the names of God. This may be a holdover from an earlier time when Jewish people were not allowed to have Mezuzahs. The letters of the cryptogram themselves comprise a religiously significant "divine name" which Orthodox belief holds keeps the forces of evil in check.

In the 19th century, the personal advertisements section in newspapers would sometimes be used to exchange messages encrypted using simple cipher schemes. Kahn (1967) describes instances of lovers engaging in secret communications enciphered using the Caesar cipher in The Times. Even as late as 1915, the Caesar cipher was in use: the Russian army employed it as a replacement for more complicated ciphers which had proved to be too difficult for their troops to master; German and Austrian cryptanalysts had little difficulty in decrypting their messages.

Caesar ciphers can be found today in children's toys such as secret decoder rings. A Caesar shift of thirteen is also performed in the ROT13 algorithm, a simple method of obfuscating text widely found on Usenet and used to obscure text (such as joke punchlines and story spoilers), but not seriously used as a method of encryption.

The Vigenère cipher uses a Caesar cipher with a different shift at each position in the text; the value of the shift is defined using a repeating keyword. If the keyword is as long as the message, chosen random, never becomes known to anyone else, and is never reused, this is the one-time pad cipher, proven unbreakable. The conditions are so difficult they are, in practical effect, never achieved. Keywords shorter than the message (e.g., "Complete Victory" used by the Confederacy during the American Civil War), introduce a cyclic pattern that might be detected with a statistically advanced version of frequency analysis.

In April 2006, fugitive Mafia boss Bernardo Provenzano was captured in Sicily partly because some of his messages, written in a variation of the Caesar cipher, were broken. Provenzano's cipher used numbers, so that "A" would be written as "4", "B" as "5", and so on.

In 2011, Rajib Karim was convicted in the United Kingdom of "terrorism offences" after using the Caesar cipher to communicate with Bangladeshi Islamic activists discussing plots to blow up British Airways planes or disrupt their IT networks. Although the parties had access to far better encryption techniques (Karim himself used PGP for data storage on computer disks), they chose to use their own scheme instead (implemented in Microsoft Excel) "because 'kaffirs', or non-believers, know about it [ie, PGP] so it must be less secure".

Sources :

1. http://www.cs.trincoll.edu/~crypto/historical/caesar.html

2. http://en.wikipedia.org/w/index.php?title=Caesar_cipher&printable=yes

Simple Substitution Cipher

Introduction

The simple substitution cipher is a cipher that has been in use for many hundreds of years. It basically consists of substituting every plaintext character for a different ciphertext character. It differs from Caesar cipher in that the cipher alphabet is not simply the alphabet shifted, it is completely jumbled. The simple substitution cipher offers very little communication security, and it will be shown that it can be easily broken even by hand, especially as the messages become longer (more than several hundred ciphertext characters).

Example

Here is a quick example of the encryption and decryption steps involved with the simple substitution cipher. The text we will encrypt is 'defend the east wall of the castle'.
Keys for the simple substitution cipher usually consist of 26 letters (compared to the caeser cipher's single number). An example key is:

plain alphabet : `abcdefghijklmnopqrstuvwxyz`

cipher alphabet: `phqgiumeaylnofdxjkrcvstzwb`

An example encryption using the above key:

plaintext : `defend the east wall of the castle`

ciphertext: `giuifg cei iprc tpnn du cei qprcni`

It is easy to see how each character in the plaintext is replaced with the corresponding letter in the cipher alphabet. Decryption is just as easy, by going from the cipher alphabet back to the plain alphabet. When generating keys it is popular to use a key word, e.g. 'zebra' to generate it, since it is much easier to remember a key word compared to a random jumble of 26 characters. Using the keyword 'zebra', the key would become:

cipher alphabet: `zebracdfghijklmnopqstuvwxy`