POLYALPHABETIC CIPHERS

Substitution is made from many alphabet sequencies:

1. A set of related monoalphabetic substitution rules is used
2. A key determines what particular rule is chosen for a given transformation

VIGENERE CIPHER

The best known and one of the simplest is Vigenere cipher. The **Vigenère cipher** is a method of encryption invented by Giovan Batista Belaso and described in his 1553 book *La cifra del. Sig. Giovan Batista Belaso*. It was misattributed to Blaise de Vigenère in the 19th century, and given his name (http://en.wikipedia.org/wiki/Vigen%C3%A8re_cipher).

In this scheme, the set of related monoalphabetic substitution rules consists of the 26 Caesar ciphers, with shifts from 0 to 25. Each cipher is denoted by a key letter, which is the ciphertext letter that substitutes for the plaintext letter a. Thus, a Caesar cipher with a shift 3 is denoted by the key value d.

A matrix known as Vigenere tableau is used:

VIGENERE CIPHER (CONT 1)

	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

Each of the 26 ciphers is laid out horizontally, with the key letter for each cipher to the left. The encryption process:

VIGENERE CIPHER (CONT 2)

Given a key letter x and a plaintext letter y, the ciphertext letter is at the intersection of the row labeled x and the column labeled y; in this case the ciphertext letter is V. To encrypt a message, a key is needed that is as long as the message. Usually, a key is a repeating keyword. For example, if the keyword is *deceptive,* the message “we are discovered save yourself” is encrypted as follows:

Key: dec e p t i v e de c e p t i ve d e c e p t i v e
Plaintext: wea r e d i s c ov e r e d s av e y o u r s e l f
Ciphertext: ZICVTWQNGRZGVTWAVZHCQYGLMGJ

Decryption is equally simple. The key letter identifies the row. The position of ciphertext letter in that row determines the column, and the plaintext is at the top of that column.

In spite of using multiple alphabets, some frequency information is preserved in Vigenere ciphertext.

Let’s sketch a method of breaking this cipher.

Suppose that the opponent believes that the ciphertext was encrypted using either monoalphabetic substitution or a Viginere cipher. A simple test can be made to make a determination. If a monoalphabetic substitution is used, then the statistical properties of the ciphertext should be the same as that of the language of the plaintext. If, on the other hand, a Viginere cipher is suspected, then progress depends on determining the length of the keyword, as it will be seen in a moment. How keyword length can be determined? If 2 identical sequences of the plaintext letters occur at a distance of integer multiple of the keyword length, they will generate identical ciphertext sequences. In our example, 2 instances of the sequence “red” are separated by 9 character positions. Consequently, in both cases, r is encrypted using key letter e, e is encrypted using key letter p, and d is encrypted using key letter t. Thus, in both cases ciphertext is VTW. Analyst may make assumption, that keyword length is either 3, either 9. Having long enough messages, cryptanalyst can determine keyword length definitely by finding common factor of all displacements of such sequences.

If keyword length is N, then the cipher consists of N monoalphabetic substitution ciphers. For example, with the keyword DECEPTIVE, the letters in positions 1, 10, 19, and so on, are all encrypted with the same monoalphabetic cipher. Thus, we can use the known frequency characteristics of the plaintext language to attack each of the monoalphabetic ciphers separately.

The periodic nature of the keyword can be eliminated by using a non-repeating keyword that is as long as message. Vigenere proposed in 1585

VIGENERE CIPHER (CONT 3)

(http://encyclopedia.thefreedictionary.com/Blaise%20de%20Vigen%E8re) what is referred to as autokey system, in which a keyword is concatenated with the plaintext itself to provide a running key. For example,

Key: dec e p t i v e we are d i s c o ver e d s av
Plaintext: wea r e d i s c o v ered s av e you r s e l f
Ciphertext: ZICVTWQNGKZEIIGASXSTSLVVWLA

Even this scheme is vulnerable to cryptanalysis, because the key and the plaintext share the same frequency distribution of letters, a statistical technique can be applied.

The ultimate defense against such a cryptanalysis is to choose a keyword that is as long as the plaintext and has no statistical relationship to it. Such a system was introduced by an AT&T engineer Gilbert Vernam in 1918. His system works on binary data rather than letters. The system can be expressed succinctly as follows:

$$c_i = p_i \oplus k_i ,$$

where Ci- ith binary digit of ciphertext, Pi – of the plaintext, Ki – of the key, $\oplus$ - exclusive or (XOR) operation

Decryption is made by

$$p_i = c_i \oplus k_i$$

Keyword here is long enough but repeating. It can be broken with the use of known plaintext sequences.

ONE-TIME PAD

An US Army Signal Corps Captain, Joseph Mauborgne, in 1918, proposed an improvement (http://en.wikipedia.org/wiki/One-time_pad) to Vernam cipher that yields the ultimate in security. He suggested using of a random key that was truly as long as the message, with no repetitions. Such a scheme, known as one-time pad, is unbreakable. It produces random output that bears no statistical relationship to the plaintext. Because the ciphertext contains no information whatsoever about the plaintext, there is no way to break the code.

Let's consider example. Suppose that we are using a Vigenere scheme with 27 characters in which the 27th character is the space character, but with a one-time key that is as long as the message. Thus, Vigenere tableau must be expanded to 27x27. Consider the ciphertext:

ANKYODKYUREPFJBYOJDSPLREYIUNOFDOIUERFPLUYTS

We now show 2 different decryptions using 2 different keys:

Ciphertext: ANKYODKYUREPFJBYOJDSPLREYIUNOFDOIUERFPLUYTS
Key: px l mv ms yd o f t y rv zwc t nl e b n ecv g du p a hf zz l mn y ih
Plaintext: mr mu s t ar d with th e can d l est i ck i n th e h a ll

ONE-TIME PAD (CONT 1)

and

Ciphertext: ANKYODKYUREPFJBYOJDSPLREYIUNOFDOIUERFPLUYTS
Key: mf u g p mi yd g axg ou f hkl l l mhs qd qo g t e wbqf q yov u h wt
Plaintext: mi s s s c ar l e t wi t h t he kn if e i n t he l ib r a r y

If cryptanalyst would manage to find these keys, he will not be able to decide which key is correct, and which plaintext is correct. Randomness of the key leads to randomness of the ciphertext, and therefore cryptanalyst can't use patterns or regularities to attack the ciphertext, the code is unbreakable.

But in practice, one-time pad has 2 fundamental difficulties:

- there is the practical problem of making large quantities of random keys. Any heavily used system might require millions of random characters on a regular basis. Supplying truly random characters in this volume is a significant task
- even more daunting is the problem of key distribution and protection. For every message to be sent, a key of equal length is needed by both sender and receiver. Thus, a mammoth key distribution problem exists.

TRANSPOSITION TECHNIQUE

Another approach to enciphering is usage of transpositions, or permutations on the plaintext letters. The simplest such cipher is the rail fence technique, in which the plaintext is written down as a sequence of diagonals and then read off as the sequence of rows. For example, to encipher the message "meet me after the toga party" with a rail fence of depth 2, we write

m e m a t r h t g p r y
e t e f e t e o a a t

The encrypted message is
MEMATRHTGPRYETEFETEOAAAT

A more complex scheme is to write the message in a rectangle, row by row, and read the message off, column by column, but to permute the order of columns. The order of columns then becomes the key to the algorithm. For example,

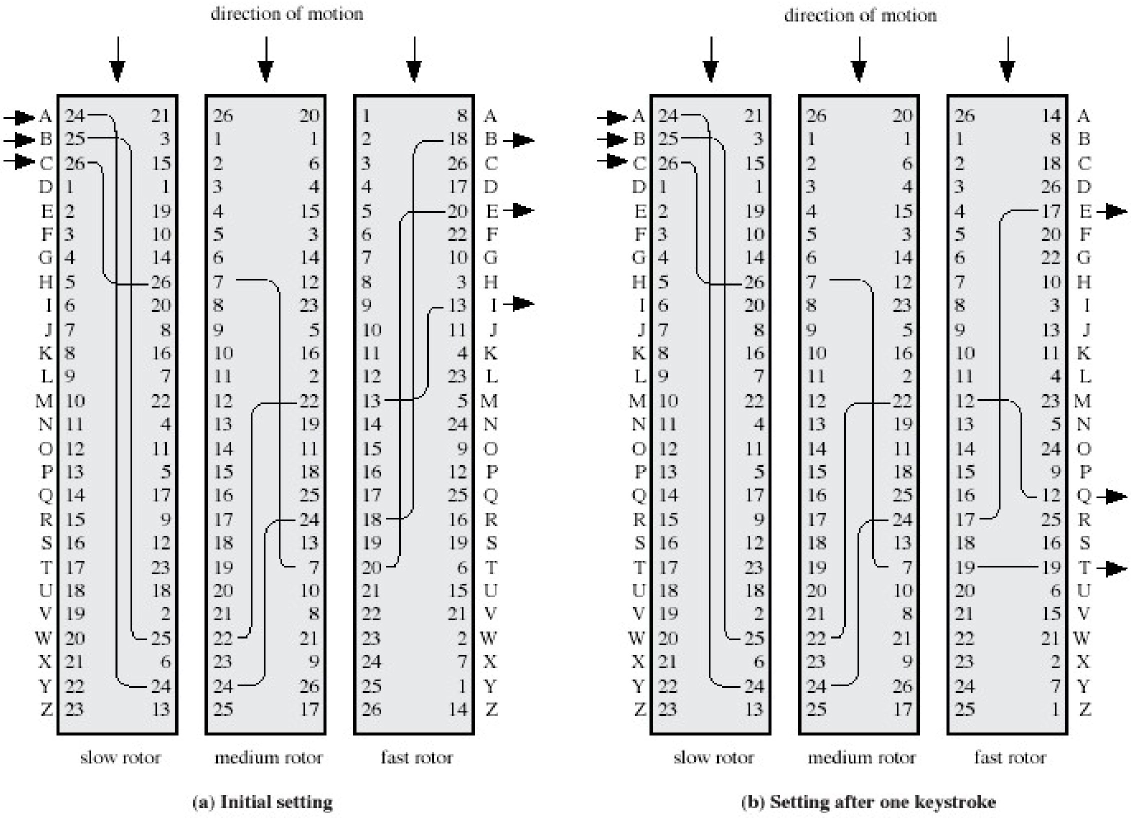
Key: 4 3 1 2 5 6 7
Plaintext: a t t a c k p
o s t p o n e
d u n t i l t
w o a m x y z
Ciphertext: TTNAAPTMTSUOAODWCOIXKNLYPETZ

A pure transposition cipher is easily recognized because it has the same letter frequencies as the original plaintext.

The transposition cipher can be made more secure by performing more than 1 transposition

ROTOR MACHINES

Several stages of encryption (substitution, transposition) can produce more secure algorithm. Before the introduction of DES, the most important application of multiple stages of encryption was a class of systems known as rotor machines (invented by two Dutch Naval officers Teo A. van Hangel and R.P.C.Spengler in 1915, http://encyclopedia.thefreedictionary.com/Rotor+cypher+machine).



(a) Initial setting

(b) Setting after one keystroke

Figure 2.7 Three-Rotor Machine With Wiring Represented by Numbered Contacts

The machine consists of a set of independently rotating cylinders through which electrical pulses can flow. Each cylinder has 26 input and 26 output pins, with internal wiring that connects each input pin to a unique output pin. For simplicity, only 3 of the internal connections in each cylinder are shown. If we associate each input and output pin with a letter of the alphabet, then a single cylinder defines a monoalphabetic substitution. For example, if an operator depresses the key for letter A, an electric signal is applied to the 1st pin of the 1st cylinder and flows to the 25th output pin.

Consider a machine with a single cylinder. After each input key is depressed, the cylinder rotates one position, so that the internal connections

ROTOR MACHINES (CONT 1)

are shifted accordingly. Thus, a different monoalphabetic cipher is defined. After 26 letters of the plaintext, the cylinder would be back to the initial position. Thus, we have a polyalphabetic substitution algorithm with a period of 26.

The power of the rotor machine is in the use of multiple cylinders, in which the output pins of one cylinder are connected to the input pins of the next. In Fig. 2.7, plaintext letter A (1st pin) is routed through 3 cylinders to the output 2nd pin (ciphertext letter B).

With multiple cylinders, the one farthest from the operator input rotates one pin position with each keystroke. The right half of Fig. 2.7 shows the system configuration after a single keystroke. For every complete rotation of the outer cylinder, the middle cylinder rotates one pin position. Finally, for every complete rotation of the middle cylinder, the inner cylinder rotates one pin position. This is similar to odometer, or to arrows in watches. The result is that there 26x26x26=17576 different substitution alphabets used before the system repeats. The significance of the rotor machine today is that it points the way to the most widely used cipher – DES.

STEGANOGRAPHY

The methods of steganography conceal the existence of the message, whereas the methods of cryptography render the message unintelligible to outsiders by various transformations of the text.

Use of:

The 1st letters of some text spells out the hidden message

Character marking (by pencil, pin punctures) of necessary letters in the text

Invisible inks

Typewriter correction ribbon: print between lines printed with a black ribbon – visible only under a strong light

Least significant bit of each 24-bit pixel

Steganography requires a lot overhead to hide a relatively few bits of information. Once the system is discovered, it becomes worthless. But a message can be at first encrypted, then hidden.

The advantage of steganography is that it can be employed by parties wanting to hide the fact of their secret communication