What is Vigenère Square Encryption?

To answer that question, let's look at what is probably the most simplest and famous type of encryption technique. The Caesar shift substitution cipher. A Caesar substitution is very straight forward. As the name suggests, it is a substitution cipher, which means that every character of the original text is substituted for from the cipher text (the re-ordered text). How is the reference cipher text obtained? That's where the shift comes in. Rolling the plain-text alphabet by a fixed shift of one or more characters gives you the corresponding Caesar shift cipher text. For example, shifting the plain-text English alphabet by 23 characters, gives the following combination.

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
Cipher	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

Using this shift, we simply substitute a character from the cipher text for the corresponding plain-text character.

Now, why is this not a great way to encipher messages? Well, this type of enciphering can be broken by simple frequency analysis. The way that approach works is by determining the frequencies of each character in the ciphered text and then comparing the distribution to the frequencies of the standard plain-text English alphabet.

An improvement over this technique would be to encipher each character with a different mono-alphabetic cipher. If a Caesar shift of 13 was selected for the first character, then a Caesar shift of 20 could be used to encipher the second and so on. But a simply random choice of shifts would be too difficult to remember. Enter the Vigenère square. This technique was first described by Giovan Battista Bellaso in 1553 but was misattributed to Blaise de Vigenère.

The way the cipher works is very simple. First, you decide on a key that will be used to encrypt the plain-text or decrypt the enciphered text. Let us consider any word/phrase to be the key for simplicity's sake although this is not the only way to implement it. The next step involves using a Vigenère square which looks like the following.

	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

What you would now do is look at the first character of your key, say “J” in “JUICYSMOOYAY”. Then, look up “J” in the left column of the Vigenère square. Look up the plain-text alphabet in the top row of the Vigenère square. The element in the row corresponding to “J” and in the column corresponding to the plain-text alphabet would be the substitution character. Each row is essentially a mono-alphabetic Caesar shift of n, where n starts from 0. The first row is a zero shift row.

Now, to de-crypt it you follow essentially the same process. The message receiver knows the key told to him/her by the message sender before hand. The message receiver then looks up the rows corresponding to the characters in the keyword and determines the plain-text equivalent.

The beauty of this technique is that messages encrypted in this way do not lend themselves to frequency analysis. Since a different mono-alphabetic substitution cipher is used for each character, the same character in the plain-text, say “e” might get encrypted as different characters in the encrypted text. Frequency analysis can only work if we are able to match the frequency of occurrence of cipher characters and compare it to the standard frequency distribution.

So is there a way to crack the impregnable Vigenère cipher?

Well yes, Charles Babbage the British cryptographer (and you probably know him as computation daddy) first cracked this cipher and his method was actually quite straight-forward.

I’ll try to keep the explanation concise but for the nitty-gritty details you can refer to the comments in the code itself.

Once you have the encrypted message with you, you first look for repeated groups of encrypted characters, maybe something like “JGHF” or “GUUBK” or stuff like that. How could these patterns have been repeated if the substitution cipher kept changing according to the key?

There are only two possibilities. Since the key keeps rotating among the rows of the square, either one entire rotation of “JUICYSMOOYAY” coincided with the repetition of the same plain-text word leading to the same encrypted string, or just the right substitution ciphers were chosen for the perfect combination of plain-text characters that resulted in the same encrypted string representing two entirely different strings from the plain-text. Although the latter is possible it is highly improbable.

Let’s forget “JUICYSMOOYAY” for the time being and imagine the keyword was something shorter like “GREEN” and say with the repetitions, you can now count the character gaps between each occurrence. Say the first encrypted string repeats twice with a gap of 10 characters, a second encrypted string repeats with a gap of 20 characters and the third encrypted string repeats with a gap of 15 characters. This implies that the keyword had to be of a length that was a common factor between these three gap values. The only common factor between these three numbers (aside from 1 of course) is 5. Therefore, we can conclude that the keyword must be five characters long.

Armed with the information that the keyword is five characters long, we can conclude that the same five rows have been repeatedly used to encipher the plain-text. If the same five rows have been repeatedly used, then it follows that the 1^{st}, 6^{th}, 11^{th} and so characters must have been encrypted with the same substitution shift. Similarly, the 2^{nd}, 7^{th}, 12^{th} and so characters must have been encrypted with the same substitution cipher corresponding to the second character of the keyword. Separating the cipher-

text characters into the groups corresponding to each possible substitution cipher, we can then perform frequency analysis on the groups separately to identify the shift. Each shift would then give us the character in the keyword that was used to encrypt the message.

If you thought this wasn't a very good explanation (and I kind of second that), you have to check out "The Code" by Simon Singh. The book is a great read and I was inspired to try my hand at it after reading the book. He does a great job of explaining it very simply. You could also probably find some information on Wikipedia and other online resources.

Sample Cipher Text and Solution

Cipher Text

KQOWEFVJPUJUUNUKGLMEKJINMWUXFQMKJBGWRLFNFGHUDWUUMBSVLPSNCMUE
KQCTESWREEKOYSSIWCTUAXYOTAPXPLWPNTCGOJBGFQHTDWXIZAYGFFNSXCSEYN
CTSSPNTUJNYTGGWZGRWUUNEJUUQEAPYMEKQHUIDUXFPGUYTSMTFFSHNUOCZGM
RUWEYTRGKMEEDCTVRECFBDJQCUSWVBPNLGOYLSKMTEFVJJTWWMFMWPNMEMT
MHRSPXFSSKFFSTNUOCZGMDOEOYEEKCPJRGPMURSKHFRSEIUEVGOYCWXIZAYGOSA
ANYDOEOYJLWUNHAMEBFELXYVLWNOJNSIOFRWUCCESWKVIDGMUCGOCRUWGNMA
AFFVNSIUDEKQHCEUCPFCMPVSUDGAVEMNYMAMVLFMAOYFNTQCUAFVFJNXKLNEI
WCWODCCULWRIFTWGMUSWOVMATNYBUHTCOCWFYTNMGYTQMKBBNLGFBTWOJFT
WGNTEJKNEEDCLDHWTVBUVGFBIJGYYIDGMVRDGMPLSWGJLAGOEEKJOFEKNYNOLR
IVRWVUHEIWUURWGMUTJCDBNKGMBIDGMEEYGUOTDGGQEUJYOTVGGBRUJYS

Solution

SOUVENTPOURSAMUSERLESHOMMESDEQUIPAGEPRENNENTDESALBATROSVASTESOIS
EAUXDESMERSQUISUIVENTINDOLENTSCOMPAGNONSDEVOYAGELENAVIREGLISSANT
SURLESGOUFFRESAMERSAPEINELESONTILSDEPOSESSURLESPLANCHESQUECESROISD
ELAZURMALADROITSETHONTEUXLAISSENTPITEUSEMENTLEURSGRANDESAILESBLA
NCHESCOMMEDESAVIRONSTRAINERACOTEDEUXCEVOYAGEURAILECOMMEILESTGAU
CHEETVEULELUINAGUERESIBEAUQUILESTCOMIQUEETLAIDLUNAGACESONBECAVEC
UNBRULEGUEULELAUTREMIMEENBOITANTLINFIRMEQUIVOLAITLEPOETEESTSEMBLA
BLEAUPRINCEDESNUEESQUIHANTELATEMPETEETSERITDELARCHERBAUDELAIREEXI
LESURLESOLAUMILIEUDESHUEESLEMOTPOURETAGEQUATREESTTRAJANSESAILESDE
GEANTLEMPECHENTDEMARCHER

The above de-crypted piece of text is a French poem L'Albatros by Charles Baudelaire.

Images from Wikipedia.