The first row shows the plaintext character to be encrypted. The first column contains the characters to be used by the key. The rest of the tableau shows the ciphertext characters.To find the ciphertext for the plaintext "she is listening" using the word "*PASCAL*" as the key, we can find "s" in the first row, "*P*" in the first column, the cross section is the ciphertext character "H". We can find "h" in the first row and "*A*" in the second column, the cross section is the ciphertext character "H". We do the same until all ciphertext characters are found.

Cryptanalysis of Vigenere Ciphers Vigenere ciphers, like all polyalphabetic ciphers, do not preserve the frequency of characters. However, Eve still can use some techniques to decipher an intercepted ciphertext. The cryptanalysis here consists of two parts: finding the length of the key and finding the key itself.

1. Several methods have been devised to find the length of the key. One method is discussed here. In the so-called **Kasiski test,** the cryptanalyst searches for repeated text segments, of at least three characters, in the ciphertext. Suppose that two of these segments are found and the distance between them is d. The cryptanalyst assumes that $d|m$ where m is the key length. If more repeated segments can be found with distances $d_1, d_2, \ldots, d_n$, then

$$gcd\,(d_1, d_2, \ldots, d_n) \mid m$$

This assumption is logical because if two characters are the same and are $k \times m$ ($k = 1, 2, \ldots$) characters apart in the plaintext, they are the same and $k \times m$ characters apart in the ciphertext. Cryptanalyst uses segments of at least three characters to avoid the cases where the characters in the key are not distinct. Example 3.19 may help us to understand the reason.

The Index of Coincidence (IC) method is often used to confirm the m value determined by the **Kasiski test**. The Index of Coincidence is defined as follows:

Definition The Index of Coincidence of $x = x_1x_2\ldots x_n$, which is a string of length n formed by the alphabets A, B, …, Z, is defined as the probability that the random elements of x are the same. Thus if the frequencies of A, B, …, Z in x are denoted by the $f_0, \ldots, f_{25}$,

$$I_c(x) = \frac{\sum \binom{f_i}{2}}{\binom{n}{2}} = \frac{\sum f_i(f_i - 1)}{n(n-1)} \approx \sum \left(\frac{f_i}{2}\right)^2 = \sum p_i^2$$

The Index of coincidence is an invariant for any shift cipher. This is because in a shift cipher, the individual probabilities will get permuted but the sum of the squares of the probabilities will remain constant, thus keeping the IC value invariant. For standard English language text, the value of IC is approximately 0.065. However, if all the letters are equally likely then the IC value is $26(1/26)^2 \approx 0.038$. Since these two values are quite far apart, the IC serves as an important tool to "distinguish" between English text and a random string of English alphabets. This fact is used in the following discussion.

Now, we shall discuss how the Index of Coincidence method can be used to check the m value reported by the Kasiski test for a *Vignere cipher*.

Using the m value of the Kasiski test, we arrange the given alphabetic string $y = y_1\ldots y_n$ into m substrings as follows:

$$Y_1 = y_1 y_{m+1} y_{2m+1} \cdots$$
$$Y_2 = y_2 y_{m+2} y_{2m+2} \cdots$$
$$\cdots$$
$$Y_m = y_m y_{2m} y_{3m} \cdots$$

If the value of m reported by Kasiski test is correct, each substring Y_i, $1 \le i \le m$ is a shift cipher which has been shifted by a key K_i. Hence the expected value of $I_c(Y_i)$ is about 0.065. However, if the guess of m is incorrect, each substring is a random string and thus the IC value is about 0.038. Thus we can confirm the value of m reported by the Kasiski test.

Next we investigate a method to actually determine the key $K = (k_1, \ldots, k_m)$.

For this we need the concept of *Mutual Index of Coincidence* (MI) between two alphabetic strings x and y.

Definition Suppose $x = x_1 x_2 \ldots x_n$ and $y = y_1 y_2 \ldots y_n$, are two alphabetic strings. Then the Mutual Index of Coincidence between x and y is the probability that a random element of x is equal to that of y. Thus if the probabilities of A, B, ... are $f_0, f_1, \ldots, f_{25}$ and $f'_0, f'_1, \ldots, f'_{25}$ respectively in x and y, then:

$$MI_c(x, y) = \frac{\sum_{i=0}^{25} f_i f_i'}{nn'}$$

Consider Table 3.4 containing the alphabets and their corresponding probability distributions.

A	B	...	Z
p0	p1	...	p25

Imagine that due to a key K_i being used as a key in a shift cipher, the corresponding probability distribution is as shown in Table 3.5.

$A + k_i$	$B + k_i$	...	$Z + k_i$
P0	p1	...	p25

Now what is the probability that in the cryptogram a character is A? If the letters A, ..., Z are numbered from 0, …, 25 then a letter denoted by a number say j in the unencrypted text thus becomes $j + k_i$. Thus when $j + k_i$ is A in the ciphertext, we may write numerically $j + k_i = 0 \pmod{26}$, or $j = -k_i \pmod{26}$

Hence the corresponding probability of A in the encrypted text is $p_j = p_{-ki}$. Note that the value in the suffix is modulo 26.

Thus if we consider two strings x and y, which have been shifted by k_i and k_j respectively, the probability that both characters in x and y are A is $p_{-ki} p_{-kj}$. Likewise the probability that both the characters are B is $p_{1-ki} p_{1-kj}$ and so on.

$$MI_c(x, y) = \sum_{h=0}^{25} p_{h-k_i} p_{h-k_j} = \sum_{h=0}^{25} p_h p_{h+k_i-k_j}$$

1. For two strings, x and y ciphered using keys k_i and k_j the value of $MI_c(x,y)$ depends on the difference $k_i - k_j \pmod{26}$.

2. A relative shift of s yields the same value as 26-s. This is left as an exercise to the reader.

When $k_i - k_j = 0$, the value of MI_c is maximum and is equal to 0.065. However for other values, the estimate is comparatively less and ranges from 0.032 to 0.045 on an average.

So in order to find the actual key, we divide the given string of encrypted characters into m rows as described before. Each row is a shift cipher, which has been shifted by a key say, k_i. Thus for each row we find the Mutual Index of Coincidence with respect to an unencrypted English text. We compute the MI values by varying the keys, k_i from 0 to 25. The values for which the MI values become close to 0.065 will indicate the correct key, k_i. This process is repeated for the m rows to obtain the entire key.

Example 3.19 Let us assume that the intercepted text is as follows:

LIOMWGFEGGDVWGHHCQUCRHRWAGWIOWQLKGZETKKMEVLWPCZVGTHVTSGXQOVGCSVETQLTJSUMV-
WVEUVLXEWSLGFZMVVWLGYHCUSWXQHKVGSHEEVFLCFDGVSUMPHKIRZDMPHHBVWVWJWIXGFWLTSH-
GJOUEEHHVUCFVGOWICQLTJSUXGLW

Kasiski test for repetition of three character segments yields the results as shown in Table 3.4.

Table 3.4 *Kasiski test for Example 3.19*

String	*First Index*	*Second Index*	*Difference*
QLT	65	165	100
LTJ	66	166	100
TJS	67	167	100
JSU	68	168	100
SUM	69	117	48
VWV	72	132	60

The greatest common divisor is thus 4, thus suggesting that the key length is a multiple of 4. We try confirm this guess by the Index of Coincidence test.

We divide the ciphertext into 4 rows as shown below. We also mention the corresponding Index of Coincidence values. The high values of the IC confirms the key length reported in the Kasiski test.

1st string : LWGWCRAOKTEPGTQCTJVUEGVGUQGECVPRPVJGTJEUGCJG
IC = 0.067677

2nd string : IGGGQHGWGKVCTSOSQSWVWFVYSHSVFSHZHWWFSOHCOQSL
IC = 0.074747

3rd string: OFDHURWQZKLZHGVVLUVLSZWHWKHFDUKDHVIWHUHFWLUW
IC = 0.070707

4th string: MEVHCWILEMWVVXGETMEXLMLCXVELGMIMBWXLGEVVITX
IC = 0.076768

Then we perform the Mutual Index of Coiucidence to obtain the actual key value. Running the test, we obtain that the key value is CODE, and the corresponding plaintext is:

JULIUSCAESARUSEDACRYPTOSYSTEMINHISWARWHICHISNOWREFERR
EDTOASCAESARCIPHERITISASHIFTCIPHERWITHTHEKEYSETTOTHREEE
ACHCHARACTERINTHEPLAINTEXTISSHIFTERTHREECHARACTERSOCRE
ATEACIPHERTEXT

Note that the plaintext makes sense and hence we believe the decryption is correct. We format the obtained as follows:

Julius Caesar used a cryptosystem in his wars, which is now referred to as Caesar cipher. It is anadditive cipher with the key set to three. Each character in the plaintext is shifted three characters to create ciphertext.