i	$\varphi(i)$	i	$\varphi(i)$	i	$\varphi(i)$	i	$\varphi(i)$
0	0.0469	7	0.0461	13	0.0505	19	0.0312
1	0.0393	8	0.0194	14	0.0561	20	0.0287
2	0.0396	9	0.0286	15	0.0215	21	0.0526
3	0.0586	10	0.0631	16	0.0306	22	0.0398
4	0.0259	11	0.0280	17	0.0386	23	0.0338
5	0.0165	12	0.0318	18	0.0317	24	0.0320
6	0.0676					25	0.0443

Figure 10–2 The value of $\varphi(i)$ for $0 \le i \le 25$ using the model in Figure 10–1.

EXAMPLE: Using Konheim's model of single-character frequencies [1092, p. 16], the most likely keys (in order) are $i = 6$, $i = 10$, $i = 14$, and $i = 3$. Konheim's frequencies are different than Denning's, and this accounts for the change in the third most probable key.

A variant of the shift cipher, called an *affine cipher*, uses a multiplier in addition to the shift. Exercise 4 examines this cipher.

10.2.2.1 Vigenère Cipher

The shift cipher maps every character into another character in one alphabet. Such a cipher is a *monoalphabetic* cipher. As noted above, it preserves the statistics of the underlying message, which a cryptanalyst can use to decipher the message.

A *polyalphabetic* cipher uses multiple alphabets to generate the ciphertest, thereby obscuring the statistics. The Vigenère cipher is such a cryptosystem. In it, the key is a sequence of letters. The key letters are applied to successive plaintext characters, and when the end of the key is reached, the key starts over. The length of the key is called the *period* of the cipher. Figure 10–3 shows a *tableau*, or table, to implement this cipher efficiently. Because this requires several different key letters, this type of cipher is called *polyalphabetic*.

EXAMPLE: The first line of a limerick is enciphered using the key "BENCH," as follows:

Key	`B ENCHBENC HBENC HBENCH BENCHBENCH`
Plaintext	`A LIMERICK PACKS LAUGHS ANATOMICAL`
Ciphertext	`B PVOLSMPM WBGXU SBYTJZ BRNVVNMPCS`

For many years, the Vigenère cipher was considered unbreakable. Then a Prussian cavalry officer, Major Kasiski, noticed that repetitions occur when characters of the key appear over the same characters in the ciphertext. The

	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

Figure 10–3 The Vigenère tableau.

number of characters between the repetitions is a multiple of the period. From this observation, he developed an effective attack.

EXAMPLE: Let the message be "THE BOY HAS THE BAG" and let the key be "VIG." Then

Key	`VIGVIGVIGVIGVIG`
Plaintext	`THEBOYHASTHEBAG`
Ciphertext	`OPKWWECIYOPKWIM`

In the ciphertext, the string "OPKW" appears twice. Both are caused by the key sequence "VIGV" enciphering the same ciphertext, "THEB." The ciphertext repetitions are nine characters apart. As Figure 10–4 shows, the lower this value, the less variation in the characters of the ciphertext and, from our models of English, the longer the period of the cipher.

Period	Expected IC	Period	Expected IC	Period	Expected IC
1	0.0660	7	0.0420	50	0.0386
2	0.0520	8	0.0415	60	0.0385
3	0.0473	9	0.0411	70	0.0384
4	0.0450	10	0.0408	80	0.0384
5	0.0436	20	0.3940	90	0.0383
6	0.0427	30	0.0389	99	0.0383
		40	0.0387		

Figure 10–4 Indices of coincidences for different periods.

The first step in the Kasiski method is to determine the length of the key. The *index of coincidence* (IC) measures the differences in the frequencies of the letters in the ciphertext. It is defined as the probability that two letters randomly chosen from the ciphertext will be the same. The lower this value, the less variation in the characters of the ciphertext and, from our models of English, the longer the period of the cipher.

Let F_c be the frequency of cipher character c, and let N be the length of the ciphertext. Then the index of coincidence IC can be shown to be (see Exercise 6)

$$IC = \frac{1}{N(N-1)} \sum_{i=0}^{25} F_i(F_i - 1)$$

We examine the ciphertext for multiple repetitions and tabulate their length and the number of characters between successive repetitions. The period is likely to be a factor of the number of characters between these repetitions. From the repetitions, we establish the probable period, using the index of coincidence to check our deduction. We then tabulate the characters for each key letter separately and solve each as a shift cipher.

EXAMPLE: Consider the Vigenère cipher

```
ADQYS MIUSB OXKKT MIBHK IZOOO EQOOG IFBAG KAUMF
VVTAA CIDTW MOCIO EQOOG BMBFV ZGGWP CIEKQ HSNEW
VECNE DLAAV RWKXS VNSVP HCEUT QOIOF MEGJS WTPCH
AJMOC HIUIX
```

Could this be a shift cipher (which is a Vigenère cipher with a key length of 1)? We find that the index of coincidence is 0.0433, which indicates a key of around length 5. So we assume that the key is of length greater than 1, and apply the Kasiski method. Repetitions of length 2 are likely coincidental, so we look for repetitions of length 3 or more:

Letters	**Start**	**End**	**Gap length**	**Gap length factors**
OEQOOG	24	54	30	2, 3, 5
MOC	50	122	72	2, 2, 2, 3, 3

The longest repetition is six characters long; this is unlikely to be a coincidence. The gap between the repetitions is 30. The next longest repetition, "MOC," is three characters long and has a gap of 72. The greatest common divisor of 30 and 72 is 6. So let us try 6.

To verify that this is reasonable, we compute the index of coincidence for each alphabet. We first arrange the message into six rows, one for each alphabet:

```
A I K H O I A T T O B G E E E R N E O S A I
D U K K E F U A W E M G K W D W S U F W J U
Q S T I Q B M A M Q B W Q V L K V T M T M I
Y B M Z O A F C O O F P H E A X P Q E P O X
S O I O O G V I C O V C S C A S H O G C C
M X B O G K V D I G Z I N N V V C I J H H
```

We then compute the indices of coincidence for these alphabets:

Alphabet #1: IC = 0.0692	Alphabet #4: IC = 0.0562
Alphabet #2: IC = 0.0779	Alphabet #5: IC = 0.1238
Alphabet #3: IC = 0.0779	Alphabet #6: IC = 0.0429

All indices of coincidence indicate a single alphabet except for the indices of coincidence associated with alphabets #4 (period between 1 and 2) and #6 (period between 5 and 6). Given the statistical nature of the measure, we will assume that these are skewed by the distribution of characters and proceed on the assumption that there are 6 alphabets, and hence a key of length 6.

Counting characters in each column (alphabet) yields

Row	*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*
#1	3	1	0	0	4	0	1	1	3	0	1	0	0	1	3	0	0	1	1	2	0	0	0	0	0	0
#2	1	0	0	2	2	2	1	0	0	1	3	0	1	0	0	0	0	0	1	0	4	0	4	0	0	0
#3	1	2	0	0	0	0	0	0	2	0	1	1	4	0	0	0	4	0	1	3	0	2	1	0	0	0
#4	2	1	1	0	2	2	0	1	0	0	0	0	1	0	4	3	1	0	0	0	0	0	0	2	1	1
#5	1	0	5	0	0	0	2	1	2	0	0	0	0	0	5	0	0	0	3	0	0	2	0	0	0	0
#6	0	1	1	1	0	0	2	2	3	1	1	0	1	2	1	0	0	0	0	0	0	3	0	1	0	1

An unshifted alphabet has the following characteristics ("L" meaning low frequency, "M" meaning moderate frequency, and "H" meaning high frequency):

H M M M H M M H H M M M M H H M L H H H M L L L L L

We now compare the frequency counts in the six alphabets above with the frequency count of the unshifted alphabet. The first alphabet matches the characteristics of the unshifted alphabet (note the values for "A," "E," and "I" in particular). Given the gap between "B" and "I," the third alphabet seems to be shifted with "I" mapping to "A." A similar gap occurs in the sixth alphabet between "O" and "V," suggesting that "V" maps to "A." Substituting into the ciphertext (lowercase letters are plaintext) produces

```
aDiYS riUkB OckKl MIghK aZOto EiOOl iFtAG paUeF
VatAs CIitW eOCno EiOOl bMtFV egGoP CneKi HSseW
nECse DdAAa rWcXS anSnP HheUl QOnoF eEGos WlPCm
aJeOC miUaX
```

In the last line, the group "aJe" suggests the word "are." Taking this as a hypothesis, the second alphabet maps "A" into "S." Substituting back produces

```
aliYS rickB Ocksl MIghs aZOto miOOl intAG paceF
Vatis CIite eOCno miOOl butFV egooP Cnesi HSsee
nECse ldAAa recXS ananP Hhecl QOnon eEGos elPCm
areOC micaX
```

The last block suggests "mical," because "al" is a common ending for adjectives. This means that the fourth alphabet maps "O" into "A," and the cipher becomes

```
alimS rickp Ocksl aIghs anOto micOl intoG pacet
Vatis qIite ecCno micOl buttV egood Cnesi vSsee
nsCse ldoAa reclS anand Hhecl eOnon esGos eldCm
arecC mical
```

In English, a "Q" is always followed by a "U," so the "I" in the second group of the second line must map to "U." The fifth alphabet maps "M" to "A." The cipher is solved:

```
alime rickp acksl aughs anato mical intos pacet
hatis quite econo mical butth egood onesi vesee
nsose ldoma recle anand thecl eanon essos eldom
areco mical
```

With proper spacing, capitalization, and punctuation, we have

A limerick packs laughs anatomical
Into space that is quite economical.
    But the good ones I've seen
    So seldom are clean,
And the clean ones so seldom are comical.

The key is "ASIMOV."

The Vigenère cipher is easy to break by hand. However, the principles of attack hold for more complex ciphers that can be implemented only by computer. A good example is the encipherments that several older versions of WordPerfect used [171, 173]. These allowed a user to encipher a file with a password. Unfortunately, certain fields in the enciphered file contained information internal to WordPerfect, and these fields could be predicted. This allowed an attacker to derive the password used to encipher the file, and from that the plaintext file itself.

10.2.2.2 One-Time Pad

Repetitions provide a means for the cryptanalyst to attack the Vigènere cipher. The *one-time pad* is a variant of the Vigenère cipher with a key that is at least as long as the message and is chosen at random, so it does not repeat. Technically, it is a threshold scheme (see Section 16.3.2), and is provably impossible to break [240] (see also Section C.3.3, "Perfect Secrecy").

The weakness of the one-time pad is that the key must never be used more than once.

EXAMPLE: In 1943, the U.S. Army's Signal Intelligence Service began to examine messages sent from Soviet agents in the United States to Moscow. These messages were encoded using a complex cipher that was based on a one-time pad, which in this context was a set of pages of random number groups. This in theory made the messages unbreakable. But sometimes the manufacturers of these pads reused pages. Taking advantage of this duplication, cryptanalysts in the Signal Intelligence Service and, later, the U.S. National Security Agency, were able to decipher many of the messages sent between 1943 and 1980, providing insight into Soviet espionage of that time.

10.2.3 Data Encryption Standard

The Data Encryption Standard (DES) [2146] is one of the most important symmetric cryptosystems in the history of cryptography. It provided the impetus for