PART FOUR

Polyalphabetic Substitution Systems

CHAPTER 8

PERIODIC POLYALPHABETIC SUBSTITUTION SYSTEMS

Section I
Characteristics of Periodic Systems

8-1. Types of Polyalphabetic Systems

All the substitution systems explained up to this point are monoalphabetic systems. Whether they deal with one letter at a time or several, whether they have one cipher equivalent for each plaintext letter or more than one, they are still systems with only one alphabet. The constant feature that makes a system monoalphabetic is that a given ciphertext value always translates into the same plaintext value. In polyalphabetic systems, a given ciphertext value changes its plaintext meaning.

a. Most polyalphabetic systems are monographic; they encipher a single letter at a time. Polygraphic polyalphabetics are possible, but have little practical military value.

b. A typical polyalphabetic system will use from 2 to 26 different alphabets. Polyalphabetic systems which repeat the same set of alphabets over and over again in the same sequence are known as periodic systems. Polyalphabetic systems which do not keep repeating the same alphabets in the same order are known as aperiodic systems. Periodic systems, because of their regular repeating keys, are generally less secure than aperiodic systems. Aperiodic systems, on the other hand, are generally more difficult to use, unless the encipherment is done automatically by a cipher machine or computer.

c. The classic types of polyalphabetic systems use a set of alphabets, such as the 26 alphabets pictured in Figure 8-1. Figure 8-1, known as a Vigenere square, includes all possible alignments of a direct standard alphabet. Mixed alphabets can also be used in such a square. If all 26 alphabets are used, any letter can equal any other letter. There are necessarily three elements to the encryption process with polyalphabetic ciphers, which the square and the accompanying examples illustrate. The plaintext letters are listed across the top of the square. The cipher equivalents are found in the 26 sequences below. The final element is the key that designates which alphabet is used at any given time. The key letter is found on the left side of the square. The first example in Figure 8-1 shows the use of a repeating key based on a keyword. Since the same key is repeated over and over again, the resulting system is periodic. The second example uses a nonrepeating key based on a quotation. Since this key does not repeat, it is an aperiodic system. Note that the reuse of the same alphabets does not constitute a repeating key. For the system to be classified as periodic, the same alphabets must be reused over and over again in the same sequence.

Plaintext

Key	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

PERIODIC

Plain: repor tatze rotwo twoze rotom orrow
Key: RIFLE RIFLE RIFLE RIFLE RIFLE RIFLE
Cipher: IMUZV KIYKI IWYHS KETKI IWYZQ FZWZA

APERIODIC

Plain: mount ainpa ssesb locke dbyhe avysn owfal llast night
Key: FOURS COREA NDSEV ENYEA RSAGO OURFO REFAT HERSB ROUGH
Cipher: RCOEL CWETA FVWWW PBAOE UTYNS OPPXB FAKAE SPRKU EWANA

Figure 8-1. Use of Vigenere square.

d. Another way to picture the same system as the first example in Figure 8-1 is shown below. In this case, instead of using the complete alphabet square, only the alphabets actually used are shown. These alphabets are used repeatedly to produce the same results. In this example, the key is expressed in terms of the number of the cipher sequence used, instead of by the repeating key letters.

p:	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
C1:	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
C2:	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
C3:	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
C4:	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
C5:	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

Plaintext:	repor	tatze	rotwo	twoze	rotom	orrow
Key:	12345	12345	12345	12345	12345	12345
Ciphertext:	IMUZV	KIYKI	IWYHS	KETKI	IWYZQ	FZWZA

e. Another type of polyalphabetic system does not use multiple alphabets in the classic sense, but instead enciphersa message in a single alphabet. Then it applies either a repeating key or nonrepeating key to the first encipherment to create a polyalphabetic. One method of applying a polyalphabetic key to a monoalphabetic encipherment is to use a numeric system and arithmetically add a key to it. For example, here is a dinomic system, which has been further enciphered by a repeating numeric additive. The first encipherment is labeled I, for intermediate cipher, and the second encipherment is labeled C. The 8-digit repeating key is labeled K. Modulo 10 arithmetic is used (paragraph 5-3f(1)).

	0	1	2	3	4	5	6	7	8	9
3	m	u	r	p	h	y	s	l	a	w
6	b	c	d	e	f	g	i	j	k	n
9	o	q	t	v	x	z	.	,	?	/

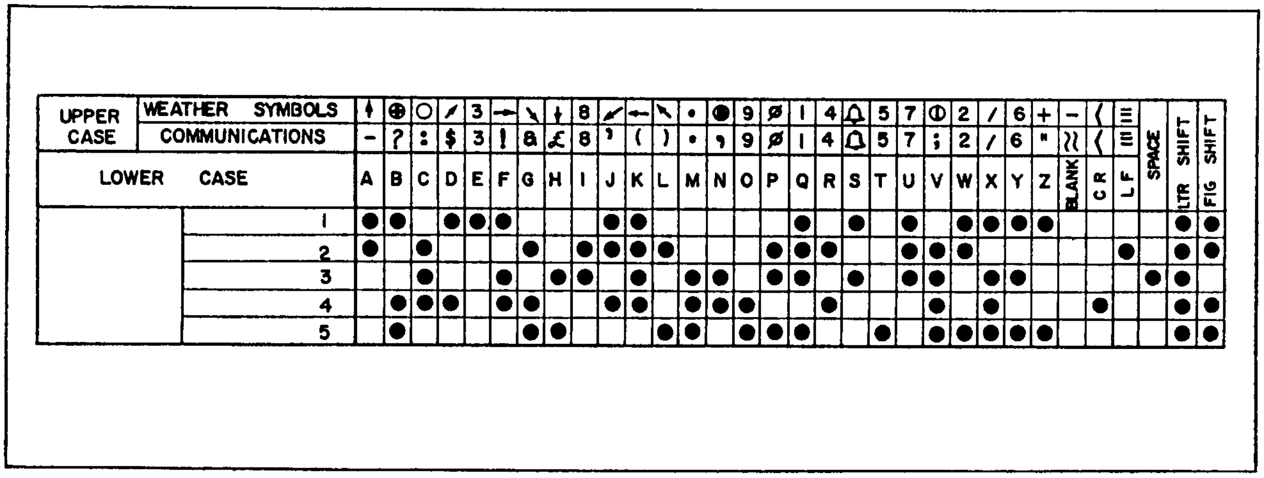
p:	a t	t a	c k	a t	z e	r o	n i	n e	h u	n d	r e	d .
I:	3892	9238	6168	3892	9563	3290	6966	6963	3431	6962	3263	6296
K:	4209	9336	4209	9336	4209	9336	4209	9336	4209	9336	4209	9336
C:	7091	8564	0367	2128	3762	2526	0165	5299	7630	5298	7462	5522

f. Another approach to applying a polyalphabetic key begins with the built-in encoding system used by teleprinters or computers. Paragraph 8-2 shows examples of these.

8-2. Machine Based Polyalphabetics

When text is sent electronically by radio or wire, some form of coding must be used. The earliest system of coding for electronic transmission was Morse code, which is still used widely today. When teleprinters took their place in communications, a new binary type of coding system was devised, which can be handled by machine more easily than Morse code can. Any binary coding system uses only two characters, which can be represented electronically as a signal pulse or no signal pulse, high voltage or low voltage, or one frequency or another frequency. Which of these approaches is used depends on the equipment in use and is not our concern here. We are concerned with how the two binary characters, whatever their electronic origin, are combined to represent alphabetic, numeric, and special characters, and how they may further be encrypted. Various notations have been used to represent the two binary characters—Xs and 0s, 1s and 0s, +s and -s, or Ms (for marks) and Ss (for spaces). We will use 1s and 0s in this text, but you should be aware that you may see other notations elsewhere, particularly in older literature.

a. **The Baudot Code.** Teleprinter systems generally use a 5-digit binary code known originally as the Baudot code. There are 32 possible combinations of 5 digits, which are not enough for the letters, numbers, and printer control characters needed for communications. The number of possible characters is approximately doubled by the use of upper and lower shift characters, similar to the shift key on a typewriter, giving all characters two alternate meanings except the shift characters themselves and the space character. There are still not enough characters for upper and lower case letters, so all traffic passed by such teleprinter systems use capital letters only. The standard international teleprinter code is shown in Figure 8-2. Each dot represents a 1 and each space represents a 0. Other codes are also used besides the one shown.



UPPER CASE	WEATHER SYMBOLS	↑	⊕	○	↗	3	→	↘	↓	8	↙	←	↖	•	●	9	Ø	I	4	🔔	5	7	⦶	2	/	6	+	−	⟨	≡	SPACE	LTR SHIFT	FIG SHIFT
	COMMUNICATIONS	−	?	:	$	3	!	&	£	8	'	(	)	.	,	9	Ø	I	4	🔔	5	7	;	2	/	6	"	≀	⟨	≡			
LOWER CASE		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	BLANK	CR	LF			
	1	●	●		●	●	●				●	●						●		●		●		●	●	●	●					●	●
	2	●		●				●		●	●	●	●				●	●	●			●	●	●						●		●	●
	3			●			●		●	●		●		●	●		●	●		●		●	●		●	●					●	●	
	4		●	●	●		●	●			●	●		●	●	●			●				●		●				●			●	●
	5		●					●	●				●	●		●	●	●			●		●	●	●	●	●					●	●

Figure 8-2. International teleprinter code.

The binary digits themselves are known as bauds—a term derived from the Baudot code. The terminology has carried over into modern computer. systems as well. Polyalphabetic keys, also in 5-digit binary form, are easily applied to coded text electronically by baud addition. An example of this process is shown below. Although other rules are also possible, the addition of key and plaintext bauds is usually accomplished by the rule, *Like values sum to 0; unlikes sum to 1.* (In computer logic, this would be called an exclusive OR, or XOR operation.)

Plaintext:	e	n	e	m	y
Bauded plain:	10000	00110	10000	00111	10101
Key:	**01010**	**11010**	**10100**	**01110**	**10110**
Bauded cipher:	11010	11100	00100	01001	00011
Ciphertext:	J	U	(space)	L	O

One advantage of this rule of addition is that adding the same key to the ciphertext produces the plaintext again.

b. **Computer Codes.** Communications between computers use more than 5 digits. Typical computer codes use either 7- or 8-binary digits (bits), giving a range of 128 characters or 256 characters. These permit upper and lower case letters, a full range of punctuation marks and special characters, and a number of codes to control printers and communications devices as well. With the 8-bit, 256 character set, graphics may also be enabled to permit transmitting pictures as well as text. The most common standard for the first 128 characters, whether 7-bit or 8-bit, is the American standard code for information interchange (ASCII) standard, which you can find in many computer manuals. Encipherment and decipherment can be accomplished in 7- and 8-bit operation just as was shown for 5-digit teleprinter operations. The more complex systems are far beyond the scope of this manual, but simple repeating key systems can be solved using the techniques discussed here. One problem that computer codes present is that less than half of the possible 7-bit characters are letters and numbers, and many of them stand for printer control codes that do not print out as characters normally. Working with binary numbers themselves is unwieldy, but any 7- or 8-bit value can be represented by two hexadecimal (base 16) arithmetic digits. Hexadecimal arithmetic is not explained here, but explanations are available in many computer manuals and texts, if needed. Hexadecimal and binary numbers are also explained in Army Correspondence Course Program Subcourse SA0709.

Section II
Identifying Periodic Systems

8-3. Analysis of Repeated Ciphertext

Polyalphabetic systems normally have very flat frequency counts. The phi IC is normally close to the random expectation of 1.00. Since other systems, including variant multiliterals and aperiodic systems, also can produce flat frequency counts, this is not enough to identify a system as periodic. The key to identifying a system as periodic is to recognize through repeated ciphertext that a repeating key is used.

a. Repeated ciphertext can occur in two ways. Whenever the same plaintext is enciphered by the same keys, the ciphertext will also repeat. Such repeats are called causal repeats. The second way that ciphertext can repeat is by pure chance. Different plaintext enciphered with different keys will sometimes produce short ciphertext repeats. Causal repeats are much more likely to occur than accidental repeats, particularly if they are longer than two or three characters. The example below, repeated from Section I, shows how causal repeats occur.

```
 Plaintext: repor tatze rotwo twoze rotom orrow
       Key: 12345 12345 12345 12345 12345 12345
Ciphertext: IMUZV KIYKI IWYHS KETKI IWYZQ FZWZA
```

The plaintext words *ZERO* and *TWO* both occur twice. The repeated *ZEROs* lined up with the same alphabets, producing a ciphertext repeat. The repeated *TWOs* lined up with different alphabets and did not produce a ciphertext repeat.

b. Whenever causal repeats occur, the distance between them must be a multiple of the period length. In the example above, the two *ZEROs* occurred 10 letters apart. Note that the distances are counted from the first letter of one repeat to, but not including, the first letter of the second repeat. If the distance was not a multiple of the period five, the ciphertext repeat would not have occurred.

c. The distance between causal repeats is a multiple of the period length. Given a cryptogram of unknown period that includes ciphertext repeats, the period can be determined, or at least narrowed down, by analyzing the distances between repeats. The period must be a factor of the distance. The factors of a number are all the numbers which divide evenly into that number. When there is more than one repeat, the period must be a common factor of all such distances. For example, if a cryptogram has repeats that are 28, 35, and 42 letters apart, the only number that evenly divides all the distances is 7. The period must be 7. Utility tables showing common factor numbers are in Appendix E.

d. Here is a more complex example. Suppose a cryptogram suspected of being periodic includes the following repeats.

Repeat	Distance
GXKLRYPDL	84
ZBHHNST	90
XTVTB	36
SRM	35

The next step after determining the distances is to list the factors for each repeat, as shown below.

Repeat	Distance	Factors
GXKLRYPDL	84	2, 3, 4, 6, 7, 8, 12
ZBHHNST	90	2, 3, 5, 6, 9, 10
XYVTN	36	2, 3, 4, 6, 9, 12
SRM	35	5, 7

No numbers evenly divide the distances between all the repeats. In such cases, either the system was not a periodic system, or one or more of the repeats is accidental. In this problem, the SRM repeat is probably accidental, because it is the shortest. Discarding the SRM repeat from consideration, the remaining repeats all have common factors of 2, 3, and 6. Where more than one factor is possible, it is generally safest to assume the largest. If the period is actually 3, for example, it will reveal itself by repeated alphabets as the cryptogram is solved.

8-4. Analysis by Frequency Counts

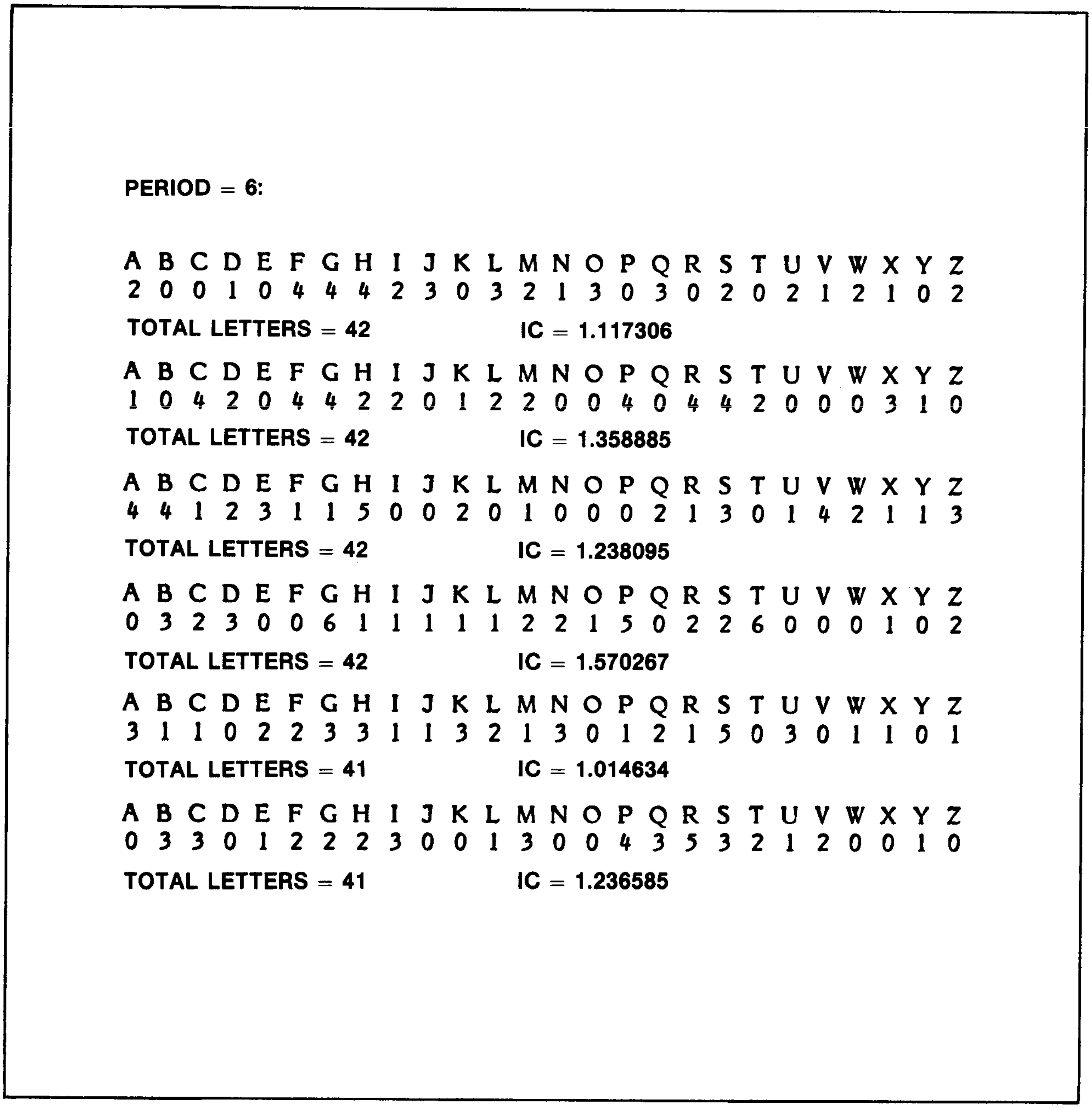
Periodic systems can be identified even when there are no repeated words in the text. Causal single-letter ciphertext repeats will still occur and significantly outnumber the accidental single-letter repeats.

a. To find the causal single-letter repeats, take frequency counts for each alphabet according to its position in the suspected repeating cycle. If the period is incorrect, the separate frequency counts will remain flat. If the period is correct, the separate frequency counts will be as rough as plaintext on the average. Recognizing when a count is rough or flat is difficult by eye, particularly with anything but very long cryptograms, but the phi test performed on each separate alphabet gives a reliable indication. Taking separate frequency counts by position for each suspected period and then calculating phi tests on each is a laborious and time-consuming process by hand. It can be done when necessary, but it is best performed by computer support. Figures 8-3, 8-4, and 8-5 show computer generated output for suspected periods of 6, 7, and 8 for the following cryptogram.

```
LPADW GUGHG ETZHV KSRQS ACNPJ   GHTHH QCKGS CHHRB HMDIH HMCJM
EXEVH LVPQS OCHPK MZYBZ SMMPF   TLBGF KRAEA FBMHQ IXSZC PGAQT
KPLPS GXIVX BGFRI TSTGF SPYNS   SNTAL SIOSC MJRMI ZSICF RQTUV
HLVPQ SOCHP KQFDW SFRAK MILRG   GECAU HFEGN YXXZO GLGMZ DUHUC
XGRIL SARZQ FDWBB PSRUD UGJGD   JNTWF BTABQ SVBGF WRDPP BFRGN


 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
10 11 11 8 6 13 20 17 9 5 7 9 11 6 4 14 10 13 19 10 7 7 5 7 3 8

TOTAL LETTERS = 250                     MONOGRAPHIC IC = 1.098474
```

b. The average ICs for each period in Figure 8-3 and 8-4 are flat, The average IC for a period of 8 in Figure 8-5 is much higher than the other two. This clearly shows that the period of 8 is more likely correct than periods of 6 and 7.

c. The computer program used to generate these examples is listed in Appendix F. It is written in GW BASIC, and is readily adaptable to many different computers.

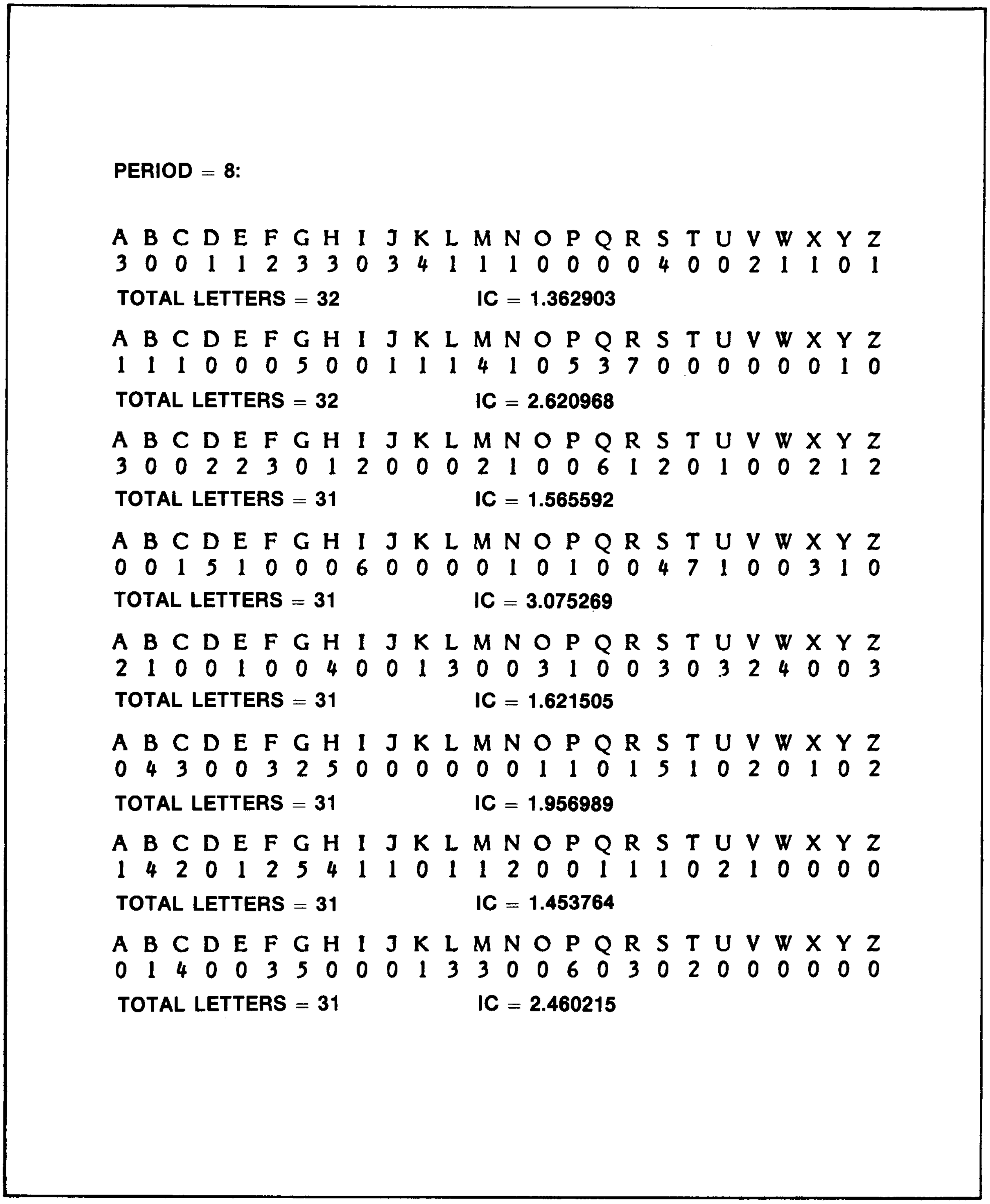


```
PERIOD = 6:

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
2 0 0 1 0 4 4 4 2 3 0 3 2 1 3 0 3 0 2 0 2 1 2 1 0 2
TOTAL LETTERS = 42          IC = 1.117306

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 0 4 2 0 4 4 2 2 0 1 2 2 0 0 4 0 4 4 2 0 0 0 3 1 0
TOTAL LETTERS = 42          IC = 1.358885

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
4 4 1 2 3 1 1 5 0 0 2 0 1 0 0 0 2 1 3 0 1 4 2 1 1 3
TOTAL LETTERS = 42          IC = 1.238095

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 3 2 3 0 0 6 1 1 1 1 1 2 2 1 5 0 2 2 6 0 0 0 1 0 2
TOTAL LETTERS = 42          IC = 1.570267

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
3 1 1 0 2 2 3 3 1 1 3 2 1 3 0 1 2 1 5 0 3 0 1 1 0 1
TOTAL LETTERS = 41          IC = 1.014634

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 3 3 0 1 2 2 2 3 0 0 1 3 0 0 4 3 5 3 2 1 2 0 0 1 0
TOTAL LETTERS = 41          IC = 1.236585
```

Figure 8-3. Frequencies, period 6.

```
PERIOD = 7:

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
2 2 2 2 1 3 4 1 1 0 0 1 1 1 1 1 3 1 2 2 1 2 0 1 0 1
TOTAL LETTERS = 36            IC = .784127
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 0 1 1 2 4 3 4 1 1 3 1 1 1 0 2 0 1 4 2 2 0 0 1 0 0
TOTAL LETTERS = 36            IC = 1.155556
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
2 2 1 2 0 1 1 2 3 0 1 1 4 2 1 2 2 1 2 0 1 1 2 1 0 1
TOTAL LETTERS = 36            IC = .7428572
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
2 1 5 1 2 0 2 2 1 2 1 1 1 0 0 2 1 3 1 2 1 1 1 0 2 1
TOTAL LETTERS = 36            IC = .8666667
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 2 0 1 0 2 2 2 0 0 1 3 2 2 1 1 1 3 4 1 1 1 2 3 1 0
TOTAL LETTERS = 36            IC = .9079365
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 2 2 0 0 2 4 4 1 1 0 0 1 0 0 4 1 2 2 2 0 2 0 1 0 3
TOTAL LETTERS = 35            IC = 1.22353
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
2 2 0 1 1 1 4 2 2 1 1 2 1 0 1 2 2 2 4 1 1 0 0 0 0 2
TOTAL LETTERS = 35            IC = .9176471
```

Figure 8-4. Frequencies, period 7.



```
PERIOD = 8:

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
3 0 0 1 1 2 3 3 0 3 4 1 1 1 0 0 0 0 4 0 0 2 1 1 0 1
TOTAL LETTERS = 32           IC = 1.362903
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 1 1 0 0 0 5 0 0 1 1 1 4 1 0 5 3 7 0 0 0 0 0 0 1 0
TOTAL LETTERS = 32           IC = 2.620968
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
3 0 0 2 2 3 0 1 2 0 0 0 2 1 0 0 6 1 2 0 1 0 0 2 1 2
TOTAL LETTERS = 31           IC = 1.565592
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 0 1 5 1 0 0 0 6 0 0 0 0 1 0 1 0 0 4 7 1 0 0 3 1 0
TOTAL LETTERS = 31           IC = 3.075269
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
2 1 0 0 1 0 0 4 0 0 1 3 0 0 3 1 0 0 3 0 3 2 4 0 0 3
TOTAL LETTERS = 31           IC = 1.621505
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 4 3 0 0 3 2 5 0 0 0 0 0 0 1 1 0 1 5 1 0 2 0 1 0 2
TOTAL LETTERS = 31           IC = 1.956989
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 4 2 0 1 2 5 4 1 1 0 1 1 2 0 0 1 1 1 0 2 1 0 0 0 0
TOTAL LETTERS = 31           IC = 1.453764
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 4 0 0 3 5 0 0 0 1 3 3 0 0 6 0 3 0 2 0 0 0 0 0 0
TOTAL LETTERS = 31           IC = 2.460215
```

Figure 8-5. Frequencies, period 8.