19 Lorenz Cipher Machine

During the Second World War, the codebreakers at Bletchley Park devised a variety of techniques that enabled the Allies to break the major codes used by the Germans. Not only was this hugely significant in helping the Allies to win the war, but it also led on to whole new branches of statistical analysis and to the development of the electronic computer.

In this unit we will illustrate how one of the important ciphers used by the Germans was broken. The cipher is the *Lorenz* cipher which was based on the use of five bit binary numbers.

ENCIPHER

The code used to convert letters to binary numbers is given below.

A	11000	B	10011	C	01110	D	10010	E	10000	F	10110	G	01011
H	00101	I	01100	J	11010	K	11110	L	01001	M	00111	N	00110
O	00011	P	01101	Q	11101	R	01010	S	10100	T	00001	U	11100
V	01111	W	11001	X	10111	Y	10101	Z	10001				

There were also six special symbols used to control the automatic printers – these are given below.

3	00010	4	01000	8	11111	9	00100	+	11011	/	00000

These special symbols did not have their usual meanings; the only one to be used in messages is '9' which means 'a space' (between words).

Clearly, if this was the code, it would not take too long to break it (letter frequency, for example, could be used). But the code-writers also used a 'key' letter to add to each plaintext letter in order to disguise it.

We add the binary representation in a special way, namely

$0 + 0 = 0$	$0 + 1 = 1$
$1 + 0 = 1$	$1 + 1 = 0$

Example

Encipher the plaintext letter J using the key letter B.

Solution

$$\begin{array}{rl} & \text{J} \Rightarrow 11010 \\ + & \text{B} \Rightarrow 10011 \\ \hline & \text{L} \Leftarrow 01001 \end{array}$$

Exercise 1

Using the key letter B, *encipher the plaintext letters* A, B, C, D *and* E.

This code it still relatively easy to break so it is further complicated by the adding of a different key letter to each letter; that is, the sender and receiver of the message have a sequence of (secret) key letters.

Example

Encipher the message HELP using the key sequence ABCD.

Solution

Plaintext message	H E L P	$\Rightarrow$	00101	10000	01001	01101
Key sequence	A B C D	+	11000	10011	01110	10010
	Q O M 8	$\Leftarrow$	11101	00011	00111	11111

Of course, the German cipher system would never have used a *simple* key sequence like this.

Exercise 2

Using the given key sequence HBVQZM, *encipher* LONDON.

DECIPHER

If you know the key sequence, deciphering is easy. Because of the form of addition used, the operation is what we call s*elf-inverse*; that is, you do the same thing again.

Example

Decipher QOMB using the key sequence ABCD.

Solution

Q O M 8	$\Rightarrow$	11101	00011	00111	11111
A B C D	+	11000	10011	01110	10010
H E L P	$\Leftarrow$	00101	10000	01001	01101

Exercise 3

Decipher your coded message in Exercise 2, using the same key sequence.

Activity 1

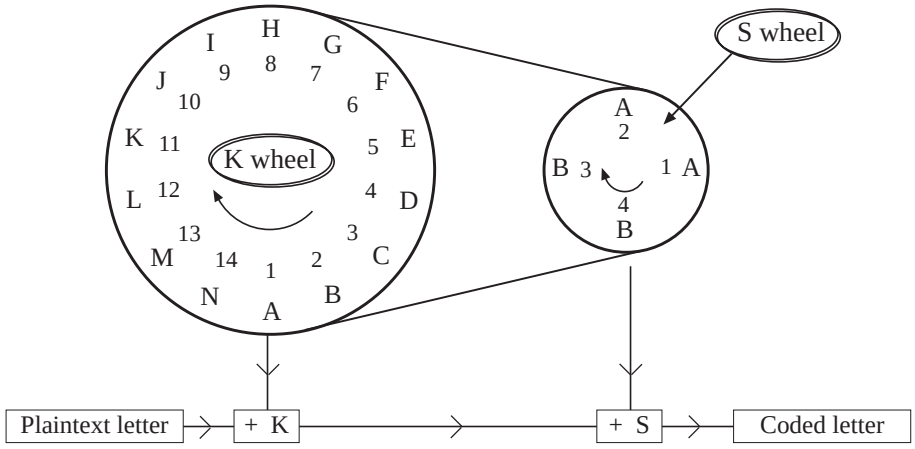
How many codes can be formed using 5 bit binary numbers?

Why (mathematically) must all these codes be used?

Whilst the binary addition used here is straightforward, it also becomes rather tedious to keep looking up the codes for the letters, etc. The 32×32 matrix in Appendix 1 shows the additions for the full set of characters used.

Simplified model for the Lorenz cipher machine

In fact, the actual Lorenz cipher machine had twelve cipher wheels. We will use a much simplified version with two wheels.



The first wheel, the K wheel, has 14 positions (numbered 1 to 14) and the second, the S wheel, has just 4 positions (1 to 4). In the position shown in the diagram, the machine would first add A to the plaintext letter, then add B to the result to give the final cipher for the letter. For the next letter, though, the wheels on the machine would move one position clockwise, so that it would add B and then A.

Example

Encipher THE with the starting position shown in the diagram.

Solution

Using the table in the Appendix,

$$T + A + B = T + G = R$$
$$H + B + A = H + G = C$$
$$E + C + A = E + F = N$$

So the message would read RCN.

Exercise 4

Using the starting position 5 on the K wheel and 2 on the S wheel, encipher

S E C R E T M E S S A G E

(Use 9 here to represent a space between the words.)

As with our first example, deciphering is easy if you know the keys; it is just the reverse of enciphering.

Example

Decipher R C N using the same key sequences as before.

Solution

We now have two key sequences (from the K and S wheels)

$$R + A + B = R + G = T$$
$$C + B + A = C + G = H$$
$$N + C + A = N + F = E$$

to recover the message 'T H E' .

Exercise 5

Decipher the coded message

U Y F X 9 4 L F V T 8 B Q Z

found in Exercise 4.

So deciphering is straightforward if you know the *two* key sequences. For our example there are $14 \times 4 = 56$ possibilities, but for the actual Lorenz cipher machine, there are about 16 million million million possibilities! This was the challenge for the codebreakers!

Breaking this cipher

The Wartime success at Bletchley Park in breaking this cipher depended upon the fact that most of the German plaintext messages contained many pairs of repeated characters. (There were certain technical reasons why the Germans adopted this practice.) The methods used at Bletchley Park were based on statistical evidence derived from the cipher messages. Probability theory had a prominent role as there was no certainty that the initial results relating to the keys were correct but only levels of probability that they were so.

We will illustrate the technique with the following short message in which the words are separated by double spaces represented by pairs of 9s.

99HERE99IS99A99TEST99MESSAGE99FOR99YOU99TO99TRY99OUT99

Exercise 6

With the starting position K = 7, S = 3, *show that the first enciphered characters are*

U D Z D M R + J

In fact, the fully enciphered message is

UDZDMR+JMSDC+TXUVQMYEDE8LWOKUD3TMK+G4UDC3NXWKOBYEFURWH

In what follows, we will represent a sequence of such characters by **Z**. The method used to decipher this message when the starting positions of cipher wheels K and S are not known, is as follows.

1. In **Z**, the coded message, add together the 1st and 2nd characters, then add together the 2nd and 3rd characters, and so on, and call this new sequence $\Delta\mathbf{Z}$ (the 'delta Z' sequence).

2. For each starting position, in turn, of the K wheel, take the key sequence **K** and also determine $\Delta\mathbf{K}$.

3. Add $\Delta\mathbf{Z}$ to $\Delta\mathbf{K}$ to form a new sequence (one for each starting position of the K wheel).

4. Note the number of '/'s in this sequence.

5. The position of **K** which gives the most '/'s in $\Delta\mathbf{Z} + \Delta\mathbf{K}$ is likely to be the actual starting position of the K wheel (the theory for this is given in Appendix 2).

We will see how this works for just the first 8 letters of this message; note that in coded form the letters are

U D Z D M R + J

We will follow this process through with $K = 1$.

1.

Z	=	U	D	Z	D	M	R	+	J
$\Delta\mathbf{Z}$	=	C	O	O	Y	P	Z	T	

2. Here $K = 1$, so we first find the **K** sequence.

K	=	A	B	C	D	E	F	G	H
$\Delta\mathbf{K}$	=	G	Q	U	3	N	Q	C	

3.

$\Delta\mathbf{Z}$	=	C	O	O	Y	P	Z	T
$\Delta\mathbf{K}$	=	G	Q	U	3	N	Q	C
$\Delta\mathbf{Z}+\Delta\mathbf{K}$	=	H	K	8	X	G	I	V

4. There are no '/'s in this sequence for $\Delta\mathbf{Z} + \Delta\mathbf{K}$.

Exercise 7

Follow the same procedure with the starting position $K = 7$.

How many '/'s do you obtain in $\Delta\mathbf{Z} + \Delta\mathbf{K}$ *?*

Of course, this is getting rather tedious to complete, particularly if we use the full message.

The printout copied below gives the results for all 14 starting positions of K.

K = 1
$\Delta\mathbf{Z} + \Delta\mathbf{K}$ = HK8XGIVSH3G++4/CCASAIGWDMSUUHP898AKR3RG3HVHJTD/A4O/XG

K = 2
$\Delta\mathbf{Z} + \Delta\mathbf{K}$ = B8TBE84GXZDS9QFVEUVGGSGTK+OLBIT/9GWYEWDPJJB+8F+GVUD9D

K = 3
$\Delta\mathbf{Z} + \Delta\mathbf{K}$ = DTH4OAXW94QGZGXZSMUISNAAZ9F8DDH+XINMO/QDVIDH+P4IECTQQ

K = 4
$\Delta\mathbf{Z} + \Delta\mathbf{K}$ = IHK+9MHRQM3KMRLYVS+BNYTXCZ/KIFK4EBSSJV3MWPIT/KVB3QAD3

K = 5
$\Delta\mathbf{Z} + \Delta\mathbf{K}$ = 4KPU+YFBDAX4NSPCUB9TYIC4+MT/4PPVVTMPYEXZAB4JBWETZ9XPX

K = 6
$\Delta\mathbf{Z} + \Delta\mathbf{K}$ = BPROLNVUPPTLAEFQ+IFDIOZQPN89BKREQDK3RHTENXBLSN3D4G4AT

K = 7
$\Delta\mathbf{Z} + \Delta\mathbf{K}$ = /RYZJ8/OA+/XUGHJ9KHGOU9GIA+8/YW3CGZQ8B/C3I/CGSZGMSQC/

K = 8
$\Delta\mathbf{Z} + \Delta\mathbf{K}$ = MYM3OE8FCJKBMA3HFPU9ULDRDU/IMNMZX9C4LDKRW8MZWM49ATGVK

K = 9
$\Delta\mathbf{Z} + \Delta\mathbf{K}$ = AMS+IVR/V9J4S8QXHSB+L8BSFMBGASS4A++K4IJZRAAORKM+PXRZJ

K = 10
$\Delta\mathbf{Z} + \Delta\mathbf{K}$ = RSPSBJUTZ/T+B/V9U+IC8KPEPSSSRMPMMCP8F4T3PMREBZAC+FSYT

K = 11
$\Delta\mathbf{Z} + \Delta\mathbf{K}$ = WP3GNIQ8Y+DUIDUOB9WAK/LGKBGNWK3ADAITDBDHDYWLUCPAJ4ECD

K =12
$\Delta\mathbf{Z} + \Delta\mathbf{K}$ = /3QKEPO+C4YOKTHDIZVW/9DAWIWY/ZQP9WDHL/YJ/N/NO++W9IGQY

K = 13
$\Delta\mathbf{Z} + \Delta\mathbf{K}$ = VQ44ZBM/QVRZPARPWMCM98T8NKRIVC4+HMFKJMR4B8VWFPJM/XAJR

K = 14
$\Delta\mathbf{Z} + \Delta\mathbf{K}$ = E4KLVXUBJEA3SXYAVNEO8IN/SPBOE+KJ+OPPQAA+REEI/I9O+98HA

Activity 2

Count the number of '/'s for each of the 14 positions.

What do you conclude?

So the evidence would point to $K = 7$ (although $K = 12$ might also be the answer) but this is now relatively straightforward.

We use $K = 7$, and try each possible position of the S wheel. For example, using again only the first 8 characters of the code, and setting $S = 1$ gives, for deciphering,

	U	D	Z	D	M	R	+	J	
+	G	H	I	J	K	L	M	N	(K wheel)
+	A	A	B	B	A	A	B	B	(S wheel)
	V	V	C	+	T	+	V	V	

This does not seem to be a meaningful message!

Activity 3

Complete the same process with $S = 2, 3$ and 4. What can you deduce?

Even with the vastly simplified Lorenz cipher it should be evident to you that the task of finding the sequences of $\Delta\mathbf{Z} + \Delta\mathbf{K}$ characters by hand would be extremely long and tedious.

The real Lorenz cipher messages contained several thousand characters and this would have made the task very much longer. A further extension arose from the fact that the number of possible wheel starting positions was also much greater than for the simplified version of the machine. (It was often necessary to take into account over 1000 possible positions at one time.)

The combination of these two factors made it impossible to find all of the sequences of $\Delta\mathbf{Z} + \Delta\mathbf{K}$ characters by hand in a realistic time. This difficulty was overcome by using a very fast electronic machine called 'Colossus' to carry out the task. The machine was designed and constructed in 1943 and became operational at Bletchley Park in January 1944. It was the first digital computer in the world.

Appendix 1

Teleprinter addition table

	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	9	8	+	4	3	/
A	/	G	F	R	4	C	B	Q	S	3	N	Z	8	K	+	Y	H	D	I	W	9	X	T	V	P	L	U	M	O	E	J	A
B	G	/	Q	T	O	H	A	F	8	L	P	J	S	Y	E	K	C	W	M	D	V	U	R	9	N	3	X	I	4	+	Z	B
C	F	Q	/	U	K	A	H	G	3	S	E	M	L	4	P	O	B	9	J	V	D	T	X	W	+	8	R	Z	Y	N	I	C
D	R	T	U	/	3	9	W	X	K	4	I	+	Y	S	Z	8	V	A	N	B	C	Q	G	H	M	O	F	P	L	J	E	D
E	4	O	K	3	/	N	+	Y	U	R	C	W	X	F	B	Q	P	J	9	Z	I	8	L	M	H	T	S	V	G	A	D	E
F	C	H	A	9	N	/	Q	B	J	I	4	8	Z	E	Y	+	G	U	3	X	R	W	V	T	O	M	D	L	P	K	S	F
G	B	A	H	W	+	Q	/	C	M	Z	Y	3	I	P	4	N	F	T	8	R	X	9	D	U	K	J	V	S	E	O	L	G
H	Q	F	G	X	Y	B	C	/	L	8	+	I	3	O	N	4	A	V	Z	9	W	R	U	D	E	S	T	J	K	P	M	H
I	S	8	3	K	U	J	M	L	/	F	D	H	G	R	V	T	Z	N	A	P	E	O	Y	+	W	Q	4	B	X	9	C	I
J	3	L	S	4	R	I	Z	8	F	/	9	B	Q	U	W	X	M	E	C	+	N	Y	O	P	V	G	K	H	T	D	A	J
K	N	P	E	I	C	4	Y	+	D	9	/	X	W	A	Q	B	O	S	R	8	3	Z	M	L	G	V	J	T	H	F	U	K
L	Z	J	M	+	W	8	3	I	H	B	X	/	C	V	R	9	S	O	Q	4	Y	N	E	K	U	A	P	F	D	T	G	L
M	8	S	L	Y	X	Z	I	3	G	Q	W	C	/	T	9	R	J	P	B	N	+	4	K	E	D	F	O	A	U	V	H	M
N	K	Y	4	S	F	E	P	O	R	U	A	V	T	/	H	G	+	I	D	M	J	L	8	Z	B	X	3	W	Q	C	9	N
O	+	E	P	Z	B	Y	4	N	V	W	Q	R	9	H	/	C	K	L	X	3	8	I	J	S	F	D	M	U	A	G	T	O
P	Y	K	O	8	Q	+	N	4	T	X	B	9	R	G	C	/	E	M	W	I	Z	3	S	J	A	U	L	D	F	H	V	P
Q	H	C	B	V	P	G	F	A	Z	M	O	S	J	+	K	E	/	X	L	U	T	D	9	R	4	I	W	3	N	Y	8	Q
R	D	W	9	A	J	U	T	V	N	E	S	O	P	I	L	M	X	/	K	G	F	H	B	Q	8	+	C	Y	Z	3	4	R
S	I	M	J	N	9	3	8	Z	A	C	R	Q	B	D	X	W	L	K	/	Y	4	+	P	O	T	H	E	G	V	U	F	S
T	W	D	V	B	Z	X	R	9	P	+	8	4	N	M	3	I	U	G	Y	/	Q	C	A	F	S	E	H	K	J	L	O	T
U	9	V	D	C	I	R	X	W	E	N	3	Y	+	J	8	Z	T	F	4	Q	/	B	H	G	L	P	A	O	M	S	K	U
V	X	U	T	Q	8	W	9	R	O	Y	Z	N	4	L	I	3	D	H	+	C	B	/	F	A	J	K	G	E	S	M	P	V
W	T	R	X	G	L	V	D	U	Y	O	M	E	K	8	J	S	9	B	P	A	H	F	/	C	I	4	Q	N	3	Z	+	W
X	V	9	W	H	M	T	U	D	+	P	L	K	E	Z	S	J	R	Q	O	F	G	A	C	/	3	N	B	4	I	8	Y	X
Y	P	N	+	M	H	O	K	E	W	V	G	U	D	B	F	A	4	8	T	S	L	J	I	3	/	9	Z	R	C	Q	X	Y
Z	L	3	8	O	T	M	J	S	Q	G	V	A	F	X	D	U	I	+	H	E	P	K	4	N	9	/	Y	C	R	W	B	Z
9	U	X	R	F	S	D	V	T	4	K	J	P	O	3	M	L	W	C	E	H	A	G	Q	B	Z	Y	/	+	8	I	N	9
8	M	I	Z	P	V	L	S	J	B	H	T	F	A	W	U	D	3	Y	G	K	O	E	N	4	R	C	+	/	9	X	Q	8
+	O	4	Y	L	G	P	E	K	X	T	H	D	U	Q	A	F	N	Z	V	J	M	S	3	I	C	R	8	9	/	B	W	+
4	E	+	N	J	A	K	O	P	9	D	F	T	V	C	G	H	Y	3	U	L	S	M	Z	8	Q	W	I	X	B	/	R	4
3	J	Z	I	E	D	S	L	M	C	A	U	G	H	9	T	V	8	4	F	O	K	P	+	Y	X	B	N	Q	W	R	/	3
/	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	9	8	+	4	3	/

Appendix 2

Theoretical approach

The task is to find the correct starting position for the K wheel by counting the occurrence of '/'s in the sequences of characters formed by $\Delta\mathbf{Z} + \Delta\mathbf{K}$, for all the possible starting positions of this wheel.

There are two distinct situations to consider:

a) When the assumed starting position is *wrong*.

If

$$\Delta\mathbf{Z} + \Delta\mathbf{K} = /$$

then

$$\Delta\mathbf{Z} = \Delta\mathbf{K}.$$

The random probability that $\Delta\mathbf{Z} = \Delta\mathbf{K}$, written $p(\Delta\mathbf{Z} = \Delta\mathbf{K})$ is

$$\frac{1}{32}\ (\approx 0.031)$$

b) When the assumed starting position is *correct*.

Then

$$Z = P + K + S$$

so that

$$\Delta\mathbf{Z} = \Delta\mathbf{P} + \Delta\mathbf{K} + \Delta\mathbf{S}$$

and hence

$$\Delta\mathbf{Z} + \Delta\mathbf{K} = \Delta\mathbf{P} + \Delta\mathbf{S}$$

So if

$$\Delta\mathbf{Z} + \Delta\mathbf{K} = /$$

then

$$\Delta\mathbf{P} + \Delta\mathbf{S} = /$$

and hence

$$\Delta\mathbf{P} = \Delta\mathbf{S}$$

With the S wheel in use

$$S = A\ A\ B\ B\ A\ A\ B\ B \text{ ... etc.}$$

then

$$\Delta\mathbf{S} = /\ G\ /\ G\ /\ G\ /\ \text{ ... etc.}$$

Hence

$$p(\Delta\mathbf{S} = /) = \frac{1}{2}$$

Appendix 2 Theoretical approach *(continued)*

From the message (by counting the repeats) for the sequence of characters in $\Delta\mathbf{P}$, the probability of a / character,

$$p(\Delta\mathbf{P} = /) = \frac{2}{9}$$

so that

$$p(\Delta\mathbf{S} = / \text{ and } \Delta\mathbf{S} = /) = \frac{1}{2} \times \frac{2}{9} = \frac{1}{9} \ (\approx 0.111)$$

But if $\Delta\mathbf{P} \neq /$

then

$$\Delta\mathbf{P} = \Delta\mathbf{S}$$

is only satisfied when

$$\Delta\mathbf{P} = \Delta\mathbf{S} = \text{G}$$

$$p(\Delta\mathbf{S} = \text{G}) = \frac{1}{2} \text{ and } p(\Delta\mathbf{P} = \text{G}) = \frac{1}{32}$$

so that

$$p(\Delta\mathbf{S} = \text{G} \text{ and } \Delta\mathbf{P} = \text{G}) = \frac{1}{2} \times \frac{1}{32} = \frac{1}{64} \ (\approx 0.016)$$

So we deduce that

$$p(\Delta\mathbf{P} + \Delta\mathbf{S} = /) = 0.111 + 0.016 = 0.127$$

Given that in our example there are 54 characters in the message, when the K wheel is at any one of the wrong starting positions, the expected number of '/'s in the sequence $\Delta\mathbf{Z} + \Delta\mathbf{K}$ is

$$53 \times \frac{1}{32} \ (\approx 1.6)$$

When the K wheel is at the correct starting position, the expected number of '/'s in the sequence $\Delta\mathbf{Z} + \Delta\mathbf{K}$ is

$$53 \times 0.127 \ (\approx 7)$$

[**Note**: Almost all messages containing at least 50 characters and having characteristics similar to the given sample can be broken by this statistical method.]