

**University of Southampton**  
**in association with**  
**EPSRC, IBM, and EducationGuardian.co.uk**  
**proudly presents the second**  
**National Cipher Challenge**  
Starting October 1<sup>st</sup> 2003  
<http://www.maths.soton.ac.uk/cipher>



**Teachers' notes to accompany the lesson packs**

# On substitution ciphers

Graham A. Niblo  
University of Southampton National Cipher Challenge 2003

*These notes form a brief introduction to substitution ciphers, to accompany the lesson plans provided with the University of Southampton National Cipher Challenge, 2003. We would like to thank Hugh Evans of Sholing Technology College for his assistance in the design of these teaching materials.*

## Caesar shift ciphers

The easiest method of enciphering a text message is to replace each character by another using a fixed rule, so for example every letter a may be replaced by D, and every letter b by the letter E and so on.

Applying this rule to the previous paragraph produces the text

WKH HDVLHVW PHWKRQ RI HQFLSKHULQJ D WHAW PHVVDJH LV WR  
UHSODFH HDFK FKDUDFWHU EB DQRWKHU XVLQJ D ILAHG UXOH, VR  
IRU HADPSOH HYHUB OHWWHU D PDB EH UHSODFHG EB G, DQG  
HYHUB OHWWHU E EB WKH OHWWHU H DQG VR RQ.

(Note the convention that ciphertext is written in capital letters, while plaintext is usually lowercase.)

Such a cipher is known as a shift cipher since the letters of the alphabet are shifted round by a fixed amount, and as a Caesar shift since such ciphers were used by Julius Caesar. To decode a Caesar shift it is enough to work out the amount of shift, which can be done, for example, by discovering which character has replaced the letter e. In the example above we might guess that the three letter word starting the sentence is the and therefore that the letter e has been replaced by H. A quick check shows that the Caesar shift by 3 does indeed encode the word the as WKH, and it is easy to complete the decryption.

In fact there are only 26 Caesar shift ciphers (and one of them does nothing to the text) so it is not too hard to decipher the text by brute force. We can try each of the shifts in turn on the first word of the cipher text until we discover the correct shift. This process can be simplified by using a cipher wheel, a simple mechanical device which allows one to generate each of the Caesar shift ciphers, and to encode or decode messages using it. At the back of this pack you will find a sheet, which can be photocopied onto thin card in order to make a cipher wheel. Cut out the two discs, and fasten through their centres with a paper fastener to make the wheel. Use the convention that you read plaintext on the outer rim of the wheel and cipher text from the smaller wheel.

## Keyword substitution ciphers

To increase the difficulty of deciphering the text we need a richer family of ciphers. A good example is furnished by the keyword cipher. In this we design an encryption table by choosing a keyword or phrase, which is used to jumble the alphabet as follows:



Write down the phrase, with no spaces between the letters, and omitting any repeated character. So if the phrase is “The Simpsons” we write down THESIMPON. Now we continue to go round the alphabet until every letter appears exactly once, and write the list under the standard alphabet:

abcdefghijklmnopqrstuvwxyz  
THESIMPONQURUVWXYZABCFGJKL

Of course if the key phrase is carefully chosen (for example “The quick brown fox jumps over the lazy dog”) we may not need to complete the list as above, but such a choice is not necessary. The number of such ciphers is  $26!$ , or approximately  $10^{27}$ , and brute force cannot be used to attack the problem. However an attack is possible.

Consider the text

VEP HYXHLVHTP MO AWFJYFLT H RFNEPS HJNEHAPV FL VEFU ZHC  
FU VEHV FV FU PHUC VM KPKMSFUP VEP IPCZMSY MS IPCNESHUP,  
HLY EPLRP VEP RFNEPS HJNEHAPV. VEFU FU FKNMSVHLV, APRHWUP  
FO VEP UPLYPS EHU VM IPPN VEP RFNEPS HJNEHAPV ML H NFPRP  
MO NHNPS, VEP PLPKC RHL RHNWSP VEP NHNPS, YFURMXPS VEP  
IPC, HLY SPHY HLC RMKKWLF RHVFMLU VEHV EHXP APPL PLRSCNVPY  
ZFVE FV. EMZPXPS FO VEP IPC RHL AP RMKKFVVPY VM KPKMSC FV  
FU JPUU JFIPJC VM OHJJ FLVM PLPKC EHLYU.

As before we notice that the first word has three letters and, since it occurs several times, may well be the word the. This gives a strong hint that the letter e is enciphered as the letter P in the keyphrase cipher. Of course other three letter words are possible, *e.g.*, and or but. Nonetheless a quick check shows us that the letter P is the most common letter in the enciphered text, just as e is the most common letter in English so it is reasonable to assume that the correct decryption translates P to e. This also suggest that V stands for t and E for h, allowing us to begin to decipher the text. We will use the convention that uppercase letters denote enciphered letters and lowercase denotes plaintext characters:

the HYXHLtHtE MO AWFJYFLT H RFNheS HJNhhAet FL thFU ZHC  
FU thHt Ft FU eHUC tM KeKMSFue the IeCZMSY MS IeCNhSHUe,  
HLY heLRe the RFNheS HJNhhAet. thFU FU FKNMStHLt, AeRHWUe  
FO the UeLYeS hHU tM Ieen the RFNheS HJNhhAet ML H NFRe  
MO NHNeS, the eLeKC RHL RHNtWSe the NHNeS, YFURMXeS the  
IeC, HLY SeHY HLC RMKKWLF RHtFMLU thHt hHXe AeEL eLRSCNteY  
ZFth Ft. hMZeXeS FO the IeC RHL Ae RMKKFtteY tM KeKMSC Ft  
FU JeUU JFIeJC tM OHJJ FLtM eLeKC hHLYU.

Reading carefully we see the single letter word H, and the four letter word thHt in the first line, and guess that H enciphers the letter a. Making that replacement we get:

the aYXaLtaTe MO AWFJYFLT a RFNheS aJNhaAet FL thFU ZaC  
FU that Ft FU eaUC tM KeKMSFUe the IeCZMSY MS IeCNhSaUe,  
aLY heLRe the RFNheS aJNhaAet. thFU FU FKNMStaLt, AeRaWUe  
FO the UeLYeS haU tM Ieen the RFNheS aJNhaAet ML a NFere  
MO NaNeS, the eLeKC RaL RaNtWSe the NaNeS, YFURMXeS the  
IeC, aLY SeaY aLC RMKKWLFratFMLU that haXe AeEL eLRSCNteY  
ZFth Ft. hMZeXeS FO the IeC RaL Ae RMKKFtteY tM KeKMSC Ft  
FU JeUU JFIeJC tM OaJJ FLtM eLeKC haLYU.

Now the two 2 letter words Ft FU are probably “it is” meaning that F enciphers “i” and U enciphers “s”. Hence we get:

the aYXaLtaTe MO AWiJYiLT a RiNheS aJNhaAet iL this ZaC  
is that it is easC tM KeKMSise the IeCZMSY MS IeCNhSase,  
aLY heLRe the RiNheS aJNhaAet. this is iKNMStaLt, AeRaWse  
iO the seLYeS has tM Ieen the RiNheS aJNhaAet ML a NieRe  
MO NaNeS, the eLeKC RaL RaNtWSe teh NaNeS, YisRMXeS the  
IeC, aLY SeaY aLC RMKKWLiRatiMLs that haXe AeEL eLRSCNteY  
Zith it. hMZeXeS iO the IeC RaL Ae RMKKitteY tM KeKMSC it  
is Jess JiIeJC tM OaJJ iLtM eLeKC haLYs.

Continuing with appropriate guesses (haXe = have, easC = easy and so on) we decipher the text to get the following extract from Simon Singh’s excellent history of codes and ciphers, *The Code Book*:

“The advantage of building a cipher alphabet in this way is that it is easy to memorise the keyword or keyphrase, and hence the cipher alphabet. This is important, because if the sender has to keep the cipher alphabet on a piece of paper, the enemy can capture the paper, discover the key, and read any communications that have been encrypted with it. However if the key can be committed to memory it is less likely to fall into enemy hands.”



## Frequency analysis

A more methodical attack is frequency analysis. One counts the number of occurrences of each character in the cipher text and compares it with an expected frequency for the standard English alphabet. In the cipher text above a character count gives us the following table of occurrences:

a	b	c	d	e	f	g	h	i	j	k	l	m
7	0	12	0	26	27	0	32	6	9	11	20	18
n	o	p	q	r	s	t	u	v	w	x	y	z
16	5	55	0	14	17	2	17	35	4	4	11	4

Compare this to a table of expected frequencies, taken from Simon Singh's "The Code Book":

a	b	c	d	e	f	g	h	i	j	k	l	m
8.2	1.5	2.8	4.3	12.7	2.2	2.0	6.1	7.0	0.2	0.8	4.0	2.4
n	o	p	q	r	s	t	u	v	w	x	y	z
6.7	7.5	1.9	0.1	6.0	6.3	9.1	2.8	1.0	2.4	0.2	2.0	0.1

Using this and information about common one, two and three letter words we have enough to begin to tackle the cipher.

## Disguising the word structure

The chink in the armour of our ciphers so far has been the preservation of word structure. This allows one to spot common words. In order to avoid such weakness cryptographers usually remove punctuation and block the characters together in groups of four or five, so our previous cipher text looks like

```
VEPHY XHLVH TPMOA WFJYF LTHRF NEPSH JNEHA PVFLV EFUZH
CFUVE HVFVF UPHUC VMKPK MSFUP VEPIP CZMSY MSIPC NESHU
PHLYE PLRPV EPRFN EPSHJ NEHAP VVEFU FUFKN MSVHL VAPRH
WUPFO VEPUP LYPSE HUVMI PPNVE PRFNE PSHJN EHAPV MLHNF
PRPMO NHNPS VEPPL PKCRH LRHNV WSPVE PNHNP SYFUR MXPSV
EPIPC HLYSP HYHLC RMKKW LFRHV FMLUV EHVEH XPAPP LPLRS
CNVPY ZFVEF VEMZP XPSFO VEPIP CRHLA PRMKK FVVPY VMKPK
MSCFV FUJPU UJFIP JCVMO HJJFL VMPLP KCEHL YU
```

Usually the length of the text groups doesn't matter, however, in analysing a Vigenère cipher (see below) a carelessly chosen block length may make the length of the keyword more apparent, since it can reveal the repetitions more easily.

To attack cipher text that has been grouped in this way we have to work with letters not words. To do so we use the frequency analysis described above, together with a little judgement (or luck!). The process can be hard, but wars have been won or lost on the back of it, and so have fortunes.

“It was hard going, but Jericho didn’t mind. He was taking action, that was the point. It was the same as code-breaking. However hopeless the situation, the rule was always to do *something*. No cryptogram, Alan Turing used to say, was ever solved by simply staring at it.” From *Enigma*, by Robert Harris.

## Affine shift ciphers

Despite the advantages for an agent in using keyword substitution ciphers most modern ciphers are automated and rely on a mathematical encryption algorithm. Indeed the Caesar shift cipher can be viewed as just such a cipher:

We start by encoding each letter by its numerical position in the alphabet:

a	b	c	d	e	f	g	h	i	j	k	l	m
1	2	3	4	5	6	7	8	9	10	11	12	13
n	o	p	q	r	s	t	u	v	w	x	y	z
14	15	16	17	18	19	20	21	22	23	24	25	26

Next we shift the alphabet by adding 3 to each position:

a	b	c	d	e	f	g	h	i	j	k	l	m
4	5	6	7	8	9	10	11	12	13	14	15	16
n	o	p	q	r	s	t	u	v	w	x	y	z
17	18	19	20	21	22	23	24	25	26	1	2	3

Of course  $24+3 = 27 \neq 1$ , but here we are carrying out modular arithmetic, familiar as clock arithmetic, so that when we reach 26 we continue from 1.

Finally we replace the numbers with the letters they stand for:

a	b	c	d	e	f	g	h	i	j	k	l	m
d	e	f	g	h	i	j	k	l	m	n	o	p
n	o	p	q	r	s	t	u	v	w	x	y	z
q	r	s	t	u	v	w	x	y	z	a	b	c

This recovers the cipher table constructed in lesson plan 1 for the Caesar shift by 3.

There is a convenient shorthand for the Caesar shift by  $n$ , given by  $x \rightarrow x+n$ . It is confusing since here we are using  $x$  to stand for the position of a letter, and  $n$  to stand for the shift amount, *i.e.*,  $x$  and  $n$  are each one of the values  $1 \dots 26$ . It is clear that since the shift is defined by the integer  $n$  there are only 26 Caesar shift ciphers.

There is a bigger class of shift ciphers which can be written in these terms known as the affine shift ciphers, and they exploit the fact that we can multiply as well as add integers in modular arithmetic. It is slightly complicated to set up formally but rather easy to do in practice so we will work through an example.



### The affine shift $x \rightarrow 3x+5$

We start as before with the position table, but this time instead of replacing a position  $x$  with the number  $x+3$  we will replace it by the number  $3x+5$ , where this number is interpreted appropriately. So for example  $2 \rightarrow 3 \cdot 2+5 = 11$ , while  $8 \rightarrow 3 \cdot 8+5 = 29$  which is interpreted as 3 ( $29=26+3$ ). Whenever the result of the computation is larger than 26 we keep subtracting 26 until it becomes smaller. More formally we compute  $3x+5$  and then take the remainder after division by 26. This yields the table:

a	b	c	d	e	f	g	h	i	j	k	l	m
8	11	14	17	20	23	26	3	6	9	12	15	18
n	o	p	q	r	s	t	u	v	w	x	y	z
21	24	1	4	7	10	13	16	19	22	25	2	5

And from this we recover the encryption table as given on the handout for lesson 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
H	K	N	Q	T	W	Z	C	F	I	L	O	R	U	X	A	D	G	J	M	P	S	V	Y	B	E

The affine shift ciphers can also be written in a shorthand form  $x \rightarrow ax+b$  and the Caesar shift ciphers are special cases of the affine shift ciphers with  $a=1$ .

Now notice that in both the Caesar shift  $x \rightarrow x+3$  and the affine shift  $x \rightarrow 3x+5$  the letter y is enciphered as B, since  $25+3 = 28 = 26+2$ , and  $3 \cdot 25+5 = 80 = 3 \cdot 26+2$ . It follows that two different affine shift ciphers can encrypt a letter in the same way, so it is no longer sufficient to discover the letter substituting for e in order to decipher the message. Since there are two degrees of freedom in our choice of cipher we might hope that deciphering two letters is sufficient, and it is, since, if we know two values of the expression  $ax+b$  we can solve the two corresponding simultaneous equations to find  $a$  and  $b$ .

We may be more familiar with this exercise when solving the equations over the real numbers, but the same method works for modular arithmetic, with the caveat that in general we cannot divide. This caveat has an interpretation in cryptography. In order for the rule  $x \rightarrow ax+b$  to define a cipher it had better be the case that each of the numbers  $1 \dots 26$  appears exactly once in the list of numbers  $ax+b$  as  $x$  ranges from 1 to 26. If we choose  $a$  carelessly (so that we can't divide by  $a \bmod 26$ ) this might not be the case.

For example the rule  $x \rightarrow 2x$  tries to encipher both m and z as Z, since  $2 \cdot 13 = 26$  and  $2 \cdot 26 = 52$  both of which are equal to 26 modulo 26. Such an encryption cannot easily be deciphered since the recipient of the message is unable to determine whether the sender intended Z to be read as m or z.

From a mathematician's point of view the enciphering rule defines a function from the alphabet to itself, and this needs an inverse if the cipher is to be decipherable in a deterministic way. In other words the number theory function  $x \rightarrow ax+b$  needs to have an inverse in mod 26 arithmetic. It is a fact from elementary number theory that it will have such an inverse if and only if  $a$  is coprime to 26, that is, their only common divisor is 1.

There are 12 numbers less than 26 and coprime to it (those odd numbers not divisible by 13) so we have 12 possible choices of the number  $a$ , and 26 choices for the number  $b$ , yielding 312 affine shift ciphers. This makes a brute force attack, without frequency analysis, less practical than the much simpler situation for Caesar shift ciphers.

## Polyalphabetic ciphers

The main weakness allowing us to tackle a substitution cipher is the irregularity in the distribution of letters in English text. Other languages demonstrate similar (though language specific) irregularities and you can find frequency tables for them on the web.

In order to remove this weakness from a cipher it is necessary to disguise the frequencies of letters in the plaintext and the easiest way to do this is by using a polyalphabetic cipher. In such a cipher each plaintext letter may be encoded in more than one way so that, for example, the letter e may be enciphered as both X and G within the ciphertext. One problem with this approach is that if X and G both encode for e we don't have enough letters left to encode the other 25 letters. One elegant solution to this problem is the famous French cipher known as the Vigenère cipher.

In a Vigenère cipher ANY letter might be encoded by any other; a given Vigenère cipher uses a subset of the 26 possible Caesar shift ciphers. Of course for a genuine recipient to have any hope of deciphering the message there has to be a way to determine for each cipher character which of the shifts has been used. The answer to this tricky problem is to choose a sequence of them known to both parties but to no-one else.

So the two parties might agree to use shifts of 22, 9, 7, 5, 14, 5 18, and 5 in that order and to continue repeating the pattern for the entire text: 22, 9, 7, 5, 14, 5 18, 5, 22, 9, 7, 5, 14, 5 18, 5, 22, 9 *etc.*

In order to decode the cipher text the recipient shifts the first cipher character back by 22, the second back by 9 and so on to recover the cipher text. Of course the question remains how one can memorise the correct sequence, but here we borrow an idea from the keyword cipher. The shift numbers 1, ..., 26 are taken to stand for the alphabet a, ..., z, and then the pattern 22, 9, 7, 5, 14, 5 18, 5 spells the word vigenere.



To set up a Vigenère cipher the two parties agree in advance to use the shift pattern encoded by some agreed keyword or phrase; in our previous Golden Jubilee Cipher challenge we used a Vigenère cipher based on the keyword GOLD, so characters were shifted in turn by 7, 15, 12, 4. Such a cipher is very hard to crack.



The method we recommend is due to Babbage and Kasiski who independently discovered it, and is based on the regularity of the repetition. An analysis of repeated strings of letters is used to try to determine the length of the keyword, and once this is done a standard frequency analysis is applied to each part of the ciphertext encoded by a single cipher. A very good account of Babbage-Kasiski deciphering can be read in chapter 2 of Simon Singh's *The Code Book*.

# Cryptography Lesson Plan 1

**Class: Cracking the Caesar shift ciphers.**

**Resources:**

- Leaflet “On substitution ciphers”.
- Two handouts each with a plaintext and a cipher table
- Teachers’ solutions for the handouts.
- One OHP slide with cipher text to crack, and partial decrypt and solution.

**Starter: (10 minutes approximately) Uses handouts for Groups A and B**  
*Encryption exercise – split the class into groups A and B. Give each group the enclosed text to encipher using the given code. Encourage accuracy AND secrecy! Answers enclosed with handouts.*

**Main activity: (40 minutes approx) Uses OHP**

- *Introduce the idea of a substitution cipher in general and the Caesar shift in particular.*
- *Suggest trial and error as a deciphering technique.*
- *Work through a very simple Caesar shift (by 3).*
- *Split the class again, swap over the ciphertexts from the starter exercise and get them to tackle them.*

**Plenary (approx 10 minutes)**

*Discuss how to make the code harder to crack using a rule that is harder to determine, but remark on the need for an easy to remember rule (stressed agents must remember it and can’t write it down!) Mention “keyword” substitution.*

**Handout for lesson 1.**

**GROUP A**

**Code: Caesar shift by 2**



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
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

**Plaintext**

There were plenty of schools in the world, but they were all run either by the various churches or the Guilds. Miss Butts objected to churches on logical grounds and deplored the fact that the only Guilds that considered girls worth educating were the Thieves and the Seamstresses. It was a big and dangerous world out there and a girl could do worse than face it with a sound knowledge of geometry and astronomy under her bodice.  
From “Soul Music” by Terry Pratchett.

## Handout for lesson 1.

### GROUP B

Code: Caesar shift by 4

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
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

The four houses are called Gryffindor, Hufflepuff, Ravenclaw and Slytherin. Each house has its own noble history and each has produced outstanding witches and wizards. While you are at Hogwarts, your triumphs will earn your house points, while any rule-breaking will lose house points. At the end of the year the house with the most points is awarded the House Cup, a great honour. I hope each of you will be a credit to whichever house becomes yours.

From “Harry Potter and the Philosopher’s Stone” by J.K. Rowling.

**Teachers' solutions to encryption challenge**



**Ciphertext A**

VJGTG YGTG RNGPVA QH UEJQQNU KP VJG YQTNF, DWV VJGA YGTG  
CNN TWP GKVJGT DA VJG XCTKQWU EJWTEJGU QT VJG IWKNFU.  
OKUU DWVVU QDLGEVGF VQ EJWTEJGU QP NQIKECN ITQWPFU CPF  
FGRNQTF VJG HCEV VJCV VJG QPNA IWKNFU VJCV EQPUKFGTGF  
IKTNU YQTVJ GFWECVKPI YGTG VJG VJKGXGU CPF VJG  
UGCOUVTGUUGU. KV YCU C DKI CPF FCPIGTQWU YQTNF QWV VJGTG  
CPF C IKTN EQWNF FQ YQTUG VJCP HCEG KV YKVJ C UQWPF  
MPQYNGFIG QH IGQOGVTA CPF CUVTQPQOA WPFGT JGT DQFKEG.

**Ciphertext B**

XLI JSYV LSYWIW EVI GEPPIH KVCJJMRHSV, LYJJPITYJJ,  
VEZIRGPEA ERH WPCXLIVMR. IEGL LSYWI LEW MXW SAR RSFPI  
LMWXSVC ERH IEGL LEW TVSHYGIH SYXWXERHMRK AMXGLIW ERH  
AMDEVHW. ALMPI CSY EVI EX LSKEVXW, CSYV XVMYQTLW AMPP  
IEVR CSYV LSYWI TSMRXW, ALMPI ERC VYPI-FVIEOMRK AMPP PSWI  
LSYWI TSMRXW. EX XLI IRH SJ XLI CIEV XLI LSYWI AMXL XLI  
QSWX TSMRXW MW EAEVHIH XLI LSYWI GYT, E KVIEX LRSYV. M  
LSTI IEGL SJ CSY AMPP FI E GVIHMX XS ALMGLIZIV LSYWI  
FIGSQIW CSYVW.

## OHP Slide for lesson 1

### Ciphertext

*WKH HDVLHVW PHWKRG RI HQFLSKHULQJ D WHAW PHVVDJH LV WR  
UHSODFH HDFK FKDUDFWHU EB DQRWKHU XVLQJ D ILAHG UXOH, VR  
IRU HADPSOH HYHUB OHWWHU D PDB EH UHSODFHG EB G, DQG HYHUB  
OHWWHU E EB WKH OHWWHU H DQG VR RQ.*

**Partial decrypt:** Guess that the first word is “the” so that t is enciphered as W, h as K and e as H. This suggests a shift by 3:

the eDVLeVt PethRG RI eQFLSheULQJ D  
teAt PeVVDJe LV tR UeSODFe eDFh  
FhDUDFteU EB DQRtheU XVLQJ D ILAeG  
UXOe, VR IRU eADPSOe eYeUB OetteU D  
PDB Ee UeSODFeG EB G, DQG eYeUB OetteU  
E EB the OetteU e DQG VR RQ.

The word teAt could be tent, test or text, with text fitting with the shift by 3; the word OetteU which occurs twice, would decipher to “letter” confirming our guess.

### Code: Caesar shift by 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
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

### Plaintext

The easiest method of enciphering a text message is to replace each character by another using a fixed rule, so for example every letter a may be replaced by d, and every letter b by the letter e and so on.

## Cryptography Lesson Plan 2

**Class:** Cracking keyword substitution ciphers – emphasises letter frequency analysis and team work.

**Resources:**

- Leaflet “On substitution ciphers”.
- OHP 1 containing ciphertext
- OHP 2 Containing expected frequency table and incomplete actual frequencies.
- Handout summarising details of deciphering technique.
- OHP 3 With further thoughts on disguising the text.



**Starter: (10 minutes approximately) (Uses OHP 1)**

*Split the class into teams and get them to count the letter frequencies in the ciphertext. Emphasise the need for speed and accuracy. Maybe set the scene as a race against time.*

**Main activity: (30 minutes approx) (Uses OHP 1 and OHP 2 and handout)**

- *Introduce the idea of a keyword cipher to make encryption more secure and more memorable (see “On substitution ciphers”).*
- *Discuss the hunt for common words and letters and introduce frequency analysis – show a table of common frequencies and check it against the examples in lesson 1.*
- *Discuss the speed improvements given by parallel processing of the text. Split into 26 teams to do a frequency analysis of the given ciphertext on OHP 1. {It may be worth remarking that standard computer attacks on ciphers use this idea of parallel processing to speed up the attack.}*
- *Whole class session to construct frequency table, compare with expected frequencies (computed from percentages) and identify the letters “e” and “t”.*

**Plenary (20 minutes approx) (Uses OHP 2 and, time permitting OHP 3)**

*Draw together the intelligence gained by the groups and crack the cipher together. (You may wish to give out the handout summarising the technique after completing the exercise.)*

*If time permits (OHP 3):*

- *Discuss how to make the code harder to crack by disguising the letter groups.*
- *Remark that the frequency table can mislead for non-standard or foreign language texts! Examine the extract from the book “A Void” by Georges Perec.*

## Ciphertext

VEP HYXHLVHTP MO AWFJYFLT H RFNEPS HJNEHAPV  
FL VEFU ZHC FU VEHV FV FU PHUC VM KPKMSFUP  
VEP IPCZMSY MS IPCNESHUP, HLY EPLRP VEP  
RFNEPS HJNEHAPV. VEFU FU FKNMSVHLV, APRHWUP  
FO VEP UPLYPS EHU VM IPPN VEP RFNEPS  
HJNEHAPV ML H NFPRP MO NHNPS, VEP PLPKC RHL  
RHNWVSP VEP NHNPS, YFURMXPS VEP IPC, HLY  
SPHY HLC RMKKWLF RHVFMLU VEHV EHXP APPL  
PLRSCNVPY ZFVE FV. EMZPXPS FO VEP IPC RHL  
AP RMKKFVVPY VM KPKMSC FV FU JPUU JFIPJC VM  
OHJJ FLVM PLPKC EHLYU.





**OHP Slide 2 for lesson 2**  
**Occurrences 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

**Expected Frequency table**

a	b	c	d	e	f	g	h	i	j	k	l	m
8.2	1.5	2.8	4.3	12.7	2.2	2.0	6.1	7.0	0.2	0.8	4.0	2.4
n	o	p	q	r	s	t	u	v	w	x	y	z
6.7	7.5	1.9	0.1	6.0	6.3	9.1	2.8	1.0	2.4	0.2	2.0	0.1

This table was taken from “The Code Book” by Simon Singh, and gives expected frequencies as a percentage. To accurately compare it to the actual frequencies above you should compute the actual frequencies as percentages.

**Actual Frequencies as percentages**

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

## Handout for lesson 2

STAGE 1 – P is the commonest letter in the ciphertext so could stand for e - maybe the first word is the:

the HYXHLtHtE MO AWFJYFLT H RfNheS HJNhHAet FL thFU ZHC  
FU thHt Ft FU eHUC tM KeKMSFUE the IeCZMSY MS IeCNhSHUE,  
HLY heLRe the RfNheS HJNhHAet. thFU FU FKNMStHLt, AeRHWUe  
FO the UeLYeS hHU tM Ieen the RfNheS HJNhHAet ML H NFeRe  
MO NHNeS, the eLeKC RHL RHntWSe the NHNeS, YFURMXeS the  
IeC, HLY SeHY HLC RMKKWLFrHtFMLU thHt hHXe AeEL eLRSCNteY  
ZFth Ft. hmZeXeS FO the IeC RHL Ae RMKKFtteY tM KeKMSC Ft  
FU JeUU JFIeJC tM OHJJ FLtM eLeKC hHLYU.

STAGE 2 We see the single letter word H, and the four letter word thHt in the first line - guess that H encodes the letter a.

the aYXaLtaTe MO AWFJYFLT a RfNheS aJNhaAet FL thFU ZaC  
FU that Ft FU eaUC tM KeKMSFUE the IeCZMSY MS IeCNhSaUE,  
aLY heLRe the RfNheS aJNhaAet. thFU FU FKNMStaLt, AeRaWUe  
FO the UeLYeS haU tM Ieen the RfNheS aJNhaAet ML a NFeRe  
MO NaNeS, the eLeKC RaL RaNtWSe the NaNeS, YFURMXeS the  
IeC, aLY SeaY aLC RMKKWLFraTfMLU that haXe AeEL eLRSCNteY  
ZFth Ft. hmZeXeS FO the IeC RaL Ae RMKKFtteY tM KeKMSC Ft  
FU JeUU JFIeJC tM OaJJ FLtM eLeKC haLYU.

STAGE 3 The two 2 letter words Ft FU are probably it is meaning that F encodes i and U encodes s:

the aYXaLtaTe MO AWiJYiLT a RiNheS aJNhaAet iL this ZaC  
is that it is easC tM KeKMSe the IeCZMSY MS IeCNhSase,  
aLY heLRe the RiNheS aJNhaAet. this is iKNMStaLt, AeRaWse  
iO the seLYeS has tM Ieen the RiNheS aJNhaAet ML a NieRe  
MO NaNeS, the eLeKC RaL RaNtWSe teh NaNeS, YisRMXeS the  
IeC, aLY SeaY aLC RMKKWLiRatiMLs that haXe AeEL eLRSCNteY  
Zith it. hmZeXeS iO the IeC RaL Ae RMKKitteY tM KeKMSC it  
is Jess JiIeJC tM OaJJ iLtM eLeKC haLYs.

STAGE 4: haXe = have, easC = easy and so on - we get the following extract from Simon Singh's excellent history of codes and ciphers, *The Code Book*:

"The advantage of building a cipher alphabet in this way is that it is easy to memorise the keyword or keyphrase, and hence the cipher alphabet. This is important, because if the sender has to keep the cipher alphabet on a piece of paper, the enemy can capture the paper, discover the key, and read any communications that have been encrypted with it. However if the key can be committed to memory it is less likely to fall into enemy hands."

**OHP Slide 3 for lesson 2**



**Obscuring a substitution cipher**

1. We can disguise the word structure by regrouping the letters into blocks:

VEPHY XHLVH TPMOA WFJYF LTHRF NEPSH JNEHA  
PVFLV EFUZH CFUVE HVFVF UPHUC VMKPK MSFUP  
VEPIP CZMSY MSIPC NESHU PHLYE PLRPV EPRFN  
EPSHJ NEHAP VVEFU FUFKN MSVHL VAPRH WUPFO  
VEPUP LYPSE HUVMI PPNVE PRFNE PSHJN EHAPV  
MLHNF PRPMO NHNPS VEPPL PKCRH LRHNV WSPVE  
PNHNP SYFUR MXPSV EIPIC HLYSP HYHLC RMKKW  
LFRHV FMLUV EHVEH XPAPP LPLRS CNVPY ZFVEF  
VEMZP XPSFO VEPIP CRHLA PRMKK FVVPY VMKPK  
MSCFV FUJPU UJFIP JCVMO HJJFL VMPLP KCEHL YU

2. We can distort the frequency table – this text was adapted for last years cipher challenge!

Augustus, who has had a bad night, sits up blinking and purblind. Oh what was that word (is his thought) that ran through my brain all night, that idiotic word that, hard as I'd try to pin it down, was always just an inch or two out of my grasp - fowl or foul or Vow or Voyal? - a word which, by association, brought into play an incongruous mass and magma of nouns, idioms, slogans and sayings, a confusing, amorphous outpouring which I sought in vain to control or turn off but which wound around my mind a whirlwind of a cord, a whiplash of a cord, a cord that would split again and again, would knit again and again, of words without communication or any possibility of combination, words without pronunciation, signification or transcription but out of which, notwithstanding, was brought forth a flux, a continuous, compact and lucid flow: an intuition, a vacillating frisson of illumination as if caught in a flash of lightning or in a mist abruptly rising to unshroud an obvious sign - but a sign, alas, that would last an instant only to vanish for good.

From "A Void" by Gilbert Adair. The letter "e" does not appear even once in the book!

## Cryptography Lesson Plan 3

**Class:** Affine shift ciphers – emphasises clock arithmetic and gives more practice at frequency analysis.

**Resources:**

- Leaflet “On substitution ciphers”.
- OHP 1, giving partial encryption table for the  $3x+5$  affine shift cipher together with teachers’ solution.
- OHP 2-4, with cipher text to crack, method and solution.

**Starter: (10 minutes approximately) Uses handout**

*Complete the encryption table on the OHP (the affine shift cipher  $x \rightarrow 3x+5$  is discussed in the teachers’ notes).*

*Encourage them to try to spot the pattern and guess the rule which should be concealed.*

**Main activity: (40 minutes approx) Uses OHP**

- *Introduce the class of affine shift ciphers mentioning “clock arithmetic” mod 26*
- *Show them that the cipher table arises from the affine shift  $x \rightarrow 3x + 5$ .*
- *Discuss the fact that you only need to know the value of two letters to deduce the affine shift (solving two simultaneous equations mod 26).*
- *Use frequency analysis and modular arithmetic to decipher an affine shifted text together or in groups.*

**Plenary (approx 10 minutes)**

*Discuss generalisations to modular arithmetic mod  $n$ .*

OHP slide 1 for lesson 3.



Spot the pattern?

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
H	K	N					C																		

$$x \rightarrow 3x + 5$$

1	2	3	4	5	6	7	8	9	10	11	12	13
8	11	14	17	20	23	26	3	16	9	12	15	18
14	15	16	17	18	19	20	21	22	23	24	25	26
21	24	1	4	7	10	13	16	19	22	25	2	5

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

## OHP Slide 2 for lesson 3

### Ciphertext

LMYFU BKUUS DDYFA XWCLA OLPSF AOLMJ FASDS  
NSFGJ FAOEL SOMYT DJLAX EMHJM BFMIB JUMIS  
HFSUL AXUBA FKJAM XLSKF FKXWS DJLSO FGBJM  
WFKIU OLFMX MTMWA OKTTG JLSXL SKFFK XWSDJ  
LSIZG TSXWJ LJLSX LSUMF JSDJL SIZGH FSQYS  
XOGLS DMMDT SDMXJ LSBAT SMHBK BSFLS BFMCT  
SDKFM YXDJL SLYJM ZTANA MYUXM CJMCL MCKUT  
MMEAX WKJLA IKXDC LMCKU XJJLA UCKUC LKJAJ  
LKDZS SXTAE SHMFJ SXAXJ SFIAX KZTSI MXJLU  
TKUJG SKFXM CMXDS FLSLK DWMXS IKDJL SOLMF  
YUTAX SMHIS KXAXW TSUUT SJJSF UDKXO SDZSH  
MFSLA USGSU ZYJJL SGCSF SXMJI SKXAX WTSUU  
JLSGC SFSTM KDSDC AJLJL SIMUJ NAJKT ISKXA  
XWAIK WAXKZ TSAHM XTGLS OMYTD HAXDA JZYJC  
LSFSC KUJLS BKJJS FXCLS FSCKU JLSBK JJSFX  
CLSFS CKUJL SBKJJ SFXHF MISXA WIKZG FMZSF  
JLKFF AU

### Occurences table:

A	B	C	D	E	F	G	H	I	J	K	L	M
34	12	19	23	4	41	12	10	16	50	38	46	42
N	O	P	Q	R	S	T	U	V	W	X	Y	Z
3	11	1	1	0	78	22	28	0	13	41	11	11



**OHP Slide 3 for lesson 3**

**Use frequency analysis to guess that S enciphers for e, and J for t.**

This tells us that for an affine shift cipher

$$x \rightarrow ax + b$$

$$a.5 + b = 19 \quad (e \rightarrow s)$$

$$a.20 + b = 10 \quad (t \rightarrow j)$$

Solving mod 26 we see that  $15a \equiv -9 \pmod{26}$ . Now 7.15 is congruent to 1 mod 26 since  $7.15 = 105 = 104 + 1 = 4.26 + 1$ . It follows that  $7.15a \equiv 7.-9$ , or  $a$  is congruent to  $-63$ .

Now  $-63 = -52 - 11$ , so  $a$  is congruent to  $-11$ , or equivalently to  $15 \pmod{26}$ . Hence  $a = 15$ . Now from  $a.5 + b = 19$  we get  $75 + b$  is congruent to 19, or  $b$  is congruent to  $-56 \pmod{26}$ .

Since  $-56 = -2.26 - 4$ ,  $b$  is congruent to  $-4 \pmod{26}$  so  $b = 22$ .

To check this  $20a + b = 300 + 22 = 322 = 12.26 + 10$ , so  $a.20 + b = 10$  as required. So the affine shift is  $x \rightarrow 15x + 22$  and the decrypt is given by the inverse function  $y \rightarrow 7(y - 22)$

[It might look strange but “dividing by 15” is the same as multiplying by 7 in mod 26 arithmetic.]

Equivalently the decryption is achieved by the affine shift  $y \rightarrow 7y + 2$ .

**Encryption table:**

a	b	c	d	e	f	g	h	i	j	k	l	m
I	P	W	D	K	R	Y	F	M	T	A	H	O
n	o	p	q	r	s	t	u	v	w	x	y	z
V	C	J	Q	X	E	L	S	Z	G	N	U	B

## Decrypt

hours passe dduri ngwhi chjer ichot riede  
veryt rickh ecoul dthin kofto promp tsome  
fresh inspi ratio nhear range dthec rypto  
grams chron ologi cally thenh earra ngedt  
hemby lengt hthen hesor tedth embyf reque  
ncyhe doodl edont hepil eofpa perhe prowl  
edaro undth ehuto blivi ousno wtowh owasl  
ookin gathi mandw howas ntthi swasw hatit  
hadbe enlik efort enint ermin ablem onths  
lasty earno wonde rheha dgone madth echor  
uslin eofme aning lessl etter sdanc edbef  
orehi seyes butth eywer enotm eanin gless  
theyw erelo adedw ithth emost vital meani  
ngima ginab leifo nlyhe could findi tbutw  
herew asthe patte rnwhe rewas thepa ttern  
where wasth epatt ernfr omeni gmaby rober  
tharr is