4 Cryptography

```
LOBBY VULTM XYQBB UWZGY QVTTB RYVZU VQZEB XZDNV KKQHI BKBHO UBWBU ZDLQY
ZWBRZ WBYQV TPBTT HLBNV YQZDY YZZPD RQYOZ DMKBV YVTRZ WBWDT VULHU HGGVU
BRZWB TZVYV THTVP EKBTD MTYVY DYVZU RZWBT VURBV YVTEO BTBUY BWVUM KZRFT
ZGGVI BVYTQ HOWBO YZWBR ZWBYQ BGVOT YEBOT ZUVUY QBRKH TTNQZ WBRZW BTVYH
UWUZY VGVBT PBMXB PHVKL BYTHT EBRVH KEOVC B
```

A message such as this is meant to appear as a meaningless jumble to most anyone who reads it. It is a simple message coded in a simple way. In this section, we shall learn different ways to code messages, and some approaches used to decode them.

Cryptography is a discipline which concerns itself with communication secrecy. Coded messages have long been used by businesses, governments and the military, and for obvious reasons. If you want to send a message to a friend or partner, you do not want it to understood by everyone who intercepts that message. Even before the advent of the computer, and the possibility of sending (and intercepting) electronic messages, it was well understood that some coding device was necessary to insure secrecy from "the enemy" who might intercept messages.

With the advent of the computer, issues of secrecy have come closer to home, because electronic interception (snooping) has become commonplace, and much easier. When ordering on the Internet, and you give your credit card number, you do not want any intruder to find it. So responsible companies who do business on the Internet must use very sophisticated coding systems to keep this information secret from intruders. In this section we consider some of the methods used to code and decode messages.

For simplicity in our treatment, we shall start by sending with messages using the 26 letters A through Z. Thus, we omit small letters, punctuation, spaces, and even numerals. It is convenient to assign numbers to these letters, and the most natural numbering is given in the following table.

Letter	A	B	C	D	E	F	G	H	I	J	K	L	M
Number	1	2	3	4	5	6	7	8	9	10	11	12	13

Letter	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Number	14	15	16	17	18	19	20	21	22	23	24	25	0

Note that we have assigned Z the number 0, rather than 26. This is because we shall work modulo 26–with remainders when we divide by 26–so it is most convenient to use the numbers from 0 through 25. (Computer scientists usually start labeling A as 0, etc. but we have avoided this, since most of us have learned that A is the first letter, B the second, etc.) It will be useful to learn this table. I find the simplest way is to know that the numbers 5, 10, 15, 20, and 25 correspond to E, J, O, T, and Y. If you know your alphabet, the rest is easy.

Shift Substitutions. A very simple coding device is to replace every letter by the letter two positions after it. Thus, A is replaced by C, B by D, and so on. When we come to Y, we replace it by A (going around the circle). Similarly, replace Z by B. Thus, the message

STAY RIGHT THERE FOR NOW

will be coded as

UVCA TKIJV VJGTG HQT PQY

or

UVCAT KIJVV JGTGH QTPQY

since it is customary in coding to give the message in groups of 5 letters. (The spaces between words may give clues as the possible meanings.) Such a code was used by Julius Caesar, and so it has a long history.

The person who receives the message UVCAT KIJVV JGTGH QTPQY and who knows the coding device will replace each letter by the letter two before it. The result is of course:

STAYR IGHTT HEREF ORNOW

The words can now easily be separated and the message is decoded.

The original message is called the *plaintext*; the coded version is called the *ciphertext*. The process of changing plaintext into ciphertext is called *coding* or *encryption*. The process of changing ciphertext back into plaintext is called *decoding* or *decryption*. Encryption and decryption can be done if you are privy to the method used. If you have intercepted a message in ciphertext, and you want to find its plaintext meaning, but you don't know the decryption method, you must use the art of *cryptanalysis* to find the plaintext. This will be considered in what follows.

The encryption above can be given by a simple mathematical formula. Coding A as C, B as D, etc. is described numerically as coding 1 as 3, 2 as 4, 3 as 5, etc. Just add 2. If the plaintext number is p and the corresponding ciphertext number is c, then this code is simply $c = p + 2$. However, this formula falls apart for X (24), Y (25) since $24 + 2 = 26$, and $25 + 2 = 27$. But the formulas will be valid if we work with the remainders when divided by

26–that is, congruence mod 26. Thus the correct encryption formula is

$$c \equiv p + 2 \bmod 26 \tag{3}$$

The congruence (3) can be solved for p:

$$p \equiv c - 2 \bmod 26 \tag{4}$$

One of the necessary components of an encryption method is that a ciphertext should be easily coded and decoded by anyone who knows the method, and very difficult by someone doesn't. Formulas (3) and (4) are simple schemes from this point of view. From our point of view, equation (4) is the decryption equation, since it converts ciphertext into plaintext, while equation (3) is the encryption equation. The conditions $0 \leq p \leq 25$ and $0 \leq c \leq 25$ are needed in order that p and c can be assigned a letter of the alphabet.

We can illustrate these equations as follows: If the ciphertext letter is F, then we have $c = 6$, since F is the 6th letter of the alphabet. Using equation (4), we find $p \equiv c - 2 \bmod 26$, or $p \equiv 4 \bmod 26$. So the plaintext must be D, the fourth letter of the alphabet. On the other hand, if the ciphertext was A, we would find $p \equiv c - 2 \bmod 26$, or $p \equiv 1 - 2 = -1 \bmod 26$. To convert $-1 \bmod 26$, we just add 26 (which is, after all, 0 mod 26) to find $-1 \equiv 25 \bmod 26$, and so the plaintext letter is Y. Of course we can do this more easily without a formula: it amount to shifting two places to the left, and "going around a circle." But in our later examples, this will be difficult to do without a formula.

An encryption based on a formula of the type

$$c \equiv p + a \bmod 26 \tag{5}$$

is called a *shift code* because we can think of shifting the alphabet over a spaces to do the coding. Because of this equation, it is also called an *additive* code or cipher. What makes it relatively easy to decipher is that we only have to know how one letter is enciphered to figure out the rest. Equivalently, we need to know the value of a. Here is an example.

Example. Sally has intercepted the message

YMJDF WJMJW J

and believes that it is a ciphertext based on a shift code. Furthermore, she believes that the letter E is enciphered as J, because there are so many J's in the ciphertext. Help her to decipher the message.

Method. The conversion formula is of the form $c \equiv p + a \bmod 26$. We know that when $p = 5$ (plaintext E), we have $c = 10$ (ciphertext J). Thus, $10 \equiv 5 + a \bmod 26$. So $a \equiv 5$

mod 26, and the encryption formula is $c \equiv p + 5 \bmod 26$. The ciphertext is thus obtained by shifting 5 to the right, and so deciphering shifts 5 to the left: $p \equiv c - 5 \bmod 26$. Now work letter by letter. Y's number is 25. Here $c = 25$ so $p \equiv 25 - 5 = 20$, so the plaintext corresponding to Y is T. The letter D, has $c = 4$, so here $p \equiv 4 - 5 = -1 \equiv 25$ (adding 26 to -1). Thus the plaintext for D is Y. Proceeding this way, we decode this messages as THEYA REHER E, or adjusting spaces: THEY ARE HERE.

For additive ciphers given by Equation (5), the entire code is known, once we know the single number a. This number is called the *key* for this coding. Obviously, when you know the key, coding and decoding is simple. Another usage, will come in handy, is to let the key be the ciphertext corresponding to the plaintext A. So for a shift of $a = 2$, we can speak of the key 2, or equivalently, the key C. Similarly, a shift code with key letter G will have $a = 6$.

While shift codes can be easily use for coding, they have a severe limitation. Since they are all of the form $c \equiv p + a \bmod 26$, there are only 25 encipherings of this type available corresponding to $a = 1$ to 25. (Of course, $a = 0$ amounts to no coding at all, or the trivial coding.) If you know that a shift code is being used, all you would have to do is to try the 25 different values of a and the decoding of a message can be done. This can be done fairly easily by hand. And with the right computer program, all 25 possible decodings of a message can be done almost instantaneously.

All shift codes are given on page 51. It is convenient to use this table for shift codes. The first line, the usual alphabet, is the plaintext line. Each line following represents a different shift a, for $a = 1$ to 25. For example, the line starting with D is the shift code with key letter D, corresponding to $a = 3$.

We can rework Example 1 easily using this table. Since E is enciphered as J. look under E until you find J. This will be in the row starting with F. So the code used had key F (A→F). Now take the message YMJDF WJMJW J and read these letters on row F. The plaintext will be right above each letter in the first row. This easily yields THEYA REHER E as before and with a lot less fuss.

Affine Codes. An *affine code*, described in what follows, is a coding scheme where the letters are more mixed up than in a shift code. For simplicity, we shall illustrate using a truncated alphabet of 10 letters A through J. Suppose the plaintext A is coded as the ciphertext D. But now we code B as G[1] (3 beyond D) as in the diagram:

Plain	A	B	C	D	E	F	G	H	I	J
Cipher	D	G								

Now continue by assigning C to J (three beyond G: H, I, J), D to C (three beyond J: A, B,

[1] In a shift code, we would code B as E, C as F, etc.

C)[2], E to F (3 beyond C: D, E, F), etc., completing the table:

Plain	A	B	C	D	E	F	G	H	I	J
Cipher	D	G	J	C	F	I	B	E	H	A

We can decode in this theoretical example by reversing the cipher and the plaintext rows. For simplicity, we can alphabetize the cipher row to obtain:

Cipher	A	B	C	D	E	F	G	H	I	J
Plain	J	G	D	A	H	E	B	I	F	C

Reading directly from the first table, we can code the message BID HIGH into GHC EHBE. Reading from the second table, see if you can decode the message GDCCH JF.[3]

How is this done algebraically with the full 26 letter alphabet? Suppose we use the same scheme for 26 letters, coding A as D, B as G (3 beyond D), C as J (3 beyond G) etc. In this case, the formula relating p (plaintext) and c (ciphertext) is

$$c \equiv 3p + 1 \bmod 26, 0 \leq c \leq 25 \tag{6}$$

Note how the factor 3 in this formula has the intended effect. The following table shows this:

p	1	2	3	...
$c = 3p + 1$	4	7	10	...

The factor 3 increases each value of c by 3 for every one increase in p. The constant 1 was chosen so that $p = 1$ (A) corresponded to $c = 4$ (D).

Let's use this formula to code the message GOOD LUCK. The work is done in the following table. The table has only one of the O's in this message.) Use a calculator to do this calculation. It can also be done on a spreadsheet.

plaintext	p	$c = 3p + 1$	c mod 26	ciphertext
G	7	22	22	V
O	15	46	20	T
D	4	13	13	M
L	12	37	11	K
U	21	64	12	L
C	3	10	10	J
K	11	34	8	H

[2]After J comes A. We go around in a circle

[3]You should get BAD DICE.

The enciphered message is thus VTTMK LJH.

How do we decipher a message on the full alphabet which uses this code? We can do as we did for the letters A through J above, by finding a complete deciphering table. However, we shall show how this can be done more directly, using algebraic methods. Assuming the affine code $c \equiv 3p + 1 \bmod 26$ was use, let's *decipher* the ciphertext:

FLCCP QMPCQ TR

Start with Equation (6) $c \equiv 3p+1 \bmod 26$ and solve for p. We can't divide by 3 (no fractions allowed here) so we use the nice trick of *multiplying* by 9.[4] This leads to $9c \equiv 27p + 9 \bmod 26$. Reducing mod 26, this gives $9c \equiv p + 9 \bmod 26$, or

$$p \equiv 9c - 9 \bmod 26. \tag{7}$$

We now decode the message using this formula. We work letter by letter as follows:

ciphertext	c	$9c - 9$	$9c - 9 \bmod 26$	plaintext
F	6	45	19	S
L	12	99	21	U
C	3	18	18	R
P	16	135	5	E
Q	17	144	14	N
M	13	108	4	D
T	20	171	15	O
R	18	153	23	W

Since the ciphertext was FLCCP QMPCQ TR, the plaintext message is therefore SURRE NDERN OW. Restoring correct spacing, the message is SURRENDER NOW. (In the lab, this will also be accomplished on a computer.)

In general, we define an *affine code* using the formula

$$c \equiv ap + b \bmod 26 \tag{8}$$

Here, b can be any integer between 10 and 25, inclusive. But some care must be taken choosing the coefficient a. For example, we would certainly not choose $a = 0$, since this would give $c \equiv b \bmod 26$, so all letters would be enciphered by the same letter. In that case, of course, decoding is impossible. But the value $a = 2$ would be equally out of the question. For example, if we had $c \equiv 2p + 5 \bmod 26$, then both $p = 0$ and $p = 13$ would both give the value $c = 5$. In coding terms, we would encode both Z and M as E. So decoding E would be

[4]The idea is to get the coefficient of p congruent to 1, mod 26.

ambiguous. In this case, each of the odd numbered letters, A, C, E, etc. as ciphertext, would have two corresponding plaintext letters, while the even numbered letters B, D, etc would have none. The general rule is that we must choose a so that it has no common factors with 26: GCD(a,26)=1. Thus there are the following 12 possibilities for a:

$$a = 1,\ 3,\ 5,\ 7,\ 9,\ 11,\ 15,\ 17,\ 19,\ 21,\ 23,\ 25$$

If formula (8) is used as a code, it would be natural to call the pair (a, b) as the key. Another possibility is to use the ciphertext of the plaintext AB. Thus, the formula $c \equiv 5p + 7$ would have key (5,7) or key LQ. Do you see why?

Example. The message

YDCUZYP RLX AFOZMFSFA

is received. It is known that the cipher used was an affine cipher with equation $c \equiv 5p + 7$ mod 26. What was the original message?

Method: We solve $c \equiv 5p + 7$ mod 26 for p. Multiply by 5.[5] Working mod 26, this gives $5c \equiv 25p + 35 \equiv -p + 9$. Solving, $p \equiv -5c + 9$ mod 26.

For example, to decode the first letter Y, we first find its cipher number $c = 25 \equiv -1$. Therefore $p \equiv -5c + 9 = (-5)(-1) + 9 = 14$. So the plaintext letter is N. Continuing in this fashion, the reader is welcome to discover the full text.

As we have seen, the algebraic problem of decoding an affine code involves *solving a linear congruence.* This is an equation of the form $ax \equiv b$ mod n. In all of our problems, we shall assume that a and n have no common factor. So, as above when $n = 26$, we arrange for a to be between a and 25, odd, and not 13. We *cannot* solve the congruence $ax \equiv b$ mod 26 by dividing by a, since we only use integers when working mod 26, and do not allow fractions. However, a systematic way of solving the equation $ax \equiv b$ mod 26 is to *multiply* this equation by some number b for which $ab \equiv 1$ mod 26. Such a number b is called the *inverse* of a mod 26. For example, the inverse of 3 mod 26 is 9, since $3 \cdot 9 = 27 \equiv 1$ mod 26. A quick check shows the following table of inverses mod 26 for the above 12 numbers.

Number	1	3	5	7	9	11	15	17	19	21	23	25
Inverse	1	9	21	15	3	19	7	23	11	5	17	25

There is a handy way of remembering this table. Listing all the numbers with their inverses, we have (1,1), (3,9), (5,21), (7,15), (11,19), (17,23), and (25,25). Recalling that $-c \equiv$

[5] In this case, the idea is to get the coefficient of p congruent to -1 mod 26.

$26 - c$ mod 26, these can be written as (1,1), (3,9), (5,−5), (7,15), $(-15, -7)$, $(-9, -3)$, and $(-1, -1)$. So a short list is simply:

$$(1,1),\ (3,9)\ (5, -5),\ (7, 15)$$

Here, we simply have to remember that we can multiply each term of the pair by -1 to get another pair. And remember the numbers in the pairs are inverses: the product of the two numbers in a pair is 1, mod 26. In what follows, we shall refer to this short list.

Example: An affine code is given by the equation $c \equiv 21p + 17$ mod 26. Solve for p to find the decoding formula.

Method: Write the equation as $c \equiv -5p - 9$ mod 26. (Subtraction 26 keeps the numbers the same mod 26.) Since $(5, -5)$ is in the short list, 5 and -5 are inverses, and we multiply by 5. This gives $5c \equiv p - 45$ mod 26, or $p \equiv 5c + 45 \equiv 5c + 19$.

Example: An affine code is given by the formula $c \equiv 11p + 2$ mod 26. A cipher letter is P. What is the corresponding plaintext letter?

Method: Since $11 \equiv -15$ mod 26, we'll use the short list couple (7,15), but in the form $(-7, -15)$. The original equation is $c \equiv -15p + 2$ mod 26. So multiply by -7 to get $-7c \equiv p - 14$ mod 26, or $p \equiv -7c + 14$ mod 26. In this example, the cipher letter was P, and so $c = 16$. Therefore $p \equiv -7c + 14 = -98$ mod 26. Since $-98 \equiv 6$ mod 26[6], we have $p = 6$, and so the plaintext letter is F.

Example: Solve for x: $13x \equiv 5$ (mod 100)

Method: We don't have a short list for inverses mod 100, so we proceed as follows. Multiply by 7, bringing the coefficient of x to 91, or -9. This gives $91x \equiv 35$ (mod 100), or $-9x \equiv 35$ (mod 100). Now multiply by 11 to get $-99x \equiv 385$ (mod 100), or $x \equiv 85$ (mod 100).

The appendix on page 49 gives a very effective algorithm for computing the inverse of any number modulo another, provided the GCD of these numbers is 1. A spreadsheet method will be given in he lab.

It is worth mentioning at this time that we have used congruences to nicely mix up the letters of the alphabet. This is yet another example of a mathematical curiosity put to a practical use.

General Substitution Codes.

The additive and, more generally, affine codes each give simple ways to rearrange the letters

[6]We added $104 = 4 \cdot 26$ to -98 to get 6.

of the alphabet. If an affine code is used, it would be of the form $c \equiv ap + b \bmod 26$. As we have seen, there are 12 possibilities for a and 26 possibilities for b, giving only $12 \cdot 26 = 312$ possible affine codes[7]. This seems like a lot, but using a computer, it would not be difficult to test every one of them.

A *substitution code* is one where each of the letters from A to Z has a different letter as ciphertext. Briefly, the letters are rearranged in some manner in order to obtain the code. The rearrangement can be quite arbitrary, in contrast to the orderly manner an affine code is generated. for example, the following is a random arrangement of the letters, leading to a simple substitution code:

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

The decoding scheme works backwards from ciphertext to plaintext, as follows:

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

If we had a message and coded it as

ZRVR EX Q XRDPRDSR. BRPX XRR GIO ARSIAR EP.

you could easily decode it, using this key. Try it.

Because there is no easy way to remember such a substitution code, it would seem to be quite hard to break.[8] However, the down side is that the recipient must have the entire scheme, and not be content with a simple key. Yet, codes are broken. We now briefly consider how this is done.

Cryptanalysis Cryptanalysis is the art of breaking codes. It is an art which is just about as old as the art of creating codes. How is a code broken? And how does one create a code which is difficult to break? During World War II, code messages to German submarines were transmitted electronically and were capable of being intercepted. Since these messages could give important information as to the whereabouts of these submarines, it was a matter of life and death for the Allies to decode these, and it was equally important for the Germans to have an unbreakable code. One of the important German codes (the "Enigma" code) was in fact broken with the help of a large team of mathematician and cryptologist.

Newspapers often have cryptanalysis features which challenge the reader to decode a message.

[7] Actually 259, because the case $a = 1$ and $b = 0$ is not technically a code, since it doesn't change a single letter.

[8] That is, to find the coding scheme if you had a coded message.

These all involve substitution codes, which might be randomly generated, and so cannot be as easily broken as an affine code. How can these be cracked?

Based on observations and careful analysis of texts, it has been observed that for an average English text, the letter E is repeated by far the most times, and that after that, in order, the letters T, A, O, I, N, S, H, R are repeated with high frequencies. The following interesting table[9] gives the approximate relative frequencies of the 26 letters. Of course, these are statistical averages, and we can never expect exactly these results in any specific text.

Group	Letter	Frequency of Group(%)	Range of individual frequencies
1	E	12.7%	12+%
2	T,A,O,I,N,S,H,R	56.9%	6%-9%
3	D,L	8.3%	4%
4	C,U,M,W,F,G,Y,P,B	19.9%	1.5%–3%
5	V,K,J,X,Q,Z	2.2%	Less than 1%

Thus, for a large enough ciphertext using a substitution code, one way of breaking the code would be to make a *frequency count* of its letters, expecting the most frequent, (or next most frequent) letter to be E, and the next frequent letters to be in the second group. For example, in the first paragraph of this section (which was written innocently without regard to which letters were being used) exactly 468 letters were used. The eight letters used most, with the percentage of times they were used were as follows:

Letter	Count	Percentage
E	65	13.9
S	42	9.0
O	39	8.3
T	38	8.1
N	37	7.9
I	30	6.4
A	28	6.0
R	26	5.6

Note that E (group 1) led the list, and the remaining 7 were all in group 2. This observation gives us a method to decipher a ciphertext: Simply count frequencies and see what happens if the letter with the largest frequency decodes as E, or perhaps T.[10]

Let us illustrate how a frequency count can help decode a message. For simplicity, we assume that the code is a shift code. The message is

[9]Taken from *Cryptology*, by Albrecht Beutelspacher, The Mathematical Association of America, 1994.

[10]Poe's short story, *The Gold Bug* uses this method as a crucial part of the story. In this story, E is said to have the largest frequency, but T comes much lower than second in his list.

BPQAK WLMQA DMZGM IAGBW JZMIS

A frequency count gives the following table:

Letter	Count	Percentage
M	4	16
A	3	12
B,G,I,Q,W,Z	2	8

We suspect that E is coded as M, the most frequent letter. There's no guarantee, of course, because the size of the message is relatively small. However, if this so, and we have a shift code of the form $c = p + a$, we should have plaintext E ($p = 5$) correspond to ciphertext M ($c = 13$. Thus, $13 = 5 + a$, and the key is $a = 8$, or key letter I. To decode, we use $p = c - a = c - 8$. The decoded message is

THISC ODEIS VERYE ASYTO BREAK

or simply THIS CODE IS VERY EASY TO BREAK. Frequency analysis saved us from trying the 25 different possible shift codes. If $a = 8$ didn't work we would next have tried $a = -4$ (in order to have E be coded as A) in the hope that this worked.

It is possible to use more than letter frequency to decode substitution codes. For a long code, we can count frequency of *pairs* or *triples* of letters. Thus, the pair TH is common in English (THE, THERE, THEY), as well as HE (as in THE, THERE, SHE), and high frequency of pairs can be further used to break the code. We shall not do this in these notes.

Frequency counts of texts, as well as coding and decoding affine or shift codes can be easily done on the computer, and programs to do this will be available in the lab.

Because of frequency considerations of this sort, few people would want to use a substitution code for a relatively long and important message. It would be fairly easy to decipher. Another problem with substitution codes is that if they are created randomly, the receiver of the ciphertext must know the entire code, and this can be cumbersome. Compare this with an affine code, in which the receiver must know only two numbers (the values a and b).

Vigenère Ciphers.[11] When using a shift code, the person who is receiving the message must know the shift in order to decode. This could be any single letter as indicated above. For example, the letter J would correspond to the shift code in which plaintext A corresponds to ciphertext J, and which corresponds to $a = 9$. This in turn corresponds to the row J on

[11]Named after a French diplomat.

page 51. A Vigenère cipher uses shift codes, but varies the shift according to a *key word* which is known by the two parties to the message. For example, suppose a key word is SCHEME. This means that the shift code corresponding to S, C, H, E, etc. are used, one after the other, and repeated until the message is sent. For example, let's encipher the message:

CODING IS LOTS OF FUN

or, less prosaically,

CODIN GISLO TSOFF UN

We write the keywords over the message, and the ciphertext below:

```
SCHEM ESCHE MESCH EM (Key Word)
CODIN GISLO TSOFF UN (plaintext)
UQKMZ KAUSS FWGHM YZ (ciphertext)
```

This is easily done using the table on page 51. For example, the first letter C of the plaintext message is coded using line S on the table, and this is seen to be U. Continuing, the encrypted message is UQKMZ KAUSQ FWGHM WZ. To decode, it is necessary to know the key word SCHEME. Once this is known, we can decode by working backwards:

```
SCHEM ESCHE MESCH EM (Key Word)
UQKMZ KAUSS FWGHM YZ (ciphertext)
CODIN GISLO TSOFF UN (plaintext)
```

For example, the first letter U, as the ciphertext is found on line S (of the key word), and its plaintext letter C is found above it on the first (plaintext) line. This process continues until the full message is decoded. (In the lab, we will be able to do this directly on the computer.)

A Vigenère cipher is not a substitution scheme. For example, in the above message, the letter O (of "coding") was coded as Q, but the letter O (of "of") was encrypted as G.

The reader should practice, by decoding the message

YAMO OAG ZAMXAUMXZ EEXT YOKL OOFWXIVHFEW ATAG PF AVAGAESK IL

It is a Vimenère code and the keyword is MATH.

Block Substitutions.

One way of overcoming the frequency count is to group the letter of a plaintext into blocks of two or more, and then use some substitution method on the blocks. For example, suppose we

work with the 26 letters again, and take the plaintext "BEWARE THE IDES OF MARCH":

BEWARETHEIDESOFMARCH

Now break into blocks of two to get

BE WA RE TH EI DE SO FM AR CH

Take the original numerical codes A→1, B→ 2, and write this plaintext numerically using two digits–B→ 02, E→ 05, W→23, etc.:

0205 2301 1805 2008 0509 0405 1915 0613 0118 0308

We now have 4 digit numbers representing these pairs, and it is much more difficult to analyze the frequency of these pairs even for a long message. But how do we code 4 digit numbers all numbered 2525 or under? We take some number over 2525, say 2600, and then use an affine substitution modulo this number. For example, we might try

$$c \equiv 17p + 1531 \bmod 2600 \tag{9}$$

As in the simple affine substitution discussed before it is essential that the coefficient of p and the modulus 2600 have GCD 1. For 2600, this means we must avoid any a which is divisible by 2, 5, or 13. We now encode this message. (We used a spreadsheet to do the computation.)

p	$17p + 1531$	$c \equiv 17p + 1531 \bmod 2600$
205	5016	2416
2301	40648	1648
1805	32216	1016
2008	35687	1867
509	10184	2384
405	8416	616
1915	34086	286
613	11952	1552
118	3537	937
308	6767	1567

Putting in initial 0's to keep four digits, the ciphertext is

2416 1648 1016 1867 2384 0616 0286 1552 0937 1567

All this is fine for the person doing the coding. But how about decoding? Presumably, a stranger picking up this message will get nowhere with it.[12] But what about our friend who knows the basic code $c \equiv 17p + 1531 \bmod 2600$. In order to decode, it will be necessary to find the *inverse* of 17 mod 2600 to solve for p. It turns out that 153 is the inverse, since $17 \times 153 = 2601 \equiv 1 \bmod 2600$.[13] Therefore, we can solve equation (9) by multiplying by 153. This gives

$$\begin{gathered} c \equiv 17p + 1531 \bmod 2600 \\ 153c \equiv p + 153 \cdot 1531 \equiv p + 234243 \equiv p + 243 \bmod 2600 \\ p \equiv 153c - 243 \bmod 2600 \end{gathered}$$

For example the first block in our ciphertext is $c = 2416$. Computing, using $p \equiv 153c - 243 \bmod 2600$, we get

$$p \equiv 153(2416) - 243 = 369405 \equiv 205 \bmod 2600$$

Of course, 205 (0205) converts back into the letter pair BE, and this is the beginning of the deciphering process.

In practice, this coding and decoding will be done on the computer, and we will set this up on the lab computer. We will only consider grouping into blocks of two, but we will also vary the coding formula, and even the modulus 2600 which was chosen arbitrarily. If the number of letters in the plaintext is odd, the last block will have only one letter. In this case, we simply add the letter X or Z to the end to make it a block of two.

We have only scratched the surface of this important topic of cryptography. Methods of coding are being developed up to the present time (as well as methods to break the code.) This section showed how number theory, especially the theory of congruences, can be of use in this process. Once again, a mathematical subject, once deemed to be interesting but utterly impractical, has become an important tool in a very practical situation.

[12] We could confound an interceptor by grouping this code in groups of 5. Thus 24161 64810 ...

[13] See page 49 for a simple way to compute the inverse of a mod b.

Appendix: How to find the inverse of a mod b.

We can find the inverse of a modulo b by using a small extension of the Euclidean algorithm which was used to computer GCD of a and b. Unlike the Euclidean algorithm, in this procedure the quotients are also used. We illustrate first with a simple example.

Problem: Find the inverse of 17 mod 37.
We are looking for a solution of the congruence $17x \equiv 1 \bmod 37$. We introduce another column in the Euclidean algorithm. The computation is as follows:

Larger Number	Smaller Number	Quotient	Remainder	0 1
37	17	**2**	3	$-2 = \mathbf{-2 \cdot 1 + 0}$
17	3	**5**	2	$11 = \mathbf{-5 \cdot -2 + 1}$
3	2	**1**	1	$-13 = \mathbf{-1 \cdot 11 + -2}$
2	1	2	0	

The mysterious last column is found as follows. Start by putting 0 and 1 on top of this column. Then for each entry in this last column, after each computation of quotient and remainder, compute

Minus the quotient, times the number on top of this entry, plus the number two up from it.

The calculations are shown in the table. The number opposite remainder 1 is the required inverse of 17 mod 37. In this case, the answer is -13, and we can easily verify that $-13 \cdot 17 \equiv 1 \bmod 37$.

We illustrate with another example, in which the computations are left to you. Check the above rule to see how this works!

Problem: Solve the congruence $173x \equiv 1 \bmod 2600$.
Method: The computation is as follows:

Larger Number	Smaller Number	Quotient	Remainder	0 1
2600	173	**15**	5	-15
173	5	**34**	3	511
5	3	**1**	2	-526
3	2	**1**	1	1037

The solution is $x \equiv 1037 \bmod 2600$. For example, the number 511 in the last column, was obtained by multiplying -34 (the negative of the quotient) by -15 (the number above it) and adding 1 (the number above that.)

Problem: A message is coded using blocks of 2, and the affine substitution $c \equiv 173p + 329 \bmod 2600$. What is the decoding formula? How do you decode the message 1202 554?

Method: To solve this congruence, simply multiply by the inverse of 173 mod 2600. We found this to be 1037 in the above problem. Multiplying, we get $1037c \equiv 1037 \cdot 173p + 1037 \cdot 329 \bmod 2600$. Since $1037 \cdot 329 = 341173 \equiv 573$, this congruence becomes $1037c \equiv p + 573 \bmod 2600$. Solving for p, we find $p \equiv 1037c - 573 \bmod 2600$ or $p \equiv 1037c + 2027 \bmod 2600$.

To decode the message, we use the formula $p \equiv 1037c + 2027 \bmod 2600$ for $c = 1202$ and for $c = 554$. The results are given in the following table:

c	$1037c + 2027$	mod 2600	Block equivalent
1202	1248501	501	EA
554	576525	1925	SY

The message is simply decoded as EASY.

These computations will also be done in the lab using the computer.

Table of Shift Substitutions

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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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

To use the above table, treat the first row, starting with A as the plaintext row. Each of the other 25 rows is the ciphertext corresponding to the letter above it on the first row. For example, if (plaintext) C is coded as (ciphertext) R in a shift code, we look under the C in the first row until we find R. This is in row P, and this determines the shift code. We then use row A plaintext and row P as ciphertext, and this will allow quick coding and decoding.

Exercises on cryptography

1. Using the shift code $c \equiv p + 4$, encipher the message: LEAVE NOW.

2. Using the shift code $c \equiv p + 18$, encipher the message: CODING IS FUN.

3. Using the shift code $c \equiv p - 7$, encipher the message: JUST DO IT.

4. Using the affine code $c \equiv 3p + 7$, encipher the message: DYLAN IS THE GREATEST.

5. Using the affine code $c \equiv 5p - 1$, encipher the message: ARE WE HAVING FUN YET.

6. Using the affine code $c \equiv -p + 7$, encipher the message: MATH CLASSES ARE THE BEST.

7. Make a table listing plaintext letters A...Z, and corresponding ciphertext letters for the affine code $c \equiv -p + 7$. Describe the table in words.

8. The coded message AOPZP ZUAOHY KHAHS S was received. It is known that the code was a shift code with equation $c = p + 7$. Decode the message. Be sure to put in appropriate spaces.

9. The coded message MXOQCY IFUDT YDWIE CKSXJ YCUED JXYI was received. It is known that the code was a shift code with equation $c = p - 10$. Decode the message. Be sure to put in appropriate spaces.

10. For the affine code $c = 3p + 11$, what is the ciphertext letter associated with the plaintext letter O? With the plaintext letter F?

11. For the affine code $c = 3p + 11$, what is the plaintext letter associated with the ciphertext letter O? With the ciphertext letter F?

12. For the affine code $c = 5p - 7$, what is the ciphertext letter associated with the plaintext letter O? With the plaintext letter F?

13. For the affine code $c = 5p - 7$, what is the plaintext letter associated with the ciphertext letter O? With the ciphertext letter F?

14. The coded message IYBFB FVPII BQVIP MBTLF was received. It is known that the code was an affine code with equation $c = 3p + 1$. Decode the message. Be sure to put in appropriate spaces.

15. The coded message TNFZS SDFNS ZBSZW SOXJI YDVN was intercepted. It is known that the code was an affine code with equation $c = 9p - 5$. Decode the message. Be sure to put in appropriate spaces.

16. Code the message YOUR QUEEN IS IN DANGER, using a Vigenère cipher with key word ENOUGH.

17. Code the message THE BEATLES LIVE, using a Vigenère cipher with key word YEAH.

18. The message QRSGA LVVGE XKMUA IRAA was sent to Erich, and electronically intercepted by a student. It was guessed that this is a Vigenère cipher, probably with key word ERICH. Test this and see if you can decode this message.

19. The message BAEPI HYDTS NFSEH EHCIX EUNPG SWTTN NTHTM was sent to Paul, and electronically intercepted by a student. It was guessed that this is a Vigenère cipher, probably with key word PAUL. Test this and see if you can decode this message.

20. Using the algorithm on page 49 solve the congruence $151x \equiv 1 \bmod 2301$.

21. Find the inverse of 323 mod 2600.

22. Using the affine block substitution $c \equiv 49p + 301 \bmod 2600$, code the message DO IT NOW.

23. Using the affine block substitution $p \equiv 179c + 1515 \bmod 2600$, code the message NO WAY.

24. The message 0830 1430 2345 is received, and it is known to be an affine block substitution with formula $c \equiv 17p + 345 \bmod 2600$. What is the plaintext message?

25. The message 0761 1310 2012 is received, and it is known to be an affine block substitution with formula $c \equiv 349p + 1712 \bmod 2600$. What is the plaintext message?

26. Solve the congruence for p: $c \equiv 3p - 1 \bmod 26$.

27. Solve the congruence for p: $c \equiv 7p + 3 \bmod 26$.

28. Solve the congruence for p: $c \equiv 3p - 1 \pmod{50}$.

29. Solve the congruence for p: $c \equiv 11p + 2 \pmod{52}$.

The following are optional problems, not covered in the notes.

30. Find a number x such that $x \equiv 7 \bmod 11$ and $x \equiv 3 \bmod 7$.

31. The integers x and y satisfy the congruences:
$$3x + 4y \equiv 9 \pmod{100}; \; 5x + 9y \equiv 11 \pmod{100}.$$
Find possible integer solutions x and y.