Math 135 (Summer 2006)
Vigenère Keyword Cipher

The Vigenère square cryptosystem is an example of a **polyalphabetic substitution**. That is, different letters in the plaintext are encrypted with different substitution alphabets.

Recall that the numerical equivalents of the letters are as follows:

A	B	C	D	E	F	G	H	I	J	K	L	M
0	1	2	3	4	5	6	7	8	9	10	11	12
N	O	P	Q	R	S	T	U	V	W	X	Y	Z
13	14	15	16	17	18	19	20	21	22	23	24	25

Secret key: correspondents agree on a keyword.

To encrypt: Write the keyword repeatedly alongside the plaintext, convert both the plaintext and the keyword letters to their numerical equivalents (0 for A, 25 for Z) and add them modulo 26.

Example: If the keyword is `WIND` and the plaintext is `GO AHEAD MAKE MY DAY`, then the ciphertext is

plain	G	O	A	H	E	A	D	M	A	K	E	M	Y	D	A	Y
x	6	14	0	7	4	0	3	12	0	10	4	12	24	3	0	24
key	W	I	N	D	W	I	N	D	W	I	N	D	W	I	N	D
k	22	8	13	3	22	8	13	3	22	8	13	3	22	8	13	3
$(x+k)$ MOD 26	2	22	13	10	0	8	16	15	22	18	7	15	20	11	13	1
cipher	C	W	N	K	A	I	Q	P	W	S	R	P	U	L	N	B

To decrypt: Write the keyword repeatedly alongside the ciphertext, convert both the ciphertext and the keyword letters to their numerical equivalents, and subtract them modulo 26.

Example: If the keyword is `NUMBER` and the plaintext is `GBATI NUIOB RTBOZ UEEQN TPWVJ BADEE G`, then the encipherment is

cipher	G	B	A	T	I	N	U	I	O	B	R	T	B	O	Z	U	E	E	Q
y	6	1	0	19	8	13	20	8	14	1	17	19	1	14	25	20	4	4	16
key	N	U	M	B	E	R	N	U	M	B	E	R	N	U	M	B	E	R	N
k	13	20	12	1	4	17	13	20	12	1	4	17	13	20	12	1	4	17	13
$(y-k)$ MOD 26	19	7	14	18	4	22	7	14	2	0	13	2	14	20	13	19	0	13	3
plain	T	H	O	S	E	W	H	O	C	A	N	C	O	U	N	T	A	N	D

cipher	N	T	P	W	V	J	B	A	D	E	E	G
c	13	19	15	22	21	9	1	0	3	4	4	6
key	U	M	B	E	R	N	U	M	B	E	R	N
k	20	12	1	4	17	13	20	12	1	4	17	13
$(c-k)$ MOD 26	17	7	14	18	4	20	7	14	2	0	13	19
plain	T	H	O	S	E	W	H	O	C	A	N	T

And the joke is: There are three kinds of mathematicians, ...

Observation: Using modular arithmetic, it is easy to encrypt and decrypt messages using the Vigenère square. Notice that if the keyword is of length k, then every kth letter is enciphered with the same shift substitution.

Property: Vigenère encipherments with longer keywords tend to even out the distribution of letters in the ciphertexts. Thus the statistics in the underlying plaintext are obscured.

Cryptanalysis of a Vigenère Enciphered Text

Example: Suppose that the ciphertext is

```
CTMYR DOIBS RESRR RIJYR EBYLD IYMLC CYQXS RRMLQ FSDXF OWFKT CYJRR IQZSM X
```

and it is known that the keyword is a three letter English word (i.e., $k = 3$).

The basic idea is that three monoalphabetic shifts are used to get

```
C _ _ Y _ _ O _ _ S _ _ ...
_T _ _ R _ _ I _ _ R _ ...
_ _ M _ _ D _ _ B _ _ E ...
```

That is, if we write the ciphertext in three columns, we see that every letter in the first column is the result of a shift by an amount corresponding to the first letter in the keyword, every letter in the second column is the result of a shift by an amount corresponding to the second letter in the keyword, and similarly, every letter in the third column is the result of a shift by an amount corresponding to the third letter in the keyword.

```
CTM
YRD
OIB
SRE
SRR
RIJ
YRE
BYL
DIY
MLC
CYQ
XSR
RML
QFS
DXF
OWF
KTC
YJR
RIQ
ZSM
X
```

Determine the likely shift values and try out the corresponding possible keywords.

let#1	freq#1	let#2	freq#2	let#3	freq#3
C	2	T	2	M	2
Y	3	R	4	D	1
O	2	I	4	B	1
S	2	Y	2	E	2
R	3	L	1	R	3
B	1	S	2	J	1
D	2	M	1	L	2
M	1	F	1	Y	1
X	2	X	1	C	2
Q	1	Y	2	Q	2
K	1	J	1	S	1
Z	1			F	2

If E $\mapsto$ Y, then shift is $24 - 4 = 20$ so key letter is U.
If T $\mapsto$ Y, then shift is $24 - 19 = 5$ so key letter is F.
If N $\mapsto$ Y, then shift is $24 - 13 = 11$ so key letter is L.
etc.

Continuing in this way, we find the most likely first, second, and third letters.

likely#1	likely#2	likely#3
U	N	N
F	Y	Y
L	E	E
K	D	D
H	A	A
Q	J	J
Y	R	R
G	Z	Z

Therefore, some likely keywords are: FAD, FAN, FAR, FED, FEN, LEE, LEA, KEN, KEY, HER, GAY, ...

Knowledge that the key"word" is a real three-letter English word vastly reduces the amount of work from trying out all three-letter strings, or all two–, four–, five–, ... letter strings.

cipher	CTMYRDOIB...
key	FEDFEDFED...
plain	XJTN...

cipher	CTMYRDOIBSRE...
key	KEYKEYKEYKEY...
plain	SPOONFEEDING...

And so we find: SPOONFEEDING IN THE LONG RUN TEACHES US NOTHING BUT THE SHAPE OF THE SPOON.