Breaking a Vigenère Cipher

Introduction

We are presented with the following substitution cipher:

```
ANYVG YSTYN RPLWH RDTKX RNYPV QTGHP HZKFE YUMUS AYWVK ZYEZM EZUDL
JKTUL JLKQB JUQVU ECKBN RCTHP KESXM AZOEN SXGOL PGNLE EBMMT GCSSV
MRSEZ MXHLP KJEJH TUPZU EDWKN NNRWA GEEXS LKZUD LJKFI XHTKP IAZMX
FACWC TQIDU WBRRL TTKVN AJWVB REAWT NSEZM OECSS VMRSL JMLEE BMMTG
AYVIY GHPEM YFARW AOAEL UPIUA YYMGE EMJQK SFCGU GYBPJ BPZYP JASNN
FSTUS STYVG YS
```

Monoalphabetic or Polyalphabetic?

Our first goal is to determine if it is monoalphabetic or polyalphabetic. We first do a frequency count:

a	14	g	12	l	13	q	5	v	10
b	8	h	8	m	16	r	11	w	9
c	7	i	5	n	13	s	18	x	7
d	5	j	11	o	4	t	15	y	16
e	22	k	14	p	13	u	14	z	11
f	6								

Notice all the letters appear several times, and the frequency does not vary much. This suggests a polyalphabetic substitution cipher. The IC is 0.041 for this cipher, so this suggests one also, and indicates the period is rather long (5 characters or more.)

First we seek to establish the period. We do a Kasiski examination, and write down all repetitions and how far apart they occur:

repetitions	first	next	interval	factors
YVGYS	3	283	280	2, 2, 2, 5, 7
STY	7	281	274	2, 137
GHP	28	226	198	2, 3, 3, 11
ZUDLJK	52	148	96	2, 2, 2, 2, 2, 3
LEEBMMTG	99	213	114	2, 3, 19
SEZM	113	197	84	2, 2, 3, 7
ZMX	115	163	48	2, 2, 2, 2, 3
GEE	141	249	108	2, 2, 3, 3, 3

The common factor to these is 2. But 2 occurs whenever the period is even, and is probably too short, so let us look at other factors. Possibilities are 3 (7 out of 8 intervals), 6 (6 out of 8), 4 (5 out of 8), 12 (4 out of 8), 5 (1 out of 8), 7, 8, 9, 14, 16, and 28 (2 out of 8), and all others in 1 out of 8. 3 is probably too short, and 4 and 12 make the repetition of `LEEBMMTG` accidental, which is very unlikely. So the period is probably 6.

Working from this, we do a frequency count for each of the 6 alphabets The following table summarizes the counts:

	alphabets					
letter	#1	#2	#3	#4	#5	#6
a	2	4		1	4	2
b		2	1	1	2	2
c		1			3	3
d					1	4
e	1		2	6	9	4
f				4	1	1
g	2		5	4	1	
h	3		1	1	2	
i		1	1		3	
j	6	3		1		1
k	3	1	10			
l	3	2	2		2	4
m	3	10	2			1
n		2	2	5	3	1
o			2	2		
p		3	1	1		8
q		3		2		
r			1	5	3	2
s	6		1	2	6	2
t	1	4	5	1		4
u	5	1	2	3	3	
v	2	3	2	2		
w	5	4				
x		2	2			3
y	2	1	4	2	3	5
z	4	1	2	4		

To check ourselves, we compute the IC for each of the 6 alphabets:

```
#1    0.064 #3    0.071 #5    0.069
#2    0.070 #4    0.054 #6    0.067
```

These indicate each of the 6 alphabets is monoalphabetic, so we are probably on the right track.

Now notice the counts for each alphabet. They look like those expected of English, only shifted. For example, in alphabet 1, notice the long gap between N and R, which is surrounded by many letters in the ranges J to M and S to W. The normal alphabet profile has a similar feature, the gap being from V to Z, and the surrounding letters being R to U and A to E. This indicates that the cipher is a shifted one, and that S may be A. A similar gap (from D to H) occurs in the frequency chart of the second alphabet, so following the same reasoning, I is probably A. Substituting the resulting characters, we obtain:

```
ifYVG YalYN RPtoH RDTsp RNYPd iTGHP prKFE YceUS AYenK ZYEhe EZUDt
bKTUL rdKQB JciVU ECstN RCTph KESXu sZOEN apGOL PofLE EBueT GCSan
MRSEh eXHLP sbEJH TchZU EDecN NNRes GEEXa dKZUD tbKFI XplKP IAheX
FACeu TQIDc oBRRL blKVN AroVB REioT NSEhe OECSa nMRSL reLEE BueTG
AYdaY GHPme YFARe sOAEL chIUA YgeGE EMriK SFCom GYBPr tPZYP rsSNN
FalUS STgnG YS
```

From this point on, we can simply guess. The HE in the second group of the second line suggests the E in alphabet #6 is really a T; trying that out, and assuming again a shifted alphabet, we get:

```
ifYVG nalYN RetoH RDisp RNYed iTGHe prKFE nceUS AnenK ZYthe EZUst
bKTUa rdKQB yciVU ErstN RCiph KESmu sZOEc apGOL eofLE EqueT GChan
MRSth eXHLe sbEJH ichZU EsecN NNges GEEma dKZUs tbKFI mplKP IpheX
FAreu TQIsc oBRRa blKVN proVB RtioT NSthe OECha nMRSa reLEE queTG
AndaY GHeme YFAge sOAEa chIUA ngeGE EbriK SFrom GYBer tPZYe rsSNN
ualUS SignG YS
```

In line 6, group 5, AND suggests AND; also, note that in group 8 of line 1, the three letters NCE suggext that the preceding one is A or E. Given these, most likely the fifth alphabet is unshifted, so:

```
ifYVg nalYN retoH Rdisp RNyed iTGhe prKFe nceUS anenK Zythe EZust
bKTua rdKQb yciVU erstN Rciph KEsmu sZOec apGOl eofLE equeT Gchan
MRsth eXHle sbEJh ichZU esecN Nnges GEema dKZus tbKFi mplKP ipheX
Fareu TQisc oBRra blKVn proVB rtioT Nsthe OEcha nMRsa reLEe queTG
andaY Gheme YFage sOAea chIUa ngeGE ebriK Sfrom GYber tPZye rsSNn
ualUS signG Ys
```

In line 1, group 6, we see `HE` again. Guess that the preceding letter, `G`, represents `T`; if so, and if the alphabet is shifted, the N should be `A`. We confirm this by looking in groups 2 and 3 on line 1.; group 3 begins with `RE`, which suggests `ARE`, and indeed group 2 ends in `N`. Substituting,

```
ifYig nalYa retoH edisp Rayed iTthe prKse nceUf anenK mythe Emust
bKgua rdKdb yciVh erstN eciph Krsmu sZbec apGbl eofLr equeT tchan
Mesth eXule sbEwh ichZh esecN anges Grema dKmus tbKsi mplKc ipheX
sareu Tdisc oBera blKin proVo rtioT asthe Orcha nMesa reLre queTt
andaY theme Ysage sOnea chIha ngeGr ebriK ffrom Glber tPmye rsSan
ualUf signG ls
```

At this point the message can be read off:

```
ifsig nalsa retob edisp layed inthe prese nceof anene mythe ymust
begua rdedb yciph ersth eciph ersmu stbec apabl eoffr equen tchan
nesth erule sbywh ichth esech anges arema demus tbesi mplec ipher
sareu ndisc overa blein propo rtion asthe ircha ngesa refre quent
andas theme ssage sinea chcha ngear ebrie ffrom alber tjmye
rsman ualof signa ls
```

It is a Vigenère cipher; the keyword is `SIGNAL`.

```
If signals are to be displayed in the presence of an enemy, they
must be guarded by ciphers.  The ciphers must be capable of frequent
changes.  The rules by which these changes are made must be simple.
The ciphers are undiscoverable in proportion as their changes are
frequent and as the messages in each change are brief. From Albert
J. Meyers' Manual of Signals.
```