Keyword Ciphers

Look at the following key for a minute.

```
abcdefghijklmnopqrstuvwxyz
KPFHIGLDEXCVTOUBJQZMRNAYSW
```

Now, cover up the key and write it from memory.

Unless you have a remarkable memory, it is unlikely that you would be able to remember and recreate the plaintext-ciphertext correspondence for this simple substitution key; it is likely you and the receiver would have to write down the key. Having to write down the key jeopardizes key security. Key security is enhanced if the key need not be written. (Of course, torturing the person who has memorized the key might be an effective method for obtaining the key.)

Affine ciphers (including Caesar ciphers and multiplicative ciphers) have memorable keys, but the number of keys is small and, although they might not be easy to spot, patterns are introduced into single letter frequencies. Another scheme that uses a memorable key for a simple substitution cipher is called the keyword cipher. Its key is an easily memorized word or phrase. Here is how one common version of the keyword cipher works.

Cryptography

The key has two parts – a word or phrase and a letter of the alphabet.

1. Select a keyword or phrase.

Northern Kentucky University

and a keyletter

j

2. Reading from left to right, write the word or phrase without duplicating letters.

NORTHEKUCYIVS

3. Underneath the plaintext alphabet, beginning with the keyletter, write, letter for letter, the keyphrase with the duplicate letters removed.. After the last keyphrase letter, write the so far unused letters of the alphabet in their usual order – cycling back to the beginning.

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

Here is a message encrypted using a simple substitution cipher and the key given above.

```
ivhhd agyix paecm vypmg ijrpi lxrpd kgcoi empyl
czjpi xpicg qqzle hlpci gzhqz yxrzh oyaxz lxcgm
zgipm qceti xpwpc tghxz wxlet tghmz hjpcr zhevi
ieixp sgcze vygct dlett ghmy
```

Frequency analysis shows the following single-letter frequencies.

A	111
B	
C	11111111111
D	111
E	111111111
F	
G	111111111111
H	1111111111
I	1111111111111
J	111
K	1
L	1111111
M	1111111
N	
O	11
P	1111111111111
Q	1111
R	1111
S	1
T	1111111
U	
V	1111
W	11
X	1111111111
Y	1111111
Z	11111111111

The frequencies show the peaks and valleys that should be expected with a simple substitution cipher, but the distribution of those peaks and valleys do not correspond to plaintext or to a Caesar shift.

Here are the disjoint cycles for this key:

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

If our keyword were computer science and keyletter g, our key would be:

```
abcdefghijklmnopqrstuvwxyz
QVWXYZCOMPUTERSINABDFGHJKL
```

Notice that once the keyword ends, there are strings of letters that are nearly in alphabetical order – especially near the end of the alphabet.

Here is even a worse case.

If our keyword were bad and keyletter were a, we would have the key:

```
abcdefghijklmnopqrstuvwxyz
BADCEFGHIJKLMNOPQRSTUVWXYZ
```

Many letters are left fixed by this key.

Shifting the keyletter helps

```
abcdefghijklmnopqrstuvwxyz
WXYZBADCEFGHIJKLMNOPQRSTUV
```

but there is still an extremely long string of ciphertext letters in alphabetical order. This key is very similar to a Caesar cipher with additive key 4. we would prefer to have a keyword that resulted in a more random arrangement of ciphertext letters.

Using the Pattern in the Key for Cryptanalysis

So, we can create a memorable key for a simple substitution cipher.

But, there is a trade off. As we did for affine ciphers, we have created a memorable key at the expense of having created a pattern – we have placed a pattern in the key. We noted in the keys above that keys may have long strings of ciphertext letters in alphabetical order. This pattern can be exploited by the cryptanalyst.

Before doing a complete cryptanalysis of a ciphertext message, we will consider how using the pattern in the key can help the cryptanalyst.

Assume that we have a ciphertext message that we suspect was encrypted with a keyword cipher and that we been able to partially solve it. Assume that we have recovered this much of the key.

```
abcdefghijklmnopqrstuvwxyz
  Z EN   Y B FG   JLM   R V
```

Recall how the key is created for a keyword cipher. Z is unlikely to be in the keyword; so, we might suspect that the keyword begins immediately after Z. FG might indicate that the keyword has ended and we are seeing the "unused letters of the alphabet in their usual order." Looking back a few letters, it seems reasonable to assume that Y ends the keyword. If we are correct in assuming that Y ends the keyword, then plaintext k must correspond to ciphertext A. Put that in place.

```
abcdefghijklmnopqrstuvwxyz
  Z EN   YAB FG   JLM   R V
```

E appears in the keyword; so, the letter between B and F is either C or D, which means that either C or D is in the keyword. Plaintext p must correspond to ciphertext H and q to I. u corresponds to O, v corresponds to P, and w corresponds to Q. Two of STU must be in the keyword. a must correspond to W, and b to X.

Here is what we suspect.

```
abcdefghijklmnopqrstuvwxyz
WXZ EN   YAB FGHIJLMOPQR V
```

and K is in the keyword, and either C or D is in the keyword, and two of STU are in the keyword.

```
_ E N _ _ _ Y
```

Doesn't _ E N suggest KEN? If so, the keyword might be KENTUC(K)Y.

The pattern in the keyword helped us complete the key.

Cryptanalysis Using Known Plaintext

Here is a ciphertext message:

```
PYWPX  FBWAK  PVGUF  DPXFB  PHQDE  WXXWK  DPYDL
HQZLY  QFLPY  JQDQU  LABWK  JBWWC
```

Cryptanalysis of any ciphertext message begins with frequency analysis.

A	11
B	1111
C	1
D	11111
E	1
F	1111
G	1
H	11
I	
J	11
K	111
L	1111
M	
N	
O	
P	1111111
Q	11111
R	
S	
T	
U	11
V	1
W	1111111
X	1111
Y	1111
Z	1

The frequencies have peaks and valleys; so, we suspect that a simple substitution cipher was used. The arrangement of the peaks and valleys suggests that a Caesar cipher was not used.

We have a crib. We suspect that the plaintext message contains the name `Thomas Jefferson`. Because this appears to be a simple substitution cipher, we expect to find a ciphertext pattern corresponding to

```
_ _ 1 _ _ 2 _ 3 4 4 3 _ 2 1 _
t h o m a s j e f f e r s o n
```

We search the ciphertext for such a pattern, and we find one.

PYWPX FBWAK PVGUF DPXFB **PHQDE** **WXXWK** **DP**YDL
HQZLY QFLPY JQDQU LABWK JBWWC

From the crib we obtain a substantial amount of the plaintext and a substantial amount of the key.

```
oneof  the r  o   t  so th  omasj  effer  sons
PYWPX  FBWAK  PVGUF  DPXFB  PHQDE  WXXWK  DPYDL

ma  n  at on   asa     her   hee
HQZLY  QFLPY  JQDQU  LABWK  JBWWC


abcdefghijklmnopqrstuvwxyz
Q   WX B E  HYP  KDF
```

Look at the partial key. It appears to have been created by a keyword. X_B and the placement of Y suggests that Y is in the keyword, that XZB appears as a ciphertext string, and that the keyword begins with B. DF suggests that the keyword ends with K. The keyword appears to be

```
B _ E _ _ H Y P _ _ K
```

If we are correct about our assumptions, A and C must be in the keyword, and two of RSTUV are in the keyword.

From

```
t_on
FLPY
```

it is reasonable to expect that plaintext i corresponds to ciphertext L.

Then the key becomes

```
abcdefghijklmnopqrstuvwxyz
Q   WX BLE  HYP  KDF
```

P _ _ K knowing that a must be in the keyword suggests PARK.

```
abcdefghijklmnopqrstuvwxyz
Q   WX BLE  HYPARKDF
```

The keyword is likely to be BLETCH(LE)YPARK. That is the correct key.

Cryptanalysis Using a Ciphertext Attack

Here is a ciphertext message:

```
DWTRF  TCPBX  CGITT  KPNME  TGUPC  WXUDR  XMWTB
JWTBT  DWTTI  DXBTP  EMWPQ  TDXCC  WXUDT  S
```

We will gather more than just single-letter frequencies. Here is its frequency analysis.

Single-letter frequencies

A	
B	1111
C	11111
D	111111
E	11
F	1
G	11
H	
I	11
J	1
K	1
L	
M	111
N	1
O	
P	11111
Q	1
R	11
S	1
T	1111111111111
U	111
V	
W	1111111
X	111111
Y	
Z	

The most frequent bigraphs

> WT 4 times
>
> UD, DX, DW, MW, TB, TT, TD, XC, XU, BT, CW two times each.

The most frequent trigraphs

> CWX, WXU, DWT, WTB, XUD twice each.

Because there are peaks and valleys in the single-letter frequencies, it is likely that we are dealing with a simple substitution cipher. It does not appear to be a Caesar cipher. As we collect more information, we will try to determine from partial keys whether we are dealing with a keyword cipher. If it appears to be a keyword cipher, we will use what we know about the construction of those keys to help with the cryptanalysis.

Because T is the most frequent ciphertext letter, DWT is one of the most frequent trigraphs, D is a frequent ciphertext letter, and DW is a frequent bigraph; ciphertext DWT likely corresponds to plaintext the.

```
the    e         ee          e      h  t     he
DWTRF  TCPBX  CGITT  KPNME  TGUPC  WXUDR  XMWTB

 he e  thee   t  e     h    et     h  te
JWTBT  DWTTI  DXBTP  EMWPQ  TDXCC  WXUDT  S
```

```
abcdefghijklmnopqrstuvwxyz
    T  W           D
```

Notice

```
efgh
T  W
```

This suggests that f corresponds to U and g corresponds to V and that this is part of the long string of letters at the end of the alphabet that often appear in alphabetical order in a keyword cipher.

```
abcdefghijklmnopqrstuvwxyz
    TUVW           D
```

We will assume that XYZ do not appear in the keyword so the key looks like

```
abcdefghijklmnopqrstuvwxyz
    TUVWXYZ        D
```

and the ciphertext looks like

```
the    e   i     ee         e      hift   i he
DWTRF  TCPBX  CGITT  KPNME  TGUPC  WXUDR  XMWTB

 he e  thee   ti e     h    eti    hifte
JWTBT  DWTTI  DXBTP  EMWPQ  TDXCC  WXUDT  S
```

Notice

```
thee   ti e
DWTTI  DXBTP
```

`e_ti_e` might be `entire`.

If so, the message and key look like

```
the    e  ri    nee         e      hift   i her
DWTRF  TCPBX  CGITT  KPNME  TGUPC  WXUDR  XMWTB

 here  theen  tire     h    eti    hifte
JWTBT  DWTTI  DXBTP  EMWPQ  TDXCC  WXUDT  S
```

```
abcdefghijklmnopqrstuvwxyz
    TUVWXYZ  I   B D
```

From the partial key, it appears that `s` corresponds to `C`. The keyword is likely to be between `Z` and `B`: _ _ `I` _ _ _ .

```
the    es ri  s nee         e   s  hift   i her
DWTRF  TCPBX  CGITT  KPNME  TGUPC  WXUDR  XMWTB

 here  theen  tire     h    etiss  hifte
JWTBT  DWTTI  DXBTP  EMWPQ  TDXCC  WXUDT  S
```

The last letter in the ciphertext message `S` seems to correspond to `d`.

s nee
CGITT suggests that o corresponds to G.

hift   i her
WXUDR   XMWTB suggests that c corresponds to R and p corresponds to M.

If we are correct, the partial key now looks like

```
abcdefghijklmnopqrstuvwxyz
  RSTUVWXYZ  IGM BCD
```

and the keyword is _ _ I G M _.

ENIGMA seems to leap out, and it is the keyword.

In this example we were able to use information gained from frequency analysis, from the partial decrypts and from the partial keys to cryptanalyze the message.

Brute Force

If the key consists of one dictionary word, it is possible to find it by brute force using a dictionary attack – use a computer to try decrypting the message using each word in a dictionary as the possible key word and each letter of the alphabet as the keyletter. It is not elegant, but it works. This argues for using key phrases rather than keywords.

Exercises

1. Construct a plaintext-ciphertext correspondence for a keyword cipher with keyphrase THE CODEBREAKERS BY DAVID KAHN and keyletter r. Write the disjoint cycle representation for this key. Using this key, encrypt the following message:

```
The computer has in no way conferred total victory upon
cryptanalysis.
```

2. The following message has been enciphered with a keyword cipher with keyword CRYPTOLOGICAL and keyletter e. Decrypt it.

```
ZTRRT   CPCGG   IVAVA   ZICFG   CPVMC   WCXBI   CRVIB
KHVHX   FSDJB   YFVDP   CFH
```

3. Often using a keyword to construct a substitution alphabet leaves the end of the alphabet unchanged; this can be a weakness. Design a keyword that avoids this weakness. Write the disjoint cycle representation for the key.

4. Find a memorable keyphrase that will scramble the entire alphabet; i.e., no letter is enciphered as itself. Write the disjoint cycle representation for the key.

5. Find a memorable keyphrase that does not have long strings of ciphertext letters in alphabetical order. Write the disjoint cycle representation for the key.

6. Here is a partial key for a keyword cipher. Complete the key.

```
abcdefghijklmnopqrstuvwxyz
N QT     Z AC    B     JL
```

7. Here is a partial key for a keyword cipher. Complete the key.

```
abcdefghijklmnopqrstuvwxyz
 ST    Z     AYCDF I   MO
```

8. Here is a ciphertext message. It is known that the message was encrypted with a keyword cipher. It is suspected that the word `confidential` appears in the plaintext message. Decrypt the message. Recreate the key.

```
INXBC  IXHIC  THHOH  IBXLT  GGOKT  EHTIU  EXIVN
EXPDT  GSLXG  XVCBY  OWXBI  OTEBC  IICUX  EXYIE
POBZT  UCJI
```

9. Here is a ciphertext message. It is known that the message was encrypted with a keyword cipher. Decrypt the message. Recreate the key.

```
JTON  QUJ  QAJSPKIW  XOKNP  KDJ  CSGBSP  YKOPKN
BITOIQOW  QAO  PBHLGO  YBLAON  PXPQOH  QAKQ
EOKNP  ABP  IKHO.
```

10. Here is a ciphertext message. It is known that the message was encrypted with a keyword cipher. Decrypt the message. Recreate the key.

```
RPMHM  AMHMJ  YWZMX  UFRQY  XUNYH  FVVRP  MKQNN
QJIVR  QMURP  MAFHG  YXIUX  QOPRU  PQNRZ  FDWMX
RURPM  UZMJQ  FVYZM  MFRYH  GYXIU  FXKYL  MHRQW
M
```

11. The American Cryptogram Association describes four types of keyword ciphers – K1, K2, K3, and K4. What we have described in this section and will continue to use as our version of the keyword cipher is type K2 – the plaintext alphabet is normal, and the ciphertext alphabet contains the key.

Here are the other three types:

K1: The plaintext alphabet contains the key, and the ciphertext alphabet is normal. Here is their example:

```
poultryabcedfghilkmnqsvwxz
RSTUVWXYZABCEDFGHIJKLMNOPQ
```

K3: Both alphabets are keyed with the same key. Here is their example:

```
conquestabdfghijklmprvwxyz
HIJKLMPRVWXYZCONQUESTABDFG
```

K4: Both alphabets are keyed; each is keyed with a different word. Here is their example:

```
shoptalkbcdefgijmnqruvwxyz
VWXYZJUPITERABCDFGHKLMNOQS
```

11a. Construct a key of type K1 using the keyword geheimschreiber. Write the disjoint cycle representation for this key.

11b. Construct a key of type K3 using the keyword sturgeon. Write the disjoint cycle representation for this key.

11c. Construct a key of type K4 using the keywords porpoise and TUNNY. Write the disjoint cycle representation for this key.

11d. Is one of the four types of keys better? Explain your response in terms of both the use of the key by authorized users and attempts to cryptanalyze the cipher.

.

12. If a message is first encrypted with a keyword cipher having keyword TURING and keyletter e and then encrypted again with a keyword cipher having keyword WELCHMAN and keyletter m, a simple substitution cipher results. Determine the key. Does this key provide more security than encryption only once with a keyword cipher? Write the disjoint cycle representation for each of the three ciphers.

13. How many possible keyword cipher keys are possible? Recall that a keyword cipher is a simple substitution cipher.