USING MAPLETS IN TEACHING CRYPTOLOGY

Rick Klima
Department of Mathematical Sciences
Appalachian State University
342 Walker Hall
Boone, North Carolina 28608
klimare@appstate.edu

Neil Sigmon
Department of Mathematics
Radford University
212 Walker Hall
Radford, Virginia 24142
npsigmon@radford.edu

A method for disguising information so that ideally it cannot be understood by anyone but the intended recipient of the information is called a *cipher*. *Cryptanalysis* refers to the process of an unintended recipient of disguised information attempting to remove the disguise and understand the information, and successful cryptanalysis is called *breaking* a cipher.

When a cipher is used to exchange information, the undisguised information is called the *plaintext*, and the disguised information the *ciphertext*. The process of converting from plaintext to ciphertext is called *encryption*. Upon receiving a ciphertext, the recipient must remove the disguise, a process called *decryption*. To be able to effectively encrypt and decrypt messages, correspondents must typically share knowledge of a secret *key*, which is used in applying the cipher. More specifically, the key for a cipher is information usually known only to the originator and intended recipient of a message, which is used by the originator to encrypt the plaintext, and by the recipient to decrypt the ciphertext.

Substitution Ciphers

One common and popular type of cipher for newspaper games and puzzle books is a *substitution* cipher. In simple substitution ciphers, users agree upon a rearrangement, or *permutation*, of the alphabet letters, yielding a collection of correspondences to be used for converting plaintext letters into ciphertext letters. This rearrangement of the alphabet letters is often called the *cipher alphabet*. To say that the cipher alphabet is a permutation means that each possible plaintext letter in the original alphabet is paired with one and only one possible ciphertext letter, and vice versa. With more sophisticated substitution ciphers, messages and cipher alphabets can include numbers, punctuation marks, or mixtures of multiple characters.

One way to form a substitution cipher is to just use a random cipher alphabet. A problem with this is that the cipher alphabet may then be cumbersome for users to keep a record of. For instance, users wishing to use a substitution cipher with a random cipher alphabet would most likely have to keep a written record of the alphabet. One solution to this problem is for users to use a *keyword* in forming the cipher alphabet.

For *simple* keyword substitution ciphers, users agree upon one or more keywords for the cipher. Spaces and duplicate letters in the keyword(s) are removed, and the resulting letters are then listed in order as the ciphertext letters that correspond to the initial plaintext letters. The remaining alphabet letters not included in the keyword(s) are then listed in the natural order to correspond to the remaining plaintext letters.

Example 1. Consider a simple keyword substitution cipher with the keywords `TIM FREEMAN`. Removing the space and duplicate letters in these keywords gives `TIMFREAN`, and yields the following cipher alphabet.

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

□

Example 1 reveals a problem with simple keyword substitution ciphers. Often with this type of cipher, notably when the keyword does not contain any letters near the end of the alphabet, the last several correspondences in the cipher alphabet are letters corresponding to themselves. Such correspondences are called *collisions*, and can make a cipher more vulnerable to cryptanalysis. Keyword *columnar* substitution ciphers can help to alleviate this problem.

For keyword columnar substitution ciphers, users again agree upon one or more keywords, and remove spaces and duplicate letters in the keyword(s). The resulting letters are then listed in order in a row, with the alphabet letters not included in the keyword(s) listed in order in successive rows of the same size (except possibly the last row) beneath the keyword letters. The cipher alphabet is then obtained by taking the columns of the resulting array of letters in order starting from the left, and placing these columns as rows under the plaintext letters.

Example 2. Consider a keyword columnar substitution cipher with the keywords `TIM FREEMAN`. Removing the space and duplicate letters again gives `TIMFREAN`, and placing these letters in a row, with the remaining alphabet letters listed in order in successive rows, yields the following array.

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

Transcribing this array by columns starting from the left yields the following cipher alphabet.

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

□

Cryptanalysis of Substitution Ciphers

Considering the number of possible cipher alphabets, substitution ciphers seem impossible to break. With 26 letters, there are more than 4×10^{26} possible cipher alphabets. To test them all would be infeasible. However, as it turns out, most substitution ciphers are fairly easy to break through the use of *frequency analysis*.

In languages like English, it is known that certain letters and combinations of letters occur more often than others. In ordinary English, the letters that naturally occur the most often are, in order, E, T, A, O, I, N, and S. The frequency with which each of the 26 letters in our alphabet occurs in ordinary English is shown in Table 1.

Letter	Frequency	Letter	Frequency
A	8.17%	N	6.75%
B	1.49%	O	7.51%
C	2.78%	P	1.93%
D	4.25%	Q	0.10%
E	12.70%	R	5.99%
F	2.23%	S	6.33%
G	2.02%	T	9.06%
H	6.09%	U	2.76%
I	6.97%	V	0.98%
J	0.15%	W	2.36%
K	0.77%	X	0.15%
L	4.03%	Y	1.97%
M	2.41%	Z	0.07%

Table 1. Letter frequencies in ordinary English

Common digraphs (letter pairs), trigraphs (letter triples), and repeated letters in ordinary English are also known. The most common digraphs are TH, ER, ON, AN, RE, HE, IN, ED, and ND. The most common trigraphs are THE, AND, THA, ENT, ION, TIO, FOR, NDE, HAS, and NCE. The most common repeated letters are LL, EE, SS, TT, OO, MM, and FF. For a thorough analysis of common letter sequences in ordinary English, see [3].

For a ciphertext formed using a substitution cipher, with a sufficient number of ciphertext letters and the spacing between words in the plaintext preserved, frequency analysis can usually be used to break the cipher.

Example 3. Consider the following ciphertext, which was formed using a substitution cipher.

```
WZIS VZIL VRRQ VZI CRAEVZ TGISYGISV, M WTJ JMFISV
BIOTAJI M YRS'V YITF YEAPJ. WZIS VZIL VRRQ VZI JMNVZ
TGISYGISV, M QIHV UAMIV BIOTAJI M QSRW M'G MSSROISV.
```

```
WZIS VZIL VRRQ VZI JIORSY TGISYGISV, M JTMY SRVZMSP
BIOTAJI M YRS'V RWS T PAS. SRW VZIL'KI ORGI CRE VZI
CMEJV TGISYGISV TSY M OTS'V JTL TSLVZMSP TV TFF.
```

The frequency with which each letter occurs in this ciphertext is shown in the following table.

Letter:	I	S	V	R	T	M	Z	J	Y	G	A	O	W
Count:	33	28	28	17	17	16	15	11	11	10	7	7	7
Letter:	L	Q	E	F	P	B	C	H	K	N	U	D	X
Count:	6	5	4	4	4	3	3	1	1	1	1	0	0

Table 2. Letter frequencies in a ciphertext

Based on the frequency and locations of the letter I in the ciphertext, it seems likely that this letter corresponds to E in the plaintext. In addition, the trigraph VZI occurs in the ciphertext eight times, four of these as a single word. Since the trigraph that occurs with the highest frequency in ordinary English is THE, it seems reasonable to suppose that VZI in the ciphertext corresponds to THE in the plaintext. The validity of this is reinforced by the fact that it causes the second most common letter in the ciphertext, V, to correspond to the second most common letter in ordinary English, T. Note also that the one letter words M and T both occur in the ciphertext. In ordinary English, the most common one letter words are A and I. The fact that the ciphertext also contains the word M'G suggests it is likely that the ciphertext letter M corresponds to I in the plaintext, and consequently that the ciphertext letter T corresponds to A in the plaintext.

Next, note that the repeated letters RR occur three times in the ciphertext, each in the middle of the word VRRQ, which suggests that the ciphertext letter R corresponds to a vowel in the plaintext. The repeated vowels most likely to occur in ordinary English are EE and OO, and since a ciphertext letter has already been assigned to the plaintext letter E, it seems reasonable that the ciphertext letter R corresponds to the plaintext letter O. In addition, the fact that each time VRRQ occurs in the ciphertext it is followed by THE in the plaintext suggests that the ciphertext word VRRQ corresponds to TOOK in the plaintext. Thus, we will assign the ciphertext letter Q to the plaintext letter K. Also, note that the third most common letter in the ciphertext is S. Based on the positions of S in the ciphertext, it appears likely that the ciphertext letter S corresponds to a consonant in the plaintext. Since we have already assigned a ciphertext letter to the most common consonant in ordinary English, T, it seems reasonable to assume that the ciphertext letter S corresponds to the second most common consonant in ordinary English, N. The following shows the part of the cipher alphabet we have assigned so far.

```
Plain:   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
Cipher:  T _ _ _ I _ _ Z M _ Q _ _ S R _ _ _ _ V _ _ _ _ _ _
```

The following shows the complete ciphertext, with the part of the plaintext given by the plain/cipher letter correspondences that we have determined provided above the ciphertext letters.

```
 HEN THE  TOOK THE  O  TH A EN  ENT, I  A   I ENT
WZIS VZIL VRRQ VZI CRAEVZ TGISYGISV, M WTJ JMFISV

 E A  E I  ON'T  EA       .  HEN THE  TOOK THE  I TH
BIOTAJI M YRS'V YITF YEAPJ. WZIS VZIL VRRQ VZI JMNVZ

A EN  ENT, I KE T   IET  E A  E I KNO  I'  INNO ENT.
TGISYGISV, M QIHV UAMIV BIOTAJI M QSRW M'G MSSROISV.

 HEN THE  TOOK THE  E ON  A EN  ENT, I  AI  NOTHIN
WZIS VZIL VRRQ VZI JIORSY TGISYGISV, M JTMY SRVZMSP

 E A  E I  ON'T O N A   N. NO  THE ' E  O E  O  THE
BIOTAJI M YRS'V RWS T PAS. SRW VZIL'KI ORGI CRE VZI

 I  T A EN  ENT AN  I  AN'T  A  AN THIN  AT A  .
CMEJV TGISYGISV TSY M OTS'V JTL TSLVZMSP TV TFF.
```

The first two words in the plaintext now appear to be WHEN and THEY. Other resulting apparent words in the plaintext are DON'T, INNOCENT, NOTHING, and ANYTHING. With just a little more thought, the plain/cipher letter assignments can be completed, yielding the following full plaintext.

```
WHEN THEY TOOK THE FOURTH AMENDMENT, I WAS SILENT
BECAUSE I DON'T DEAL DRUGS. WHEN THEY TOOK THE SIXTH
AMENDMENT, I KEPT QUIET BECAUSE I KNOW I'M INNOCENT.
WHEN THEY TOOK THE SECOND AMENDMENT, I SAID NOTHING
BECAUSE I DON'T OWN A GUN. NOW THEY'VE COME FOR THE
FIRST AMENDMENT AND I CAN'T SAY ANYTHING AT ALL.
```

□

The reason we were able to break the cipher in Example 3 relatively easily is because a sufficient numbers of ciphertext letters corresponded to plaintext letters that occur frequently in ordinary English. Having the punctuation and spacing between words preserved made it easier to break as well. A ciphertext with a smaller number of letters or in which punctuation and spacing had been removed could have been much more difficult to cryptanalyze. Substitution ciphers in which entire plaintext words are replaced with numbers or words (known as *nomenclators*) can also be more difficult to break, as can ciphers in languages in which letter frequencies are different from those in English. However, history has shown that most substitution ciphers are insecure and can be broken through persistence.

Transposition Ciphers

Transposition ciphers differ from substitution ciphers in that plaintext letters are not encrypted by being replaced by other letters, but rather by being rearranged according to some rule agreed upon by the two parties wishing to exchange the message. That is, to form the ciphertext for a transposition cipher, the plaintext letters are rearranged in some manner, as opposed to being replaced by other letters.

For *columnar* transposition ciphers, users agree upon some prescribed number of columns, and then the plaintext letters (with spaces and punctuation removed) are used to form an array of letters, similar to the array used in keyword columnar substitution ciphers, with this prescribed number of columns. The ciphertext is obtained by taking the columns of the resulting array in some specified order, and placing the letters in these columns in a row. For *simple* columnar transposition ciphers, the ciphertext is obtained by taking the columns of the array in order, starting from the left, and placing the letters in these columns in a row.

Example 4. Consider a simple columnar transposition cipher with nine columns. To use this cipher to encrypt the plaintext `ALL IN THE FAMILY WAS A CLASSIC AND SOMETIMES CONTROVERSIAL TV SHOW`, we begin by using these plaintext letters to form the following array.

```
A  L  L  I  N  T  H  E  F
A  M  I  L  Y  W  A  S  A
C  L  A  S  S  I  C  A  N
D  S  O  M  E  T  I  M  E
S  C  O  N  T  R  O  V  E
R  S  I  A  L  T  V  S  H
O  W
```

To form the ciphertext, we transcribe this array by columns starting from the left. Thus, the ciphertext is `AACDS ROLML SCSWL IAOOI ILSMN ANYSE TLTWI TRTHA CIOVE SAMVS FANEE H`.

□

A problem with simple columnar transposition ciphers is that ciphertexts are always obtained by taking the columns of the array in order starting from the left. This can make a cipher more vulnerable to cryptanalysis. *Keyword* columnar transposition ciphers can help to alleviate this problem.

For keyword columnar transposition ciphers, users agree upon one or more keywords, and remove spaces in the keyword(s). However, unlike for keyword substitution ciphers, for keyword transposition ciphers duplicate letters are not removed from the keyword(s). The number of columns in the array is then equal to the number of keyword letters, with the keyword letters placed in order as labels on the columns, and the ciphertext obtained

by taking the columns of the array (not including the keyword letter labels) in alphabetical order by the keyword letter labels, and placing the letters in these columns in a row. If the keyword(s) contain any duplicate letters, then columns with identical keyword letter labels are taken in order starting from the left.

Example 5. Consider a keyword columnar transposition cipher with the keyword BUNKER. To use this cipher to encrypt the plaintext ALL IN THE FAMILY WAS A CLASSIC AND SOMETIMES CONTROVERSIAL TV SHOW, we begin by using these keyword and plaintext letters to form the following array. The numbers above the keyword letter labels indicate the order in which the columns should be taken to form the ciphertext.

1	6	4	3	2	5
B	U	N	K	E	R
A	L	L	I	N	T
H	E	F	A	M	I
L	Y	W	A	S	A
C	L	A	S	S	I
C	A	N	D	S	O
M	E	T	I	M	E
S	C	O	N	T	R
O	V	E	R	S	I
A	L	T	V	S	H
O	W				

Thus, the ciphertext is AHLCC MSOAO NMSSS MTSSI AASDI NRVLF WANTO ETTIA IOERI HLEYL AECVL W.

□

Cryptanalysis of Transposition Ciphers

Because keyword columnar transposition ciphers need not take the columns of the cipher array in order, cryptanalysis can be more difficult than for simple columnar transposition ciphers. The cryptanalysis process can be simplified, however, with the knowledge of a crib (that is, a known part of the plaintext) that is longer than the keyword(s).

Example 6. Consider the ciphertext AHLCC MSOAO NMSSS MTSSI AASDI NRVLF WANTO ETTIA IOERI HLEYL AECVL W, which was formed using a keyword columnar transposition cipher, and suppose we have the crib THE FAMILY. (That is, suppose we know THE FAMILY is part of the corresponding plaintext.) To try to decrypt this message, in the hope that our crib is longer than the keyword(s) for the cipher, we will start by assuming there are exactly eight letters in the keyword(s). If there were eight letters in the keyword(s), then the array would have eight columns, and the crib would appear in these columns in the following form.

```
T  H  E  F  A  M  I  L
Y
```

Thus, the digraph TY would have to appear in the ciphertext. However, TY does not appear in the ciphertext, and so there are not exactly eight letters in the keyword(s). So next we will assume there are exactly seven letters in the keyword(s). If there were seven letters in the keyword(s), then the array would have seven columns, and the crib would appear in these columns in the following form.

```
T  H  E  F  A  M  I
L  Y
```

However, the digraphs TL and HY do not both appear in the ciphertext, and so there are not exactly seven letters in the keyword(s). (Although neither digraph appears in the ciphertext, either one failing to appear would be enough to indicate this.) So we will assume there are exactly six letters in the keyword(s), in which case the array would have six columns, and the crib would appear in these columns in the following form.

```
T  H  E  F  A  M
I  L  Y
```

Since the digraphs TI, HL, and EY all appear in the ciphertext, it is likely that there are exactly six letters in the keyword(s) and six columns in the array. Because the ciphertext contains 56 letters, the first two columns of this array will contain 10 letters, and the remaining four columns will contain nine letters. Thus we will split the ciphertext into blocks of nine letters each, which we label as follows.

$$\underbrace{\texttt{AHLCCMSOA}}_{1}\ \underbrace{\texttt{ONMSSSMTS}}_{2}\ \underbrace{\texttt{SIAASDINR}}_{3}\ \underbrace{\texttt{VLFWANTOE}}_{4}\ \underbrace{\texttt{TTIAIOERI}}_{5}\ \underbrace{\texttt{HLEYLAECV}}_{6}\ \texttt{LW}$$

Next, we will arrange these blocks as columns in an array in the only way in which the known crib and digraphs TI, HL, and EY all line up correctly. This yields the following.

5	1	6	4	3	2
		H	V	S	O
T	A	L	L	I	N
T	H	E	F	A	M
I	L	Y	W	A	S
A	C	L	A	S	S
I	C	A	N	D	S
O	M	E	T	I	M
E	S	C	O	N	T
R	O	V	E	R	S
I	A				

When reading across the rows of this array from the top, the letters begin to form sensible English starting with the columns labeled 1 and 6. Thus, it is likely that these columns are the first two in the original cipher array, and would therefore be the two columns that contain 10 letters instead of nine. So we will split the ciphertext into blocks again, using 10 letters in the blocks labeled 1 and 6, and nine letters in the rest.

$$\underbrace{\texttt{AHLCCMSOAO}}_{1}\ \underbrace{\texttt{NMSSSMTSS}}_{2}\ \underbrace{\texttt{IAASDINRV}}_{3}\ \underbrace{\texttt{LFWANTOET}}_{4}\ \underbrace{\texttt{TIAIOERIH}}_{5}\ \underbrace{\texttt{LEYLAECVLW}}_{6}$$

Arranging these blocks as columns in the same order as in the previous array, with the block labeled 5 moved from the front of the array to the end, yields the following.

1	6	4	3	2	5
A	L	L	I	N	T
H	E	F	A	M	I
L	Y	W	A	S	A
C	L	A	S	S	I
C	A	N	D	S	O
M	E	T	I	M	E
S	C	O	N	T	R
O	V	E	R	S	I
A	L	T	V	S	H
O	W				

Thus, the plaintext is ALL IN THE FAMILY WAS A CLASSIC AND SOMETIMES CONTROVERSIAL TV SHOW.

□

Using Maplets in Teaching Cryptology

Several years ago we were invited to create a new general education mathematics course for the Honors Academy at Radford University. Wanting to create a multidisciplinary course that would demonstrate some interesting mathematical applications and also be accessible and intriguing to students with a wide variety of interests and backgrounds, we decided on a course in cryptology. Designed for students with only a basic understanding of algebra, statistics, and number theory at the secondary level, the course has been one of the most popular offerings at the Honors Academy.

When deciding on material for the course, since we expected most of our students to come from nontechnical fields, our goal was to choose topics that would be easy to understand, showed the importance of cryptology in both cultural and historical contexts, and demonstrated some stimulating but relatively simple mathematical applications. A lesser goal was for students to be motivated to study the subject further and perhaps even consider careers in mathematics or the sciences.

In order to include substantive examples of the techniques presented in the course, we incorporated the mathematics software package Maple. However, the use of Maple was met with only limited success, since students had to develop a rudimentary working knowledge of Maple syntax. For some students, especially from nontechnical fields with little or no background in programming or computer syntax, this proved too daunting a task. We alleviated this problem through the use of Maplets, which employ (and need) the Maple engine, but operate exclusively within windows that are simple to use and require no background in computers. By simply typing information in textboxes and using buttons and drop-down menus in Maplet windows, our students were able to complete complicated tasks with relative ease. We produced these Maplets with written code in Maple, but because our students only needed to use the end product, they did not need to have even a rudimentary working knowledge of Maple syntax.

Our Maplets and other course materials eventually evolved into a book [1], which includes presentations of a variety of types of ciphers and other cryptographic techniques and systems, including our sample presentations of substitution and transposition ciphers in this paper, most supplemented with Maplets. Due to space limitations, it is not possible for us to demonstrate any of these Maplets here. However, all of our Maplets, including those for implementation and cryptanalysis of substitution and transposition ciphers, can be downloaded from [2], and examples with screenshots and accompanying discussion for all of our Maplets can be found in [1].

A complete list of the types of ciphers and other cryptographic techniques and systems presented and supplemented with Maplets in [1] includes substitution ciphers (with cryptanalysis), Playfair ciphers, transposition ciphers (with cryptanalysis), ADFGVX ciphers, the Enigma machine, the Navajo code, shift and affine ciphers (with cryptanalysis), Alberti ciphers, Vigenère ciphers (with cryptanalysis), the Friedman and Kasiski tests, Hill ciphers (with cryptanalysis), RSA ciphers (with cryptanalysis), the Diffie-Hellman key exchange, ElGamal ciphers (with cryptanalysis), stream ciphers, the Advanced Encryption Standard, digital signatures, hash functions, and public-key infrastructures. Also included in [1] are basic introductions to combinatorics, modular arithmetic, probability, matrices, the Euclidean algorithm, modular exponentiation, ASCII, primality testing, integer factorization, discrete logarithms, and number base conversions, some supplemented with Maplets, and all presented to a general audience.

References

[1] R. Klima and N. Sigmon. *Cryptology, Classical and Modern, with Maplets.* Taylor & Francis/CRC Press, Boca Raton, FL, 2012.

[2] R. Klima and N. Sigmon. Textbook homepage for *Cryptology, Classical and Modern, with Maplets*. Available at http://www.radford.edu/~npsigmon/cryptobook.html, 2012.

[3] P. Wiedman. Cryptopop's Hints. Available at http://www.freewebs.com/gidusko/cryptopop/, 2007.