Polyalphabetic Substitution Ciphers

Last Revised – May 19, 2006

Mordecai Golin
golin@cs.ust.hk
www.cs.ust.hk/˜golin

These slides are based on Chapter 2 of
“Cryptological Mathematics”
by Robert Edward Lewand.

Until now we've concentrated on
Monoalphabetic Substitution Ciphers.

These are ciphers in which a plaintext letter gets mapped to exactly one ciphertext letter and it gets mapped to the same ciphertext letter each time. So, there is a *one-one mapping* between plaintext and ciphertext letters.

Monoalphabetic Substitution Ciphers can be attacked using *frequency analysis*.

In this section we will discuss the more sophisticated Polyalphabetic Substitution Ciphers in which

- a plaintext letter can be mapped to more than one ciphertext letter and

- a ciphertext letter can represent more than one plaintext letter.

With Polyalphabetic Substitution Ciphers it would be possible to have

```
meet me at the usual place at eight oclock
sss sss sss sss sss sss sss sss sss sss sss s
```

This would definitely be harder for the bad-guy to *break* but how would it be possible for the intended recipient to *decipher* this code?

First Attempt: Disguising Frequencies

We will associate with each letter A...Z a ***unique subset*** of the numbers $\{00, 01, \ldots, 99\}$ such that

- all subsets are distinct
  (they don't share any numbers)
- Every number is in some subset
- the number of elements in subset associated with a letter is the integer closest to $100\times$ freq of letter.

Enciphering rule: replace a plaintext letter by *randomly* choosing *one* of the numbers in its set.

Deciphering rule: Replace each two digit number with the letter whose set it's in.

Example

Letter	Subset of S
a	15, 33, 37, 55, 57, 72, 91, 96
b	24
c	03, 39, 67
d	04, 43, 61, 88
e	08, 12, 20, 46, 47, 59, 64, 79, 81, 85, 90, 94, 97
f	40, 48
g	29, 53
h	05, 16, 30, 42, 69, 99
i	14, 45, 50, 60, 73, 82, 93
j	11
k	77
l	01, 26, 71, 98
m	34, 87
n	06, 17, 22, 31, 49, 58
o	02, 10, 41, 51, 66, 75, 83
p	13, 18
q	36
r	21, 25, 65, 68, 92, 95
s	00, 28, 52, 63, 74, 78
t	07, 19, 23, 35, 38, 54, 70, 84, 89
u	09, 32
v	44
w	56, 80
x	86
y	62, 76
z	27

For the cipher on the previous page the plaintext

Meet me at the usual place at eight o'clock

could be replaced by *both*

872079383408377089421209523296011871333946570759505316191003267567777

and

340808388781151970059409003255011326153985578497452969894103018339777

Our One-to-Many Cipher

Good points: Makes frequency analysis attacks much more difficult since every ciphertext number has (almost) the same frequency.

Bad points: Key is *very* large and hard to communicate.

We will now see the Vigenère square, a polyalphabetic cipher with an easy to remember key.

The Vigenère square

	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
a	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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

The Vigenère square

The Vigenère square was invented in 1586 by Blaise de Vigenère.

The Vigenère poylalphabetic substitution cipher works as follows:

- Decide on a keyword and write it down
- Write the plaintext, character for character, along the keyword, repeating the keyword as many times as necessary.
- Replace each letter of the plaintext message with the letter that lies in the intersection of the column indexed by the plaintext letter and the row indexed by the corresponding keyword letter.

Note that what is really happening is that the keyword letter is designating a Caesar cipher to apply to the plaintext letter.

Keyword	Plaintext	Ciphertext
c	m	O
h	o	V
a	l	L
o	l	Z
s	y	Q
c	w	Y
h	i	P
a	l	L
o	l	Z
s	n	F
c	e	G
h	v	C
a	e	E
o	r	F
s	b	T
c	r	T
h	e	L
a	a	A
o	k	Y
s	t	L
c	h	J
h	i	P
a	s	S

Example with keyword “JOKE”

A mathematician, a physicist, and an engineer are each asked to prove the assertion that all odd numbers greater than one are prime. The mathematician says, “three is prime, five is prime, seven is prime, and by mathematical induction, all odd numbers greater than one are prime.” The physicist says “three is prime, five is prime, seven is prime, nine is an experimental error, eleven is prime, and so, yes, all odd numbers greater than one are prime.” The engineer says, “three is prime, five is prime, seven is prime, nine is prime, eleven is prime, thirteen is prime, fifteen is prime....”

JAKXQ SWECW MMJBK TQMCM LWCXJ BNEWS
XKRBO IAOBI NOMLJ GUIMH YTACF ICVOE
BGOVC WYRCV KXJZV SMRXY VPOVB UBIJH
OVCVK RXBOE ASZVR AOXQS WECVO QJHSG
ROXWJ MCXQF OIRGZ VRAOJ RJOMB DBMVS
CIESX MBDBM VSKRM GYFHA KXQSW ECWME
UWXHD QDMXB KPUCN HWIWF NFCKA SKXNF
DLJBY RNOBI YFSQN HRIYV IWRQS WCGKC
BHRVN SSWYF SQNTS ZNWCT AWWIB SFIWW
CTAWW IWWXI RGKRN LZIAW WIWHK PNFBS
ASVIE SXMBD BMVSK RMGYC NGKPU CNHWI
WFNFC KASKX NFDLJ BYRNO BIYFS QNHRI
NBQMW SOVBO IWCVB INWCT AWWIO WFIRG
ZVRAO WNJOR RGZVR AORRB OMBDB MVSOP
NJORR GZVRA OXQWB XNSXM BDBMV SPMOH
OIWWC TAWWI

Frequency Analysis will no longer work (why?).
How can this be broken?

There are repeated ciphertext strings!! Why?
Can that help us?

JAKXQ	SWECW	MMJBK	TQMCM	LWCXJ	BNEWS
XKRBO	IAOBI	NOMLJ	GUIMH	YTACF	ICVOE
BGOVC	WYRCV	KXJZV	SMRXY	VPOVB	UBIJH
OVCVK	RXBOE	ASZVR	AOXQS	WECVO	QJHSG
ROXWJ	MCXQF	OI**RGZ**	**VRAO**J	RJOMB	DBMVS
CIESX	MBDBM	VSKRM	GYFHA	KXQSW	ECWME
UWXHD	QDMXB	KPUCN	HWIWF	NFCKA	SKXNF
DLJBY	RNOBI	YFSQN	HRIYV	IWRQS	WCGKC
BHRVN	SSWYF	SQNTS	ZN**WCT**	**AWWI**B	SFIW**W**
CTAWW	**I**WWXI	RGKRN	LZIAW	WIWHK	PNFBS
ASVIE	SXMBD	BMVSK	RMGYC	NGKPU	CNHWI
WFNFC	KASKX	NFDLJ	BYRNO	BIYFS	QNHRI
NBQMW	SOVBO	IWCVB	IN**WCT**	**AWWI**O	WFI**RG**
ZVRAO	WNJOR	**RGZVR**	**AO**RRB	OMBDB	MVSOP
NJOR**R**	**GZVRA**	**O**XQWB	XNSXM	BDBMV	SPMOH
OIW**W**C	**TAWWI**				

How can we get repeated ciphertext?
Does it correspond to repeated plaintext?
Yes and no!!

Suppose we try to encode
"...on a plane. The plane is due..."
using keywords water and milk.

Keyword	w	a	t	e	r	w	a	t	e	r	w	a	t	e	r	w	a	t	e	r	w
Plaintext	o	n	a	p	l	a	n	e	t	h	e	p	l	a	n	e	i	s	d	u	e
Ciphertext	K	N	T	T	C	W	N	X	X	Y	A	P	E	E	E	A	I	L	H	L	A

Keyword	m	i	l	k	m	i	l	k	m	i	l	k	m	i	l	k	m	i	l	k	m
Plaintext	o	n	a	p	l	a	n	e	t	h	e	p	l	a	n	e	i	s	d	u	e
Ciphertext	A	V	L	Z	X	I	Y	O	F	P	P	Z	X	I	Y	O	U	A	O	E	Q

For water the repeated plaintext "plane" is not repeated in the ciphertext. For milk it is. Why?

This is because in the second case the two copies of plane appeared under the *same* key letters, while in the first case it didn't.
In other words, in the second case, the distance between the starts of the two copies of plane was a multiple of the keyword size, while in the first case, it wasn't.

The Kasiski Test: If a string of characters appears repeatedly in a polyalphabetic ciphertext message then it is possible (although not certain) that the distance between the occurances is a multiple of the length of the keyword (the longer the repeated string of characters, the more likely it is).

In our case the Kasiski Test suggests that the keyword is of length 2 or 4.

Repeated String	Position of first letter in string	Distance between pairs of occurrences of the string	Prime factorization of distance between pairs of occurrences of the string
WCTAWWI	258		
	270	12	$2^2 3$
	378	108	$2^2 3^3$
	454	76	$2^2 19$
RGZVRA	133		
	389	256	2^8
	401	12	$2^2 3$
	425	24	$2^3 3$
MBDBMVS	144		
	156	12	$2^2 3$
	308	152	$2^3 19$
	412	104	$2^3 13$
	440	28	$2^2 7$

We will not go into further details on how to break the Viggenere cipher except to note that once we know the keyword size is m we know to split up the cipher-text into m alternating Caesar ciphers and can use variants of frequency analysis.

For more details please see the instructor or class web page for references.