COMP3260 Data Security

Lecture 4

Prof Ljiljana Brankovic

Lecture Overview

1. Transposition Ciphers
2. Breaking Transposition Ciphers
3. Substitution Ciphers
4. Breaking Substitution Ciphers
5. Homophonic ciphers - Beale ciphers

Classical Ciphers

- Chapter 3 textbook “Classical Encryption Techniques”

- These lecture notes (based on the text, “Cryptography and Data Security” by D. Denning [2], lecture notes by M. Miller and other sources)

  Note that in-text references and quotes are omitted for clarity of the slides. When you write as essay or a report it is very important that you use both in-text references and quotes where appropriate.

Ciphers

Classical ciphers fall into one of the following categories:

- transposition ciphers, where the characters in the plaintext are simply rearranged

- substitution ciphers, where each character (or a group of characters) is substituted by another character (or a group of characters); substitution ciphers can be divided into:
  - monoalphabetic
  - homophonic
  - polyalphabetic
  - polygrams

Transposition Ciphers

Transposition ciphers rearrange characters according to some scheme often using some geometric figure. Recall that to encipher, we need an enciphering algorithm and an enciphering key. The ‘figure’ and the ‘writing-in’ and ‘talking-off’ methods correspond to enciphering algorithm, while some parameter that determines the figure corresponds to the enciphering key.

Example 1. *Plaintext: DISCONCERTED COMPOSER*

```
D       O       R       C       O
  I   C   N   E   T   D   O   P   S   R
    S       C       E       M       E
```

Ciphertext: DORCOICNETDOPSRSCEME

The algorithm: arrange letters of the plaintext in in rail-like way and read off by rows

The key: the ‘rail’ depth (in this case 3).

Columnar Transposition

Columnar transposition:

- plaintext is written into a matrix by rows
- ciphertext is obtained by taking off the columns in some order

Example 2: Using 6 columns, the plaintext SYDNEY OLIMPIC GAMES is written by rows as

S	Y	D	N	E	Y
O	L	Y	M	P	I
C	G	A	M	E	S

If the columns are taken off in the order 6-5-2-4-1-3 the resulting ciphertext is YISEPEYLGNMMSOCDYA.

Periodic Transpositions

Every transposition cipher is a ***permutation*** of the plaintext with some ***period*** d. The period of the permutation can be as long as the message but usually it is shorter. Why?

Let Z_d be the set of integers $\{1,2,\dots,d\}$ and let $f : Z_d \rightarrow Z_d$ be a permutation over Z_d. Then the key is f. To encipher, successive blocks of d characters are permuted according to f.

A plaintext message $M = m_1 \dots m_d m_{d+1} \dots m_{2d} \dots$ is enciphered as

$$E_{k(M)} = m_{f(1)} \dots m_{f(d)} m_{f(d+1)} \dots m_{f(2d)} \dots$$

Decipherment uses the inverse permutation.

Example 3. Suppose $d = 6$ and f is the permutation

i	1	2	3	4	5	6
$f(i)$	6	5	4	3	2	1

Then the plaintext

SYDNEY OLYMPIC GAMES

is enciphered as

YENDYSIPMYLOSEMAGC.

Periodic permutation ciphers can be implemented efficiently on a computer.

Breaking Transposition Ciphers

To recognise that a ciphertext was produced by a transposition cipher: Compare the relative frequencies of the letters in the ciphertext with the expected frequencies for the plaintext.

Transposition ciphers are broken by anagramming (the process of restoring a disarranged set of letters into their original positions).

Tables of frequency distributions for diagrams and trigrams are used in the anagramming process.

Frequency Distribution of Letters in English Text

Char	Percent	
A	8.0	****************
B	1.5	***
C	3	******
D	4.0	********
E	13.0	**************************
F	2.0	****
G	1.5	***
H	6.0	************
I	6.5	*************
J	0.5	*
K	0.5	*
L	3.5	*******
M	3.0	*******
N	7.0	**************
O	8.0	****************
P	2.0	****
Q	0.2	
R	6.5	*************
S	6.0	************
T	9.0	******************
U	3.0	******
V	1.0	**
W	1.5	***
X	0.5	*
Y	2.0	****
Z	0.2	

English Diagrams

The most frequent pairs of letters (diagrams) in English on a relative scale of 1 to 10:

Diagram	Frequency	Diagram	Frequency
TH	10.00	HE	9.05
IN	7.17	ER	6.65
RE	5.92	ON	5.70
AN	5.63	EN	4.76
AT	4.72	ES	4.24
ED	4.12	TE	4.04
TI	4.00	OR	3.98
ST	3.81	AR	3.54
ND	3.52	TO	3.50
NT	3.44	IS	3.43
OF	3.38	IT	3.26
AL	3.15	AS	3.00

English Trigrams

The most frequent trigrams in English:

ENT
ION
AND
ING
IVE
TIO
FOR
OUR
THI
ONE

Unicity Distance of a Permutation Cipher

How much ciphertext is needed to break a permutation cipher with period d ? Unicity distance of a permutation cipher with period d :

$$N = \frac{H(K)}{D} = \frac{\log_2(d\,!)}{D}$$

Sterling's approximation for large d: $d\,! \approx (\frac{d}{e})^d\sqrt{2\pi d}$.

Then $\log_2(d\,!) \approx d\log_2(\frac{d}{e})$ and $N = \frac{d\log_2(\frac{d}{e})}{3.2} = 0.3\,d\log_2(\frac{d}{e})$

Example 4: If the period is $d = 27$, then $\frac{d}{e}$ is about 10 and $\log_2(\frac{d}{e})$ is about 3.2 so $N = 27$.

The following table shows the period and the associated Unicity distance.

d	3	4	5	6	7
N	0.122804	0.66877	1.31885	2.05608	2.86579

Substitution Ciphers

Substitution ciphers can be divided into:

- monoalphabetic
- homophonic
- polyalphabetic
- polygrams

A monoalphabetic substitution cipher replaces each character of the plaintext alphabet A with the corresponding character of the ciphertext alphabet C. Usually C is a simple rearrangement of the lexicographic order of the characters in A.

Substitution Ciphers

Suppose A is a n-character alphabet $\{a_0, a_1, \ldots, a_{n-1}\}$.

Then C is a n-character alphabet $\{f(a_0), f(a_1), \ldots, f(a_{n-1})\}$, where $f : A \to C$ is a one-to-one mapping of each character of A to the corresponding character of C.

To encipher, simply rewrite the message using the corresponding characters of the ciphertext language:

$$E_k(M) = f(m_1)f(m_2)\ldots$$

Example 5.

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

Such a ciphertext alphabet is called a ***keyword mixed alphabet***. In the example above the key of the cipher is $SYDNEY\ OLYMPIC\ GAMES$. The repeated letters in the key are dropped and after the key the remaining letters appear in alphabetic order.

The message $M = DOWN\ ELEVATOR$ is encrypted as

$$E_k(M) = NFVB\ EGEUSRFK$$

Substitution Ciphers

Ciphers based of **shifted alphabets** shift the letters of the alphabet by k positions to the right, modulo the size of the alphabet:

$$f(x) = (x + k) \bmod n$$

where n is the size of the alphabet A, x denotes a letter of A by its position, and k is the key.

More complex transformations use multiplication:

$$f(x) = kx \bmod n$$

where k and n are relatively prime so that the mapping is one-to-one. Here k is the key.

Example 6. If $k = 9$ and $f(x) = kx \bmod n$

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

Affine Transformations

Affine transformation combines addition with multiplication to get

$$f(x) = (xk_1 + k_0) \, mod \, n$$

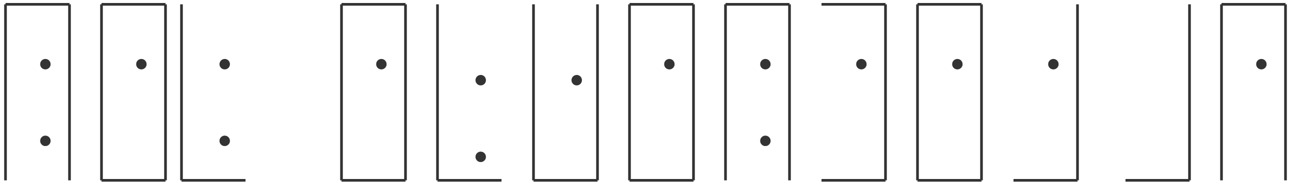
where k_1 and n are relatively prime.

In general, we can have polynomial transformations of any degree t:

$$f(x) = (x^t k_t + x^{t-1}k_{t-1} + \ldots + xk_1 + k_0) \, mod \, n$$

Note: Using nonstandard ciphertext alphabets doesn't increase the difficulty of breaking the cipher.

Example 7. A Churchyard cipher engraved on a tombstone in Trinity Churchyard, New York, 1794:



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 similar cipher was also engraved on a tombstone in St. Paul's Churchyard, New York, in 1796. The first published solution to this cipher appeared in the New York Herald in 1896 - over 100 years later.

Why did it take so long to break this cipher?

Breaking Substitution Cipher

Example 8. Find the number of letters needed to break general substitution alphabets of size n.

The number of possible keys is n! (that is the number of ways of arranging the n letters of the alphabet).

If all keys are equally likely then the unicity distance is

$$N = H(K) / D = (\log_2 n!) / D$$

For English, $N = (\log_2 26!) / 3.2 = 88.4 / 3.2 = 27.6$

That means that usually at least 28 letters are needed to break these ciphers. That explains the difficulty in solving the Churchyard ciphers (only about 15 characters).

Breaking Substitution Cipher

Ciphers based on polynomial transformations have smaller unicity distances.

For shifted alphabets the number of possible keys is only 26 and the unicity distance is

$$N \cong (\log_2 26\ )\ /\ 3.2 \cong 1.5$$

Simple substitution ciphers are easy to break in a ciphertext only attack using single letter frequency analysis: comparing the letter frequencies in a given ciphertext with the expected frequencies to match the ciphertext letters with the plaintext letters.

Diagram and trigram distributions can also be used.

Ciphers based on shifted alphabets are extremely easy to break because each ciphertext letter is a constant distance from its corresponding plaintext letter.

Breaking Substitution Cipher

Ciphers based on affine transformations

$$f(x) = (xk_1 + k_0) \, mod \, n$$

are more difficult to break BUT if a set of t correspondences between plaintext letters m_i and ciphertext letters c_i , $1 \leq i \leq t$, are known (or suspected) then it may be possible to find k_1 and k_0 by solving the following system of equations:

$$(m_1 k_1 + k_0) \, mod \, n = c_1$$

.

.

.

$$(m_t k_1 + k_0) \, mod \, n = c_t$$

… … . .

Breaking Substitution Cipher

Example 9. Suppose we have the following possible correspondences.

Plaintext	$E(4)$	$J(9)$	$N(13)$
Ciphertext	$K(10)$	$T(19)$	$V(21)$

That gives the equations

$$(4k_1 + k_0) \, mod \, 26 = 10$$
$$(9k_1 + k_0) \, mod \, 26 = 19$$
$$(13k_1 + k_0) \, mod \, 26 = 21$$

The solutions of the first two equations is $k_1 = 7$ and $k_0 = 8$. Note that we must check that the third equation is also satisfied. What would it mean if the third equation is not satisfied?

Note that in general we may need more than 2 equations to solve for k_0 and k_1, as equations of the form $ak \, mod \, 26 = c$ have multiple solutions when a divides 26.

Breaking Substitution Cipher

Cryptanalysis of a general simple substitution cipher:

- Brute force attacks: try all 26! decipherments - if 1 decipherment per microsecond, it would take more that 10^3 years!

- Instead use a single letter frequency analysis - diagram and trigram distributions are also helpful.

Homophonic ciphers

A ***homophonic substitution cipher*** maps each character x of the plaintext alphabet into a set of ciphertext elements $f(x)$ called ***homophones***.

A plaintext message $M = m_1m_2 \ldots$ is enciphered as $C = c_1c_2 \ldots$ where each c_i is picked at random from the set of homophones $f(m_i)$.

Example 10: Suppose that the English letters are enciphered as integers between 0 and 99. The number of integers assigned to a letter is proportional to the relative frequency of the letter. No integer is assigned to more than one letter.

Homophonic ciphers

Letters	Homophones
A	17 19 34 41 56 60 67 83
I	08 22 53 65 88 90
L	03 44 76
N	02 09 15 27 32 40 59
O	01 11 23 28 42 54 70 80
P	33 91
T	05 10 20 29 45 58 64 78 99

One possible encipherment of the message

M= P L A I N P I L O T is

C= 91 44 56 65 59 33 08 76 28 78

Homophonic ciphers

The first known Western use of homophonic cipher appears in correspondence between the Duchy of Mantua and Simeone de Crema in 1401. Multiple substitutions were assigned only to vowels.

Homophonic ciphers can be much more difficult to break than simple substitution ciphers, especially when the number of homophones assigned to a letter is proportional to the relative frequency of the letter. The relative frequency distribution of the ciphertext symbols will be nearly flat. Other statistical properties may be used to break the cipher (e.g., diagram distributions).

The more homophones available, the stronger the cipher. If each ciphertext symbol appears at most once in the ciphertext, the cipher is unbreakable.

Beale ciphers

Thomas Jefferson Beale left 3 ciphers ($B1$, $B2$ and $B3$) about the treasure he buried in Virginia around 1820. The second cipher was broken by James Ward in 1880 and it describes the treasure and says that the first cipher contains directions to the location where the treasure was buried.

The second cipher $B2$ is a homophonic substitution cipher which uses as a key the Declaration of Independence, where the words are consecutively numbered. Each letter in the plaintext is enciphered with a number of some word starting with that letter. For example, letter W was enciphered with the numbers $1, 19, 40, 66, 72, 290$ and 459.

Beale ciphers

The first 107 words of the Declaration of Independence

(1) When, in the course of human events, it becomes necessary
(11) for one people to dissolve the political bands which have
(21) connected them with another, and to assume among the Powers
(31) of the earth the separate and equal station to which
(41) the Laws of Nature and of Nature's God entitle them,
(51) a decent respect to the opinions of mankind requires that
(61) they should declare the causes which impel them to the
(71) separation. We hold these truths to be self -evident; that
(81) all men are created equal, that they are endowed by
(91) their Creator with certain unalienable rights; that among
(99) these are Life, Liberty, and the pursuit of Happiness.

Beale ciphers

The second cipher starts with 115 73 24 818 37 52 49 17 31 62 657 22 7 15 … which deciphers to “I have deposited…”

So far, no one has solved the first cipher. Many believe that it is a hoax. It contains 495 numbers from 1 to 2906, and DOI only has 1322 words. However, if $B1$ is deciphered using DOI, a strange sequence appears in the middle of the plaintext:

$$ABFDEFGHIIJKLMMNOHPP$$

There are 23 ‘errors’ of the kind: the first F in the above sequence is encrypted as 195 and word 194 begins with a C; similarly, the last H is encrypted as 301 and word 302 begins with O.

Higher-order homophonics

Recall that, given enough ciphertext, most ciphers are theoretically breakable because there is a single key that deciphers the ciphertext into meaningful plaintext; all other keys produce meaningless sequence of letters.

It is possible to construct higher-order homophonic ciphers where each ciphertext deciphers into more that one meaningful plaintext using different keys. For example, the same ciphertext could decipher into the following 2 different plaintexts using different keys:

THE TREASURE IS BURIED IN GOOSE CREEK
THE BEALE CIPHERS ARE A GIGANTIC HOAX

Higher-order homophonics

To construct a second-order homophonic cipher (meaning that for each plaintext there are two possible meaningful plaintexts), arrange the numbers 1 through n^2 into an $n \times n$ matrix K whose rows and columns correspond to the characters of the plaintext alphabet.

For each plaintext character a, row a of K defines one set of homophones $f_1(a)$, while column a defines another set of homophones $f_2(a)$.

A plaintext message $M = m_1 m_2 \dots$ is enciphered along with a dummy message $X = x_1 x_2 \dots$ to get ciphertext $C = c_1 c_2 \dots$, where $c_i = \mathrm{K}(m_i, x_i), i = 1,2, \dots$ That is, c_i is in row m_i and column x_i.

Higher-order homophonics

Example 11**.** Let n=5. The following is 5×5 matrix for the plaintext alphabet {E, I, L, M, S}.

	E	**I**	**L**	**M**	**S**
E	10	22	18	02	11
I	12	01	25	05	20
L	19	06	23	13	07
M	03	16	08	24	15
S	17	09	21	14	04

M = S M I L E

X = L I M E S

C = 21 16 05 19 11

Next Week

1. Polyalphabetic substitution ciphers
   a) Vigenere cipher
   b) Beaufort Cipher
   c) Variant Beaufort Cipher
2. Breaking periodic polyalphabetic ciphers
   a) Index of Coincidence
   b) Kasiski Method
3. Running Key ciphers, Rotor Machines and One-Time Pads
4. Polygram Substitution Ciphers - Playfair Ciphers

Chapter 3 textbook “Classical Encryption Techniques”

References

1. W. Stallings. “Cryptography and Network Security”, Global edition, Pearson Education Australia, 2016.

2. D. Denning. “Cryptography and Data Security”, Addison Wesley, 1982.

3. Bruce Schneier. “Secrecy, Security, and Obscurity”, Crypto-Gram Newsletter, May 15, 2002, http://www.schneier.com/crypto-gram-0205.html#1 last accessed on March 2014.