Breaking Ciphers

Data and Information Management: ELEN 3015

School of Electrical and Information Engineering,
University of the Witwatersrand

1. Cryptanalysis of ciphers

1.1. Monoalphabetic ciphers

- Common 2 and 3 letter words and frequencies
- Two letter combinations that start and end words
- Frequent letters in the language (ETAOIN)
- Letter frequency tables

Use frequency distribution of characters

1.2. Polyalphabetic substitution

Frequency distribution is flattened, thus frequency analysis alone won't work

Approach:

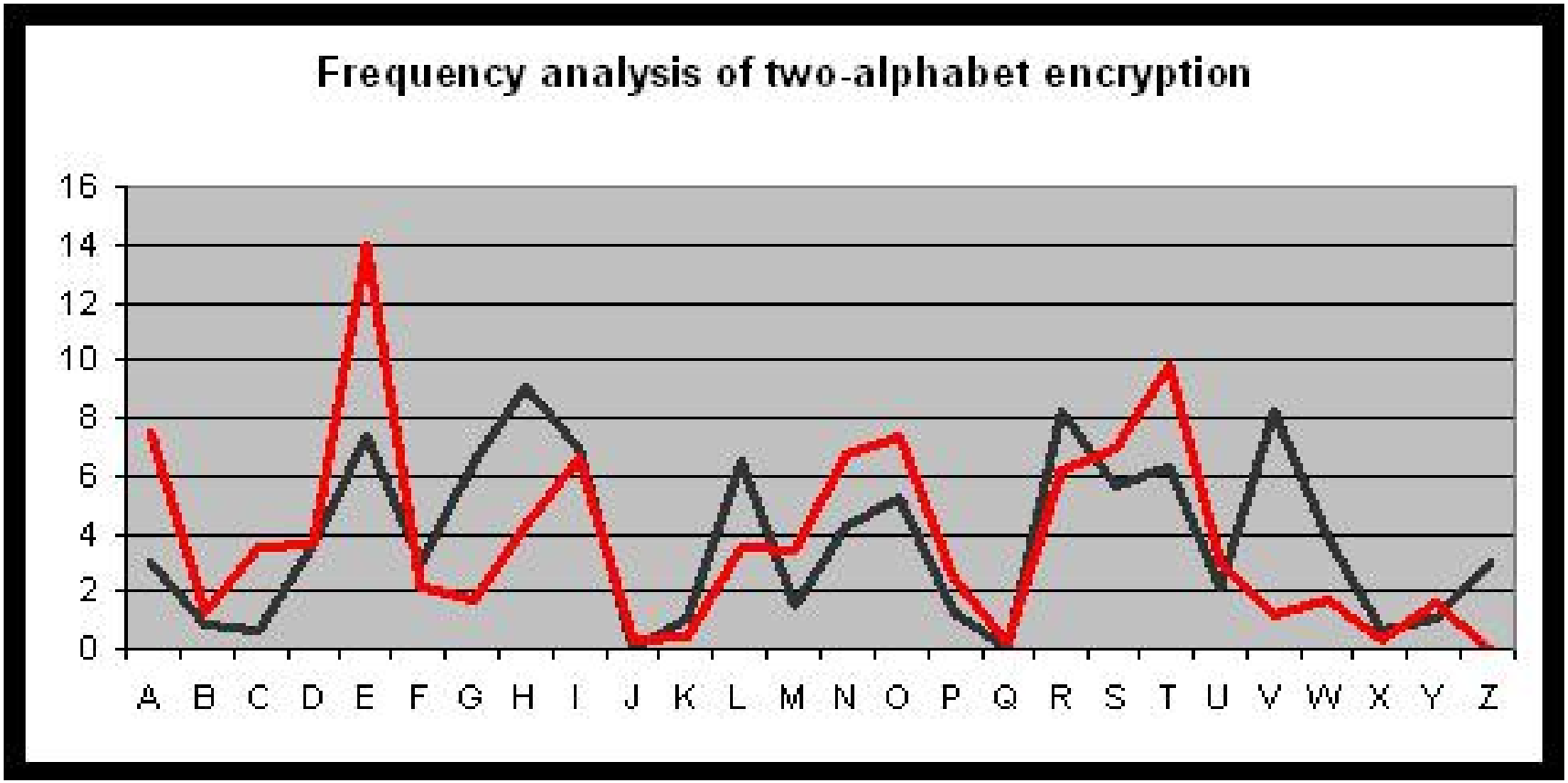
I Determine number of alphabets used

II Isolate ciphertext associated with each alphabet

Two methods:

- Method of Kasiski
- Index of Coincidence

Polyalphabetic substitution



Method of Kasiski

Theorem of Kasiski

If a message is encoded with n alphabets in cyclic rotation, and if a particular word or group of letters occurs k times in a plaintext message, it should be encoded approximately k/n times from the same (i.e. particular) alphabet

- Identify repeated pattern (3 or more characters) in ciphertext
- Note starting position of each occurrence of pattern
- Compute the differences between successive starting points
- Determine factors of each difference
- Use likely key length (factor) to separate into monoalphabetic ciphers
- Solve each monoalphabetic cipher separately and recombine text.

Method of Kasiski

Course notes - p.15

d	i	c	k	e	n	$\ldots$
I	T	W	A	S	T	$\ldots$
l	b	y	k	w	g	$\ldots$

Method of Kasiski

Starting Position	Distance from previous	Factors
20	-	-
83	63	3, 7, 9, 21, 63
104	21	3, 7, 21

Index of coincidence

Index of coincidence

Is a measure of how much variance there is in the frequency distribution for a sample text. (Course notes)

$$IC = \sum_{i=0}^{25} \frac{n_i(n_i - 1)}{N(N-1)}$$

$n_i \mid i = 0$ Number of appearances of 'a' in the text, eg. $n_0 = 2$, $i = 1 \rightarrow$ 'b'

N Total number of characters

Index of coincidence

For a perfectly flat distribution: IC $= 0.0384615$

Proof this! Hint: $n_i = N/26$

Index of coincidence

The value of IC differ for the number of alphabets used (Range from 0.068 - 0.0384)

Alphabets	1	2	3	4	Large
IC	0.068	0.052	0.047	0.044	0.038

IC is a good measure of the alphabets used when the number is small (Less than 5)

Used in conjunction with the method of Kasiski to verify the correct key length.

Creating an unbreakable cipher

Monoalphabetic and polyalphabetic ciphers are easily broken

Cracking methods all rely on some type of repetition in key

Kasiski only works if the same plaintext phrase is encrypted the same way twice

Solution?

Creating an unbreakable cipher

Monoalphabetic and polyalphabetic ciphers are easily broken

Cracking methods all rely on some type of repetition in key

Kasiski only works if the same plaintext phrase is encrypted the same way twice

Solution?
Ensure that no phrase ever gets encrypted the same way more than once by eliminating key rotation

Perfect ciphers

Known as One-Time Pad

- Use non-repeating key (series of random letters)
- Random sequence must never be used again (otherwise pattern is formed)

Vernam Cypher is based on the One-Time Pad

Random Numbers

Commonly used: telephone directory and portions of the entry's phone numbers

Computer random numbers are pseudo-random

- Large repeating pattern
- Pattern depends on the seed number given

Cracking Transpositions

Rely on regularity of English digrams and trigrams

Digrams	Trigrams
EN	ENT
RE	ION
ER	AND
NT	ING
TH	IVE
ON	TIO
IN	FOR
TF	OUR
AN	THI
OR	ONE

Cracking Transpositions

Need to determine the width of the table $\rightarrow$ distance between adjacent characters in the ciphertext

Use sliding window and digram analysis.

Cracking Transpositions

Digram analysis

1. Sequence of text is removed from the ciphertext, compared with ciphertext further on
2. Length of the sequence $<$ anticipated column width
3. Slide the text one character forward each time, calculating the mean and variance of all the letter pairs recorded.
   - The mean indicates how likely the digrams are
   - Variance indicated how likely all the digrams are
4. High mean and low variance is probably correct column width.
5. Check decrypted text. If not correct, try next highest mean and lowest variance which is reasonable.

Cracking Transpositions

tn	n	n	n	n	n
si	ti	i	i	i	i
sw	sw	tw	w	w	w
oh	sh	sh	th	h	h
ha	oa	sa	sa	ta	a
oa	ha	oa	sa	sa	ta
as	os	hs	os	ss	ss
o	ao	oo	ho	oo	so
l	l	al	ol	hl	ol
r	r	r	ar	or	hr
s	s	s	s	as	os
t	t	t	t	t	at
o	o	o	o	o	o

Attacks on Basic Ciphers

Chosen Plaintext attacks

Plaintext:

A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B
C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C

Caesar:

d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d	d
e	e	e	e	e	e	e	e	e	e	e	e	e	e	e	e	e
f	f	f	f	f	f	f	f	f	f	f	f	f	f	f	f	f

Attacks on Basic Ciphers

Chosen Plaintext attacks

Plaintext:

A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B
C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C

Polyalphabetic - Vigenere

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

Attacks on Basic Ciphers

Chosen Plaintext attacks

Plaintext:

A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B
C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C

Polyalphabetic - Key = 'dickens'

d	i	c	k	e	n	s	d	i	c	k	e	n	s	d	i	c
l	f	o	t	e	j	d	l	f	o	t	e	j	d	l	f	o
u	f	k	e	m	g	p	u	f	k	e	m	g	p	u	f	k

Attacks on Basic Ciphers

Chosen Plaintext attacks

Plaintext:

A A A A A A A A A A A A A A A A A
A A A A A A A A A A A A A A A A A
A A A A A A A A A A A A A A A A A

Ciphertext - Single Transposition Cipher

A A A A A A A A A A A A A A A A A
A A A A A A A A A A A A A A A A A
A A A A A A A A A A A A A A A A A

Attacks on Basic Ciphers

Chosen Plaintext attacks

Plaintext:

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	B
C	D	E	F	G	H	I	J	K	L	M	N	O	P

Ciphertext - Single Transposition Cipher

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

Attacks on Basic Ciphers

Adaptive Chosen Plaintext attacks

Plaintext:

X A A A A A A A A A A A A A A A A
X A A A A A A A A A A A A A A A A

Ciphertext - Single Transposition Cipher

X A A A A A A A A A A A A A A A A
X A A A A A A A A A A A A A A A A