University of South Wales

MS3S05 Cryptography
Part 1
2019/2020

Stephanie Perkins

October 14, 2019

University of South Wales

Contents

University of South Wales

Chapter 1

Introduction

Cryptography (or cryptology) refers to the art and science of designing systems in order to disguise the true meaning of a message to all but the intended recipient. Some authors consider cryptography to be only about designing cipher systems, and cryptanalysis to be about breaking these systems. We will discuss both aspects.

1.1 Alice, Bob and Eve

Alice and Bob wish to communicate over an insecure channel, such as the internet or a mobile phone. An eavesdropper, Eve, is able to see the whole communication and to inject her own messages in the channel.

Alice and Bob hence want to find a way to encode their communication so as to achieve:

- Privacy: Eve should have no information about the content of the messages between Alice and Bob;
- Authentication: Eve should not be able to impersonate Alice or Bob.

If Bob receives a message from Alice, he should be sure of the identity of the sender. If Alice receives a message from Bob she also must have the same assurance.
For this to be possible, Alice and Bob must have some secret information that Eve does not have otherwise Eve could simply run the same algorithms that Alice does, and would then be able to read the messages received by Alice over the internet or phone. She would also be able to communicate with Bob by impersonating Alice.

In the classical symmetric-key cryptography setting, Alice and Bob have agreed before on a secret key, which they use to encode and decode messages, to produce authentication information and to verify the validity of the authentication information.

In the public-key setting, Alice has a private key known only to her, and a public key known to everybody, including Eve; Bob too has his own private key and a public key known to everybody. In this setting, private and authenticated communication is possible without Alice and Bob having to agree on a shared secret key.

University of South Wales

1.2 Terminology

The message (the sequence of letters or symbols) we want to transmit is called the **plaintext**. For clarity only, I will often represent the plaintext by lowercase letters. The enciphered text (the sequence of letters or symbols that are actually transmitted) is called the **ciphertext** and is normally for clarity represented by uppercase letters. The process of transforming plaintext to ciphertext is called *enciphering* (or encrypting) and the reverse procedure ciphertext to plaintext is called **deciphering** (or decrypting).

1.3 Early History of Cryptography

Cryptography has existed for thousands of years in some form. Historical cryptography methods were symmetric-key based.

1.3.1 Early Transposition Cipher - scytale

The first transposition cipher recorded is that of a scytale (pronounced skitali). It was used by Spartan generals in the 5th century BC. The sender (Alice) and recipient (Bob) each had a cylindrical tube called a scytale of exactly the same radius. The sender would wrap a narrow ribbon of parchment around the scytale then write on it lengthways. The parchment was then unwound. It could then only be read when the parchment was again wound around a scytale of the same circumference (hopefully by the intended recipient!).

Example Encrypt "'This is the first Cryptography class of the year" using a scytale with length 7.

$$
\begin{array}{ccccccc}
T & H & I & S & I & S & T \\
H & E & F & I & R & S & T \\
C & R & Y & P & T & O & G \\
R & A & P & H & Y & C & L \\
A & S & S & O & F & T & H \\
E & Y & E & A & R &
\end{array}
$$

Encrytion is THCRAEHERASYIFYPSESIPHOAIRTYFRSSOCTTTGLH ← ciphertext

University of South Wales

1.3.2 Cryptanalysis

Suppose the following has been intercepted

FYRGYSNNIEYRCEBNAPALSEARTPACFLCOHSAU

How should Eve break the cipher?

The scytale used by Alice has a circumference which can be measured by the number of letters, Bob would also have such a scytale, but Eve does not know what size scytale to use.

36 letters 6×6?

Try columns of 6

$F \quad N \quad C$
$Y \quad N$
$R \quad I$
$E \quad E$
$Y \quad Y$
$S \quad R$

Doesn't look hopeful

Try 5 across so $36/5 = 7r1$

So 8 in first column, 7 in others.

$F \quad I \quad N \quad A \quad L$
$Y \quad E \quad A \quad R \quad C$
$R \quad Y \quad P \quad T \quad O$
$G \quad R \quad A \quad P \quad H$
$Y \quad C \quad L \quad A \quad S$
$S \quad E \quad S \quad C \quad A$
$N \quad B \quad E \quad F \quad U$
N

Final year cryptography classes can be fun

The scytale is perhaps the first documented cipher used for military use and was very secure for its time. It can be thought of as having many possible "keys", since each circumference is a key. After the demise of the Greek culture, comparably secure ciphers were not used until medieval times.

1.3.3 Early Substitution Cipher - Atbash

The Atbash cipher is a simple substitution cipher designed for the Hebrew alphabet. It consists of substituting the first letter for the last, the second for the last minus one and so on, *i.e.* the alphabet

is reversed. This is a very ancient cipher that appears to have been used in the Bible in the Old Testament book of Jeremiah (25:26 and 51:41). Unfortunately it can't be seen when looked at in the English Language. Medieval Monks were intrigued by obvious examples of cryptography in the Bible and it sparked a serious study of cryptographic methods. It does not appear that the cipher was used to conceal information in this context rather to add mystery.

The Atbash cipher for the English alphabet is:

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

Example Encrypt hello

HELLO $\rightarrow$ SVOOL

Cryptanalysis

This is obvious if it known that the cipher is Atbash.

Example Break the following cipher

GSRHRHMLGEVIBHVXIVGZGZOO

THISISNOTVERYSECRETATALL

The Atbash is a very weak cipher because it only has one possible key, and it is a simple monoalphabetic substitution cipher. However, this may not have been an issue in the cipher's time.

1.3.4 Early Substitution Cipher - Caesar Cipher

The Roman emperor Julius Caesar used a cipher in which each letter of the plaintext was replaced by the letter three places further along in the alphabet. The alphabets are

University of South Wales

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	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

Note that the cipher wraps around at the end. To use this cipher, the plaintext would be written in one row and underneath the ciphertext obtained by replacing the plain letters by the corresponding cipher letters

Plaintext	m	e	e	t		m	e		a	t		m	i	d	n	i	g	h	t
Ciphertext	P	H	H	W		P	H		D	W		P	L	G	Q	L	J	K	W

Deciphering is just the reverse process.

There is nothing special about the shift of three letters, any number is acceptable. This number is the key and must be kept secret to both sender and receiver.

Cryptanalysis

Break the following cipher:

MTBVZNHPQDINIDTZGWJFPYMNX

Shift 5
How quickly did you break this

A Caesar cipher is one of the simplest substitution encryption methods as it involves replacing each letter of the secret message with a different letter of the alphabet which is a fixed number of positions further in the alphabet.

Because each letter in the message has a direct translation to another letter, statistical analysis can be used to find the message. For example, the letter E is the most commonly used letter in the English language. Thus, if the most common letter in the ciphertext is X, it is likely that X represents E. Additionally, common bigrams and trigrams also give clues. A brute-force approach of trying all 25 possible combinations would also work to find the message.

Although nowadays this cipher appears very easy to break, in the time it was used, most people were illiterate and would not have known anything about cryptography.

1.3.5 General Simple (Mono-Alphabetic) Substitution Cipher

General Mono-alphabetic substitution ciphers involve replacing each letter of the plaintext with another letter. The substitution is fixed for each letter of the alphabet.

Example Break the following cipher, given that it is a mono-alphabetic substitution cipher

University of South Wales

V	T		Q	S	T		I	G	F	O	H	U		Z	G		R	S	G	F						T		21
																											Q		18
																											Z		16
N	G	X	S		Y	G	G	R		Q	H	R		E	K	G	Z	I	O	H	U		G	H			O		15
																											G		15
																											H		12
J	G	H	R	Q	N		W	X	Z		O	Z		V	O	K	K		W	T		Q					K		11
																											S		11
																											R		10
R	O	Y	Y	O	E	X	K	Z		G	F	T	S	Q	Z	O	G	H		Q	A						I		10
																											E		7
																											F		7
T	H	T	J	N		Q	E	Z	O	C	O	Z	N		O	H		Z	I	T							A		7
																											Y		5
																											X		5
Q	S	T	Q		I	Q	A		O	H	E	S	T	Q	A	T	R		O	H							V		5
																											N		5
																											W		4
Z	I	T		K	Q	A	Z		V	T	T	L		O	Z		A	I	G	X	K	R					U		3
																											L		3
																											J		2
W	T		Q		R	Q	S	L		H	O	U	I	Z		A	G		V	Q	Z	E	I				C		1
																											P		0
																											D		0
E	Q	S	T	Y	X	K	K	N		Y	G	S		Z	I	T		F	K	Q	H	T					B		0
																											M		0
Z	I	S	T	T		F	Q	E	L	A		V	O	K	K		W	T											
R	S	G	F	F	T	R																							

Cryptanalysis

The number of keys for such a substitution cipher based on the English Language of 26 letters is $26!$ making exhaustive search generally impractical without any further information. However, having such a large number of keys introduces its own problems for key management and in practice there would be no requirement to carry out such a large exhaustive search.

1.3.6 Super Encryption

The ciphers discussed so far are in practice easy to break. When two or more weak ciphers are combined it is possible to produce one that is much stronger than each of the original ciphers. For example, a cipher first encrypted by a substitution cipher and then encrypted by a transposition cipher will be harder to break than either of the individual ciphers.
Super encryption is an important technique used in modern ciphers.

University of South Wales

1.4 Tutorial 1 - Historical Ciphers

1.) You have received the following Ciphertexts. They are all encrypted using the historical ciphers covered so far. Find the plaintext.

(a) IVACNLEYAISODADMNEFCTISGSFHEATPEEOITALNECMEREDAOML

Scytale - I am starving please send coffee and chocolate immediately

(b) MABLPBEEUXXTLRHGVXRHNYBGWMAXDXR

Caesar - shift 7 This will be easy once you find the key

(c) ZMVZHBXRKSVIGLYIVZPWLBLFZTIVV

Atbash - an easy cipher to break do you agree

2.) Many modern day ciphers are constructed using methods similar to those discussed, where transposition and substitution are applied together or in more than one way. Design a cipher based on the simple transposition and substitution ciphers discussed. Once you have encrypted a message, give to another student to attempt the cryptanalysis of it. You may only use the types of methods discussed in this chapter.

3.) Write down appropriate mathematical statements to describe the enciphering and deciphering mechanisms for

University of South Wales

(a) Caesar Cipher

(b) Atbash Cipher

(c) Scytale Cipher

Chapter 2

The Vigenére Cipher: A Polyalphabetic Substitution Cipher

In the Caesar cipher and other monoalphabetic substitution ciphers, each letter in the cipher alphabet always represents the same letter in the plain alphabet, making them relatively easy to break. Polyalphabetic ciphers use several different cipher alphabets.

The Vigenére Cipher is a polyalphabetic substitution cipher. The method was originally described by Giovan Battista Bellaso in 1553 however, the scheme was later misattributed to Blaisede Vigenére in the 19th century, and is now widely known as the Vigenére cipher. It consists of all 26 Caesar ciphers. The cipher was broken by Babbage and Kasiski in the middle of the 19th century.
Enciphering is carried out with the aid of a table of alphabets often referred to as a Vigenére square. To encode a message, a keyword is chosen and the letters in the keyword determine which rows of the Vigenére square are used to encode each letter of the message.

	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 column labels represent the plaintext letters and the row labels represent the letters found in the keywords and the letters within the square represent the ciphertext letters.

University of South Wales

2.1 Enciphering and Deciphering

The sender and receiver must first agree on a secret key comprised of a sequence of letters. The method is demonstrated by example.

Suppose the plaintext "'agents to be recalled"' is to be enciphered using keyword "'FAMILY"'. The keyword, written as many times as necessary, is written above the plaintext. The first plaintext letter "'a"' is enciphered using keyword letter "'F"' using the Vigenére square by looking at the intersection of column "'a"' with row "'F"' to give ciphertext letter "'F"'. Similarly the second plaintext letter "'g"' is enciphered using keyword letter "'A"' to give ciphertext letter "'G"'.

F	A	M	I	L	Y	F	A	M	I	L	Y	F	A	M	I	L	Y
a	g	e	n	t	s	t	o	b	e	r	e	c	a	l	l	e	d
F	G	Q	V	E	Q	Y	O	N	M	C	C	H	A	X	T	P	B

To decipher a ciphertext, the process is applied in reverse. Ciphertext "'Q"' with keyword letter "'M"' deciphers to plaintext letter "'e"', This can be seen by scanning along row "'M" until ciphertext letter "'Q"' is found and the corresponding column label "'e" is the required plaintext letter.

University of South Wales

2.1.1 Tutorial

1.) Encrypt the phrase “i have reached the building, shall I enter " using the Vigenére cipher with keyword “HEIST".

2.) Encipher the following message using the Vigenére cipher and the keyword “ISH": “there is a secret passage behind the picture frame"

3.) Decrypt “TAX AVDAVD AV NAPG MWCT UX SWMCXLSHEL" given that the keyword is ATTACK.

2.1.2 Tutorial Solutions

1.) PLINXYIIUALHBZXIYQDWPROKAHPTAXUXMJ

2.) BZLZWPASZMUYMLWIKLIJWOQFFBZLXAJBMYMXYIEL

3.) the attack at dawn must be successful

2.2 Introduction to Cryptanalysis of the Vigenére cipher

The Vigenére cipher effectively conceals the frequency of letters. A letter count would show that there is very little difference between the number of times each letter occurs in the ciphertext. The Vigenére cipher also has an enormous number of keys. The sender and receiver can agree on any word in a dictionary, any combination of words or even invent words. A cryptanalyst cannot crack the cipher by searching all possible keys because the number of options is simply too great.

The resistance of the Vigenére cipher to frequency analysis initially earned it a reputation as an "unbreakable" cipher and indeed it would be if a random key was chosen that was the same length as the plaintext message, and the key was only ever used once and impossible to obtain by any third party. In practice this is not the case. If a key of length one is chosen then the cipher reverts to the Caesar cipher (additive or shift cipher), which we have seen is trivial to break using basic frequency analysis. If the key is longer but short compared to the size of the plaintext then during the enciphering of the plaintext the key is repeated, and exploiting the knowledge gained from this allows cryptanalysis. If the key is very long compared the plaintext (and non-repeating) then the Vigenére cipher is virtually impossible to break as the repeating key produces patterns that are critical in breaking the cipher. In order to break a Vigenére cipher it is sufficient to determine the keyword. The Vigenére cipher has been shown to be vulnerable to a number of attacks including the known plaintext attack, and those based on determining the keyword length using the **Kasiski** test and the **Friedman** test.

2.2.1 Known Plaintext Attack

If Eve possesses both the plaintext and the ciphertext she can easily determine the key.

2.2.2 Known Key Length Attack

If the length of the keyword is known then the the Vigenére cipher can be broken as multiple Caesar ciphers.

The Kasiski test involves looking for strings of characters that are repeated in the ciphertext. The strings should be as long as possible and preferably at least three characters in length for the test to be successful. The distances between consecutive occurrences of the strings are likely to be multiples

of the length of the keyword. Finding the greatest common divisor of all the distances between the repeated strings gives a strong possibility for the length of the keyword. The Friedman test consists of two equations which indicate the type of cipher used (monoalphabetic or polyalphabetic) and estimates the key length assuming a polyalphabetic cipher with a periodic key. The results of Kasiski and Friedman tests can be combined to deduce the most likely key length L. Once this length is known, the ciphertext message can be partitioned into key-length chunks of text, which are then individually susceptible to frequency analysis. A simple way is to arrange the ciphertext into L columns, with one column per letter in the key. Each column is then just a shift cipher (Caesar cipher) and so frequency analysis can be used to determine the shift (key) used. These key letters are combined to obtain the key that was used for encipherment. Once the key is known, the ciphertext can easily be deciphered.

2.2.3 Kasiski Test

The Kasiski Test can be performed in Maple using the package "stringtools" and the command "Kasiski".

Example 2.1. *The following Ciphertext has been intercepted. "IIBPM ZVVIS EDZVH ETGDI LPMCK CDFWY JEIEH CDEAW WGKLN ZUPWV XPNBZ APDSA RVTOT MRTEC OYTLN ZUZLX ISELP SJNAK VGAZS CILDA ILXCD FWMLP SRITV QTLCF RRRPN YECLN YVZTC YBRCX ESJZO TXYGG SSDSR LGZZN PVTAL YPVTA LY"*
Kasiski Analysis:

ciphertext string	*occurs at (index in string)*	*spacing (number of symbols)*	*Divisors*
ISE	*8 80*	*72*	*2 3 4 6 8 9 12 18 24 36 72*
CDFW	*25 103*	*78*	*2 3 6 13 26 39 78*
DFW	*26 104*	*78*	*2 3 6 13 26 39 78*
LNZU	*43 73*	*30*	*2 3 5 6 10 15 30*
NZU	*44 74*	*30*	*2 3 5 6 10 15 30*
LPS	*83 108*	*25*	*5 25*
PVTAL	*161 167*	*6*	*2 3 6*
VTAL	*162 168*	*6*	*2 3 6*
TAL	*163 169*	*6*	*2 3 6*

The greatest common divisor of the majority of these numbers is 6, giving a suspected key length of 6. After finding the keylength, we need to find the frequency of letters in the ciphertext to determine the actual keyword. A frequency analysis can be carried out to determine the actual keyword.

Consider just the letters that have been enciphered by the first letter of the keyword.

***I**IBPMZ **V**VISED **Z**VHETG **D**ILPMC **K**CDFWY **J**EIEHC **D**EAWWG **K**LNZUP **W**VXPNB **Z**APDSA*
*R**R**VTOTM **R**TECOY **T**LNZUZ **L**XISEL **P**SJNAK **V**GAZSC **I**LDAIL **X**CDFWM **L**PSRIT **V**QTLCF*
*R**R**RPNY **E**ECLNY **V**ZTCYB **R**CXESJ **Z**OTXYG **G**SSDSR **L**GZZNP **V**TALYP **V**TALY*

Notice V occurs almost 21% of the time in the highlighted letters so is most probably the plaintext letter e giving first keyword letter R. Decrypt all using this keyletter.

r e i m t s m t
***I** IBPMZ**V**VISEDZVHETG**D**ILPMC**K**CDFWY**J**EIEHC**D**EAWWG**K**LNZUP*
f i a a c u y e
***W**VXPNB**Z**APDSA**R**VTOTM**R**TECOY**T**LNZUZ**L**XISEL**P**SJNAK**V**GAZSC*
r g u e a n e a
***I** LDAIL**X**CDFWM**L**PSRIT**V**QTLCF**R**RRPNY**E**ECLNY**V**ZTCYB**R**CXESJ*
i p u e e
***Z**OTXYG**G**SSDSR**L**GZZNP**V**TALYP**V**TALY*

Continue similarly:

r e i m t s m t
*I **I**BPMZV**V**ISEDZ**V**HETGD**I**LPMCK**C**DFWYJ**E**IEHCD**E**AWWGK**L**NZUP*
f i a a c u y e
*W**V**XPNBZ**A**PDSAR**V**TOTMR**T**ECOYT**L**NZUZL**X**ISELP**S**JNAKV**G**AZSC*
r g u e a n e a
*I **L**DAIL**X**CDFWML**P**SRIT**V**QTLCF**R**RRPNYE**E**CLNYV**Z**TCYBR**C**XESJ*
i p u e e
*Z**O**TXYGG**S**SDSRL**G**ZZNPV**T**ALYPV**T**ALY*

Again the most common letter is V but if this is deciphered as plaintext e it gives second keyword as letter R and the following decryption which does not look plausible.
Keep trying!!

2.2.4 Friedman test (also known as Kappa test) - Index of Coincidence

Friedman (1925): "If one selects a pair of letters from a text what is the probability that the letters are equal?"

Index of Coincidence

Consider an arbitrary sequence of letters of length n. Let n_1 denote the number of a's, n_2 the numbers of b's ... n_{26} the number of z's.

What is the probability that a randomly selected pair of letters are both a's (not assuming they are consecutive)?

Choosing first letter to be a - n_1 possibilities.
Choosing second letter to be a - $n_1 - 1$ possibilities.

Since the order the letters in the pair are chosen is unimportant then the number of pairs of a's is

$$\frac{n_1(n_1-1)}{2}$$

Similarly for b's:

$$\frac{n_2(n_2-1)}{2}$$

Therefore the number of pairs that consist of equal letters (i.e. both are a's, or b's ..., or z's) is

$$\frac{n_1(n_1-1)}{2}+\frac{n_2(n_2-1)}{2}+\ldots+\frac{n_{26}(n_{26}-1)}{2}=\sum_{i=1}^{26}\frac{n_i(n_i-1)}{2}$$

So the probability of choosing at random a pair of equal letters is

$$\sum_{i=1}^{26}\frac{n_i(n_i-1)}{2}\div\frac{n(n-1)}{2}=\sum_{i=1}^{26}\frac{n_i(n_i-1)}{n(n-1)}$$

Where $\frac{n(n-1)}{2}$ is the number of ways of choosing any pair of letters in the sequence.
therefore:

$$I=\sum_{i=1}^{26}\frac{n_i(n_i-1)}{n(n-1)}.$$

2.2.5 Second Way of Finding I

In "natural" language every letter occurs with a characteristic probability p_α. Assume the letter a occurs with probability p_1, b with p_2, ..., z with p_{26}. The probability that a is in the first position we choose is also p_1, and the probability that it is in the second position is p_1. (Independence can be assumed for large bodies of text). Therefore the probability that a is in the first and second position chosen is p_1^2.

Therefore the probability that two arbitrary chosen positions contain the same letters (either both a, or both b, etc) is

$$\sum_{i=1}^{26} p_i^2$$

Observation 2.2. *For the English language*

$$\sum_{i=1}^{26} p_i^2 \approx 0.0686$$

(Therefore if pairs of letters are randomly chosen from English text, approximately 6.9% of the time the letters will be the same.)

Observation 2.3. *Imagine a purely random text, so that each letter occurs with equal probability of* $\frac{1}{26}$ *then*

$$\sum_{i=1}^{26} \left(\frac{1}{26}\right)^2 \approx 0.0385$$

which is about half that seen for standard English text.

Observation 2.4. *If the probabilities* $p_1, p_2, \ldots, p_{26}$ *are known (as for the English language) then the sum of squares is approximately equal to the Index of Coincidence, then*

$$I \approx \sum_{i=1}^{26} p_i^2.$$

Therefore $I \approx 0.0385$ is the minimum value for the Index of Coincidence, since the probabilities can never be more evenly distributed than in random text. When the text becomes irregular (i.e there are patterns as in standard English text)the value of I becomes larger.

University of South Wales

2.3 Cryptanalysis of the Vigenére cipher

In mono-alphabetic ciphers (*e.g.* Caesar) the index of coincidence is invariant under the cipher (i.e. it doesn't change) as the probability distribution is the same in the plaintext and in the ciphertext. Therefore

- If $I \approx 0.0686$ then it is likely that the ciphertext observed is from a monoalphabetic cipher.
- If $I << 0.0686$ then it is likely that the ciphertext observed is from a polyalphabetic cipher.

We use the index of coincidence to try and obtain the length of the keyword.

2.3.1 How much smaller than 0.0686 is small?

This depends on the length of the keyword. Large texts with large keywords will have small index of coincidence (tending to ≈ 0.0385), However small texts with small key words will not be sufficiently disguised and so the index of coincidence will not be very small.

Let L denote the length of the keyword. For simplicity assume all letters in the keyword are distinct. Imagine that the ciphertext has been arranged with its letters in L columns.

In the first column there appears letters in position $1, L+1, 2L+1, \ldots$ these letters will all have been enciphered using the first letter of the keyword.

Letter S_i of keyword	S_1	S_2	$\ldots$	S_L
	1	2		L
	$L+1$	$L+2$		$2L$
	$2L+1$	$2L+2$		$3L$
	$3L+1$	$3L+2$		$4L$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$

From this we determine I.

Observation 2.5. *Each column represents a monoalphabetic (Caesar) cipher. Therefore the probability that a randomly chosen pair of letters in the same column consists of equal letters is* ≈ 0.0686. *Consider randomly chosen letters from two different columns, the probability is much smaller so can assume they are equal to 0.0385.*

Suppose the number of pairs of letters that are in the same column are counted, similarly the number of pairs in different columns are counted. Let n denote the number of letters in the ciphertext.

Then each column has $\frac{n}{L}$ letters (if the text is very large then rounding errors can be ignored). There are precisely n possibilities for a randomly chosen letter. The chosen letter uniquely defines its column. In the column there are $\frac{n}{L}-1$ other letters. So therefore there are $\frac{n}{L}-1$ ways of choosing a second letter in the same column. Thus the number of pairs of letters that are in the same column is:

$$\frac{n(\frac{n}{L}-1)}{2} = \frac{n(n-L)}{2L}.$$

Since there are exactly $n-\frac{n}{L}$ letters outside the first column, then the number of paired letters in different columns is:

$$\frac{n(n-\frac{n}{L})}{2} = \frac{n^2(L-1)}{2L}.$$

Combining these observations; the expected number A of pairs of equal letters is

$$A = \frac{n(n-L)}{2L}0.0686 + \frac{n^2(L-1)}{2L}0.0385$$

Therefore the probability that a randomly chosen pair consists of equal letters is:

$$\begin{aligned}\frac{A}{\frac{n(n-1)}{2}} &= \frac{n-L}{L(n-1)}0.0686 + \frac{n(L-1)}{L(n-1)}0.0385 \\ &= \frac{1}{L(n-1)}\left[0.0301n + L(0.0385n - 0.0686)\right]\end{aligned}$$

Since the index of coincidence is approximately equal to the above:

$$I \approx \frac{0.0301n}{L(n-1)} + \frac{0.0385n - 0.0686}{n-1}$$

Rearranging:

$$\begin{aligned}(n-1)I &= \frac{0.0301n}{L} + 0.0385n - 0.0686 \\ (n-1)I - 0.0385n + 0.0686 &= \frac{0.0301n}{L}\end{aligned}$$

$$L = \frac{0.0301n}{(n-1)I - 0.0385n + 0.0686}$$

where

$$I = \sum_{i=1}^{26} \frac{n_i(n_i-1)}{n(n-1)}$$

For the ciphertext example

"IIBPM ZVVIS EDZVH ETGDI LPMCK CDFWY JEIEH CDEAW WGKLN ZUPWV XPNBZ APDSA RVTOT MRTEC OYTLN ZUZLX ISELP SJNAK VGAZS CILDA ILXCD FWMLP SRITV QTLCF RRRPN YECLN YVZTC YBRCX ESJZO TXYGG SSDSR LGZZN PVTAL YPVTA LY"

i	n_i	n_i-1	$n_i(n_i-1)$
A	8	7	56
B	3	2	6
C	10	9	90
D	8	7	56
E	9	8	72
F	3	2	6
G	6	5	30
H	2	1	2

i	n_i	n_i-1	$n_i(n_i-1)$
I	9	8	72
J	3	2	6
K	3	2	6
L	13	12	156
M	4	3	12
N	7	6	42
O	3	2	6
P	10	9	90
Q	1	0	0

i	n_i	n_i-1	$n_i(n_i-1)$
R	8	7	56
S	10	9	90
T	11	10	110
U	2	1	2
V	10	9	90
W	5	4	20
X	5	4	20
Y	8	7	56
Z	11	10	110
	172		1262

$n = 172$ and $\sum_{i=1}^{26} n_i(n_i-1) = 1262$ then

$$I \approx \sum_{i=1}^{26} \frac{n_i(n_i-1)}{n(n-1)} = 0.042907657$$

$$\begin{aligned} L &= \frac{0.0301n}{(n-1)I - 0.0385n + 0.0686} \\ &= \frac{0.0301 \times 172}{171 \times 0.0429 - 0.0385 \times 172 + 0.0686} \\ &= \frac{5.1772}{0.7825} \\ &= 6.6 \end{aligned}$$

Which together with the Kasiski test indicates that the keyword is of length 6 (rather than a multiple of 6).

Note: The Friedman test should be used with caution on small texts and small keywords as the results are unlikely to be accurate due to the sample size.

University of South Wales

2.4 The Vigenére Cipher Tutorial

1.) Consider the following ciphertext:

"WGIXF IRTNX AMWPZ GFCLN BZTEF ROOZN MAOUR TLRNO DSXJW XXDAN ZHDIX NQTTA HOGCM RWRVJ NUMYB GXAVT MGZDT EWLQS WVWTM LGBLK NRINS OZGIF BGNLM FPSQN XHVJA UFGMJ XYXUM HQSXV VZTEA BZRPT LRIJY IVNTO FYWEW UYFSE BEIAW VBIMM IGWHQ CEYTK PPIEN UDMIQ NKMTW BNIDY TGITM LCFQY FHEGP GHEWV VIQBI PWSQL ITMTP AVLZK MDMAO GXMSF BGXLS SOKDM EAGYZ AZNTG ZDHVX RAMEG QIFRE SNOOD YEQTS TLREG DIRPT APIRT BNQFE ZWAEZ MZUKD MEAVZ QLITZ GYTLN BXQVI NTKPG IEUGZ YWCZU BOWRL GXLMN VIQWM GPSQX MPWGS AMAAZ FHIHV OFEHF BGFEW NVJIH SFMJU SGYWY GRIUM RBEHC ZUBEP NUKDI GNQTF OXUMC MR"

Here are the results of the Friedman test:

i	n_i	$n_i - 1$	$n_i(n_i - 1)$
A	19		
B	16		
C	7		
D	13		
E	23		
F	19		
G	29		
H	13		

i	n_i	$n_i - 1$	$n_i(n_i - 1)$
I	28		
J	7		
K	8		
L	16		
M	30		
N	24		
O	14		
P	14		
Q	17		

i	n_i	$n_i - 1$	$n_i(n_i - 1)$
R	18		
S	16		
T	26		
U	14		
V	17		
W	21		
X	18		
Y	14		
Z	21		
	462		

Here are the results of the Kasiski Analysis:

Trigram	Frequency	Positions	Distance(s)
IRT	2	5, 317	312
LNB	2	18, 348	330
ZTE	2	21, 141	120
MAO	2	30, 252	222
TLR	3	35, 149, 305	114, 270, 156
NQT	2	55, 451	396
CMR	2	63, 459	396
BGX	2	74, 260	186
GZD	2	81, 279	198
TML	2	93, 213	120
FBG	3	109, 259, 409	150, 300, 150
PSQ	2	116, 386	270
XUM	2	132, 456	324
RPT	2	147, 312	165
TKP	2	188, 356	168
ITM	2	212, 240	28
VIQ	2	230, 380	150
QLI	2	238, 340	102
LIT	2	239, 341	102
GXL	2	261, 375	114
KDM	2	267, 333	66
DME	2	268, 334	66
MEA	2	269, 335	66
UKD	2	332, 446	114
CZU	2	367, 439	72
ZUB	2	368, 440	72

(a) Apply the Friedman test to the ciphertext to calculate I, what can you deduce from this value?

(b) Given the values of n_i when compared to the frequencies of letters in the English language does this agree with your deduction in (a)?

(c) Apply the Friedman test to the ciphertext to approximate L, and compare with the results of the Kasiski test to determine the length of the key word.

(d) Attempt to decrypt the text.

2.) Crack the following ciphertext. You should perform Kasiski and Friedman to determine the keyword length. Spaces have been left in the correct places.

"RG QELW NRLWTE FATI AUECQWPK STTZED FA JE DFB BN PNBUEC XQQE ZK BUE QNZR IY MQF LZIOVNRX IG BLPME SEWMRT WNNR ID NVSIYNBRLJ XBEAYLME TSFV NNJYPVNR BPVCS YPR MTSL BF XFV POFQL VNGJVG. WP BWHLO SWG

University of South Wales

DLWM GO NTVPETAM GHP YPVNRX EUINM IEE CJIYLJ RMEE NTUZOYUTNCPX WS EINAGEYHM. VF HJ KBUWI NYY ZZB BF EMIG WTSLBW SFVQ IY MIAD STDRR ZAME TSNA TRPFB PIED ORNEQG EEXTDR TSJ ZBOQX IAD AJMC IY FB GHP VCREC YPVNRX EUINM IEE RTQAG ZS BUE DYZNNRJ KBIYHQQEYHMF TSJ XYAYSQAGD YPR CCTAF PFWXBSPX BUE HTVQECKCY CSFQAS ZK MIEYYA JOCPQAG EMZBURM ORNPWIGIZS IAD WJIQIYL BB TSJ UBSE TCGR CJAHLEX QG WZZTQ MLPM NLW KQPTTTV JIEM QGS NTVIEYYQBNLQQGIPX IAD QTZRSPJV POYHTHSTTVF MZXB FTLQM NNO ZVCRZKQGAMQM"

3.) If the keyword of a Vigenére cipher has repeated letters, what problems might this cause when performing the Kasiski test?

4.) Write down a function that enciphers the plaintext $(x_1, \ldots, x_n)$ into the ciphertext$(y_1, \ldots, y_n)$ using a keyword $E_k = k$, where $k = (k_1, \ldots k_n)$ using the Vigenére cipher. Also write down the decipher function.

University of South Wales

2.5 The Vigenére Cipher Tutorial Solutions

i	n_i	$n_i - 1$	$n_i(n_i - 1)$
A	19	18	342
B	16	15	240
C	7	6	42
D	13	12	156
E	23	22	506
F	19	18	342
G	29	28	812
H	13	12	156

i	n_i	$n_i - 1$	$n_i(n_i - 1)$
I	28	27	756
J	7	6	42
K	8	7	56
L	16	15	240
M	30	29	870
N	24	23	552
O	14	13	182
P	14	13	182
Q	17	16	272

i	n_i	$n_i - 1$	$n_i(n_i - 1)$
R	18	17	306
S	16	15	240
T	26	25	650
U	14	13	182
V	17	16	272
W	21	20	420
X	18	17	306
Y	14	13	182
Z	21	20	420
	462		8726

Trigram	Number of Occurrences	Positions	Distance(s)	Divisors
IRT	2	5, 317	312	$1,2,3,4,6,8,12,13,24,26,39,52,78,104,156,312$
LNB	2	18, 348	330	$1,2,3,5,6,10,11,15,22,30,33,55,66,110,165,330$
ZTE	2	21, 141	120	$1,2,3,4,5,6,8,10,12,15,20,24,30,40,60,120$
MAO	2	30, 252	222	$1,2,3,6,37,74,111,222$
TLR	3	35, 149, 305	114, 270, 156	$\{1,2,3,6,19,38,57,114\}$, $\{1,2,3,5,6,9,10,15,18,27,30,45,54,90,135,270\}$, $\{1,2,3,4,6,12,13,26,39,52,78,156\}$
NQT	2	55, 451	396	$1,2,3,4,6,9,11,12,18,22,33,36,44,66,99,132,198,396$
CMR	2	63, 459	396	$1,2,3,4,6,9,11,12,18,22,33,36,44,66,99,132,198,396$
BGX	2	74, 260	186	$1,2,3,6,31,62,93,186$
GZD	2	81, 279	198	$1,2,3,6,9,11,18,22,33,66,99,198$
TML	2	93, 213	120	$1,2,3,4,5,6,8,10,12,15,20,24,30,40,60,120$
FBG	3	109, 259, 409	150, 300, 150	$\{1,2,3,5,6,10,15,25,30,50,75,150\}$, $\{1,2,3,4,5,6,10,12,15,20,25,30,50,60,75,100,150,300\}$
PSQ	2	116, 386	270	$1,2,3,5,6,9,10,15,18,27,30,45,54,90,135,270$
XUM	2	132, 456	324	$1,2,3,4,6,9,12,18,27,36,54,81,108,162,324$
RPT	2	147, 312	165	$1,3,5,11,15,33,55,165$
TKP	2	188, 356	168	$1,2,3,4,6,7,8,12,14,21,24,28,42,56,84,168$
ITM	2	212, 240	28	$1,2,4,7,14,28$
VIQ	2	230, 380	150	$1,2,3,5,6,10,15,25,30,50,75,150$
QLI	2	238, 340	102	$1,2,3,6,17,34,51,102$
LIT	2	239, 341	102	$1,2,3,6,17,34,51,102$
GXL	2	261, 375	114	$1,2,3,6,19,38,57,114$
KDM	2	267, 333	66	$1,2,3,6,11,22,33,66$
DME	2	268, 334	66	$1,2,3,6,11,22,33,66$
MEA	2	269, 335	66	$1,2,3,6,11,22,33,66$
UKD	2	332, 446	114	$1,2,3,6,19,38,57,114$
CZU	2	367, 439	72	$1,2,3,4,6,8,9,12,18,24,36,72$
ZUB	2	368, 440	72	$1,2,3,4,6,8,9,12,18,24,36,72$

1.) $I = \dfrac{8726}{462 \times (462-1)} = 0.040970598$ therefore probably polyalphabetic substitution cipher.
Yes. There is not a lot of range in the frequencies of the letters in the ciphertext.

$l = \frac{0.0301 \times 462}{(462-1)0.040970598 - 0.0385 \times 462 + 0.0686} = 11.9$ keyword length may be 12 (Friedman) or 6 (Kasiski).

Find keyword length to be 6 then use "'TLR" as "'the"' and then keyword becomes "' E—MA"' guess "'ENIGMA"'

"STARTING FROM SCRATCH AT THE BEGINNING OF THE FIRST WORLD WAR BRITAIN HAD BY WARS END BUILT A LARGE AND PROFESSIONAL CODE BREAKING ESTABLISHMENT UNDOUBTEDLY THE BEST IN THE WORLD THE ADMIRALITYS ROOM FORTY WAS IT CAME TOBE CALL ED PERFORMED FEATS OF CRYPT ANALYSIS THAT HAVE SINCE BECOME LEGENDARY READING ALMOST ALL OF GERMANYS NAVAL AND DIPLOMATIC TRAFFIC DURING THE WAR INCLUDING THE FAMOUS ZIMMERMANN TELEGRAM THAT REVEALED GERMANYS PROPOSED ALLIANCE WITH MEXICO AGAINST THE UNITED STATES AND WHOSE DISCLOSURE HELPED PROPEL AMERICA IN TO THE WAR"

Chapter 3

Rail Fence Cipher

An example of a transposition cipher is the rail fence (or zig zag) cipher. In the rail fence cipher, the plaintext is written in a zig-zag format.

Example 3.1. *Encrypt the message "YOU HAVE BEEN SPOTTED DO NOT ENTER" with four rails.*

Y	-	-	-	-	-	E	-	-	-	-	-	P	-	-	-	-	-	D	-	-	-	-	-	N	-	-	-
-	O	-	-	-	V	-	B	-	-	-	S	-	O	-	-	-	D	-	O	-	-	-	E	-	T	-	-
-	-	U	-	A	-	-	-	E	-	N	-	-	-	T	-	E	-	-	-	N	-	T	-	-	-	E	-
-	-	-	H	-	-	-	-	-	E	-	-	-	-	-	T	-	-	-	-	-	O	-	-	-	-	-	R

The encrypted message is read row by row: "YEPDNOVBSODOETUAENTENTEHETOR".

The plaintext may actually start at any point on the cycle,and key types are of the form "'number of rails, offset" i.e in "4, 0" indicates four rails and no offset. Offsets can be between 0 and $2R$-3, where R is the number of rows (rails).

Example 3.2. *Again encrypt the message "YOU HAVE BEEN SPOTTED DO NOT ENTER" this time with four rails and an offset of 1.*

-	-	-	-	-	-	V	-	-	-	-	-	S	-	-	-	-	-	D	-	-	-	-	-	E	-	-	-	-
-	Y	-	-	-	A	-	E	-	-	-	N	-	P	-	-	-	E	-	D	-	-	-	T	-	N	-	-	-
-	-	O	-	H	-	-	-	B	-	E	-	-	-	O	-	T	-	-	-	O	-	O	-	-	-	T	-	R
-	-	-	U	-	-	-	-	-	E	-	-	-	-	-	T	-	-	-	-	-	N	-	-	-	-	-	E	-

The encrypted message is "VSDEYAENPEDTNOHBEOTOOTRUETNE".

Decryption is very simple as if the key is known then the rail fence can be easily reconstructed and the plaintext read.

Example 3.3. *Decrypt the message "BNVEUOOREYETDTEOSYUOSUOB" with five rails and an offset of 2.*

Solution to Example :

-	-	-	-	-	-	-	-	B	-	-	-	-	-	-	-	N	-	-	-	-	-	-	-	V	-
-	-	-	-	-	-	-	E	-	U	-	-	-	-	-	O	-	O	-	-	-	-	-	R	-	E
-	-	Y	-	-	-	E	-	-	-	T	-	-	-	D	-	-	-	T	-	-	-	E	-	-	-
-	-	-	O	-	S	-	-	-	-	-	Y	-	U	-	-	-	-	-	O	-	S	-	-	-	-
-	-	-	-	U	-	-	-	-	-	-	-	O	-	-	-	-	-	-	-	B	-	-	-	-	-

The decrypted message is “You see, but you do not observe".

If only a small number of rails are used then the cipher can be easily broken even if the key is unknown using exhaustive search. Alternatively, if the text is of a sufficient size then frequency analysis may be used.

University of South Wales

3.1 Tutorial

1.) Encrypt the message “I CANNOT THINK OF MANY MESSAGES" using the railfence cipher with 4 rails and zero offset.

2.) Decrypt the message “IANCFNLFNRNAYIHOMIVINITGNTHTDADEFITSEAHWHNNLNESEYRHIGEICUTILTNMO" using the railfence cipher with 6 rails and an offset of 1.

3.) Find the plaintext given the ciphertext “ITAAEQTHSSHLSRIFNEUSINTIETLCEO"

3.2 Solutions

1.) ITOMECOTKFYEGSANHNMNSANIAS

2.) 64 letters Life is infinitely stranger than anything which the mind of man could invent

3.) this is the last railfence question three rails offset 2. (Look for Q and U. Q appears in the sixth position and U in the 19th position. Consider where they could fit on rails. Assume initially there are 3 rails. Consider position of Q on the rails, if no offset Q would appear on the peak, etc... Now position all characters (30 positions). If offset incorrect most of the message will make sense and so can reposition to correct offset)

Chapter 4

The Columnar Transposition Cipher

The columnar transposition cipher is another simple transposition cipher.

Example 4.1. *The key for the columnar transposition cipher is a keyword for example "crypt". Encrypt "We balance probabilities and choose the most likely".*

C	R	Y	P	T
w	e	b	a	l
a	n	c	e	p
r	o	b	a	b
i	l	i	t	i
e	s	a	n	d
c	h	o	o	s
e	t	h	e	m
o	s	t	l	i
k	e	l	y	

If the plaintext does not fit neatly into a rectangle, then the empty spaces can be padded (with an x) and the cipher is known as a regular columnar transposition. An irregular columnar transposition leaves these characters blank, this makes decryption slightly more difficult.
The columns are now read off in alphabetical order to obtain the ciphertext "'warieceokaeatnoelyenol-shtselpbidsmixbcbiaohtl".
To decipher a Columnar cipher, find the length of the columns by using the known key length and then carry out the steps in reverse.

Example 4.2. *Decrypt "wenteirocetnvxtnvoaoxhmiarsxaannccxnanhdex". using the key CIPHER. The text has been padded with x's as needed.*

Solution to Example :

42 letters. $42/6 = 7$ so 7 letters per column.

C	I	P	H	E	R
w	h	a	t	o	n
e	m	a	n	c	a
n	i	n	v	e	n
t	a	n	o	t	h
e	r	c	a	n	d
i	s	c	o	v	e
r	x	x	x	x	x

What one man can invent another can discover

During World War I, the German military used a double columnar transposition cipher (i.e one applied twice with two different keywords), but they changed the keys infrequently. The ciphertexts were regularly broken by the French, who were able to quickly find the keys once they had intercepted a number of messages of the same length, this generally took them only a few days. However, the French success became widely-known and, after a publication in Le Matin, the Germans changed to a new system in November 1914.

4.1 Cryptanalysis

Since transposition does not affect the frequency of individual symbols, simple transposition can be easily detected by the cryptanalyst by doing a frequency count. If the ciphertext exhibits a frequency distribution very similar to plaintext, it is most likely a transposition. This can then often be attacked by anagramming - sliding pieces of ciphertext around, then looking for sections that look like anagrams of English words, and solving the anagrams. Once such anagrams have been found, they reveal information about the transposition pattern, and can consequently be extended.
Simpler transpositions also often suffer from the property that keys very close to the correct key will reveal long sections of legible plaintext interspersed by unintelligible text.

University of South Wales

4.2 Tutorial

1.) Encrypt the plaintext “Encipherment is easy” with key "‘Monday.

2.) Decrypt the ciphertext “ornscoaspdeindtiyx” with key "‘Monday.

3.) Obtain the corresponding plaintext given the ciphertext “IAFCCAETAPOALCIEETKMPSMHRORRSXOOANHHNLWCUI”and the length of the key is six.

4.) Encrypt using a regular Columnar cipher the message “THE PLAN IS TO ATTACK AT DAWN TOMORROW” first using the key “Tired” then the key “sleepy” (ignore the second e).

5.) Challege: Can you break this?
"‘ilhaeohtedtpahtsstlhepetmiaveslppaednutcominthohcnctairoxthibpsghayiadihve idsttehieujeiarshdamxsaetdroyetoepewtoheisrttnponhrnereufex"'.

4.3 Solutions

1.) “pnsieaeeicmenrshty”

2.) “and so is decryption”

3.) 42 letter, so six columns, order of columns ciphertext to plaintext $(1,2,3,4,5,6) \rightarrow (2,6,3,5,1,4)$ “this is an example of how to crack a columnar cipher”

4.) First stage: “LTADOOPSTTTRHNAKWOEITANRTAOCAMW” Second stage: “ASHONCTPRWAODTNERALOTKTAWOTAITM”

5.) A regular Columnar cipher with 133 letters - note the x’s which effectively give column length. Gap of 38 = column length 19. 133 letters so need divisor $\frac{133}{19} = 7$.
This is the last cipher to be attempted this morning. I hope that you have enjoyed the historical pen and paper ciphers that we have studied so far in the module

Chapter 5

The Playfair cipher

The Playfair cipher was invented in 1854 by Charles Wheatstone the English physicist and inventor, but takes its name from Lord Playfair, who was heavily involved in promoting its use to the British government of the time. The Playfair cipher encrypts pairs of letters. The cipher was used as the encryption system by the British Army in World War 1.

5.1 Implementing the Playfair cipher

A common implementation relies on a keyword and a 5×5 grid containing letters. There are 26 letters but only 25 spaces. It is common to incorporate I and J together in one square as there are virtually no words where it is not obvious whether the required letter is I or J. (Omitting Q is also occasionally used). The key k used in encryption is a permutation of the resulting 25 letter alphabet. The method of encryption will be explained using an example.

Example 5.1. *Write down the keyword in the first few spaces of the 5×5 table (starting at the top left and working row wise). Each letter can only appear once, so subsequent occurrence of a particular letter is ignored. Assume the key comprises the keyword "GLAMORGAN" followed by all remaining letters written in alphabetical order.*

Solution to Example :

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

5.2 Encrypting a message

The plaintext needs to be split into pairs of letters but identical pairs such as "AA" or "BB" must be avoided. So initially, the plaintext is searched and in between every identical pair of letters, an "X" is

placed. An extra letter “X" may also need to be added at the end of the plaintext to ensure there are an even number of letters before enciphering begins.

Example 5.2. *Plaintext - "MEET ME AT TREFFOREST STATION".*

Solution to Example :
"MEXETMEATXTREFXFORESTSTATION" which when split into pairs gives "ME XE TM EA TX TR EF XF OR ES TS TA TI ON"

For each of the pairs, find each letter in the table above. There are three possibilities:

- If both letters are in the same row, replace them by the letters immediately to their right (if one of them is on the end of the row, wrap back around to the beginning of the row)
- If both letters are in the same column, replace them by the letters immediately below them (if one of them is on the bottom of the column, wrap back around to the top of the column)
- Otherwise, imagine drawing a rectangle which has the two letters as opposite corners, and write down the two letters that form the other corners of the rectangle (starting with the corner on the row corresponding to the first of the two letters).

ME - on different rows so imagine the rectangle with M in one corner and E in the other. The other corners of this rectangle would be G and I, so write GI.

XE similarly gives VH

TM - both in same column so look at the letters immediately below T and M giving YC

EA - rectangle, so get HG

TX - rectangle, get SY

TR - Rectangle, PC

EF - same row, take the letters to the right, get FH

XF - rectangle, WH

OR - rectangle, GD

ES - rectangle, HP

TS - same row, take letters to the right, UT

TA - rectangle SM

TI - same column, take letters below, YT

ON - rectangle, LD

So the encrypted message is

"GI VH YC HG SY PC FH WH GD HP UT SM YT LD"

NOTE: A Playfair square encrypts in the same manner if the rows and/or columns of the square are reordered. For example the square given previously could be arranged as:

Solution to Example :

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

5.3 Decrypting the Playfair cipher

Decryption is trivial if the keyword is known. Using the same grid as the sender, the ciphertext is considered pair by pair and essentially the process is carried out in reverse to work out the original pair of letters each time.

- If both letters are in the same row, replace them by the letters immediately to their left (if one of them is on the beginning of the row, wrap back around to the end of the row)
- If both letters are in the same column, replace them by the letters immediately above them (if one of them is on the top of the column, wrap back around to the bottom of the column)
- Otherwise, imagine drawing a rectangle which has the two letters as opposite corners, and write down the two letters that form the other corners of the rectangle (starting with the corner on the row corresponding to the first of the two letters).

Example 5.3. *Ciphertext: "GI VH YC HG SY PC FH WH GD HP UT SM YT LD"*

Solution to Example :

The first pair is GI. They form a rectangle, so take the corresponding letters ME

The next pair is VH. They form a rectangle, so take the corresponding letters XE

Next is YC. On the same column, so look above: the pair is TM ... and so on

5.4 Cryptanalysis of the Playfair Cipher

It is useful to note that a ciphertext formed using the Playfair cipher will always contain an even number of letters and the frequency count for the text will only have a non-zero probability for 25 letters (as J or Q are not present).

Since the Playfair cipher is a substitution cipher that operates on 2 symbols at a time rather than single symbols, the cipher hides some of the information that a frequency analysis of the cipher single text letters would provide. However, it does little to conceal the "paired letters" frequencies of the plaintext. The cipher was thought to be unbreakable for many decades but is actually not difficult to break.

One cryptanaylsis technique is to use a similar strategy to before, but instead of looking for "E" as the most common letter, look for the most common pairs (known as bigrams or digraphs) of letters in English (e.g. "ED" is very common at the end of a word). It should also be noted that a bigram and its reverse will encrypt in a similar way, so if AB encrypts as YZ then BA will encrypt to ZY so by looking for words that begin and end in reversed bigram they can be then compared with plaintext words that are similar (e.g departed, receiver). This is a good way to begin constructing the key.

A good tutorial on reconstructing the key for a Playfair cipher can be found in chapter 7, "Solution to Polygraphic Substitution Systems," of Field Manual 34-40-2, produced by the United States Army. (on blackboard)

(With modern computing power, the Playfair cipher isn't difficult to crack. It is possible now to use "brute force" algorithms, essentially checking every possibility - even with over 600 possibilities for pairs, this isn't too challenging for a computer, but back in the 19th century, without computers and by hand, this was an inconceivable task.)

University of South Wales

5.5 Tutorial exercises - Playfair cipher

1.) Encrypt “I think there is a worldwide market for maybe five computers" (this is allegedly a quote from Thomas Watson, founder of IBM, in 1943) using the Playfair cipher with keyword “DESKTOP".

2.) Decrypt “QG MX HL EC HT HG TX UM RK DN VT SY" using the keyword “Glamorgan"

3.) Given the plaintext “Before turning to those moral and mental aspects of the matter which present the greatest difficulties, let the inquirer begin by mastering more elementary problems " and the cipher text “ KL DC LR UN ST DT IQ AN SC HR DB AI RB VN GL PU BR OR QH AP OD IP ER IB QY QR HY RF FC TH RH SQ QY PR KQ LR IY RH NI MG WG FA ZB YT RH SR QY PR RG PT TM LR LA KQ DT AZ IB RN RL DT ID AS KE RS LG SQ YI AR TH CO SR DL" Determine the Playfair keyword.

4.) You have intercepted the following ciphertext which is known to have been produced using a Playfair Cipher. You also know that the fragment “ Lord Playfair who was heavily involved in promoting its use" appears somewhere in the text. Determine the corresponding message and keyword.

“WE NX QW VI BQ ZK GQ EA LT WB GB XH CE WG GB AG FE WA NG GK KH VI SV QC ZB ET LM WN HE AT WC HC GN WE NE NG DP DB EM AV BD BK CW EB FK OX NG HC QC ZW WT GT BD WC BE PH KM FV MS LK QM BA IH KQ HE SH WB EA HY DQ AQ OX SM XH FK GX MC VF AK GP KA DW NW HC WE AN QK AK OW FN XH PC PH CE FM WE AW FQ AW EA QM BA IH KQ BK ME TP NG KZ XQ WC QE KQ ON DM AW WA LC WE TN GQ EA LT WB WD WG WB WE NE NG KZ XQ AK NB BU CW HP IA WE AN QK AK OW TQ QV GB HS LM LS TQ NB NE "

5.) You have intercepted the following ciphertext which is known to have been produced using a Playfair Cipher. Determine the corresponding message and keyword.

“OX XM LF NA FV EN FY EF AU PX MR RF TF XG NA XP RG DE VY AP GP KE TF NV RF BP OQ LX YW PI PA XG TF NE VB VH DP EF AZ OX YD EN BH OQ HX TE TP DP TW UC AB IV UB RU GX NI VQ SU FV RF NV GE VQ UT PO AB OH MR HI RF NV XV XH RF DY KE VM XM BU IV RK GP AN ZX ND SA QH VQ SU FV RF NB MR MA VQ RF

DE BI VY NA RQ CK ND PA EG AC IV HX VX LD AF IV BV FY VQ RQ CK ND MO NA BL LX YW PI PA XG NA RF NA EG VG NB PO VY"

5.6 Playfair Tutorial Solutions

1.) In pairs: "IT HI NK TH ER EI SA WO RL DW ID EM AR KE TF OR MA YB EF IV EC OM PU TE RS"

Ciphertext "LK IL RE SL KN KG AH VP UI EV FK DN BQ TS DL BM QO KI DG FY TP FV CN DS QK"

2.) In pairs "pl ay fa ir is ea sy to de cr yp tx"

Plaintext "Play fair is easy to decrypt"

3.) Form plaintext pairs "be fo re tu rn in gt ot ho se mo ra la nd me nt al as pe ct so ft he ma tx te rw hi ch pr es en tx th eg re at es td if xf ic ul ti es le tx th ei nq ui re rb eg in by ma st er in gm or ex el em en ta ry pr ob le ms"

Key word is sherlock

4.) In pairs "th ep la yf ai rc ip he rw as in ve nt ed in ei gh te en fi ft yf ou rb yc ha rl es wh ea ts to ne th ex en gl is hp hy si ci st an di nv en to rb ut ta ke si ts na me fr om lo rd pl ay fa ir wh ow as he av il yi nv ol ve di np ro mo ti ng it su se to th eb ri ti sh go ve rn me nt of th et im et he pl ay fa ir ci ph er en cr yp ts pa ir so fl et te rs th ec ip he rw as us ed as th ex en cr yp ti on sy st em by th eb ri ti sh ar my in wo rl dw ar on ex"
Assume the first two letters are a typical start to a sentence so 'th' then build grid.

Plaintext "The Playfair cipher was invented in eighteen fifty four by Charles Wheatstone the English physicist and inventor but takes its name from Lord Playfair who was heavily involved in promoting its use to the British government of the time. The Playfair cipher encrypts pairs of letters. The cipher was used as the encryption system by the British Army in World War one"

Key word is Wheatstone

5.) OX XM LF NA FV EN FY EF AU PX MR RF TF XG NA XP RG DE VY AP GP KE TF NV RF BP OQ LX YW PI PA XG TF NE VB VH DP EF AZ OX YD EN BH OQ HX TE TP DP TW UC AB IV UB RU GX NI VQ SU FV RF NV GE VQ UT PO AB OH MR HI RF NV XV XH RF DY KE VM XM BU IV RK GP AN ZX ND SA QH VQ SU FV RF NB MR MA VQ RF

University of South Wales

DE BI VY NA RQ CK ND PA EG AC IV HX VX LD AF IV BV FY VQ RQ CK ND MO NA BL LX YW PI PA XG NA RF NA EG VG NB PO VY

Adventure

HO WO FT EN HA VE IS AI DT OY OU TH AT WH EN YO UH AV EX EL IM IN AT ED TH EI MP OS XS IB LE WH AT EV ER XR EM AI NS HO WE VE RI MP RO BA BL EM US TB ET HE TR UT HW EK NO WT HA TH ED ID NO TC OM ET HR OU GH TH ED OX OR TH EW IN DO WO RT HE CH IM NE YW EA LS OK NO WT HA TH EC OU LD NO TH AV EB EX EN CO NC EA LE DI NT HE RO XO MA ST HE RE IS NO CO NC EA LM EN TP OS XS IB LE WH EN TH EN DI DH EC OM EX

How often have I said to you that when you have eliminated the impossible, whatever remains, however improbable, must be the truth? We know that he did not come through the door, the window, or the chimney. We also know that he could not have been concealed in the room, as there is no concealment possible. When, then, did he come?

University of South Wales

5.7 Tutorial - Revision of Early Historical Ciphers

You have intercepted the following Ciphertexts. They are all encrypted using the historical ciphers covered so far. Find the plaintext. You should use relevant cryptanalysis' tools and perform the attack by using pen and paper. You may use a computer to carry out frequency anaylsis. Outline a rationale for all decisions made. (This is similar to Section A of your coursework)

1.) DSVM YIVZPRMT XRKSVIH GSV URIHG HGVK RH GL KVIULIN HGZGRHGRXZO ZMZOBHRH

2.) TSEQTA TRKIAN DPIPOU RDSTTA NYSLOT LHOIUE SOEUNP APEHEY AEITII LASHPU AINSIE YBASGE NARODO GEDASC ALIEYA L

3.) RWR MCI HFM TFWSRAOBG HSGH VCK OPCIH YOGWGYW RWR HVS SJWRSBQS DCWBH HC O GIPGHWHIHWCB QWDVSF

4.) VVJAMK GBJWIG KDMJVI HTJHXM XSQDGS PQJFPW VVJKVI SIJSGC QTQJXX GFXFPI VHJWGS WBYBSY NRXMSA VVFYXL GFJNWZ GFDQMX VZJIMJ HSWJRG GPJYAI GBYMIR WAGJVS HHNRIW GOHMPI VHJWSG EIWXMR VVJHMT JSWYIB VHMJZM ISSJVI EWUMIV CZXTLE UOSJRS TATZWR WAGJVS HYJDWX JSXJRH GFFSHV GQJNZI TQFSEK TSJTRE PMBTVH KBFIMG VWTSEV AOSDGS OPNSEX KCSTJA QFIXSV GJJSMR XSSYAS TRXFGV ADYFRE NMXYGE PBTYGV CQPYLI EWUMIV DMXJEV EVNSKE NZUTWW KPQJOI AGGJGE WGJYLI PIRGIV QTTUXM QBXNWW KAUQCX QCLWIE V

5.) IBHG VPIB E JAFG JDGRG LEU E KGRN YIRGX SGHJVRGR LDI RGESSN HIVSX BIJ JDABC IW EBNJDABT FIRG ABJGRGUJABT JI LRAJG EU PSEABJGMJ UI UDG ZVUJ LRIJG E SIEX IW REFYSABT BIBUGBUG WIR JDG UJVXGBJU JI XGHAPDGR EBX JDGN ESS SAKGX DEPPASN GKGR EWJGR