History of Maths and χ

where χ = Cryptography

Substitution ciphers: Ancient - Renaissance

A talk at the University of Nottingham
by Peter Rowlett

"The history of codes and ciphers is the story of the centuries-old battle between codemakers and codebreakers, an intellectual arms race that has had a dramatic impact on the course of history."
– Simon Singh, The Code Book

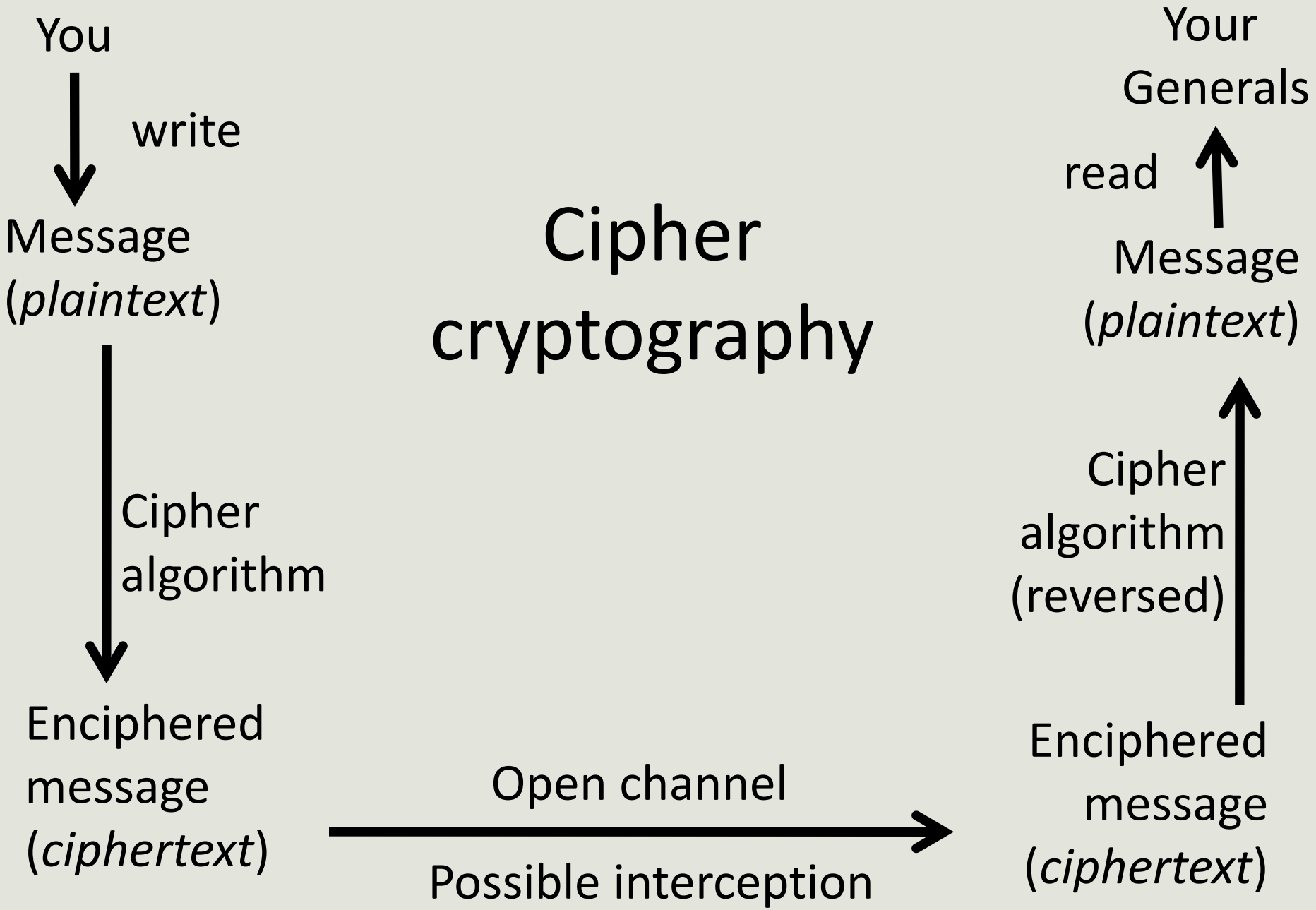
Imagine...

- You keep sending secret messages to your trusted Generals but the enemy seems to know all your plans before you realise them.
- Are messengers being intercepted on the way?
- Is one of your messengers a double agent, passing secrets to the enemy?

- You are planning a co-ordinated surprise attack and must tell your Generals without the enemy discovering your plans. How can you get the message to them so it can't be read by interceptors?
- Let’s look at cipher cryptography

Cipher cryptography

- A message, the *plaintext*, is converted through some process, the *cipher algorithm* into an enciphered form, the *ciphertext*
- The cipher algorithm is usually well known – what makes a cipher system secret is the *key*, some vital piece of information that is needed to perform the algorithm


You
write
Message
(*plaintext*)
Cipher
algorithm
Enciphered
message
(*ciphertext*)
Cipher
cryptography
Open channel
Possible interception
Your
Generals
read
Message
(*plaintext*)
Cipher
algorithm
(reversed)
Enciphered
message
(*ciphertext*)

Caesar cipher

- Famous early use of cryptography was by the Roman Emperor Julius Caesar
- Caesar cipher is a type of substitution cipher
- Cipher algorithm: each letter in the plain alphabet is replaced with the letter *n* places further on in the alphabet
- Key: *n*, the number of letters to shift

Example

- Plain letters are written in lower case and cipher letters in UPPER CASE
- Key is 3

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
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

- Write out plain message: `hello everyone`
- encipher each letter in turn by looking for the corresponding letter in the cipher translation table.
- This gives the ciphertext message:

So as long as the message recipient knows the key – how many letters you have shifted the alphabet by – they can build the cipher alphabet and decipher the message by going through the cipher algorithm in reverse.

K H O O R H Y H U B R Q H

h e l l o e v e r y o n e

Other simple substitution ciphers

- Caesar cipher has only 25 possible cipher alphabets
- Wouldn’t take long to try them all
- Other cipher systems use less regular methods for generating alphabets
- Must still have a key to generate an alphabet the recipient can reproduce

Example

- Take as your key a favourite quote.
- For example, take:

  “pure mathematics is, in its way, the poetry of logical ideas”
- First strip out repeating letters so each letter is unique

pure mathematics is, in its way,

pure*math****ics **, *n *** w*y,

the poetry of logical ideas

*** ****** of l*g**** *d***

puremathicsnwyoflgd

- Fill in this sequence as the start of your cipher alphabet.

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

- Fill up the alphabet with the letters which have not been used, in some systematic order (here we have used reverse alphabetical order)
- This cipher alphabet is less predictable than the Caesar cipher, yet it is still simple for both sender and receiver to generate, provided they know the key phrase

Your agents have intercepted an enciphered message from the enemy. Given your new knowledge of substitution ciphers, can you decipher this message without knowing the key?

Cracking substitution ciphers

- In the eighth century AD, Islamic culture entered a golden age
- The most learned society of its time
- Cryptography was routinely used for matters of state
- This led to the development of *cryptanalysis*, with scholars using a combination of mathematics, statistics and linguistics to develop techniques for deciphering messages when the key is unknown

Letter frequencies

- In studies of the text of the Qur'an, scholars had noticed that some letters appear more frequently than others
- In English the letters *e* and *t* are used much more frequently than the letters *z* and *q*, and this fact can be used to decipher messages
- This process is called *frequency analysis*

Average letter frequencies in English

Letter	Frequency
e	12.70%
t	9.06%
a	8.17%
o	7.51%
i	6.97%
n	6.75%
s	6.33%
h	6.09%
r	5.99%
d	4.25%
l	4.03%
c	2.78%
u	2.76%

Letter	Frequency
m	2.41%
w	2.36%
f	2.23%
g	2.02%
y	1.97%
p	1.93%
b	1.49%
v	0.98%
k	0.77%
j	0.15%
x	0.15%
q	0.10%
z	0.07%

Further frequency analysis

- Pairs of letters in words are most likely to be: *“ss”*, *“ee”*, *“tt”*, *“ff”*, *“ll”*, *“mm”* or *“oo”*.
- A one letter word is either *“a”* or *“I”*.
- Two letter words are commonly: *“of”*, *“to”*, *“in”*, *“it”*, *“is”*, *“be”*, *“as”*, *“at”*, *“so”*, *“we”*, *“he”*, *“by”*, *“or”*, *“on”* or *“do”*, in that order.

Further frequency analysis

- Three letter words are commonly *“the”* or *“and”*.
- The letter *h* frequently goes before *e* (as in *“he”*, *“the”*, *“then”*, etc.) but rarely goes after *e*. No other pair of letters has such an asymmetric relationship.

Further frequency analysis

- Another technique is to use a crib, which is a word or phrase you can guess will be in the message

Example

N**K**RRU N**K**X**K** OY **G** Z**K**YZ
S**K**YY**G**M**K** ZU **K**TIOVN**K**X LUX AY**K**
GY **G**T **K**D**G**SVR**K** OT **G**T **G**XZOIR**K**
LUX OYWA**G**X**K**J S**G**M**G**FOT**K**

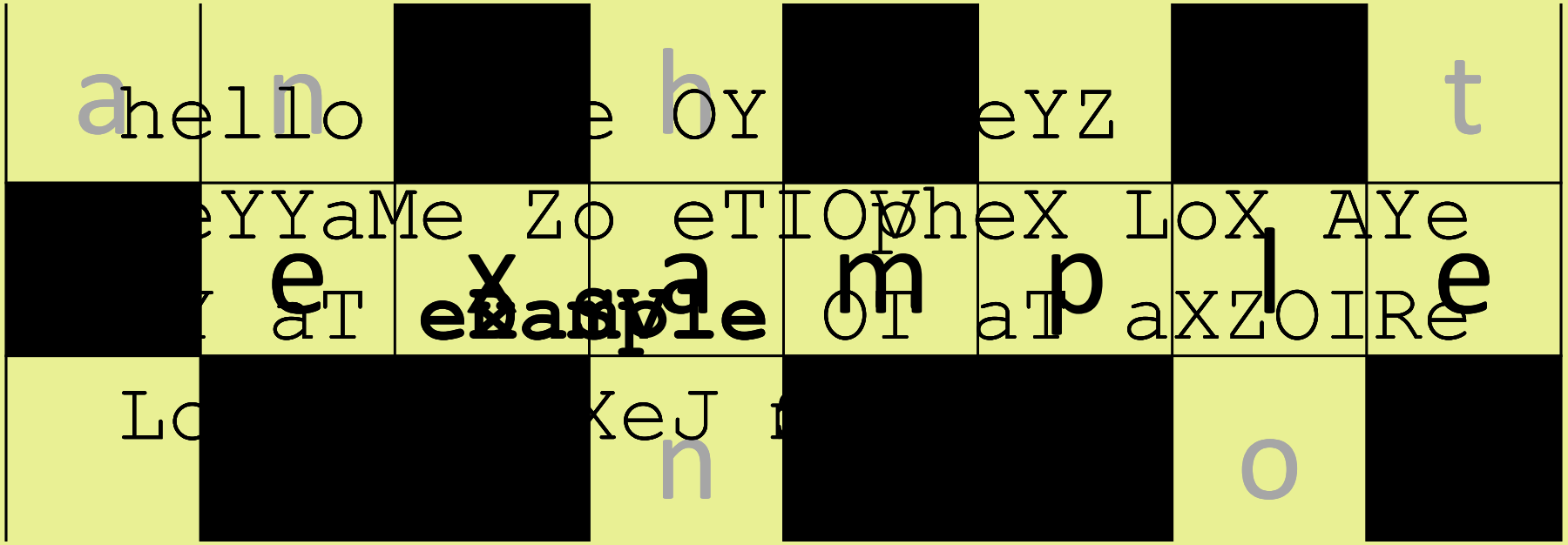
a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z
G				K																					

Example

NeRRU **Ne**Xe OY a ZeYZ
SeYYaMe ZU eTIOV**Ne**X LUX AYe
aY aT eDaSVRe OT aT aXZOIRe
LUX OYWAaXeJ SaMaFOTe

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
G				K			N																		

Example



- Notice all the letters are in alphabetical positions?

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
G				K			N				R	S		U	V								D		

Example

hello here is a test
message to encipher for use
as an example in an article
for isquared magazine

- Could this be a Caesar cipher?

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
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

Knowing the key is 6, you can now decipher future messages from your enemy. Be careful what information you act on though – if you seem too knowing they might get suspicious and change their key or algorithm!

You have discovered that your enemy is using a simple substitution cipher like your own. If you can decipher their messages using frequency analysis, they might be able to decipher yours!

Can a cipher be created to provide greater resistance to frequency analysis?

Beating frequency analysis

- During the Renaissance in Europe scholarship increased and politics became more complicated
- This contributed to the development of cryptography and cryptanalysis

Beating frequency analysis

- Methods for countering frequency analysis were developed, including:
  - Omitting spaces
  - Deliberate misspellings
  - Nulls – characters that have no meaning
  - Codes – replacing whole words or phrases with letters, words or phrases

- Such methods helped, but ultimately cryptanalysts won out and each method could be accounted for
- A better cipher was needed

Vigenère cipher

- Emerged in sixteenth century
- Uses more than one cipher alphabet and different letters are enciphered with these in turn
- The same plain letter can be enciphered and the same cipher letter deciphered in several different ways, significantly disrupting frequency analysis
- Cipher alphabets must be chosen by some systematic process

Example

- First, choose a word for your key
- Key: Choose “pauli”
- The Caesar cipher alphabets beginning with the letters of the keyword are then produced:

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
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
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
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
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
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

- Take as plaintext message: `hello`
- Cipher algorithm: encode each letter using each cipher alphabet in turn, cycling through the cipher alphabets

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
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
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
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
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
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

- *"h"* is enciphered using the *"P"* alphabet, giving *"W"*
- *"e"* is enciphered using the *"A"* alphabet, giving *"E"*
- *"l"* is enciphered using the *"U"* alphabet, giving *"F"*
- *"l"* is enciphered using the *"L"* alphabet, giving *"W"*
- *"o"* is enciphered using the *"I"* alphabet, giving *"W"*

- ciphertext message: `WEFWW`

- ciphertext message: `WEFWW`
- Notice that, crucially, we have
  - (a) enciphered the two letters *"l"* to give different cipher letters *"F"* and *"W"*;
  - and, (b) enciphered different plaintext letters *"h"*, *"l"* and *"o"* to give the same ciphertext letter *"W"*.
- Through use of multiple alphabets, the chart of letter frequencies is distorted, providing strong resistance to frequency analysis

- Vigenère is more complicated to implement than single-alphabet substitution ciphers
- This adds to the time taken to encipher and decipher messages
- It becomes worth the time and hassle if you know your enemy can decipher your simple substitution cipher messages
- For many years it had a reputation as an unbreakable cipher - but can the Vigenère cipher be broken?

Cracking Vigenère

- To find out how cryptanalysis techniques work on Vigenère ciphers you can listen to the companion podcast to this talk
- Released by the Institute of Mathematics and its Applications (IMA) through the *Travels in a Mathematical World* podcast

History of Maths and χ

- This talk is accompanied by an audio podcast and by an article in iSquared Magazine
- You can find out how to get access to other aspects of the History of Maths and x output through the website:

www.historyofmathsandx.co.uk