Code Breaking

Information thanks to Crypo Corner: http://crypto.interactive-maths.com/ and British International School, Phuket- check them out, they have a ton of interesting and useful information about codes and code breaking.

Introduction to Cyptology

Cryptology (coming from the Greek words κρυπτός (kryptos) meaning "hidden" and -λογία (-logia) denoting "study of", and hence is the study of hidden writings) is a very broad subject. In its broadest sense, it is split into two sections: Cryptography (where γράφειν (graphein) means "writing") and Steganography (where στεγανός (steganos) means "covered" or "protected").

In essence, both these strands are all about hiding messages, but they go about it in slightly different ways.

Of course, for as long as people have wanted to hide messages, there have been people that want to know what has been hidden (since if the information was not delicate, it would not need hiding in the first place). The study of breaking the code or cipher that has been used to encrypt a message is called Cryptanalysis. This is an integral part of Cryptology, as the Cryptography and Steganography would not have developed into what they are today if it were not for the people who were trying to break their codes.

Steganography

Steganography is the hiding of a message by a physi***c***al means. This has been used as a way t***o*** protect valuable information in many cultures throughout history, and there are many inventive methods that have been used. The first recor***d***ed use of St***e***ganography is detailed ***b***y He***r***odotus in his Histories from ***a***round 440BC.

In this account, we are told of a warning sent to Greece of a forthcoming attac***k***, that was wr***i***tten on wooden backi***n***g of a wax board, before the wax was poured in. Wax boards were commonly used in Ancient ***G***reece for notes, as they were reusable (by melting the wax and scrap***i***ng it flat, previou***s*** notes were removed). In essence they are the ***f***orerunner to pencils and rubbers! By ***u***sing the board beneath the wax, the warning could be sent to Greece, without their e***n***emies discovering it (unless they happened to melt and remove the wax, which they had no reason to do).

There are many other instances of Steganography being used throughout history such as:

- Writing messages on the body of a messenger - again from Herodotus, we are told of a wealthy man who tattooed a message on the scalp of one of his slaves, and allowing the hair to then grow over the message. The slave could then be sent to the recipient, who shaved off the hair to reveal the message.
- Using different font types to reveal a message in a piece of seemingly trivial text - in the paragraphs above the bold/italic letters spell "codebreaking is fun".
- Using invisible inks, which are only revealed under a special light or by heat etc.

- Messages written in Morse Code on yarn that is then knitted to make a piece of clothing.
- Written on an envelope beneath the stamp.
- In 1966, Jeremiah Denton blinked in Morse Code during a forced televised press conference by the North Vietnamese. He spelled out the word T-O-R-T-U-R-E, and confirmed for the first time that American Prisoners of War were being tortured.
- During and after World War II, many spies used microdots, which are tiny discs with tiny lettering printed on them. These are then hidden on a piece of paper (under a full stop perhaps).

These methods all have one major weakness, which is that they rely heavily on not being discovered, since if they are discovered, then the enemy can immediately read the information that was supposed to be secure.

The weakness of Steganography is tackled by Cryptography. This method of secret writing involves making systematic changes to the message that is to be sent, which can only be undone by the intended reader of the message (who is the only other person who knows the method that was used to change the message). Of course, it is often the case that people wanting to send secret messages use both Steganography and Cryptography, which makes the message even safer from detection by prying eyes, since they would first have to find it, and then they would have to decrypt the scrambled message to retrieve the actual text.

Codes and Ciphers

Cryptography is split into two ways of changing the message systematically to confuse anyone who intercepts it: these are codes and ciphers. Many people believe, and use, the word code to mean the same thing as cipher, but technically they are different.

A code is a way of changing the message by replacing each word with another word that has a different meaning. For example, "Burn the City" could become "Take the rubbish" where the word "burn" is represented by the codeword "take", and similarly for "city" and "rubbish". Using codes requires a codebook, which contains all such codewords. Considering the large number of words in most languages, this is normally quite a large book, making the use of codes rather cumbersome (it is a bit like a french dictionary, giving the translation to and from the codeword). However, they can be used to encode key words in a message. Consider the message "Kill him as soon as possible". With a simple change of a single word this becomes "Meet him as soon as possible", which may pass through security detection without being noticed. So, although potentially hard to use, a simple code can be very effective, since even if the message is intercepted, they can be used so that the code reads as an innocent or unrelated topic.

Ciphers, on the other hand, convert the message by a rule, known only to the sender and recipient, which changes each individual letter (or sometimes groups of letters). Ciphers, are significantly easier to use than codes, since the users only have to remember a specific algorithm (a mathematical word for process) to encrypt the message, and not a whole dictionary of codewords. The major setback for ciphers compared to codes is that if

someone finds a message that has been encrypted using a cipher, the output is almost certainly going to be a random string of letters or symbols, and as such the interceptor will know straight away that someone wanted to hide this message.

The task of the cryptographer is to create a system which is easy to use, both in encryption and decryption, but remains secure against attempts to break it. For this reason, many ciphers have developed over the last 4,000 years to try to stop people from discovering what it is that their secret message says. In this website we focus our attention on ciphers, since they are more interesting and more diverse than the other forms of secret writing. We will be looking at many different ciphers, and will discuss how they work as well as some history behind their invention and use.

Some kinds of ciphers:

Monoalphabetic Substitution Ciphers

Substitution ciphers are probably the most common form of cipher. They work by replacing each letter of the plaintext (and sometimes punctuation marks and spaces) with another letter (or possibly even a random symbol).

A monoalphabetic substitution cipher, also known as a simple substitution cipher, relies on a fixed replacement structure. That is, the substitution is fixed for each letter of the alphabet. Thus, if "a" is encrypted to "R", then every time we see the letter "a" in the plaintext, we replace it with the letter "R" in the ciphertext.

A simple example is where each letter is encrypted as the next letter in the alphabet: "a simple message" becomes "B TJNQMF NFTTBHF". In general, when performing a simple substitution manually, it is easiest to generate the ciphertext alphabet first, and encrypt by comparing this to the plaintext alphabet. The table below shows how one might choose to, and we will, lay them out for this example.

Plaintext 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
Ciphertext Alphabet	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

The ciphertext alphabet for the cipher where you replace each letter by the next letter in the alphabet

There are many different monoalphabetic substitution ciphers, in fact infinitely many, as each letter can be encrypted to any symbol, not just another letter.

The history of simple substitution ciphers can be traced back to the very earliest civilisations, and for a long time they were more than adequate for the purposes for which they were needed. By today's standards they are very weak, and incredibly easy to break, but they were a very important step in developing cryptography.

Atbash cipher: one of the earliest known substitution ciphers. It simply reverses the plain text alphabet. As it is so simple, and does not require a *key*, it is not very secure.

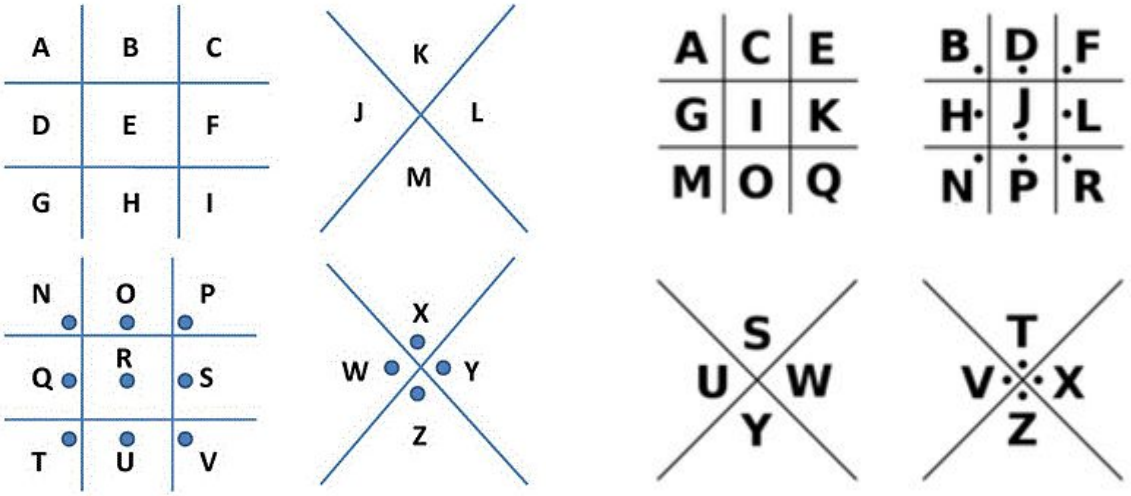
Plaintext Alphabet	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z
Ciphertext Alphabet	Z	Y	X	W	V	U	T	S	R	Q	P	O	N	M	L	K	J	I	H	G	F	E	D	C	B	A

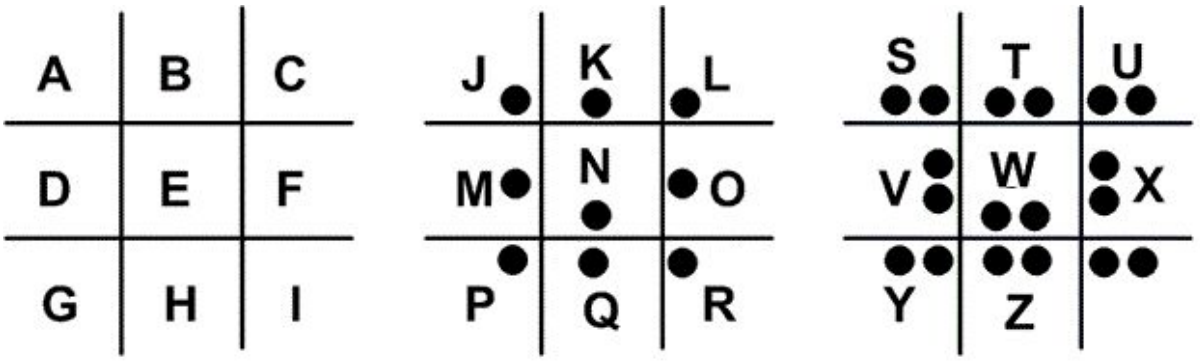
One way to make it more secure is to include numbers and punctuation marks too.

Plaintext 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	0	1	2	3	4	5	6	7	8	9
Ciphertext Alphabet	9	8	7	6	5	4	3	2	1	0	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	!	?	,	.	

Pigpen Cipher: This is another substitution cipher, but uses symbols rather than letters. The cipher has an interesting history: although its true origins are unknown, it has been used by many groups. Most notoriously, it was the cipher of choice for use by the Freemasons, a secret society in the 18th Century. In fact, they used it so much, that it is often referred to as the Freemasons Cipher. However, it was not exclusively used by them, with Union prisoners in Confederate camps using it to communicate in the American Civil War.

There are a few different versions of the cipher. Some examples:

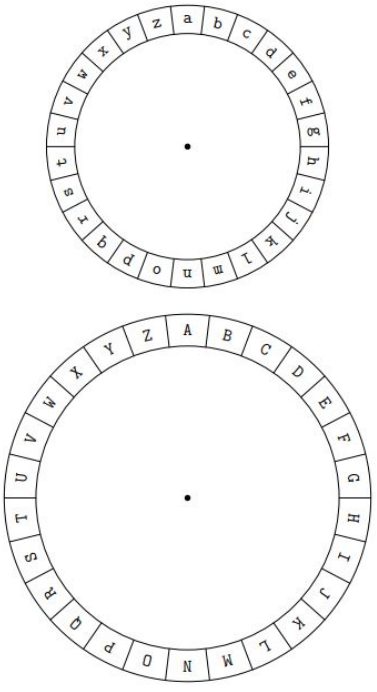




Caesar cipher: The Shift (or Caesar) Cipher is another monoalphabetic substitution cipher. Although more secure than the Atbash Cipher, it is still an easy cipher to break, especially by today's standards. Originally, it was used by Julius Caesar for sending encrypted messages to his troops. In this cipher, the alphabet is shifted by a number of letters, so a *shift* of 3 would mean that “a” would be represented by the letter that is 3 letters away from “a”, which is “D” (shown below).

Plaintext 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
Ciphertext Alphabet	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

A handy way to encrypt and decrypt using a Caesar cipher is to use a Caesar cipher wheel so you can set the *shift* to whatever value you need (there is a bigger printable version of this in the Code Breaking Tools document)



The value of the *shift* is the *key* for the encryption - you need to know (or figure out) the *key* in order to work out the message. This makes Caesar encryption more secure than the other types above (but still fairly easy to break).

Another interesting fact worth noting here is that composing multiple shifts (doing one shift and then taking the result and doing another shift) does not make the cipher any more secure. This is because a shift of a followed by a shift of b is the same as a shift of a + b (or in more concrete terms, a shift of 2 followed by a shift of 5 is identical to if we had just shifted the alphabet by 7 in the first place). Have a go to check this result.

If it is known that a Shift Cipher has been used, but the key is unknown, then it is fairly simple to break the code by a simple *brute force attack*. This simply means using a trial and error approach to attack the cipher. The main weakness of the Shift cipher is the fact that there are only 26 keys, and hence ciphertext alphabets, one of which is the identity mapping that leaves the plaintext unaltered. For this reason, the Brute Force method of attack is very effective on the Shift cipher. In its most bare form, this entails going through each key, and working out what the plaintext would be if that key had been used.

So, given the intercepted ciphertext "BMFY'X GWTBS FSI XYNHPD? F XYNHP.", where we do not know what key has been used, but we do know that a Shift Cipher has been implemented, we must first try a key of 1, then a key of 2, then a key of 3 and so on, until a plaintext that makes sense is returned. For this ciphertext we would get:

- A key of 1: “alex'w fvsar erh wxmgoc? e wxmgo.” - hmm, doesn’t really make sense…
- A key of 2: “zkdw'v eurzq dqg vwlfnb? d vwlfn.” - nope, still not right
- A key of 3: “yjcv'u dtqyp cpf uvkema? c uvkem.”
- A key of 4: “xibu't cspxo boe tujdlz? b tujdl.”
- A key of 5: “ what's brown and sticky? a stick.” - yay! Got it!

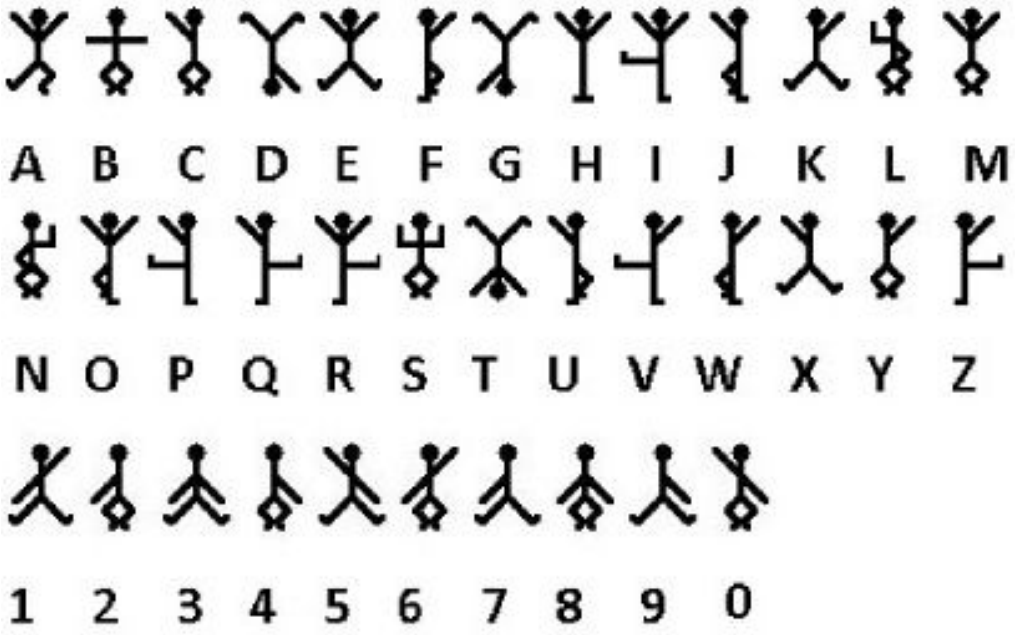
So we’ve cracked it by trial and error (and luckily the key was 5, and not 25, so we didn’t have to try too many times!)

There are two difficulties with this method. The first is that we were lucky in our example above, that we only had to do the decryption 5 times, but it is equally likely that a key of 24 could have been used, and then the time to perform the breaking of the code would be substantial. This is a problem that has been largely overcome by the invention of computers, which can perform all 26 possible calculations in a matter of seconds.

The second problem with the above method is that the message intercepted could be quite long, and hence performing each decryption could take a significant amount of time. The way around this is fairly simple, and that is to just look at the first two or three words of the intercept, and perform the calculations on these. You will still find the key, by finding the key that deciphers those words into a meaningful phrase, and can then use the key to decrypt the rest of the message as you would if you had known the key to start with.

This method of breaking the Shift Cipher is rather cumbersome, but can be useful if you know that it has been used. However, in reality it is unlikely that as an interceptor you would know which cipher has been used. There is a general method for attacking all monoalphabetic ciphers called **frequency analysis**. Check the end of this document for an explanation and example.

The Adventure of the Dancing Men
Another example of uses of ciphers is taken not from history, but popular culture. It comes in the form of a Sherlock Holmes Story, written by Sir Arthur Conan Doyle.
In this example, Sherlock Holmes is called in to help solve a mystery surrounding a piece of paper with strange stick figures on it. He uses his ingenious to both crack the code, and to send a message to the killer using his own code to entice him to come to the scene, unknowing that the message was written by Holmes.



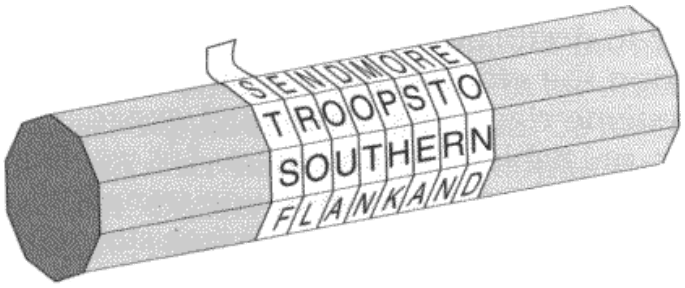
The Dancing Men were used to represent different letters. This is one very famous example of ciphers appearing in popular culture.

Transposition ciphers

Transposition Ciphers are a bit different to Substitution Ciphers. Whereas Substitution ciphers replace each letter with a different letter or symbol to produce the ciphertext, in a Transposition cipher, the letters are just moved around.
The letters or words of the plaintext are reordered in some way, fixed by a given rule (the key).

One example of a transposition cipher, is to reverse the order of the letters in a plaintext. So "a simple example" becomes "ELPMAXE ELPMIS A". Another, similar, way to encrypt a message would be to reverse the letters of each word, but not the order in which the words are written. In this case "a simple example" becomes "A ELPMIS ELPMAXE".

Another type of transposition cipher is the **Scytale**, which was an encryption device used by the Ancient Greeks and Spartans. It consisted of a polygonal rod or cylinder, around which was wrapped a piece of parchment. The sender would write the message along the faces of the rod as seen in the image below. When the parchment is removed from the Scytale, it leaves a nonsensical message going down the strip (in the case below it would read "STSF...").



When the message is received, it is wrapped around a rod of the same size and shape as the original, to reveal the original message. Clearly, you just need to get a rod of the same size, or try out a few different ones to break this code. However, its importance lies in the fact that it is one of the first uses of tools in Cryptography.

To make transposition ciphers such as the reverse ciphers above a bit more secure, it is usual to remove all punctuation marks from the plaintext first. It is quite often the case that all spaces are also removed.

Transposition ciphers can also switch around letters in different patterns (perhaps in groups of 2 letters) or changing a list of letters into a grid, and reading it in a different order (perhaps down the columns, rather than along the rows) An example:

WIEYNMMHSOEPEBATNVRNETHLEIUR

If we rewrite this as 4 lines with 7 letters in each:

W	I	E	Y	N	M	M
H	S	O	E	P	E	B
A	T	N	V	R	N	E
T	H	L	E	I	U	R

Reading down the columns we get: “What is the only even prime number?”

Vigenere encryption

The Vigenere encryption was the creation of the French diplomat, Blaise de Vigenere, 1523-1596.Here's the idea. For the given key word "FIRST", encrypt each letter of the message taken in the left-most column to the letter in the keyword-letter column. Thus, the first five letters of the message use the alphabets corresponding the the "F", "I", "R", "S", and "T" columns. So, the Vigenere code with this keyword is really five Caesar shifts used in a cyclical fashion.

Key word is "FIRST"

	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

BPFAG AMELX IKRDV ZTLK decodes to WHO INVENTED CALCULUS

With the first letter B, we look down the F column – and find B. Then look across to the far left column - this gives us W.

Next we decode P by finding it in the I column, then looking across to the far left column – this will give H etc

At the time, and for many centuries since its invention, the Vigenère Cipher was renowned for being a very secure cipher, and for a very long time it was believed to be unbreakable. It was this thought that earned it the nickname "le chiffre indéchiffrable" (French for "the unbreakable cipher"). Although this is not true (it was fully broken by Friedrich Kasiski in 1863), it is still a very secure cipher in terms of paper and pen methods, and is usable as a field cipher.

NASA, Aliens and Binary Codes from the Stars

SETI – the Search for Extra Terrestrial Intelligence – has spent the past 50 years scanning the stars looking for signals that could be messages from other civilisations. They look for non-random patterns in data strings that might suggest an advanced culture on another planet.

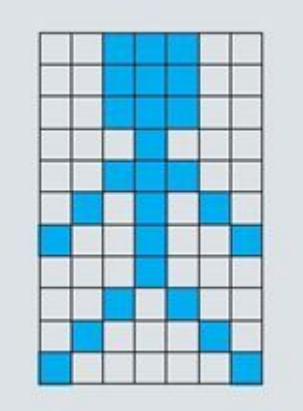
The desire to encode and decode messages is a very important branch of mathematics – with direct application to all digital communications – from mobile phones to TVs and the internet.

All data content can be encoded using binary strings. A very simple code could be to have 1 signify “black” and 0 to signify “white” – and then this could then be used to send a picture. Data strings can be sent which are the product of 2 primes – so that the recipient can know the dimensions of the rectangle in which to fill in the colours.

If this sounds complicated, an example:

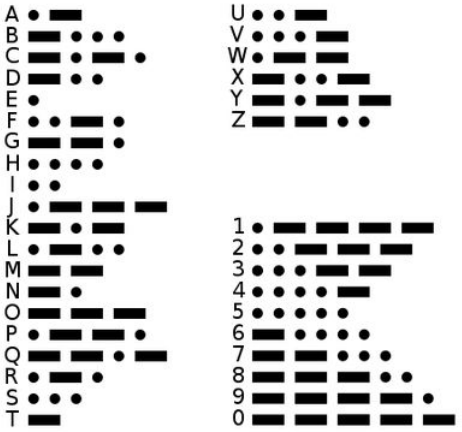
00111000011100001110000010000011100010101010010010001000001010001000101000001

If this mystery message was received from space, how could we interpret it? Well, we would start by noticing that it is 77 digits long – which is the product of 2 prime numbers, 7 and 11. Prime numbers are universal and so we would expect any advanced civilisation to know about their properties. This gives us either a 7×11 or 11×7 rectangular grid to fill in. By trying both possibilities we see that a 7×11 grid gives the message below.



Morse Code:

One of the most famous examples of a cipher in regular use is Morse Code (which is not a code, but rather a cipher). Morse Code has the benefit that it can be transmitted in several ways, such as written, by sound or by light. Each letter is replaced by a series of dots and dashes as given by the key below.



The brilliance of the Morse Code system is that it can also be transmitted as an audio signal, and was originally used to send messages via telegraph (before telephones had been invented). In this form, each dot is given by a short beep, and a dash by a longer beep (three times the length of a dot). There are various other rules as shown in the key above regarding the length of gaps and other things in transmission.
Morse code has played a pivotal role in the development of technology, specifically telecommunication. Although not secure, since the key is widely known, it is still an interesting use of ciphers in the world today (many boats and planes still use Morse Code to communicate in bad weather when voices sounds crackly), and most importantly, the development of the Code went hand in hand with other technological discoveries

Another interesting-looking activity I found, using sound to represent binary numbers, which can then be transposed into a message: http://csunplugged.org/modems-unplugged-2/ - note: I haven't tested this one out… but it looked interesting.

Frequency Analysis - the long explanation (and an example):

We have seen that there are too many possible keys to try in a brute force attack in the Mixed Alphabet Cipher, and given that we could also use symbols in our substitution, there are infinitely many different keys for a Monoalphabetic Substitution Cipher. Despite this, however, every single example of this type of cipher is easily broken, using a single method that works on all of them: Frequency Analysis.

Below we shall discuss the method for implementing Frequency Analysis, and then we shall work through an extended example, to fully appreciate how it works.

The Method

The methodology behind frequency analysis relies on the fact that in any language, each letter has its own personality. The most obvious trait that letters have is the frequency with which they appear in a language. Clearly in English the letter "Z" appears far less frequently than, say, "A". In times gone by, if you wanted to find out the frequencies of letters within a language, you had to find a large piece of text and count each frequency. Now, however, we have computers that can do the hard work for us. But in fact, we don't even need to do this step, as for most languages there are databases of the letter frequencies, which have been calculated by looking at millions of texts, and are thus very highly accurate.
From these databases we find that "E" is the most common letter in English, appearing about 12% of the time (that is just over one in ten letters is an "E"). The next most common letter is "T" at 9%. The full frequency list is given by the graph below.

Letter	Frequency
e	12.7
t	9.1
a	8.2
o	7.5
i	7.0
n	6.7
s	6.3
h	6.1
r	6.0
d	4.3
l	4.0
c	2.8
u	2.8

m	2.4
w	2.4
f	2.2
g	2.0
y	2.0
p	1.9
b	1.5
v	1.0
k	0.8
j	0.15
x	0.15
q	0.10
z	0.07

We can use this information to help us break a code given by a Monoalphabetic Substitution Cipher. This works because, if "e" has been encrypted to "X", then every "X" was an "e". Hence, the most common letter in the ciphertext should be "X".
Thus, if we intercept a message, and the most common letter is "P", we can guess that "P" was used to encrypt "e", and thus replace all the "P"'s with "e". Of course, not every text has exactly the same frequency, and as seen above, "t" and "a" have high frequencies too, so it could be that "P" was one of those. However, it is unlikely to be "z" as this is rare in the English Language. By repeating this process we can make good progress in breaking a message.
If we were to just put all the letters in order, and replace them as in the frequencies, it would likely produce gibberish. The codebreaker has to use other "personality traits" of the letters to decrypt the message. This may include looking at common pairs of letters (or digraphs): there aren't many 2 letter words; there are only a few letters which appear as doubles (SS, EE, TT, OO and FF being the most common). There are only two sensical words made of a single letter in English. Other common words also start to appear as you make some substitutions. For example "tKe" might appear frequently after making substitutions for "t" and "e". This is very likely to be "the", a very common word in English.

An example:

In this example we shall use Frequency Analysis to break the code used to encrypt the intercept given below, given that it has been encrypted with a Monoalphabetic Substitution cipher. Let's try this intercepted message:

GFS WMY OG LGDVS MF SFNKYHOSU ESLLMRS, PC WS BFGW POL DMFRQMRS, PL OG CPFU M UPCCSKSFO HDMPFOSXO GC OIS LMES DMFRQMRS DGFR SFGQRI OG CPDD GFS LISSO GK LG, MFU OISF WS NGQFO OIS GNNQKKSFNSL GC SMNI DSOOSK. WS NMDD OIS EGLO CKSJQSFODY GNNQKKPFR DSOOSK OIS 'CPKLO', OIS FSXO EGLO GNNQKKPFR DSOOSK OIS 'LSNGFU' OIS CGDDGWPFR EGLO GNNQKKPFR DSOOSK OIS 'OIPKU', MFU LG GF, QFOPD WS MNNGQFO CGK MDD OIS UPCCSKSFO DSOOSKL PF OIS HDMPFOSXO LMEHDS. OISF WS DGGB MO OIS NPHISK OSXO WS WMFO OG LGDVS MFU WS MDLG NDMLLPCY POL LYEAGDL. WS CPFU OIS EGLO GNNQKKPFR LYEAGD MFU NIMFRS PO OG OIS CGKE GC OIS 'CPKLO' DSOOSK GC OIS HDMPFOSXO LMEHDS, OIS FSXO EGLO NGEEGF LYEAGD

PL NIMFRSU OG OIS CGKE GC OIS 'LSNGFU' DSOOSK, MFU OIS CGDDGWPFR EGLO NGEEGF LYEAGD PL NIMFRSU OG OIS CGKE GC OIS 'OIPKU' DSOOSK, MFU LG GF, QFOPD WS MNNGQFO CGK MDD LYEAGDL GC OIS NKYHOGRKME WS WMFO OG LGDVS.

The first step is to find the frequency of all the letters appearing in the intercept. For this intercept we get the values given in the table below.

Ciphertext Letter	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
Frequency	5	2	26	42	23	51	67	8	33	1	35	39	35	29	85	30	14	17	88	0	17	3	16	6	10	0

If we re-order from most frequent to least:

Ciphertext Letter	S	O	G	F	D	L	K	M	I	P	N	C	E	R	U	W	Q	Y	H	X	A	V	B	J	T	Z
Frequency	88	85	67	51	42	39	35	35	33	30	29	26	23	17	17	16	14	10	8	6	5	3	2	1	0	0

Now that we have all the frequencies of ciphertext letters, we can start to make some substitutions. We see that the most common ciphertext letter is "S", closely followed by "O". From the chart and table above, we can guess that these two letters represent "e" and "t" respectively, and after making these substitutions we get:

GFe WMY tG LGDVe MF eFNKYHteU EeLLMRe, PC We BFGW PtL DMFRQMRe, PL tG CPFU M UPCCeKeFt HDMPFteXt GC tIe LMEe DMFRQMRe DGFR eFGQRI tG CPDD GFe LIeet GK LG, MFU tIeF We NGQFt tIe GNNQKKeFNeL GC eMNI DetteK. We NMDD tIe EGLt CKeJQeFtDY GNNQKKPFR DetteK tIe 'CPKLt', tIe FeXt EGLt GNNQKKPFR DetteK tIe 'LeNGFU' tIe CGDDGWPFR EGLt GNNQKKPFR DetteK tIe 'tIPKU', MFU LG GF, QFtPD We MNNGQFt CGK MDD tIe UPCCeKeFt DetteKL PF tIe HDMPFteXt LMEHDe. tIeF We DGGB Mt tIe NPHIeK teXt We WMFt tG LGDVe MFU We MDLG NDMLLPCY PtL LYEAGDL. We CPFU tIe EGLt GNNQKKPFR LYEAGD MFU NIMFRe Pt tG tIe CGKE GC tIe 'CPKLt' DetteK GC tIe HDMPFteXt LMEHDe, tIe FeXt EGLt NGEEGF LYEAGD PL NIMFReU tG tIe CGKE GC tIe 'LeNGFU' DetteK, MFU tIe CGDDGWPFR EGLt NGEEGF LYEAGD PL NIMFReU tG tIe CGKE GC tIe 'tIPKU' DetteK, MFU LG GF, QFtPD We MNNGQFt CGK MDD LYEAGDL GC tIe NKYHtGRKME We WMFt tG LGDVe.

(notice that the decoded letters have been shown in lowercase, which helps you to figure out which ones have been decoded, and which ones you still need to figure out)

We now notice that the word "tIe" is appearing frequently in the passage. In English, the most common 3 letter word is "the" and this fits with what we have already done, which suggests that "I" should be decrypted to "h".
Also, by looking at the frequencies again, we see the next most common letter is "G", which is probably one of "a", "i" or "o". We see that the third word is "tG", and the only one of these options that makes sense is "to", so we guess "G" is "o".

So now we have:

oFe WMY to LoDVe MF eFNKYHteU EeLLMRe, PC We BFoW PtL DMFRQMRe, PL to CPFU M UPCCeKeFt HDMPFteXt oC the LMEe DMFRQMRe DoFR eFoQRh to CPDD oFe LheeT oK Lo, MFU theF We NoQFt the oNNQKKeFNeL oC eMNh DetteK. We NMDD the EoLt CKeJQeFtDY oNNQKKPFR DetteK the 'CPKLt', the FeXt EoLt oNNQKKPFR DetteK the 'LeNoFU' the CoDDoWPFR EoLt oNNQKKPFR DetteK the 'thPKU', MFU Lo oF, QFtPD We MNNoQFt CoK MDD the UPCCeKeFt DetteKL PF the HDMPFteXt LMEHDe. theF We DooB Mt the NPHheK teXt We WMFt to LoDVe MFU We MDLo NDMLLPCY PtL LYEAoDL. We CPFU the EoLt oNNQKKPFR LYEAoD MFU NhMFRe Pt to the CoKE oC the 'CPKLt' DetteK oC the HDMPFteXt LMEHDe, the FeXt EoLt NoEEoF LYEAoD PL NhMFReU to the

CoKE oC the 'LeNoFU' DetteK, MFU the CoDDoWPFR EoLt NoEEoF LYEAoD PL NhMFReU to the CoKE oC the 'thPKU' DetteK, MFU Lo oF, QFtPD We MNNoQFt CoK MDD LYEAoDL oC the NKYHtoRKME We WMFt to LoDVe.

The first word is now "oFe", which when considered with the appearance of "theF", leads us to the conclusion that "F" is "n". This also fits in with the frequencies of both letters in the tables.

In the third line we see the word "Lheet", which is most likely to be "sheet", and so we replace "L" with "s". Again, the frequencies of these two letters are about right.

Now we have:

one WMY to soDVe Mn enNKYHteU EessMRe, PC We BnoW Pts DMnRQMRe, Ps to CPnU M UPCCeKent HDMPnteXt oC the sMEe DMnRQMRe DonR enoQRh to CPDD one sheet oK so, MnU then We NoQnt the oNNQKKenNes oC eMNh DetteK. We NMDD the Eost CKeJQentDY oNNQKKPnR DetteK the 'CPKst', the neXt Eost oNNQKKPnR DetteK the 'seNonU' the CoDDoWPnR Eost oNNQKKPnR DetteK the 'thPKU', MnU so on, QntPD We MNNoQnt CoK MDD the UPCCeKent DetteKs Pn the HDMPnteXt sMEHDe. then We DooB Mt the NPHheK teXt We WMnt to soDVe MnU We MDso NDMssPCY Pts sYEAoDs. We CPnU the Eost oNNQKKPnR sYEAoD MnU NhMnRe Pt to the CoKE oC the 'CPKst' DetteK oC the HDMPnteXt sMEHDe, the neXt Eost NoEEon sYEAoD Ps NhMnReU to the CoKE oC the 'seNonU' DetteK, MnU the CoDDoWPnR Eost NoEEon sYEAoD Ps NhMnReU to the CoKE oC the 'thPKU' DetteK, MnU so on, QntPD We MNNoQnt CoK MDD sYEAoDs oC the NKYHtoRKME We WMnt to soDVe.

We see the world "soDVe", which could be "solve", implying the transformations of "D" and "V" to "l" and "v" respectively.
In the second line we now have the phrase "one sheet oK so", which suggests that "K" is "r".

So now we have:

one WMY to solve Mn enNrYHteU EessMRe, PC We BnoW Pts lMnRQMRe, Ps to CPnU M UPCCerent HlMPnteXt oC the sMEe lMnRQMRe lonR enoQRh to CPll one sheet or so, MnU then We NoQnt the oNNQrrenNes oC eMNh letter. We NMll the Eost CreJQentlY oNNQrrPnR letter the 'CPrst', the neXt Eost oNNQrrPnR letter the 'seNonU' the ColloWPnR Eost oNNQrrPnR letter the 'thPrU', MnU so on, QntPl We MNNoQnt Cor Mll the UPCCerent letters Pn the HlMPnteXt sMEHle. then We looB Mt the NPHher teXt We WMnt to solve MnU We Mlso NlMssPCY Pts sYEAols. We CPnU the Eost oNNQrrPnR sYEAol MnU NhMnRe Pt to the CorE oC the 'CPrst' letter oC the HlMPnteXt sMEHle, the neXt Eost NoEEon sYEAol Ps NhMnReU to the CorE oC the 'seNonU' letter, MnU the ColloWPnR Eost NoEEon sYEAol Ps NhMnReU to the CorE oC the 'thPrU' letter, MnU so on, QntPl We MNNoQnt Cor Mll sYEAols oC the NrYHtoRrME We WMnt to solve.

In the middle of the second line we have the word "enoQRh", which is likely to be "enough", and so we have the transformations "Q" and "R" to "u" and "g" respectively, and we get:

one WMY to solve Mn enNrYHteU EessMge, PC We BnoW Pts lMnguMge, Ps to CPnU M UPCCerent HlMPnteXt oC the sMEe lMnguMge long enough to CPll one sheet or so, MnU then We Nount the oNNurrenNes oC eMNh letter. We NMll the Eost CreJuentlY oNNurrPng letter the 'CPrst', the neXt Eost oNNurrPng letter the 'seNonU' the ColloWPng Eost oNNurrPng letter the 'thPrU', MnU so on, untPl We MNNount Cor Mll the UPCCerent letters Pn the HlMPnteXt sMEHle. then We looB Mt the NPHher teXt We WMnt to solve MnU We Mlso NlMssPCY Pts sYEAols. We CPnU the Eost oNNurrPng sYEAol MnU NhMnge Pt to the CorE oC the 'CPrst' letter oC the HlMPnteXt sMEHle, the neXt Eost NoEEon sYEAol Ps NhMngeU to the CorE oC the 'seNonU' letter, MnU the ColloWPng Eost NoEEon sYEAol Ps

NhMngeU to the CorE oC the 'thPrU' letter, MnU so on, untPl We MNNount Cor Mll sYEAols oC the NrYHtogrME We WMnt to solve.

We have the word "Nount" which is could be "count" and "EessMge" which is likely to be "message", giving us that "N", "E" and "M" and "c", "m" and "a". We're getting closer!

one WaY to solve an encrYHteU message, PC We BnoW Pts language, Ps to CPnU a UPCCerent HlaPnteXt oC the same language long enough to CPll one sheet or so, anU then We count the occurrences oC each letter. We call the most CreJuentlY occurrPng letter the 'CPrst', the neXt most occurrPng letter the 'seconU' the ColloWPng most occurrPng letter the 'thPrU', anU so on, untPl We account Cor all the UPCCerent letters Pn the HlaPnteXt samHle. then We looB at the cPHher teXt We Want to solve anU We also classPCY Pts sYmAols. We CPnU the most occurrPng sYmAol anU change Pt to the Corm oC the 'CPrst' letter oC the HlaPnteXt samHle, the neXt most common sYmAol Ps changeU to the Corm oC the 'seconU' letter, anU the ColloWPng most common sYmAol Ps changeU to the Corm oC the 'thPrU' letter, anU so on, untPl We account Cor all sYmAols oC the crYHtogram We Want to solve.

It is likely that "W"->"w", "X"->"x", "Y"->"y" and "Z"->"z".
The word "occurrPng" is clearly meant to read "occurring", and it is likely that "sYmAol" is "symbol".

one way to solve an encryHteU message, iC we Bnow its language, is to CinU a UiCCerent Hlaintext oC the same language long enough to Cill one sheet or so, anU then we count the occurrences oC each letter. we call the most CreJuently occurring letter the 'Cirst', the next most occurring letter the 'seconU' the Collowing most occurring letter the 'thirU', anU so on, until we account Cor all the UiCCerent letters in the Hlaintext samHle. then we looB at the ciHher text we want to solve anU we also classiCy its symbols. we CinU the most occurring symbol anU change it to the Corm oC the 'Cirst' letter oC the Hlaintext samHle, the next most common symbol is changeU to the Corm oC the 'seconU' letter, anU the Collowing most common symbol is changeU to the Corm oC the 'thirU' letter, anU so on, until we account Cor all symbols oC the cryHtogram we want to solve.

And finally we can see that "C" is "f", "B" is "k", "U" is "d", "J" is "q" and "H" is "p".

one way to solve an encrypted message, if we know its language, is to find a different plaintext of the same language long enough to fill one sheet or so, and then we count the occurrences of each letter. we call the most frequently occurring letter the 'first', the next most occurring letter the 'second' the following most occurring letter the 'third', and so on, until we account for all the different letters in the plaintext sample. then we look at the cipher text we want to solve and we also classify its symbols. we find the most occurring symbol and change it to the form of the 'first' letter of the plaintext sample, the next most common symbol is changed to the form of the 'second' letter, and the following most common symbol is changed to the form of the 'third' letter, and so on, until we account for all symbols of the cryptogram we want to solve.

We can also now recover the key used in the encryption by putting together the ciphertext alphabet. This is useful if we have other messages intercepted from the same person, as it is likely that they will be using the same key (or a rotation of two or three keys). The key (a phrase that started the cipher) in this case was "manuscript".

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

Some Additional Resources:

British International School, Phuket - who compiled many of these challenges.

Crypto Corner is the newest and best code making and code breaking website online – it's got a huge amount of code information and also allows you to generate your own codes.

CIMT Code resources – a fantastic resource with a large number of ready made worksheets and teacher notes on lots of different codes.

Secret Codebreaker also has a lot of information about different codes

Counton website to generate different codes – generate your own codes

Nrich has a nice article about the history of codes and mathematics

NASA codes from the stars – more explanation on binary string codes.

Khan Academy code breaking videos– a large number of short videos looking at both codes through history and more modern code methods.