Preliminary Research into Ciphers

Definitions

Code – The set of letters, number, symbols etc, that are used to encrypt messages sent to others.

Cipher – A way of changing a message in order to keep it secret.

Cryptography – The process of writing or reading secret messages or codes.

Encrypt – To change information from one form to another, in order to disguise the true meaning of the message.

The Origins of Ciphers

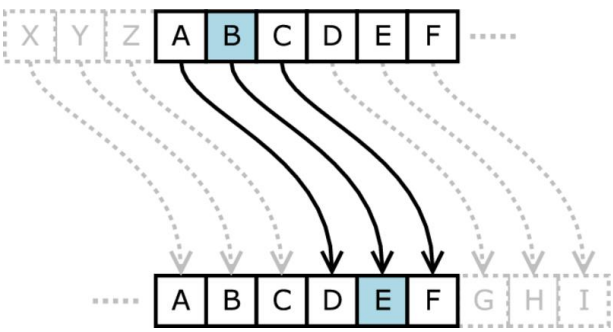
The first ciphers that we know of today originated in Ancient Egypt, created over 4,000 years ago in Menet Khufu, as an adaptation of hieroglyphics, in the tomb of Khnumhotep. Though generally agreed of as a cipher, in many ways, its purpose was in fact much different to what we might term as a cipher. It was as much to impart dignity and authority as to keep something separate. It was intended to impress the reader, as they understood the text, yet it was interesting and novel. Though this inspired the development of such ciphers, simple adaptations of the other Cyrillic scripts, as Professor Owen Lattimore of the University of Leeds states, 'literacy was always restricted to such a small minority that the mere act of putting something into writing that a mere act of putting something into writing was to a certain extent equivalent to putting it into code'. In many ways, this was a constant for the ancient world where messages would be hand delivered, and few could read it, even without a cipher applied. In many ways, the first true attempts to produce secretive codes were in India, where Artha-satra described the ways of the espionage services of India in the years prior to 500BC, where the tasks and reports were provided to superiors using code writing. In fact, Lalitavistara Sutra, which tells of the life of Gautama Buddha describes the now god-like figure's prestigious skill of encryption, while the Karma Sutra, the infamous text describing how one reaches a virtuous life, but today is known for its graphic descriptions of human sexual behaviours, describes 'secret writing' (Miecchita-Vikalpa) as the 45th skill that all women should be proficient in. At this point, the ciphers were largely restricted to substitution ciphers, where consonants and vowels were swapped, or there were random reciprocations.

Types of Ciphers

As I intend to make use of at least one cipher to ensure that the passwords stored in the system are protected and even hacking into the code would prevent anyone from accessing the secure data. Thus, one must look at the various ciphers that exist, and the various successes and flaws of the ciphers, including the feasibility of coding and how hackable they are, both by hand and by software. The complexity and thereby generally how secure the ciphers are largely also progress chronologically.

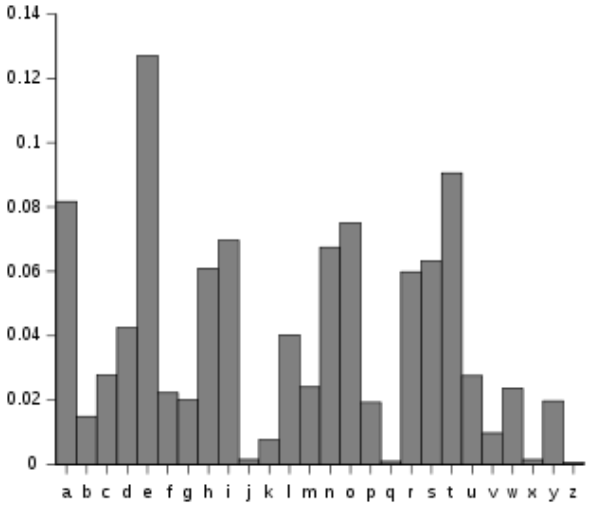
Shift Ciphers – Caesar Ciphers

One of the earliest form of ciphers, as used famously by Julius Caesar, to ensure secrecy in messages was a shift cipher where each letter of the message was moved a certain number of characters up or down the alphabet, such that a shift encryption with a 3-shift of CAT is FDX. This is in essence a very simple cipher, and relied mostly upon the reader not spending the time to break the number of possibilities. This is clearly 26,



representing the number of possible shifts that could occur, each one producing a different result. This form of cipher of represented with the figure.

Frequency Analysis



Over time, a more sophisticated form of breaking shift ciphers and in fact a number of other ciphers, was developed by the Arab polymath, Al Kindi, in his paper, 'A Manuscript on Deciphering Cryptographic Messages', by investigating the text of the Qu'ran. The process relies upon the fact that in certain languages the frequency in which certain letters turn up is characteristic. Thus, by investigating the frequency of letters in the encrypted text, one can easily see the code, as the graph would likely be shifted to the side, such that the peaks and troughs would be moved a certain amount. Thus, one can easily find the likely shifts, reducing the number of attempts required to find the exact shift used.

Substitution Ciphers

A substitution cipher is an expansion on the shift cipher, where a predesigned code (such as DFHU…) is used to define where the letters of the message to be encrypted are transposed to, such that in this example: CAB becomes HDF. The number of attempts that would be required to find the specific code in this is much higher, such that it is much harder to break by hand, in fact this is 26! (26x25x24x23x22x21…x3x2x1) = 403291461126605635584000000, since A could be replaced by 26 characters (A through Z), B by 25 (A through Z, except the replacement for A) and so forth. Though in essence the code is easy to implement, one must ensure that both parties and no one else knows the code, generally a challenge, given it should be a random string of 26 characters. Additionally, this is possible to break, using a variation of Frequency Analysis, since one can tend to find the most frequent character is likely to be an E in real life and so forth. Additionally, most systems would also make use of the fact that English and most other languages include a number of repeated words, which can be isolated to find the specific substitutions that are used in that case. Finally, it must be noted that once a few substitutions have been found, the number of possibilities to check reduces very quickly, while for many things to be readable, the entire code is unrequired. For example, it is trivial to work out that 'T?E TI?E IS ?ID?IGH?' is likely to be 'THE TIME IS MIDNIGHT' when the entire substitution code is found.

Polyalphabetic Ciphers – Vigenere Ciphers

The polyalphabetic cipher was the next step, after the Shift Cipher, where each character was shifted by a different number of position. It included a way in which the code (as a word) itself could be memorised, thus making the process of being the coder and the decipherer much easier. The process works in the following manner, let's say the code is RABBIT, and the message to be coded is HELLOMYNAMEISASHWIN:

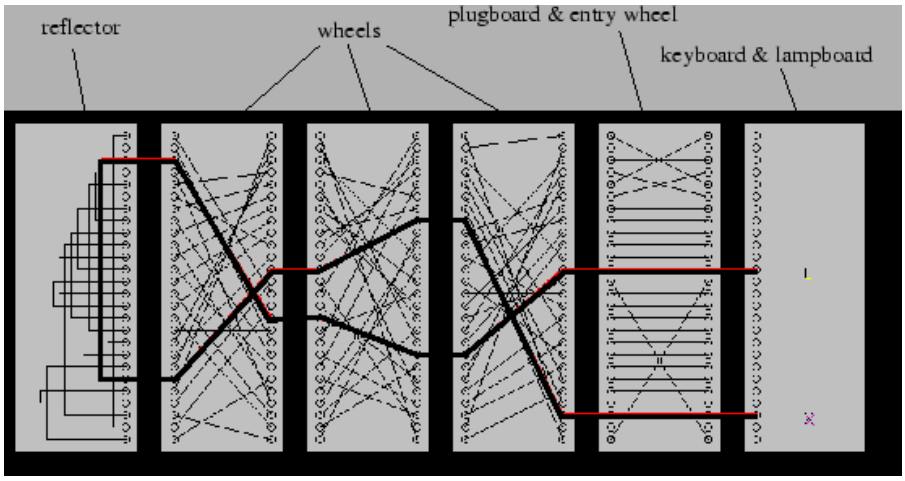
RABBIT = 18 1 2 2 9 20

	H	E	L	L	O	M	Y	N	A	M	E	I	S	A	S	H	W	I	N
	R	**A**	**B**	**B**	**I**	**T**	**R**	**A**	**B**	**B**	**I**	**T**	**R**	**A**	**B**	**B**	**U**	**T**	**R**
	8	5	12	12	15	13	25	14	1	13	5	9	19	1	19	8	24	9	14
+	18	1	2	2	9	20	18	1	2	2	9	20	18	1	2	2	9	20	18
	26	6	14	14	24	33	43	15	3	15	14	29	37	2	21	10	33	29	32
=	26	6	14	14	24	7	17	15	3	15	14	3	11	2	21	10	7	3	6
=	**Z**	**F**	**N**	**N**	**X**	**G**	**Q**	**O**	**C**	**O**	**N**	**C**	**K**	**B**	**U**	**J**	**G**	**C**	**F**

Thus, the message, when encrypted becomes: ZFNNXGQOCONCKBUJGCF, which is entirely incoherent, except to someone who is aware that the code is RABBIT, and so can easily reverse the encryption by reversing the change. Though generally simple, the cipher's security relies upon the key remaining private. For example, if Alice were sending a message to Bob, the general example in which the interceptor Eve, is attempting to be bypassed, only remains secure as long as Eve does not know it. In practise, this was used when the two people knew each other well, so were able to securely (in person) choose such a word, without others knowing it. Additionally, the cipher is still intensely vulnerable to the method of Frequency Analysis, by investigating the letters at different intervals, such that the graph of expected frequency distribution will be eventually found, thus allowing us to know the length of the code word. From here, the task is as simple as breaking a number of shift ciphers, a painful, but doable task for a person, while a trivial task for a computer. Despite the apparent and clear flaw in the system which was generally well known, the use of polyalphabetic ciphers continued for a long period, only truly being superseded at the end of the first world war, by more sophisticated systems such as Enigma. In this period, ciphers such the Viginere's cipher (used by the French in WW1), were made more complex by utilising a number of polyalphabetic ciphers, each with different code words, thus drastically increasing the number of calculations that an intercepting enemies would need to do, to break the message.

Enigma

A final adaptation of the Substitution Cipher was Enigma, the code used during the Second World War by the Nazi Germans. Though there was no repetition, the system used a systematic approach to produce the codes. Firstly, a sender would type the letters of a message into a keyboard. This would create a signal which would pass through the plug board, which systematically switched the signals to those of other letters. Then this signal would pass through a set of 3 (later became 4) rotors which interconnected letters. However, these rotors changed at intervals, with the first rotor spinning every letter, the second after 26, the 3rd after 26x26 and so forth, which changed which character linked to which. Then this now changed character was made then lit up a specific letter on the light board, under the keyboard, to be used. Though originally one would assume the key space was 263, the rotors could also be rearranged and there were over 60 rotors to choose from, thus leading to a key space of over 60x59x58x57x264 = over 5 x 1012 possible settings, far too many to check by hand. In fact, when one also included the plugboard settings, where the original simple substitution occurred, the number of keys were as follows:



$$Total\ Number\ of\ Keys\ (with\ 4\ rotors) = 26^4 \times \left(\frac{26!}{14! \times 6! \times 64}\right)$$

$$= 4.5 \times 10^{16}$$

However, the system was ultimately flawed due to one simple reason, that one letter could never encrypt to itself. This was due to the electrical wiring in the system which meant that the letter connector could not connect to itself after the reflector. This combined with the knowledge that certain crypts (keywords) such as 'Heil Hitler' would always be in the messages, thus vastly reducing the number of settings that one needed to test.

One-Time Pad

According to many, Claude Shannon is the father of modern cryptography, from here, many of the ciphers discussed from here on are still used. Shannon's work addressed the 'problems of cryptography'. He largely separated the types of cryptography into two, one where the cipher was designed to protect against hackers who have infinite resources, now known as unconditional secrecy and a second, where the cipher proytects against hackers with finite amount of resources. He also began to define the idea of 'Perfect Secrecy' the cipher text conveys 'no information about the content of the plaintext.' The manner in which this can occur is using the One-Time Pad, where frequency analysis and other cryptanalytic techniques would have no effect upon the encrypted text. The One-Time Pad is similar to the Polyalphabetic and Substitution Cipher except that the code (key) is entirely random and as long as the text to encoded, such that there is no repetition of the entire code, which left the polyalphabetic code vulnerable. Shannon definitely proved that perfect secrecy was only possible if this was true. Notice, the code is required to be perfectly random, as opposed to pseudo-random, two entirely different concepts. If one were to approach a person and ask them for four random numbers, they are more likely to choose certain numbers, in fact to a certain extent a large flaw and the reason for the failure of the Enigma System. If the code was entirely random, the code is very hard to break, as there are few ways of breaking it, lest a form of trial and error, where the number of possibilities are 26n where n is the length of the code to be encrypted.

An example of the use of a One-Time Pad

Text to be encrypted: EVEYOUCANTHEARMENOW

Code to be used: 23 19 14 4 26 16 8 8 10 13 10 26 14 20 2 13 15 4 15

	E	V	E	Y	O	U	C	A	N	T	H	E	A	R	M	E	N	O	W
	23	19	14	4	26	16	8	8	10	13	10	26	14	20	2	13	15	4	15
+	5	22	5	25	15	21	3	1	14	20	8	5	1	18	13	5	14	15	23
=	28	41	19	29	41	37	11	9	24	33	18	31	15	38	15	18	29	19	38
=	2	15	19	3	15	11	11	9	24	7	18	5	15	12	15	18	3	19	12
=	B	O	S	C	O	K	K	I	X	G	R	E	O	L	O	R	C	S	23

Though the One-Time Pad is by far the most sophisticated type of cipher that we have encountered so far, there are a few issues that it introduces. Firstly, the length of the code must be at least as long if not longer than the message that the people wish to encrypt, thus taking up a lot of length. This makes it very difficult to remember and use effectively. Additionally, the question becomes how to ensure that both parties have the same code, to decrypt and encrypt the message, since much of the communications are not in person and rely upon inherently insecure systems such as the internet. The mechanisms of fixing these issues were fundamentally fixed in the DES and Diffie-Helman Key Exchange System.

Diffie-Helman Key Exchange – Assymetric Key Exchange

The next big breakthrough came in 1976, produced by Whitfield Diffie and Martin Helman, where they designed a manner in which key's could easily be exchanged, solving one of the theoretical problems in Shannon's description of the One-Time Pad. For the first time, the two parties (Alice and Bob) never needed to come into contact for the message to be secure. It established a method of key exchange called Assymetric Key Exchange, where both parties have both a 'Private Key' and a 'Public Key'. Ensuring that Eve does not have any access to the key uses a situation known as the Discrete Logarithm Problem.

How it works:

First, we must establish the concept of Modular Arithmetic, where we alter the base of a number. For example, 3 = 1 (mod 2). This is the same as calculating the division of 3/2 and then finding the remainder of this. This is an integral part of the method, since the modular function is very easy to calculate, but very hard to reverse.

So, to start off, both Alice and Bob decide (publically) on two prime numbers, where p is a prime, and q is a generator of p. A generator is a number that when raised to whole number (integer) powers less than p never produces the same result (every modulus is equally likely). These numbers can be distributed over the internet, thus, Eve is now aware of these numbers. From here, the pair each create their own personal key (a and b) and find another number using the following formula:

$$a' = q^a \, mod(p)$$

$$b' = q^b \, mod(p)$$

From here, they transfer a' and b' over the insecure network to each other (and thereby Eve), thus allowing them to communicate using the following formula:

$$Key\ (Bob) = a'^b \, mod(p)$$

$$Key\ (Alice) = \ b'^a \, mod(p)$$

As it turns out, these two are identical, since, by substituting how these keys were produced:

$$Key\ (Bob) = q^{a^b} mod(p) = q^{ab} \, mod(p)$$

$$Key\ (Alice) = q^{b^a} mod(p) = q^{ab} \, mod(p)$$

However, despite Alice knowing a', b', p and q, finding a and b, as would be required to find the key is a challenge which requires a large amount of computing power, with computers normally taking at least a decade to solve. This is due to a problem known as the Discrete Logarithm Problem. From here, the key can be used as the encryption to encrypt and decrypt messages as required without Eve being able to decrypt them as they are sent using the insecure internet connection.

Discrete Logarithm Problem

Let g be a generator of two integers; x and p (where p is a prime);

$$Answer = \ g^x \, mod(p)$$

For example:

$$3^{29} \, mod(17) = 12$$

However, knowing this, it is hard to find 29 given all the rest of the information, in fact relying upon trial and error to solve the reverse function:

$$Answer \, mod(17) = \ log_3(12)$$

RSA

Rijndael Cipher (Advanced Encryption Standard)