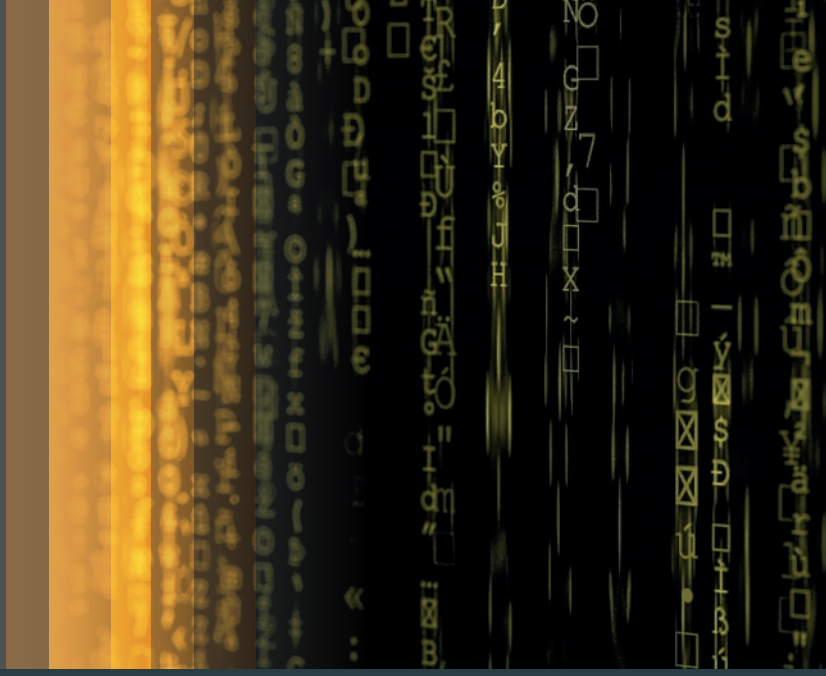
CYBER WORKS

Cryptography

Protecting the World's secrets.



Audience: Intermediate **Reading Time: 20 Mins**

Cryptography is used every day by almost everyone; and secures everything we do on the internet. But how has it developed from the days ofJulius Caesar to securing the internet today?

Key Points

- Classical ciphers are thousands of years old and can now be easily broken by hand.
- Breaking the Enigma machine saved at least 14 million lives and shortened World War II by a number of years.
- Quantum Computing aims to ensure key distribution is done securely, and ensures data is not intercepted.
- Symmetric and Asymmetric cryptography underpin internet security by encrypting data and ensuring websites are who they claim to be.
- Enigma was a mechanical cipher machine used by Germany in World War II. The code was broken by joint efforts from Britain, Poland, France and the United States.

Cryptography

Cryptography is a technique used to enable secure communication between different parties, it transforms data into another form so that only those who know how to transform the data back to the original can read it. **Plaintext** refers to the original, unaltered message. **Ciphertext** is what is created after the transformation is complete – this is called *encryption*. The reverse process, taking ciphertext and turning it into the original plaintext is called *decryption*. The different methods by which we perform these transformations are called **ciphers**, meanwhile cryptanalysis is the process of trying to break and find flaws in these ciphers.

Classical Cryptography

Classical ciphers fall into one of two categories: *substitution* ciphers or *transposition* ciphers. Substitution ciphers replace each letter in the plaintext with another in the ciphertext, meanwhile transposition ciphers keep the original letters, but rearrange the order of these in the ciphertext.

One of the simplest substitution ciphers is the **ROT13** cipher, it simply replaces each letter in the plaintext with the letter in the alphabet 13 places ahead of it; A goes to N, B to O, C to P etc. "HELLO" enciphered in ROT13 is URYYB. As the alphabet we use contains 26 characters, to decrypt ciphertext enciphered using ROT13, we simply apply ROT13 again to get back to the original. Can you work out the plaintext of EBZR? *Answer: at end of article*

Much like ROT13, the Caesar Cipher also replaces each plaintext letter with another letter further down the alphabet. However, instead of using a constant value of 13 places, the number of places changes each time it is used – this is called a key. A key details precisely how the transformation is applied. In the Caesar Cipher, the key specifies by how many places ahead in the alphabet the replacement character is. The ROT13 cipher is simply a Caesar Cipher with a key of 13.

Definitions

Plaintext - The original message that you wish to send.

Ciphertext - The message after it has been encrypted, nobody can read this message unless they have the key.

Cipher - A ciper is a method of hiding words or text with encryption by replacing original letters with other letters, numbers and symbols through substitution or transposition.

Frequency Analysis - The method to break classical ciphers by analysing the frequency of different letters in the ciphertext.

Caeser Cipher - A simple cipher created by Julius Caesar, where each letter of the alphabet in the plaintext is shifted a set number of letters in the ciphertext.

Symmetric Cipher - A cipher where the same key is used both for encryption and decryption.

Asymmetric Copher - A cipher where different keys are used for encryption and decryption.

To decrypt text enciphered with the Caesar Cipher, simply perform another Caesar Cipher with a key of 26 minus the original key. To aid with this, you can use a tool known as a cipher wheel such as this one https://bletchleypark.org.uk/blog/how-to-make-and-use-a-caesar-wheel

For example, the message ATTACK AT DAWN enciphered with a key of 9 is JCCJLT JC MJFW. Can you decrypt YNWP DRW B LJW COUH? *(Answer: at end of article)*

In both ROT13 and the Caesar Cipher, each plaintext letter always transforms to the same ciphertext letter, for example in ROT13 every H goes to U. This can make enciphering and deciphering easier, but it is also their downfall.

Frequency analysis is a type of cryptanalysis, it analyses the distribution of different letters in a piece of text. In every language, some letters occur more frequently than others. In the English language the letters E, T and A occur with the highest frequency. As both ROT13 and the Caesar Cipher always map each plaintext letter to the same ciphertext letter, these letter distributions are simply rearranged. For example, if the letter occurring most frequently in the ciphertext is P, then this likely corresponds to E, T or A. We can now try to decrypt the ciphertext with the keys 15 (if P = E), 4 (if P = T) and 11 (if P = A) and see which key fits best to unlock to ciphertext.

In addition to using the frequency distribution of single letters, we can also analyse the occurrences of sequences of letters in the same way. Pairs of letters are called *bigrams*, meanwhile *trigrams* are sequences of three letters. Another useful method is to identify single letter words or words with repeated letters in them. In English, "A" and "I" are the most common single letter words, meanwhile "ss", "ee" and "tt" are the most common sets of double letters. We can use all these methods to try and determine the key.

However, if we cannot determine the key using other methods, we can simply try to decipher the ciphertext with every possible key. This method is know as *brute force*. This is easy for ciphers with a small number of keys, especially using a computer, but will not work where there are a huge number of keys – as will become apparent later.

An improvement to substitution ciphers are *poly alphabetic* substitution ciphers – these are a form of substitution ciphers that use multiple alphabets. The Vigenère cipher is an extension of the Caesar Cipher, but it uses multiple different Caesar Ciphers with different keys. First, a keyword is used, and this determines which letters are encrypted with each key.

As a working example we will set our keyword as "LEMON" and our plaintext will be set as "ATTACKATDAWN". To encipher our plaintext, we perform a Caesar Cipher for each letter, with the key for the cipher corresponding to the next letter of our keyword. In our example, the first letter of our keyword is "L", and using the Caeser cipher means a key of 11; then applied to the first letter of the plaintext "A" gives us the letter "L". The next letter of our plaintext "T" is again enciphered with a Caesar Cipher devised from our keyword letter "E", this gives us a shift of 4, and the letter "X". As we carry on enciphering each letter with the Caesar cipher aligned with letter of the keyword, we keep looping back to the start if we reach the end of it. This leaves us with the ciphertext of LXFOPVEFRNHR.

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

Vigenère squares can help us with these transformations. A section of one is pictured above. The top row indicates the letter of the keyword, and the left-hand column indicates the plaintext letter we want to encrypt. The ciphertext letter is the letter at the intersection.

Frequency analysis is not effective against the Vigenère cipher. Multiple alphabets are used, so the mapping of plaintext letters to ciphertext letters is not constant. However, if we know the length of the key, we can determine which letters were enciphered with which alphabet. We can then use frequency analysis against each group enciphered with the same alphabet. But how do we work out the length of the key? Friedrich Kasiski in 1863 came up with the Kasiski Examination, a method used to determine the key length. If in a piece of ciphertext there is a repeated series of letters, then they will likely have been enciphered with the same key and correspond to the same plaintext, so the key is a multiple of the distance between the start of each group. By knowing the key, we can then use frequency analysis to determine the keys. For example, in the ciphertext:

NPMG FTGM ZPTT TSED GMBO PIMG LESP SPTT TSE

The sequence PTT TSE is repeated. The distance between these repetitions is 20 letters, so the key has a length of 20, 10, 5, 4, 2 or 1. Can you work out the plaintext?
(Answer: at end of article)

The Rail Fence or Zigzag Cipher is a transposition cipher. We must first choose a numeric key, which we use to encipher our plaintext message. Once we have a key, say 3, we write each letter of the plaintext diagonally downwards as many rows as the key, and then continue upwards until the top row and then go down again – like a zigzag! The ciphertext is then assembled by collecting all letters on each line together and then placing each line after each other. For example, the plaintext ATTACKATDAWN with a key of 3 will be written as:

A . . . C . . . D . . . .
. T . A . K . T . A . N
. . T . . . A . . . W . . .

These rows are then placed together and give a ciphertext of ACDTAKTANTAW. Decryption is done in reverse, but instead of collecting all letters on each row, we can simply read the zigzag to retrieve the plaintext. Cryptanalysis of the Rail Fence cipher can be awkward, so we can just try each different possible number of rails until we get reasonable plaintext. The number of rails is always between 2 and the number of letters in the ciphertext.

There are a huge number of classical ciphers that have not been mentioned so far: Playfair, Pigpen, Atbash, Affine, ADFGVX and Columnar Transposition, to name a few. Classical ciphers are not secure due to their relatively low number of keys and can mostly be broken by frequency analysis, or even brute forced by hand or a computer.

The Mechanisation of Cryptography

All classical ciphers we have seen so far have all been performed, and cracked, by hand. Cryptographic development at the start of the 19th century focussed on creating mechanical machines to avoid enciphering by hand – many of these are various forms of rotor machines. These rotor machines used multiple rotors that turn with every letter entered, enciphering each letter with a different alphabet. Fundamentally these are just mechanical machines performing poly alphabetic substitution ciphers.

One of the most well-known mechanical

cipher machines is the Enigma machine – created and used by Nazi Germany from the 1930s and used throughout the second world war. The Enigma machine was famously broken by Britain's Government Code and Cypher School (GC&CS) at Bletchley Park by mathematicians including Alan Turing, Joan Clarke, Gordan Welchman and thousands, and thousands of others. Their wartime achievements and sacrifices went unacknowledged for years until The Ultra Secret in 1974 by Frederick Winterbotham; however, many decided to never reveal their work. It is estimated that breaking the Enigma machine – and being able to read messages passed through the Enigma network – led to the shortening of the war by a number of years and saved at least 14 million lives worldwide. However, the work done at Bletchley Park – now a museum – would not have been possible without early work done by Polish mathematicians Marian Rejewski, Henryk Zygalski and Jerzy Różycki.

The Enigma machine appears just like a typewriter, it has three rotors under the keyboard with a lampboard above this and a plugboard at the front. To encipher a message, the three rotors were first set to a predetermined setting for that day; these were documented in a code book for each month – if you had the code book, you could decipher every message for that month. The plugboard also required setting up, this consisted of several leads representing letters, of which some of these were connected to each other. Once the rotors were in their correct starting position and the plugboard was set, the plaintext would be entered letter by letter. The machine would encipher the letter and light up the corresponding letter on the lampboard to indicate the ciphertext of the letter entered. One of the rotors turned with each key press, and at a predetermined point this would turn the next rotor and when that rotor had completed a full turn, the next rotor would turn. This process ensured that each letter was enciphered using a different alphabet. The result of this technique is a staggering 159 quintillion possible settings and thus alphabets – 15.9 billion billion, much too large to brute force!

Many techniques were developed to break into the Enigma, which throughout the war Germany deemed would simply be impossible. Code books containing Enigma settings were stolen, even an Enigma machine sent to the German Embassy in Poland was intercepted and copied, but there were also shortfalls in its operation. Some messages, such as weather reports, had set formats and the content of these messages could often be guessed – known as a 'known plaintext attack'! The guessed plaintext is called a *crib*. Machines called *Bombes* were designed and used at Bletchley Park to break Enigma in the later years of the war, these were huge electro-mechanical machines designed to aid cryptanalysts in determining the settings used for the Enigma.

Enigma was not the only cipher worked on at Bletchley Park, the *German Lorenz* and various Japanese ciphers were also worked on by staff at Bletchley Park. *Colossus* was a computer developed by Tommy Flowers at Bletchley Park to defeat the *German Lorenz* cipher; this work pioneered the development of modern computers.

Bletchley Park is now a museum and is open to the public and features many of the original huts that different teams worked in, alongside the original Victorian Gothic mansion. The

site is also home to the National Museum of Computing, where a Colossus machine has been rebuilt.

Modern Cryptography

Modern cryptography is a world away from its beginnings; where it now focuses on mathematics rather than language. The advent of modern cryptography was arguably in the 1970's with the Data Encryption Standard (DES). Modern cryptography falls into one of two categories: *symmetric cryptography* and *asymmetric cryptography*. *Symmetric* cryptography uses the same key for encryption and decryption, meanwhile *asymmetric* uses one key for encryption and a completely different key for decryption. All ciphers mentioned up to this point have all been symmetric, they all use the same, or a derived key, for both encryption and decryption.

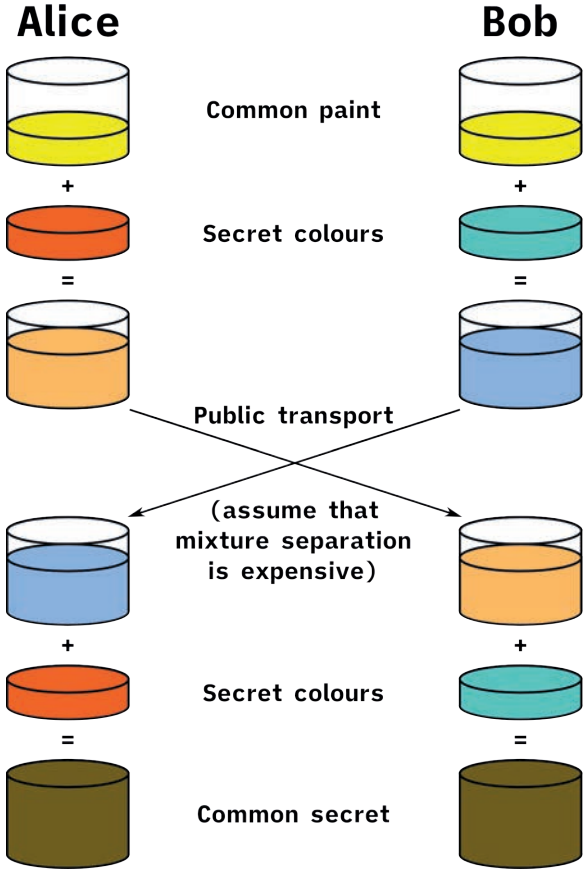
As symmetric encryption uses the same key for both encryption and decryption, this acts as a shared secret. DES was the first mainstream symmetric encryption algorithm in the 1970s. There are two types of symmetric encryption, stream ciphers and block ciphers. Stream ciphers encipher each unit of data at a time – much like classical ciphers did, taking each letter and transforming it to its ciphertext equivalent. Block ciphers, on the other hand, split the plaintext into blocks of constant sizes and encrypt each block as a unit. Commonly used symmetric ciphers are *ChaCha20, Blowfish, RC4* and *IDEA*.

Asymmetric cryptography, or public-key cryptography, uses two different keys, one for encryption (the public key) and another, separate key for decryption – the private key. The public key is meant to be open and known by anybody, hence public, while the private key should be kept secret and know only by the owner – hence private.

The *Diffie-Hellman* key exchange (sometimes Diffie-Hellman-Merkle) was one of the first uses of public-key cryptography. It enables two parties to determine a shared secret without having to transmit this shared secret.

Symmetric key cryptography requires the two parties to somehow exchange the key used for encryption and decryption at the other end, this was usually done by transferring the key using another channel; perhaps they met in person or used a courier.

The Diffie-Hellman key exchange works as shown in the image below. Alice and Bob start with the same colour paint; each mix their paint with another 'new colour', where each new colour is only known to them. They then send each other their newly mixed paint, and add their 'new colour' the received mix. Alice and Bob will now both have the same colour paint. If someone were to intercept the paint during the exchanged, they would not know the final colour without knowledge of each secret colour. Separating the paint to determine the initial colour is infeasible. Diffie-Hellman works in the same way but uses prime numbers instead of paint!



Cryptography

Symmetric, Asymmetric and Public-key cryptography systems are used everyday by almost everyone: every time you use the internet. HTTPS (Hyper Text Transport Protocol Secure) is a protocol used on the internet to access websites securely, it uses TLS (Transport Layer Security) to secure your connection to websites. It encrypts all data sent over the internet to websites and ensures that the website is who they claim to be and not an attacker pretending to be the website to steal your information. These two properties are called confidentiality and authentication.

Your browser needs to generate an encryption key to be used with a symmetric cipher that is used to secure the data sent between the browser and each website you use. It does this by using either Diffie-Hellman or a slightly different method called RSA. RSA, created by Ronald Rivest, Adi Shamir, and Leonard Adleman in the 1970s hence RSA – the first letter of each surname, performs a similar function to the Diffie-Hellman method, but works slightly differently. Once it has established a symmetric key to use, it can encrypt all data sent between your browser and the website.

The website provides authentication – proving it is the real website and not an attacker – by using a digital certificate. This certificate verifies that the website has proven itself to an external organisation – called a Certificate Authority (CA) – that they are indeed the website they claim to be. This certificate is generated using public key cryptography, the website presents its certificate to your browser and your browser then checks that the certificate is genuine and was issued to the website by the certificate authority. This process is known as signing – the certificate authority signs the certificate it gives to the website and this signature is what your browser checks.

Unbreakable Cryptography?

Quantum cryptography and quantum key distribution are both areas of research in quantum computing. In a quantum system, if a third party intercepts data sent between two parties, then the data that they view would alter what the receiver receives and the receiver would be able to determine that the data was intercepted. The quantum key distribution problem aims to securely exchange keys without these keys being able to be intercepted by anyone, the exchanged keys can then be used for symmetric ciphers.

Cryptography has a long history and has developed from linguistics and language to mathematics and physics; being an almost unrecognisable field from 100 years ago. Ciphers have developed from the simple Caesar Cipher through to the Enigma machine and public key cryptography which is essential to everyday life. Quantum cryptography aims to improve the security of exchange keys to make our data more secure.

Answers

Throughout this article there have been a few things to decipher. The answers are below:

[1] EBZR = ROME

[2] YNWP DRW B LJW COUH = PENGUINS CANT FLY.

[4] NPMG FTGM ZPTT TSED GMBO PIMG LESP SPTT TSE = CLASSICAL CIPHERS CAN BE EASY TO DECIPHER

READ MORE

1. Simon Singh. "The Code Book". https://simonsingh.net/books/the-code-book/
2. Gordan Welchman. "The Hut Six Story" ISBN13: 9780947712341
3. "Letter Frequencies in the English Language" Online at: https://www3.nd.edu/~busiforc/handouts/cryptography/Letter%20Frequencies.html [Accessed 20th September 2020]
4. "Cryptanalysis Hints" Online at: https://www3.nd.edu/~busiforc/handouts/cryptography/cryptography%20hints.html [Accessed 20th September 2020]
5. DCode. Online cryptography and cryptanalysis tools. Online at: https://www.dcode.fr/en [Accessed 20th September 2020]

About the Author

Henry Clarke is currently studying at Lancaster University. Henry has completed his BSc in Computer Science at Lancaster and is now studying his MSc in Cyber Security. Henry is part of Lancaster's Cyber Security society **LUHack**, his interests are cyber security and programming.