Cryptography

Introduction for General User

Cryptography comes from the Greek "kryptos" (hidden) and "graphos" (to write). It corresponds to text transformation in order to make it readable only by the legitimate receiver. This newsletter will illustrate the evolution of cryptography and its applications nowadays.

History

Cryptography has been used for thousands years to hide secret messages.

- **Caesar cipher**

Caesar Cipher is one of the first known ciphers. Its name came from Julius Caesar who used it to send secret messages to his generals on the front line[1]. The method used for encryption is substitution. In fact, each letter from the plaintext is shifted a certain amount of place in the alphabet between 1 and 25. If the shift is 1, A would be replaced by B, B would be replace by C…

Here is an example of cipher text:
ZHOOGRQH
(*The space between words has been removed*).

This kind of encryption is really easy to decipher. The first solution is the brute force approach which consists of trying the 25 possibilities as the shift is a number between 1 and 25. Another option is to use a frequency analysis. In fact, the letter E is the most commonly used letter in the English language. If the most common letter in a message is the letter G, it is likely that G represents the E. This technic is especially efficient for a long message. Try to decipher the message above; the solution is available at the end of the article.

The weakness of this cipher is that it depends on the system and not on a key. Once the system is known, the text can be quickly deciphered.

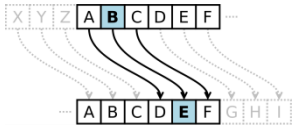

Figure 1: Caesar cipher with shift of 3[2]

- **Vigenere**

In the 16th century, Giovan Battista Bellaso created the first cypher encrypted with a key[3]. In fact, to encrypt a message a keyword has to be chosen for instance "SECRET". Ciphering a message becomes easy; the keyword has to be written under until it reaches the message length:

T H I S M E S S A G E I S E N C R Y P T E D
S E C R E T S E C R E T S E C R E T S E C R

	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

Figure 2: Vigenere table[4]

Then a table is used to encrypt the plaintext. The keystream S matches with the plaintext T. Thus, the first letter of the cipher text will be L which corresponds to the intersection between the row S (in yellow) and

the column T (in orange). When all the plaintext is encrypted, the result is the following cipher text:

LLKJQXKWCXIBKIPTVRHXGU

The method used is a polyalphabetic substitution. Even if the same letter appears twice in the plaintext, it might not have the same cipher text. In fact, in the example, the letter E is encrypted as a 'I', a 'X' and a G whereas T and H are both encrypted as 'L'.

To decrypt the cipher text the keyword is mandatory. As for encryption, the keyword is written under the cipher text until all letters match:

L L K J Q X K W C X I B K I P T V R H X G U
S E C R E T S E C R E T S E C R E T S E C R

The same table is used for the decryption. For the first letter, along the row S (in yellow) there is the letter L of the cipher text (in green) and by looking at the corresponding column (the orange one), the plaintext letter is revealed: T (in the blue box).

This cipher is much more secured than other polyalphabetic substitution because as mentioned above, the same letter can have different ciphers. The security of the process is based upon the secrecy of the key and not on the system as the previous one. Even if this system had remained unbreakable for a long time, it presents some weaknesses. In fact, the repeating nature of the key can help to guess its length. Then a frequency analysis of each block of the same length as the key can decipher the text.

- **Enigma**

In the twentieth century, substitutions could be done thanks to electrical connections. In 1918, in Berlin, Arthur Scherbius invented the Enigma machine[5]. This machine was used during the Second World War by the Germans to encrypt their communication.
The operator presses a letter on the keyboard, for instance the letter 'K'. This creates an electrical signal connected to the 'K' input on the plug board. In the plug board 'K' is wired to the letter 'T'. Then the signal passes into the static rotor and remains the same 'T'. Afterward, the signal is scrambled by three rotors among the five rotors available. Following the rotors, a reflector will turn the letter input in another and send it back among the three rotors. The three rotors work exactly in the same way on the return journey and finally, the signal reaches the static rotor again. Eventually, the plug board output is wired to a lamp corresponding to the encrypted letter. The operator writes down the cipher text and can then broadcast it.

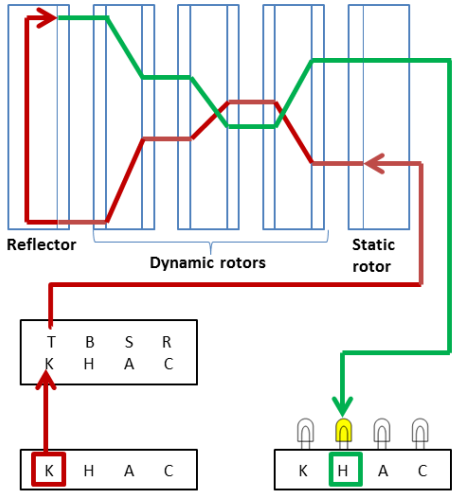

Figure 3: Enigma Machine

15 billion of billion different settings were possible to decipher a message. Thus, the Germans thought it was impossible to crack their cipher.

The primary weakness was that a letter could never be encrypted into itself. This enabled brute-force attack to guess the plug positions without trying all the options. However, Allies had to do it again every day as German modified the initial settings each day. Therefore, an automated device was created: the Bomb, and contributed to the Allies' victory.

Methods

Nowadays, encryption is wild spread and two methods are mainly used: symmetric and asymmetric encryption. For instance, Alice wants to send a message to Bob. But Alice does not want anyone else to be able to read her message. She can use different kinds of encryption.

- **Symmetric**

Alice uses a key to encrypt the message. Bob will receive the cipher text and will decrypt it using the same key as Alice. They share the same secret key.

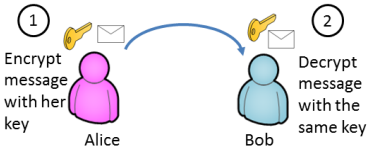

Figure 4: Symmetric encryption

This raises some issues especially for key exchange that has to be done before the exchange of the encrypted message. If someone succeed in grabbing the key during the exchange the security is not available anymore. The key length has also to be long enough to prevent people from cracking it with a brute force attack. Eventually, if Alice wants to send a message to Bob and to John she cannot use the same key and thus two keys will be needed.

DES/3-DES (Data Encryption Standard)
It is a famous symmetric-key algorithm for encryption. To encrypt a plaintext, DES groups it into 64-bit blocks which are then encrypted thanks to a 56-bit key and rounds of permutation and substitution in a block of the same length. Thanks to computational improvement, brute force attacks became feasible. Therefore, the 3-DES was created with a simple technique to increase the key length. 3-DES performs three iterations of the DES algorithm. A different key can be used with each iteration such that the total key length reached 168 bits.

AES (Advanced Encryption Standard)
It is a symmetric encryption used by the US government to protect classified documents[6]. AES encrypts data in block of 128 bits. Three keys length are available: 128, 192 or 256 bits. Depending on the size of the key a different number of rounds will be executed, respectively 10, 12 and 14. One round begins with a combination between the key and the block. Then, the output undergoes a non-linear substitution, a transposition, columns are mixed and each byte is combined again with the key. At the end a last round is done without the mix column step.

- **Asymmetric**

Asymmetric encryption uses two different keys for encryption and decryption. The issues linked to key exchange are solved. If Alice wants to send a message to Bob, they will both generate a pair of keys. They will place one of their key in a public register which will become the public key and will keep the other one secret: the private key. To send a confidential message to Bob, Alice encrypt the message with Bob's public key. The message can only be deciphered with Bob's private key. Thus, the confidentiality is preserved. Asymmetric encryption is safer but is longer than symmetric one.

To ensure the authenticity of a message, the sender can "sign" a message with its private key. He is the only one to be able to do it. The receiver will use the public key of the sender to check the message authenticity. This method is called digital signature.

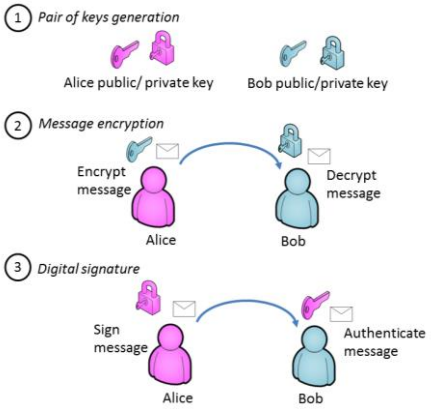

Figure 5: Asymmetric encryption

The same private key is used to decipher different messages whereas in symmetric encryption a new key has to be generated for each communication.

Number of participants	Number of symmetric key	Number of asymmetric key
2	1	4
10	45	20
100	4 950	200
1000	499 500	2 000

Table 1

Figure 6: Number of keys required depending on encryption type[7]

In order to prove the ownership of a public key, a digital certificate is provided.

RSA stands for R. Rivest, A. Shamir and L. Adleman the inventors of the algorithm[8]. The encryption is done with the public key.
Generation of the keys:

```
p and q two prime numbers
n=pq is the modulus
Φ(n)= (p-1)(q-1)
e an integer such as 0<e<Φ(n) and e is coprime with Φ(n)
d such as d*e=1 mod (Φ(n))
```

The public key is made of n and e.
The private key is made of n and d.

To encrypt the message M Alice will get the cipher text $C=M^e \bmod(n)$ that she will send to Bob. Bob can recover the plaintext with $M= C^d \bmod(n)$.

All the security of RSA relies on the confidentiality and the choice of p and q. Once at least one of these two figures is discovered, the code can be cracked.

Application To Encryption

Encryption is used in everyday life in order to protect data transmission.

Snowden reveals the NSA encryption weaknesses[9]

Edward Snowden is a former CIA employee who leaked top secret information from the U.S. National Security Agency (NSA) in 2013. The whistle-blower reveals that many encrypted communication have been cracked by the NSA to monitor citizens. For instance, data shared on Skype, even if they are encrypted, are accessible to NSA as well as data transferred by HTTPS connections. However some remain secured. Pretty Good Privacy still cannot be decrypted by NSA even if it was created in 1991. Tor, which is a network that enables anonymous communication, also remains a "major problem" for NSA.

- **Email**

Most email messages sent pass through a multitude of networks not all well-secured. Therefore, they can easily be read in plaintext by almost anyone. PGP (Pretty Good Privacy) is a software that enables email encryption[10]. It will generate a public and a private key. The private key will be directly stored in the email software whereas the public key can be published or shared with anyone. If someone tries to intercept an encrypted message it will be unreadable unless he knows the private key.

A session key can also be generated to ensure the communication between two computers. The security relies on the brevity of its use and thus the key has to be changed frequently.

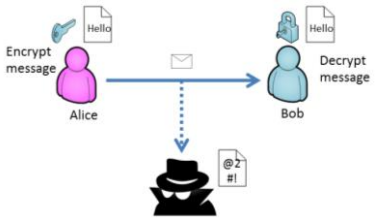

Figure 7: PGP encryption

- **Web browsing**

When a user wants to accede to a secure web page (https)[11], the browser requests a secured page; the web server sends its public key and its certificate (step 1). The browser checks that the certificates was issued by a trusted party, is still valid and related to the site. When all is correct, the browser uses the server public key to encrypt a symmetric key and sends it to the server with the encrypted URL and http data (step 2). The server uses its private key to decipher the symmetric key and then uses the symmetric key to decrypt the URL and http data (step3). The server sends back the html document and http data encrypted with the symmetric key (step 4). The browser decrypts everything and displays the information to the user (step 5).

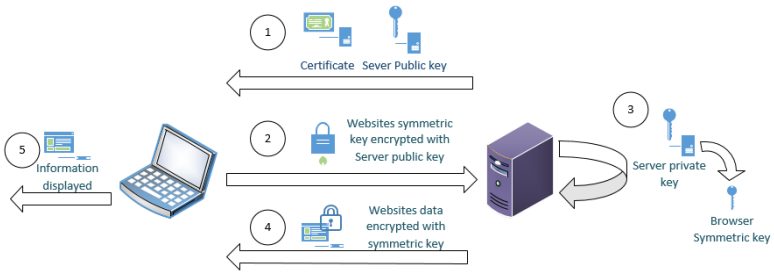

Figure 8: Web page encryption

- **Files encryption**

Files can be encrypted to prevent other users from gaining access to sensitive files.
In order to share encrypted files, the zip format can be used. .Zip format supports a symmetric encryption based on a password but it remains weak. Some new versions include AES encryption which is much more secured.
Some documents can also be encrypted with password. This kind of encryption is embedded in the software. For instance, on Word document the default encryption associated to password is AES 128-bit.

- **Octopus card**

The Octopus card uses a FeliCa smart card[12]. Since June 2011, FeliCa has been able to support AES encryption in addition to 3-DES. This algorithm is used to encrypt transactions between the card and the reader and also the reader and the controller. The encryption key is generated dynamically at each mutual authentication.

The Octopus cards and systems have never successfully been hacked yet.

Conclusion

Encryption is used everywhere to protect digital communication. Different kinds of encryption are available and can be used depending on the need. As computational capacity increases, cryptographic algorithms must continue to be improved to remain secured.

Solution of Caesar cypher: WELL DONE (the shift is 3)

Steganography

The steganography is the fact of hiding the existence of a message in order to keep it confidential whereas cryptography aims at preventing a third party from reading data. Messages can be hidden in text or even in images. For instance in a text, the first letter of each word can reveal a message. In an image, the modification of the least significant bit of each pixel is undetectable. Thus, messages can be transmitted by modifying these bits.

Steganography is now used in addition to cryptography to increase security.

References

1. "A brief history of cryptography" August 2013 Web. 20 July 2015
2. "Picture Caesar" October 2006 Web. 21 July 2015
3. "Vigenere Cipher" February 2015 Web. 20 July 2015
4. "Vigenere Table" April 2011 Web. 21 July 215
5. "The Enigma cipher machine" January 2015 Web. 21 July 2015
6. "Advanced Encryption Standard" November 2014 Web. 22 July 2015
7. "Encryption basis: How asymmetric and symmetric encryptions work" February 2010 Web. 22 July 2015
8. "RSA algorithm" November 2014 Web. 22 July 2015
9. "Inside the NSA war on Internet security" December 2014 Web. 12 August 2015
10. "How to encrypt your email" June 2006 Web. 22 July 2015
11. "How encryption secures communication on the Web" July 2015 Web. 22 July 2015
12. "Security analysis of the Octopus system" by A. Lee et al. 2007 pdf 20 July 2015

Copyright Statement