Cryptographie

IFT2105

H2012

(Alain Tapp)

Introduction

Depuis fort longtemps, les hommes ont tenté de rendre sécuritaires leurs communications confidentielles. Différentes techniques ont été utilisées.

Au début, il s'agissait seulement de cacher l'existence du message. Cette technique s'appelle la stéganographie.

Puis, des techniques de plus en plus sophistiquées furent utilisées pour rendre les messages compréhensibles seulement par leurs destinataires légitimes.

Tout au cour de l'histoire, une difficile bataille eut lieu entre les constructeurs de code (cryptographes) et ceux qui essayaient de les briser (les cryptanalystes). Il n'est toujours pas clair, même aujourd'hui, qui sera le vainqueur.

Stéganographie

Le plus ancien exemple de stéganographie a été rapporté par Hérodote. C'était lors du conflit entre la Grèce et la Perse au 5ième siècle av. J.-C.

Les Perses voulaient conquérir la Grèce et avaient préparé pendant 5 années une imposante armée. Heureusement pour les Grecques, Damaratus, un Grec exilé en Perse eu vent de ce projet.

Il inscrivit son message sur des tablettes de bois et les recouvrit de cire. Les tablettes avaient donc l'air vierges. Elles n'attirèrent pas l'attention des gardes tout au long du parcours.

Les Grecques, une fois mis au courant de l'attaque perse à venir, eurent le temps de se préparer et lors de l'attaque, ils mirent l'armé perse en déroute.

Stéganographie

Hérodote rapporte aussi l'histoire d'Histaïaeus qui, pour transmettre un message, rasa la tête de son messager et inscrivit le message sur son crane. Une fois les cheveux repoussés, le messager put circuler sans attirer l'attention.

Durant la Deuxième Guerre mondiale, les Allemands utilisaient la technique du micropoint. Il s'agit de photographier avec un microfilm le document à transmettre. La taille du microfilm était de moins d'un millimètre de diamètre. On plaçait le micropoint à la place du point final d'une lettre apparemment anodine.

En 1941, le FBI repéra le premier micropoint. De nombreux messages furent par la suite interceptés.

Chiffrement de César

Dans le célèbre film de Stanley Kubrick

2001: A Space Odyssey

un des personnages principaux est un super ordinateur appelé

HAL9000

Le film a été réalisé en 1969.

Est-ce qu'il y a un message caché dans le nom de l'ordinateur?

Chiffrement de César

Cette technique simple de chiffrement effectuant un décalage est appelé chiffrement de César.

Par exemple, avec un décalage de trois, mon nom devient

ALAIN TAPP = DODLQCWDSS

(On décale aussi les espaces…)

Cette technique de chiffrement est-elle sécuritaire?

Chiffrement de César

On intercepte le message

FAGEMYREMPURZV_EMZR_R FMNMDAZR

Essayons différents décalages…

1: **E_FDLXQDLOTQYUZDLYQZQZELMLC_YQ**

2: **DZECKWPCKNSPXTYCKXPYPYDKLKBZXP**

3… 4… 5… 6… 7… 8… 9… 10… 11… 12…

13: **TOUS_LES_CHEMINS_MENENT_A_ROME**

Clairement, le chiffrement de César n'est pas sécuritaire.

Cryptosystème a clef courte

Principe de Kerckhoff
(La cryptographie militaire 1883):

La sécurité d'un système de cryptographie ne doit pas dépendre de la préservation du secret de l'algorithme. La sécurité ne repose que sur le secret de la clef.

Substitution mono-alphabétique

Essayons autre chose.

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

TOUS_LES_CHEMINS_MENENT_A_ROME devient
FQLJRPAJRHCAE_ZJREAZAZFRDRNQEA

Le décodage devrait être plus difficile. Peut-on essayer tous les décodages possibles?

Il y a 27!=10 888 869 450 418 352 160 768 000 000 possibilités...

La substitution mono-alphabétique apparaît déjà dans le *kàma-sùtra* qui fut écrit au 5ième siècle mais qui est basé sur des écrits datant du 4ième siècle av. J.-C.

Le premier usage révélé de chiffrement par substitution dans un usage militaire est rapporté par Jules César dans *La guerre des Gaules*. César utilisait fréquemment le chiffrement et en particulier le décalage de trois caractères.

La substitution mono-alphabétique fut la technique de chiffrement la plus utilisée durant le premier millénaire. Nombreux savants de l'antiquité tenaient cette technique pour inviolable.

Ce sont les Arabes qui réussirent à briser ce code et qui inventèrent la cryptanalyse au 9ième siècle.

Exemple

BQPSNRSJXJNJXLDPCLDLPQBE_QRKJXHNKPKSJPJIKSPUN
BDKIQRBKPQPBQPZITEJQDQBTSKPELNIUNPHNKPBKPCKSS
QWKPSLXJPSNVVXSQCCKDJPBLDWPXBPSNVVXJPGKPJKDXI
PZLCEJKPGKSPSJQJXSJXHNKSPGPLZZNIIKDZKPGKSPGXV
VKIKDJKSPBKJJIKS

Comment déchiffrer ce message?

Chaque lettre est chiffrée de la même façon…
Certaines lettres sont utilisées plus souvent.

Occurrence des lettres

En français

_	19.3	L	4.7	H	0.8
E	13.9	O	4.1	G	0.8
A	6.7	D	2.9	B	0.6
S	6.3	P	2.5	X	0.4
I	6.1	C	2.4	Y	0.3
T	6.1	M	2.1	J	0.3
N	5.6	V	1.3	Z	0.1
R	5.3	Q	1.3	K	0.0
U	5.2	F	0.9	W	0.0

Dans le cryptogramme

P	14.3	D	4.6	W	1.0
K	12.8	L	4.1	U	1.0
S	9.2	V	3.1	T	1.0
J	9.2	Z	2.6	_	0.5
X	5.6	G	2.6	O	0.0
Q	5.6	C	2.6	M	0.0
N	5.6	E	2.0	F	0.0
B	5.1	R	1.5	A	0.0
I	4.6	H	1.5	Y	0.0

Remplaçons P par _ et K par E

BQ_SNRSJXJNJXLD_CLDL_QBE_QREJXHNE_ESJ_JIES_UN
BDEIQRBE_Q_BQ_ZITEJQDQBTSE_ELNIUN_HNE_BE_CESS
QWE_SLXJ_SNVVXSQCCEDJ_BLDW_XB_SNVVXJ_GE_JEDXI
_ZLCEJE_GES_SJQJXSJXHNES_G_LZZNIIEDZE_GES_GXV
VEIEDJES_BEJJIES

Remplaçons Q par A et B par L

LA_SNRSJXJNJXLD_CLDL_ALE_AREJXHNE_ESJ_JIES_UN
LDEIARLE_A_LA_ZITEJADALTSE_ELNIUN_HNE_LE_CESS
AWE_SLXJ_SNVVXSACCEDJ_LLDW_XL_SNVVXJ_GE_JEDXI
_ZLCEJE_GES_SJAJXSJXHNES_G_LZZNIIEDZE_GES_GXV
VEIEDJES_LEJJIES

Remplaçons **S** par S et **G** par D

LA_S**NR**S**JXJNJXLD**_**CLDL**_AL**E**_A**R**E**JXHN**E_ES**J**_**JI**ES_**UN**
L**D**E**I**A**R**LE_A_LA_**ZITEJ**A**D**AL**T**SE_**ELNIUN**_**HN**E_LE_**C**ESS
A**W**E_S**LXJ**_S**NVVX**SA**CC**E**DJ**_L**LDW**_**X**L_S**NVVXJ**_DE_**J**E**DXI**
_**ZLCEJ**E_DES_S**J**A**JX**S**JXHN**ES_D_**LZZNII**E**DZ**E_DES_D**XV**
VE**I**E**DJ**ES_LE**JJI**ES

Remplaçons J par T et I par R

LA_SNRSTXTNTXLD_CLDL_ALE_ARETXHNE_EST_TRES_UN
LDERARLE_A_LA_ZRTETADALTSE_ELNRUN_HNE_LE_CESS
AWE_SLXT_SNVVXSACCEDT_LLDW_XL_SNVVXT_DE_TEDXR
_ZLCETE_DES_STATXSTXHNES_D_LZZNRREDZE_DES_DXV
VEREDTES_LETTRES

Remplaçons **X** par I, **H** par Q et **N** par U

LA_SU**R**STITUTI**LD**_**CLDL**_AL**E**_A**R**ETIQUE_EST_TRES_**U**U
L**D**ERA**R**LE_A_LA_**Z**R**TE**TA**D**AL**T**SE_**EL**UR**U**U_QUE_LE_**C**ESS
A**W**E_S**L**IT_SU**VV**ISA**CC**E**D**T_L**LDW**_IL_SU**VV**IT_DE_TE**D**IR
_**ZLCE**TE_DES_STATISTIQUES_D_**LZZ**URRE**DZ**E_DES_DI**V**
VERE**D**TES_LETTRES

Remplaçons V par F et D par N

LA_SURSTITUTILN_CLNL_ALE_ARETIQUE_EST_TRES_UU
LNERARLE_A_LA_ZRTETANALTSE_ELURUU_QUE_LE_CESS
AWE_SLIT_SUFFISACCENT_LLNW_IL_SUFFIT_DE_TENIR
_ZLCETE_DES_STATISTIQUES_D_LZZURRENZE_DES_DIF
FERENTES_LETTRES

Remplaçons **R** par B et **L** par O

LA_SUBSTITUTION_**C**ONO_AL**E**_A**R**ETIQUE_EST_TRES_**U**U
LNERABLE_A_LA_**Z**R**TE**TANAL**T**SE_**E**OUR**U**U_QUE_LE_**C**ESS
A**W**E_SOIT_SUFFISA**CC**ENT_LON**W**_IL_SUFFIT_DE_TENIR
_**Z**O**CE**TE_DES_STATISTIQUES_D_O**ZZ**URREN**Z**E_DES_DIF
FERENTES_LETTRES

Finalement

LA_SUBSTITUTION_MONO_ALPHABETIQUE_EST_TRES_VU
LNERABLE_A_LA_CRYPTANALYSE_POURVU_QUE_LE_MESS
AGE_SOIT_SUFFISAMMENT_LONG_IL_SUFFIT_DE_TENIR
_COMPTE_DES_STATISTIQUES_D_OCCURRENCE_DES_DIF
FERENTES_LETTRES

Substitution+

Au lieu de faire la substitution mono-alphabétique, on peut rendre le code plus difficile à briser en faisant une substitution de mots. Chaque mot est remplacé par un nombre, d'où la nécessité d'un dictionnaire. On peut utiliser des synonymes.

Cette technique n'est pas vraiment pratique. La construction du dictionnaire est fastidieuse. Il faut se déplacer avec le dictionnaire qui pourrait être intercepté. Il est difficile de changer le code.

Substitution++

Différentes techniques peuvent être utilisées pour rendre le chiffrement par substitution plus sécuritaire tout en gardant une clef de taille raisonnable.

Premièrement, on peut utiliser des synonymes. Par exemple, la lettre E se retrouve 14% du temps et on pourrait utiliser 14 symboles différents pour représenter E et ainsi de suite pour les autres symboles.
On obtient un code de 100 symboles.

On peut aussi utiliser des blancs (symbole sans signification).

On peut coder certains mots courants par un seul symbole.

etc.…

Marie Stuart (1556)

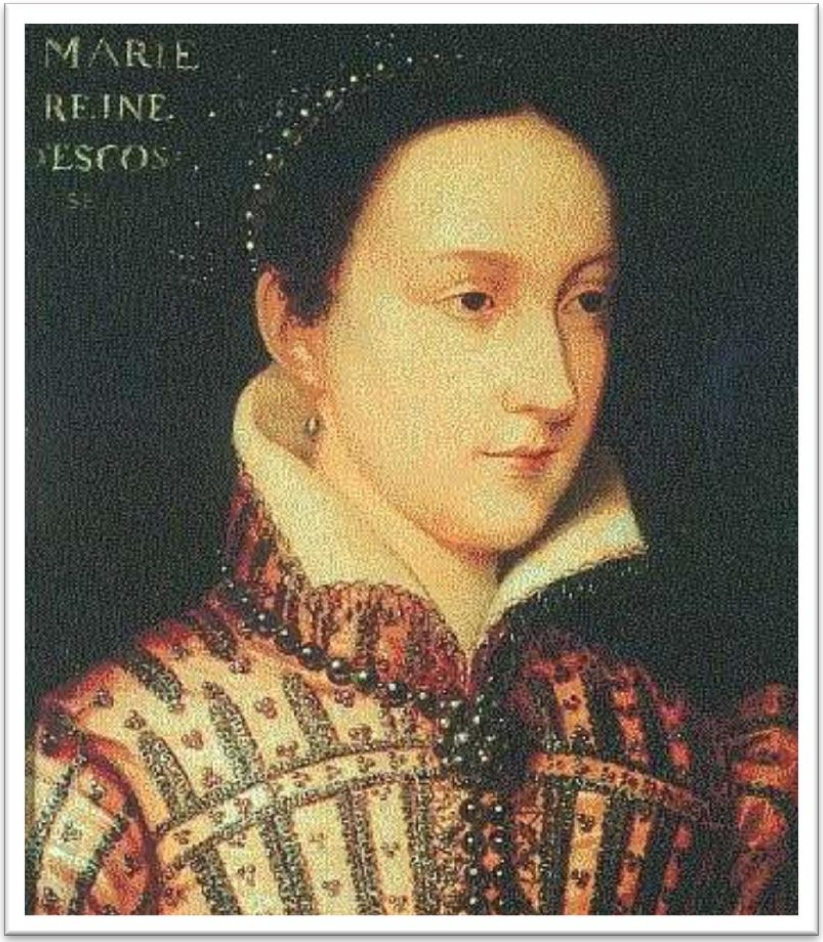
En 1586, Marie Stuart, reine d'Écosse fut jugée en Angleterre.

Elle était accusée d'avoir comploté pour assassiner la reine Elizabeth.

Le complot eut lieu durant son emprisonnement en Angleterre mais Marie utilisait le chiffrement lors de ses communications avec ses complices.

La Reine était réticente a exécuter Marie car elle était sa cousine. Le déchiffrement des lettres rendrait la preuve accablante et ne laisserait aucune chance à Marie.

Marie Stuart (1556)



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

Nulles

Dowbleth

and for with that if but where as of the from by

so not when there this in wich is what say me my wyrt

send lr̃e receave bearer I pray you Mte your name myne

Marie Stuart

Gifford transmettait secrètement les lettres de Marie mais c'était en fait un agent double et il transmettait aussi les lettres au services de renseignement de la Reine qui réussirent a briser le code utilisé par Marie.

En plus de lire toute sa correspondance et d'apprendre le contenu, ils ont falsifié un message demandant explicitement la liste des personnes impliquées.

Une fin triste pour Marie Stuart

Le chiffre indéchiffrable

Au 16ième siècle, on brisait les codes de façon routinière. La balle était dans le camp des cryptographes. ***Vigenère*** inventa un code simple et subtile. Il s'agit d'une amélioration du chiffre par décalage. On choisit un mot de code par exemple ALAIN et on l'utilise pour chiffrer.

ALAIN=1,12,1,9,14

ALAINALAINALAINALAINALAINALAINALAINALAINALAINA

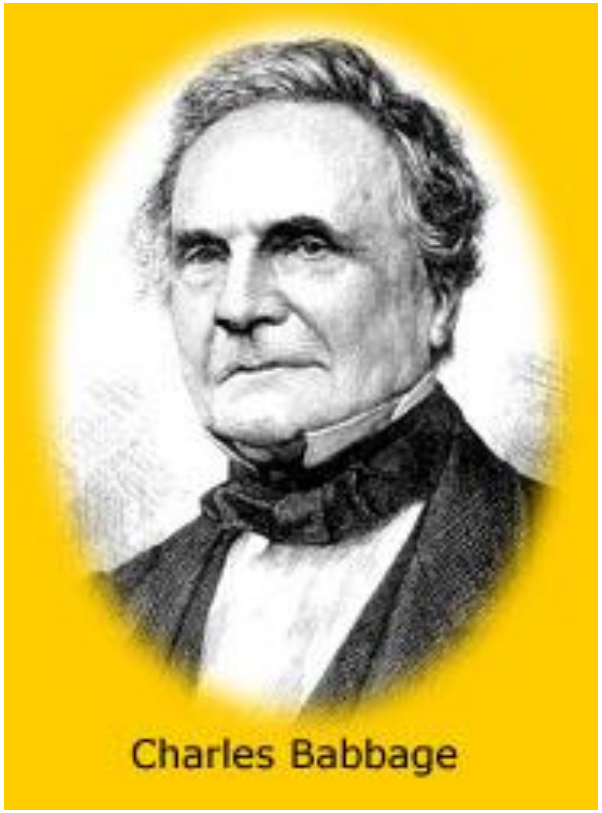
LE_CODE_DE_VIGENERE_EST_IL_INDECHIFFRABLE

MQALBEQAMSAGJPSOQSNNFDUIWMLJWRFOIRTGCBKZF

Clairement, une attaque statistique simple ne fonctionnera pas. Si le mot de code est suffisamment long (une phrase), essayer toutes les clefs est aussi impossible.

Le chiffre de Vigenère est-il indéchiffrable?

Briser le chiffre indéchiffrable!



Les cryptanalystes furent déjoués pendant près de 3 siècles par le chiffre de Vigenère.

Au 19ième siècle, Charles Babbage réussit à le briser.

La technique est relativement simple.

Exemple

```
OTDHRSIEGTD_LVISHFIESPVFLHDUOIWEGXJKLRMQHOEEEFMX
HFDVXTDQDOWZEGXNWIXNRBDRRBSED_TMDQIYLEYJCXPEIIXE
EFMXHOTFUOFFEQELHOYSHOJTLGDQDOPTQVYJXFDIHOPFCRPJ
IOVJWFSZYTIEOTDIHRSIDVIEHGXEXBDOHIDICTRKDBXEHBGT
UTDZQTDKRXWEOTDRHGWFJTDIRXXEHHVJCPWXHNDQ
```

1	2	3	4	5
9.5%	19.0%	9%	24.1% H	13.0%
	12.0%	11.7%	17.2% T	10.9%
		15.6%	27.6% D	15.2%
			22.4% E	13.0%
				17.4%
9.5%	15.5%	12.1%	22.8%	13.9%

En considérant que les caractères apparaissant le plus souvent sont soit _ ou E, on peut essayer différentes possibilités. H=E, T=E, D=_ et E=_ donne comme mot de code CODE qui permet de déchiffrer le message.

OTDHRSIEGTD_LVISHFIESPVFLHDUOIWEGXJKLRMQHOEEEFMX
HFDVXTDQDOWZEGXNWIXNRBDRRBSED_TMDQIYLEYJCXPEIIXE
EFMXHOTFUOFFEQELHOYSHOJTLGDQDOPTQVYJXFDIHOPFCRPJ
IOVJWFSZYTIEOTDIHRSIDVIEHGXEXBDOHIDICTRKDBXEHBGT
UTDZQTDKRXWEOTDRHGWFJTDIRXXEHHVJCPWXHNDQ

LE CODE DE VIGENERE PARAIT PLUS DIFFICILE A BRISER QUE LA SUBSTITUTION MONO ALPHABETIQUE IL FUT BRISE PAR BABBAGE UNE FOIS LA LONGUEUR DE LA CLEF RETROUVEE LE DECODAGE EST UN JEU D ENFANT ENCORE UNE FOIS LE MESSAGE DOIT ETRE ASSEZ LONG

Masque jetable
Un code parfait?

Chiffrer un bit d'information avec un bit de clef.

Mensonge

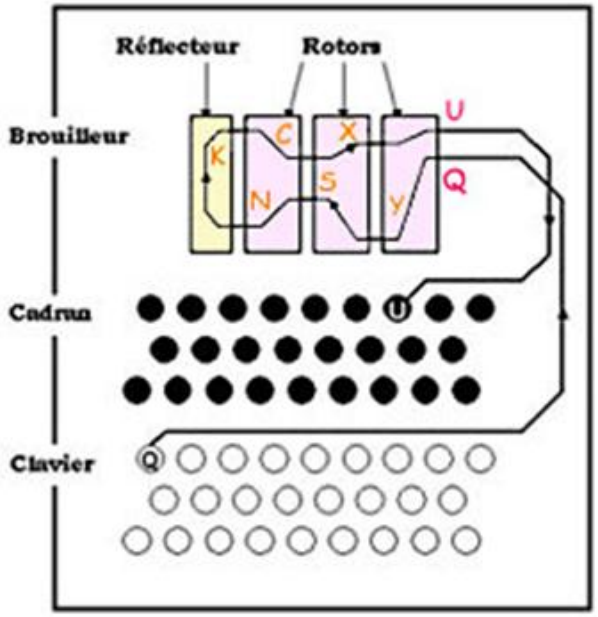
Vérité

Sécurité du masque jetable

Claude Shannon,
Communication Theory of Secrecy Systems,
Bell System Technical Journal,
vol.28(**4**), page 656–715, 1949.

ENIGMA

10 millions de milliards de clés



ENIGMA

La première version d'ENIGMA était utilisée comme suit.

Agencement des 3 rotors.
123, 132, 213, 231, 312, 321
6 possibilités.

Position des trois rotors, 3 lettres.
26x26x26=17 576 possibilités.

Connexions des fiches (6 connexions).
100 391 791 500 possibilités.

Exemple de clef: (231,DFT,AD,BE,CM,FY,UI,LP)

Nombre total de clefs:
6 * 17 576 * 100 391 791 500=10 586 916 764 424 000
10 million de milliard de possibilités…

Briser ENIGMA

Sur une période de 10 ans, les Allemands se dotèrent de plus de 30 000 machines ENIGMA.

ENIGMA est un véritable cauchemar pour les cryptanalystes.

Toute attaque statistique est inutile puisque chaque lettre du message est chiffré de façon différente.

Inutile d'essayer de deviner la clef. Il y en a trop.

La plupart des cryptanalystes abandonnèrent rapidement espoir de briser ENIGMA. Il y avait une exception. Les Polonais avaient peur d'une invasion Allemande. Pour eux, briser ENIGMA était vitale.

Briser ENIGMA

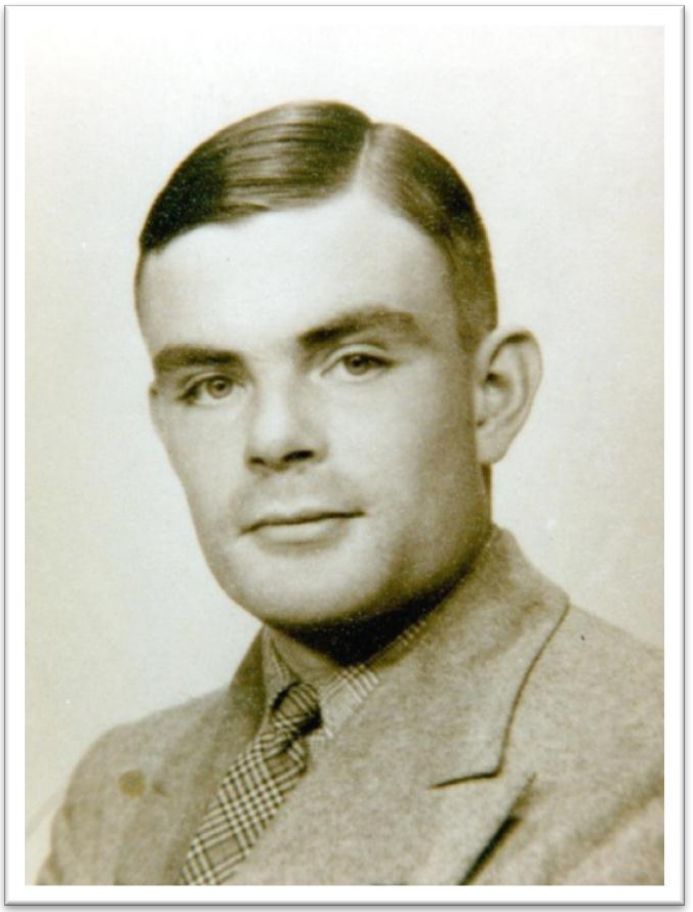
Les services de renseignement polonais ont obtenu par l'intermédiaire d'un informateur une description de la machine, ainsi que son mode d'utilisation.

Un livre de code donnait pour chaque jour la clef utilisée. Pour éviter que tous les utilisateurs d'ENIGMA utilisent la même clef, l'opérateur choisissait trois lettres au hasard qu'il chiffrait avec la clef du jour, deux fois. Ensuite la position des rotors était modifiée en fonction de ces trois lettres.

Chaque message était donc chiffré avec une clef différente.

Briser ENIGMA
1932-1944

Marian Rejewski
Polonais



Alan Turing
Britannique

Briser ENIGMA

Marian Rejewski

Le code ENIGMA fut brisé en décembre 1932 par Marian Rejewski, travaillant pour les services de renseignement polonais. A partir de 1933, les Polonais ont réussi a déchiffrer des milliers de messages allemands.

Les Polonais on réussi là ou les autres services de renseignement ont échoué.

Briser ENIGMA

La clef du succès de Marian Rejewski fut de se concentrer sur le fait que chaque message commençait par une répétition de 3 lettres.

Par exemple, pour quatre messages interceptés, on pouvait obtenir les données suivantes:

LOKRGM
MVTXZE
JKTMPE
DVYPZX

Chacun de ces chiffres dépend de l'agencement des rotors, du positionnement des fiches et bien sûr, des trois caractères choisis. Examinons la première et la quatrième lettre.

ABCDEFGHIJKLMNOPQRSTUVWXYZ
P M RX

Briser ENIGMA

Avec l'interception de plusieurs messages, on peut compléter le tableau.

```
ABCDEFGHIJKLMNOPQRSTUVWXYZ
FQHPLWOGBMVRXUYCZITNJEASDK
```

Ce tableau dépend de la clef du jour. Marian eu une intuition remarquable.

```
A-F-W-A                     3 LIENS
B-Q-Z-K-V-E-L-R-I-B         9 LIENS
C-H-G-O-Y-D-P-C     7 LIENS
J-M-X-S-T-N-U-J     7 LIENS
```

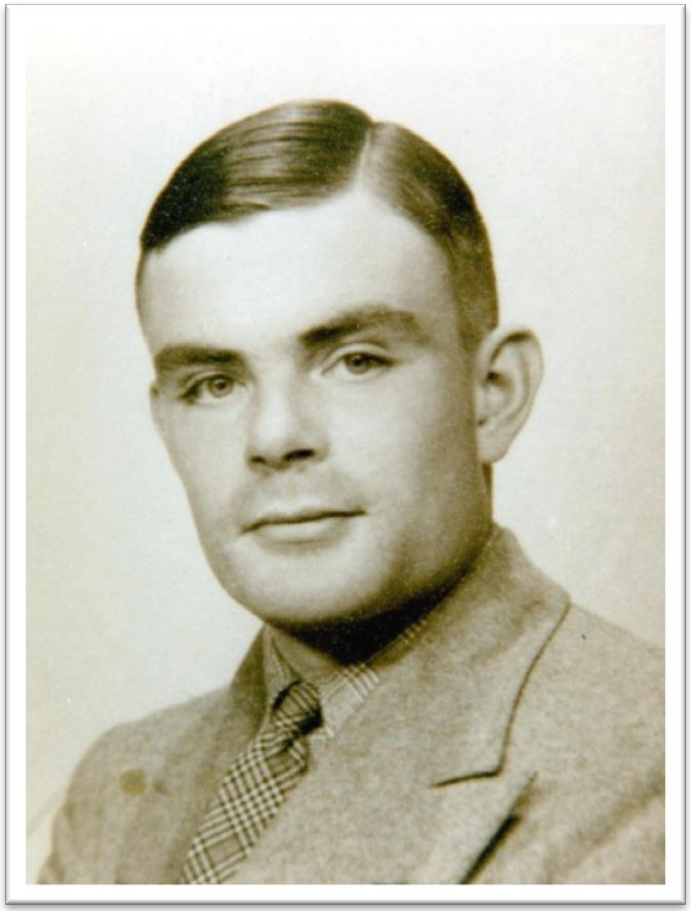
Le même exercice peut être réalisé avec les lettres numéro 2 et 5, ainsi que 3 et 6. Marian remarqua que la longueur des chaînes changeait à chaque jour. Si on change la position des fiches, les lettres des chaînes vont changer mais pas leurs longueurs. La longueur des chaînes ne dépend que de la position des rotors.

Briser ENIGMA

Il existe 6 x 26^3= 105 456 positionnements des rotors. Chacun donne lieu a une liste de chaînes avec des tailles caractéristiques. En une année, Marian réussit a construire une table de toutes les possibilités. Pour identifier la position des rotors, il suffisait d'intercepter quelques messages, calculer la longueur des chaînes, et regarder dans la table.

Il restait maintenant à trouver la position des fiches. Une fois les rotors bien positionnés, si on laisse le tableau des fiches vierge, l'opération de déchiffrement donnera un message illisible mais facile à briser. Les lettres sont simplement permutées suivant la position des fiches. Une attaque statistique trouve facilement les branchements.

ENIGMA et Turing



Alan Turing
Britannique

Un peu avant l'invasion allemande, les Polonais on dévoilé leurs techniques pour briser ENIGMA aux Britanniques. La partie n'était pas complètement gagnée. ENIGMA fut modifié durant la guerre. Des rotors furent ajoutés et à un certain moment, les Allemands ont cessé de répéter les trois lettres de la clef. Il y eut donc de courtes périodes pendant lesquelles les Alliés furent incapables de déchiffrer les messages allemands, mais des techniques de plus en plus sophistiquées et un appareillage électrique de plus en plus imposant leur permirent de déjouer les cryptographes allemands.