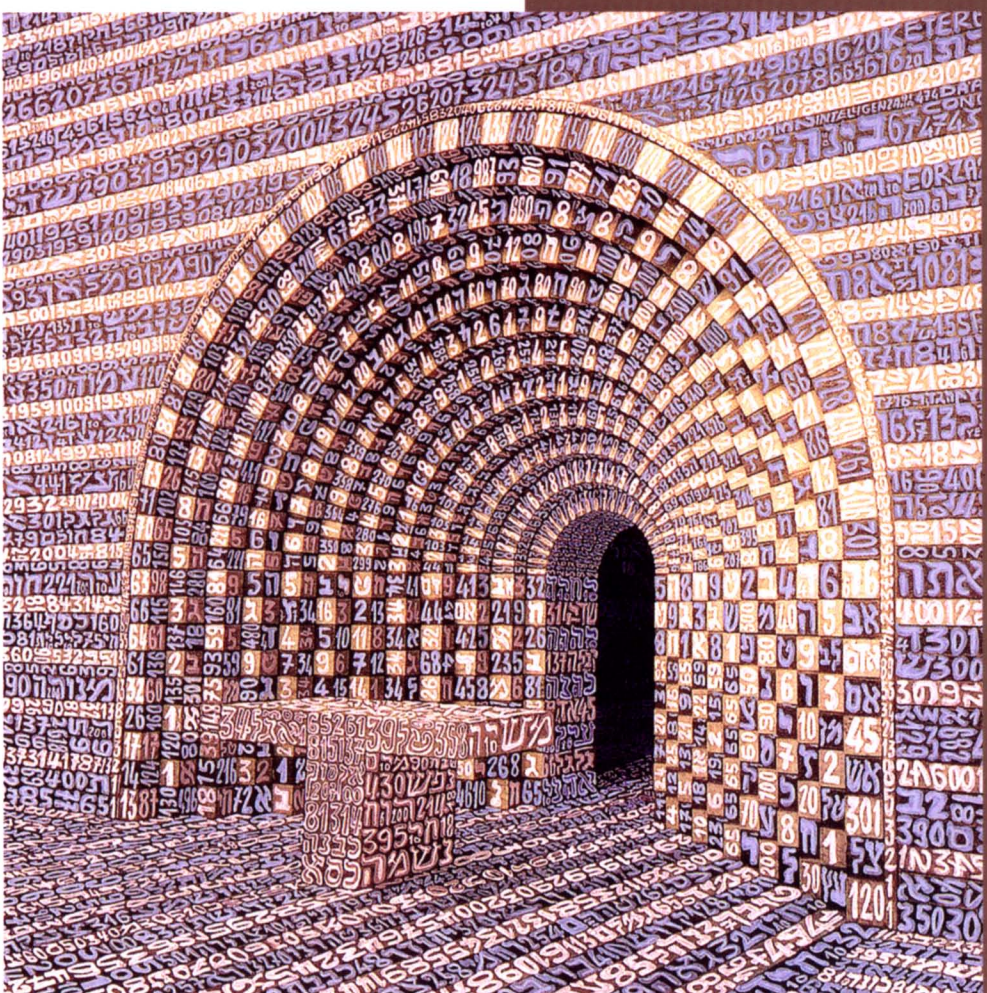
Tangente

Bibliothèque Tangente
L'aventure mathématique

Cryptographie & codes secrets



L'art de cacher

EDITIONS POLE
HS n° 26
ISSN 2263-4908

Bibliothèque Tangente

Tangente Hors-série n° 26

Cryptographie & codes secrets

L'art de cacher

Sous la direction d'Hervé Lehning

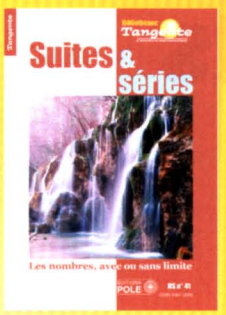
ÉDITIONS POLE

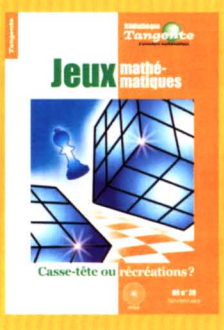
ISBN : 9782848841403 **ISSN : 2263-4908** **Commission paritaire : 1016 K 80883**

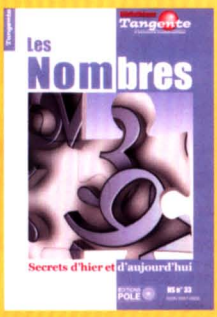
Bibliothèque Tangente

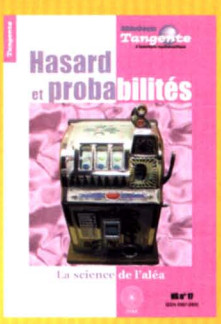
L'aventure mathématique

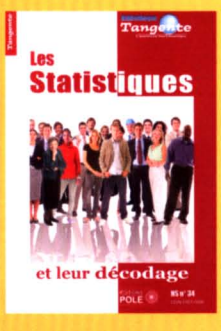
Retrouvez dans la Bibliothèque Tangente tous les secrets des chiffres et l'univers des codes.







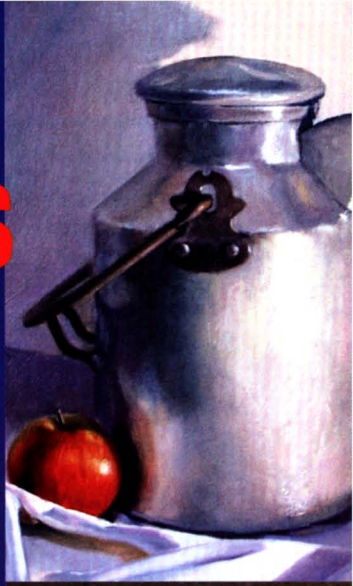




EDITIONS POLE

Cryptographie & codes secrets

Sommaire



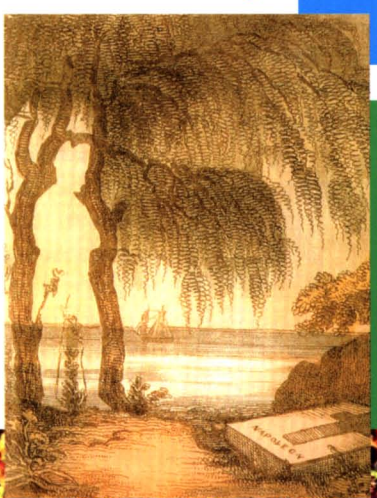
par M. Rousselet et A. Zalmanski

L'art de cacher

Des images dans une image

La stéganographie par l'image est sans doute un des procédés les plus anciens, depuis les images d'Épinal dans lesquelles Napoléon se cache dans les arbres, un crâne dans les anamorphoses d'un tableau de Holbein (*Les ambassadeurs*, 1533) ou un message écrit en Morse par les herbes aquatiques d'un paysage.

L'informatique est venue bouleverser ces procédés vieillots et puérils. Tout d'abord la découverte d'un codage permettant de cacher une forme tridimensionnelle dans une image à motifs répétitifs. Il s'agit des stéréogrammes. Il est intéressant de noter que si un algorithme permet d'automatiser le codage, seule la vision humaine est pour l'instant capable d'effectuer le décodage.



Où est Napoléon ?

Un stéréogramme.

Parler en langue étrangère

Au temps de la Seconde Guerre mondiale, les moyens usuels de chiffrement étaient longs et encombrants, quasiment inexploitables en situation de combat. Les Américains imaginèrent de communiquer dans une langue complètement étrangère pour leur ennemi : celle des Navajos. Le problème fut alors de créer des termes techniques inexistants dans la langue des Navajos tout en restant à l'intérieur de celle-ci. Ainsi, les bombes devenaient des œufs.

Messages insoupçonnés

Dans son livre *Schola Steganographica*, Gaspar Schott (1608–1666) explique comment cacher des messages dans les partitions de musique en faisant correspondre une note par lettre du message.

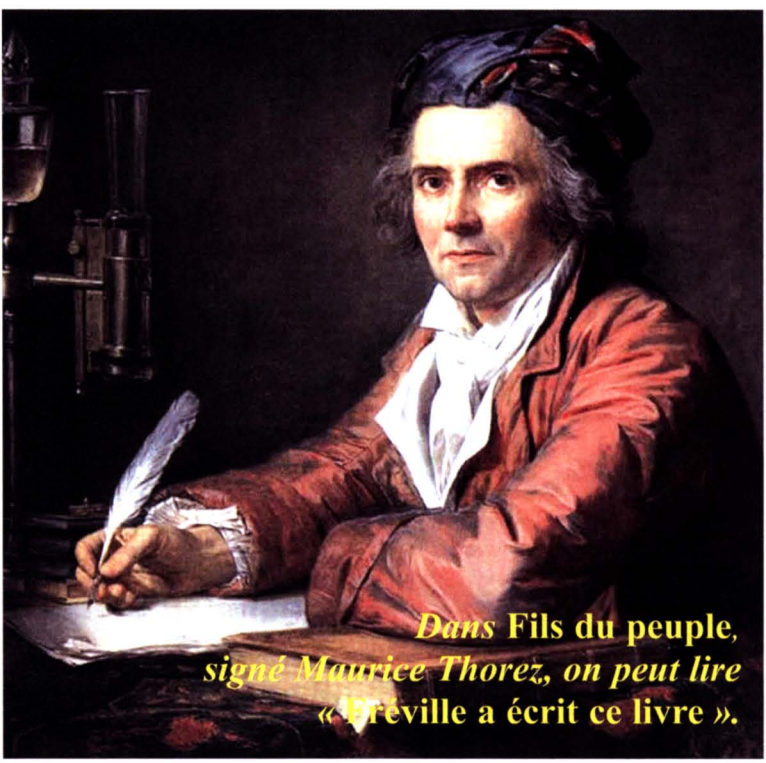
Le livre *Hypnerotomachia Poliphili* a été publié en 1499 par un auteur inconnu. Cet ouvrage qui traite des arts les plus divers – peinture, sculpture, architecture, art des jardins – contient aussi un texte caché qui révèle l'amour qui a existé entre un moine et sa maîtresse. On forme ce texte en prenant la lettrine qui débute chaque chapitre. On peut lire alors *Poliam frater Franciscus Columna peramvit* ce qui signifie « Frère Francesco Colonna aime Polia passionnément. » Des chercheurs ont attribué ce livre à Leon Battista Alberti.

par Alain Zalmanski

La stéganographie

La stéganographie, c'est l'art de cacher des messages et de rendre leur présence insoupçonnable. Acrostiche, contrepèterie, jeux de césure, sauts de mots ou de lettres... les procédés stéganographiques sont multiples et la littérature regorge de textes à double lecture.

Un message codé, par son illisibilité directe, apparaît très généralement comme tel. Il suffit de technique et de patience au curieux ou à l'indiscret pour le déchiffrer : il s'agit de cryptographie. En revanche, il est impossible de trouver un message si l'on ignore jusqu'à son existence. Cela fait longtemps que l'homme s'ingénie à dissimuler le contenu réel de messages qu'il veut transmettre au sein de documents, voire d'objets d'apparence anodine. Il s'agit de l'art de la stéganographie (du grec *steganos*, couvert et *graphein*, écriture). Avec la cryptographie, la clef de codage rend le message incompréhensible et par là même incite à la recherche d'une clef de décodage alors qu'avec la stéganographie, le message ne sera sans doute même pas soupçonné.



Dans Fils du peuple, *signé Maurice Thorez, on peut lire* « Fréville a écrit ce livre ».

La signature du nègre

La cryptographie possède un champ, limité pour l'essentiel, au milieu du renseignement et de l'espionnage en tout genre, du commerce et de la banque. La stéganographie remplit une partie de ce champ mais fleurit également dans le jeu de la communication, pamphlétaire ou amoureuse. Les pro-

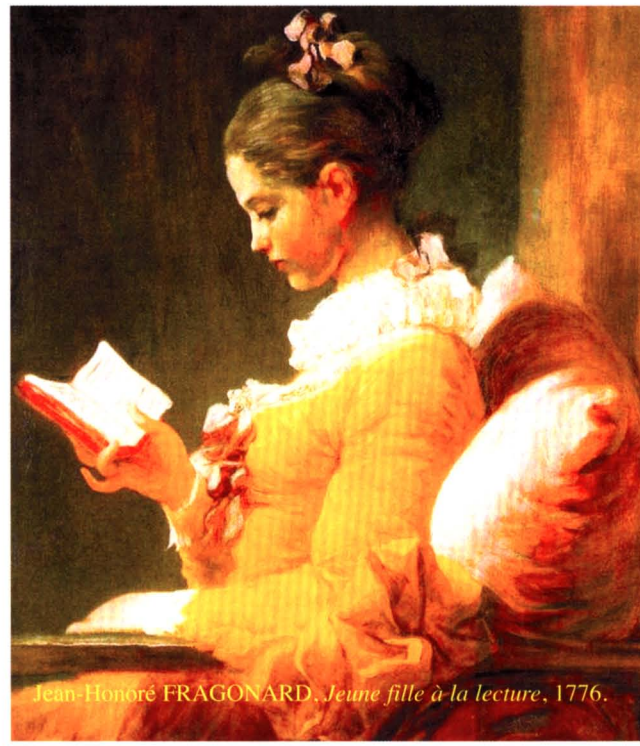
cédés en sont bien spécifiques car l'émetteur utilise souvent une clef simple, permettant une double lecture d'un texte, en prose ou en vers. L'acrostiche est également un procédé qui permet de cacher des messages ou des informations, courtes en général. Dans *Fils du peuple*, hagiographie de Maurice Thorez, publiée en 1937 signée par ce dernier mais rédigée par Jean Fréville, critique littéraire à l'*Humanité*, un indice oriente les initiés. Une description curieuse du paysage de l'adolescence de Thorez se révèle un procédé pour signer le livre :

« ferrailles rongées et verdies, informes lacis, larges entonnoirs aux escarpements crayeux, ravinés, immenses, tranchées creusées en labyrinthes, infranchissables vallonnements ravagés, embroussaillés ».

En prenant la première lettre de chaque mot, on peut lire : « *Fréville a écrit ce livre* ». Autre exemple, dû à la plume de Pierre Corneille : il se trouve dans *Horace*, aux vers 444 à 450. C'est une tirade où Horace explique qu'il est fier de devoir trucider son beau-frère Curiace. Voici la citation, je vous laisse lire l'acrostiche, peu probablement involontaire.

S'attacher au combat contre un autre soi-même,
Attaquer un parti qui prend pour défenseur
Le frère d'une femme et l'amant d'une sœur
Et rompant tous ces liens, s'armer pour la patrie
Contre un sang qu'on voudrait racheter de sa vie,
Un si grand honneur n'appartenait qu'à nous,
L'éclat de son grand nom lui fait peu de jaloux.

C'est le même Pierre Corneille qui a écrit dans *Polyeucte* : *Le désir s'accroît quand l'effet se recule*. Dans les deux cas il y a une double lecture ! L'acrostiche d'initiales a également été utilisé par des laudateurs spirituels qui



Jean-Honoré FRAGONARD, *Jeune fille à la lecture*, 1776.

Le directeur de journal genevois Charles Hubacher publia sans le savoir un poème qui disait « Hubacher crétin ».

cachaient ainsi le nom d'une ou un dédicataire ou par des pamphlétaires humoristes et farceurs qu'il n'est pas bon de quitter.

La vengeance masquée

Citons, entre autres, Willy, le premier mari de Colette, réfugié en Suisse en 1914, qui avait eu maille à partir avec un directeur de journal genevois, nommé Charles Hubacher qui le traitait de « *dévoyé du Moulin-Rouge* ». Willy répliqua de façon détournée. Il

envoya, sous un pseudonyme, le sonnet ci-dessous à un journal pacifiste français qui le publia aussitôt. Hubacher le reproduisit dans son journal sans même demander l'autorisation à l'auteur.

Hélas ! à chaque instant, le mal terrible empire !
Un cyclone de haine et de férocité
Bouleverse les champs, ravage la cité,
À flots coule le sang sous les dents du vampire
Cruauté d'autrefois ! Cet ancestral délire,
Honnis soient les bandits qui l'ont ressuscité,
Et honte à ceux dont la cruelle surdité
Refuse d'écouter la pacifique lyre.

C'est assez de combats, de furie et de deuil,
Rien ne demeurera si nul ne s'interpose
Entre les ennemis qu'enivre un même orgueil.

Toute raison à la Raison est-elle close ?
Impuissante, se peut-il que sur l'âpre écueil,
Nous laissions se briser notre nef grandiose.

On se doute de la colère de l'intéressé quand de bonnes âmes lui signalèrent qu'il avait lui-même publié :

HUBACHERCRETIN

Une autre façon de réaliser une double lecture consiste à rédiger un texte dans lequel on ne lit qu'un mot, ou une ligne sur deux ou trois, ou selon une clef particulière (voir l'encadré *Un échange épistolaire romantique*). La rédaction de tels textes, jouant sur les césures, la ponctuation ou les polysémies, constitue en soi un jeu littéraire fort apprécié (cf. encadré ci-contre).

De nombreux pamphlets, basés sur le même modèle ont traversé les siècles, circulant sous le manteau pour traîner dans la boue Napoléon sous couvert d'attaque de la Royauté ou comme sous l'Occupation, dans les années 1940 le tract apparemment collaborateur de la page 10.

J'abjure maintenant	Rome avec sa croyance
Calvin entièrement	j'ai en grande révérence
J'ai en très grand mépris	la messe et tous les saints
Et en exécration	du Pape et la puissance
De Calvin la leçon	reçois en diligence
Et ceux qui le confessent	sont heureux à jamais
Tous damnés me paraissent	le pape et ses sujets
Oui Calvin et Luther	je veux aimer sans cesse
Brûleront en enfer	ceux qui suivent la messe

La contrepèterie

Entre l'écrit et l'oral, la contrepèterie, qui est une forme phonétique d'anagramme est assurément un procédé stéganographique qui permet, sans cesse et sans fin de glisser des pans entiers et des messages particuliers, résultant de fouilles bien curieuses, dans des phrases qui peuvent être lues (ou entendues) sans malice par un œil (ou une oreille) non averti. Des spécialistes comme Luc Etienne, Jacques Antel ou Joël Martin ont érigé le contrepet en art de décaler les sons, et il n'est pas rare de voir émailler des ouvrages extrêmement sérieux de titres qui brouillent l'écoute. Tel la remarquable *Physique de la vie quotidienne* de François Graner, Springer (2003) dont les lecteurs, taupins ou agrégatifs, se réjouissent, peut-être en partie à leur insu, de tous les titres des exercices proposés, du moteur à flotte, au problème de bille infaisable ou à la pierre fine des Celtes (anagyre).

Un échange épistolaire romantique

attribué à George Sand et Alfred de Musset

Cher ami,
Je suis toute émue de vous dire que j'ai bien compris l'autre jour que vous aviez toujours une envie folle de me faire danser. Je garde le souvenir de votre baiser et je voudrais bien que ce soit une preuve que je puisse être aimée par vous. Je suis prête à montrer mon affection toute désintéressée et sans calcul, et si vous voulez me voir ainsi vous dévoiler, sans artifice, mon âme toute nue, daignez me faire visite, nous causerons et en amis franchement je vous prouverai que je suis la femme sincère, capable de vous offrir l'affection la plus profonde, comme la plus étroite amitié, en un mot : la meilleure épouse dont vous puissiez rêver. Puisque votre âme est libre, pensez que l'abandon où je vis est bien long, bien dur et souvent bien insupportable. Mon chagrin est trop gros. Accourez bien vite et venez me le faire oublier. À vous je veux me soumettre entièrement.
Votre poupée

et la réponse de Musset :

Quand je mets à vos pieds un éternel hommage,
Voulez-vous qu'un instant je change de visage ?
Vous avez capturé les sentiments d'un cœur
Que pour vous adorer forma le créateur.
Je vous chéris, amour, et ma plume en délire
Couche sur le papier ce que je n'ose dire.
Avec soin de mes vers lisez les premiers mots,
Vous saurez quel remède apporter à mes maux.

Alfred de Musset

Cette insigne faveur que votre cœur réclame
Nuit à ma renommée et répugne à mon âme.

George Sand

Visiblement, cet Alfred de Musset avait compris que, dans le message envoyé par George Sand, il fallait lire une ligne sur deux. Malheureusement pour notre sujet, il semble bien que cet échange épistolaire si plein de romantisme soit un canular de la fin du XIX^{e} siècle.

L'amour caché

L'utilisation des vers et rimes brisés permet soit une lecture continue des alexandrins, soit la lecture des hémistiches, colonne après colonne. Les deux lectures sont évidemment contradictoires. Mais il n'est pas nécessaire de travailler en vers. Les textes brisés en prose sont plus aisés à construire comme cette lettre d'un amoureux éconduit dans le style de celle attribuée à George Sand (voir l'encadré *Un échange épistolaire romantique*) :

Aimons et admirons	le chancelier Hitler
L'éternelle Angleterre	est indigne de vivre.
Maudissons, écrasons	le peuple d'outre-mer
Le Nazi sur la terre	sera seul à survivre.
Soyons donc le soutien	du Führer allemand
De ces navigateurs	la race soit maudite.
À eux seuls appartient	ce juste châtiment
La palme du vainqueur	répond au vrai mérite.

Jean-Honoré FRAGONARD, *La lettre d'amour*, 1770.

Mademoiselle,
Je m'empresse de vous écrire pour vous déclarer que vous vous trompez beaucoup si vous croyez que vous êtes celle pour qui je soupire. Il est bien vrai que pour vous éprouver, je vous ai fait mille aveux. Après quoi vous êtes devenue l'objet de ma raillerie. Ainsi ne doutez plus de ce que vous dit ici celui qui n'a eu que de l'aversion pour vous, et qui aimerait mieux mourir que de se voir obligé de vous épouser, et de changer le dessein qu'il a formé de vous haïr toute sa vie, bien loin de vous aimer, comme il vous l'a déclaré. Soyez donc désabusée, croyez-moi ; et si vous êtes encore constante et persuadée que vous êtes aimée, vous serez encore plus exposée à la risée de tout le monde et particulièrement de celui qui n'a jamais été et ne sera jamais
Votre serviteur.

La belle absente

Et que dire lorsque les lettres du message caché n'existent pas ! C'est le cas de textes rédigés selon un procédé oulipien appelé « belles absentes ». Un poème est composé d'autant de vers que le mot ou le message à trouver contient de lettres. Il est écrit à l'aide d'un alphabet simplifié (on supprime K, W, X, Y et Z). Dans chaque vers doivent apparaître, au moins une fois, toutes les lettres de cet alphabet (pangramme) sauf une : celle, qui, vers après vers, inscrit en creux dans la verti-

calité du poème le message dissimulé. Perec a su renouveler l'esprit conjoint de l'anagramme et du lipogramme pour dissimuler le nom de dédicataires !

Champ défait jusqu'à la ligne brève, (o)
J'ai désiré vingt-cinq flèches de plomb (u)
Jusqu'au front borné de ma page chétive.(l)
Je ne demande qu'au hasard cette fable en prose vague, (i)
Vestige du charme déjà bien flou qui (p)
défit ce champ jusqu'à la ligne brève. (o)

Dans ce texte Perec cache un nom en omettant dans le premier vers la première lettre du nom de l'aimée, dans le deuxième vers la deuxième lettre, etc. Saurez-vous trouver le prénom masqué par Georges Perec, dans cette lettre d'amour.

Inquiet, aujourd'hui, ton pur visage flambe.
Je plonge vers toi qui déchiffre l'ombre et
La lampe jusqu'à l'obscure frange de l'hiver :
Quêtes du plomb fragile où j'avance, masque
Nu, hagard, buvant ta soif jusqu'à accomplir
L'image qui s'efface, alphabet déjà évanoui.
L'étrave de ton regard est champ bref que je
Dois espérer, la flèche magique, verbe jeté
Plain-chant qu'amour flambant grava jadis.

A. Z.

Bibliographie

- Claude Gagnière, *Pour tout l'or des mots*, Laffont (1996)
- Michel Laclos, *Jeux de lettres, jeux de l'esprit*, Simœn, (1977)
- Jean-Paul Delahaye, *Jeux mathématiques et mathématiques des jeux*, Belin (1998)
- Georges Perec, *Beaux présents, belles absentes*, Seuil (1994)
- Simon Singh, *Histoire des codes secrets*, Lattès (1999)

Charles MEURER, *Le pouvoir de la presse*, 1902.

par Michel Rousselet

La stéganographie technique

En dehors des subtilités linguistiques, un message peut être caché par des moyens techniques : encres sympathiques, microfilms ou fichiers informatiques. Nous pénétrons ici l'art de la stéganographie technique.

La stéganographie technique est née très tôt. Ainsi, Hérodote (484–420 avant Jésus-Christ) raconte comment un certain Histiée, voulant prendre contact secrètement avec son gendre, le tyran Aristagoras de Milet, choisit un esclave dévoué, lui rasa la tête, et y inscrivit le message à transmettre. Il attendit que ses cheveux repoussent pour l'envoyer à Aristagoras avec l'instruction de se faire raser le crâne. Toujours d'après Hérodote, pour informer les Spartiates de l'attaque imminente des Perses, un certain Démarate utilisa un astucieux stratagème : il prit des tablettes, en racla la cire et grava sur le bois le message secret, puis il recouvrit les tablettes de cire. De cette façon, les tablettes, apparemment vierges, n'attirèrent pas l'attention. Au XVI[e] siècle, Giovanni Porta découvrit comment cacher un message dans un œuf dur : il suffit d'écrire sur la coquille avec une encre contenant une once d'alun (sulfate d'aluminium hydraté) pour une pinte de vinaigre (solution d'acide éthanoïque) ; la solution pénètre la coquille et dépose sur la surface du blanc d'œuf le message que l'on lira aisément après avoir épluché l'œuf.

Ainsi, la stéganographie n'est pas un gadget. Elle a de nombreux usages dans des domaines très variés, et en particulier toutes les fois que l'on veut communiquer en toute liberté même dans des conditions de censure et de surveillance. Elle permet par exemple de protéger la propriété intellectuelle (*copyright*) en cachant une signature personnelle dans un fichier. Certains nègres ont d'ailleurs utilisé cette technique pour que leur tra-

vail reste reconnaissable derrière le nom de façade de leur ouvrage.

Les encres sympathiques

À titre récréatif, proposons quelques exemples des bons vieux procédés que nos grands-parents utilisaient pour cacher leurs messages, en particulier par usage de l'encre sympathique. Proposons quelques vieilles recettes pour réaliser de bonnes sauces. Il existe deux catégories d'encres sympathiques : les liquides organiques et les produits chimiques. Les premiers deviennent visibles sous l'effet d'un léger chauffage : le lait, le citron, la sève, l'urine, entre autres, appartiennent à cette catégorie. Les produits chimiques sont invisibles une fois secs. Des caractères colorés apparaissent seulement après avoir été en contact avec un autre produit chimique appelé le réactif. Les deux techniques sont représentées dès l'Antiquité. Présentons les plus simples : celles qui utilisent des produits courants et en particulier les aliments.

Le lait constitue une excellente encre sympathique. Tout d'abord, écrivez votre message anodin sur une simple feuille de papier (assez épaisse) et ensuite tracez les mots secrets sur la feuille en utilisant un cure-dent imprégné de lait. Laissez sécher et absorbez le surplus de lait avec du sopalin. Le message inscrit au lait est alors invisible. Il suffit ensuite à votre destinataire de chauffer la feuille à l'aide d'une bougie et le message invisible écrit au lait réapparait. Vous pouvez faire de même en pressant un jus d'oignon avec quelques gouttes de citron.

Voici d'autres encres apparaissant également avec le feu : le jus de citron qui donne une couleur brune, le jus de cerise qui donne une couleur verdâtre et le

vinaigre, une couleur rouge pâle.
Parfois les encres apparaissent avec de la poudre. On peut tracer sur le papier des caractères invisibles avec tous les sucs glutineux et non colorés des fruits ou des plantes ou bien avec la bière, l'urine, le lait des animaux, et toutes les différentes liqueurs grasses ou visqueuses. Lorsque cette écriture est séchée, on répand dessus quelque poussière colorée très fine, on secoue le papier, et les caractères restent colorés. Il suffit par exemple de répandre du charbon tamisé très fin.

Le problème de la sécurité du courrier apparaît de la même façon pour les lettres électroniques. Nous retombons sur le problème général de la cryptographie ; mais on peut également cacher un message secret dans un courrier anodin (voir l'article *La stéganographie*).

Le lait constitue une excellente encre sympathique.

Cacher un message sur un site WEB

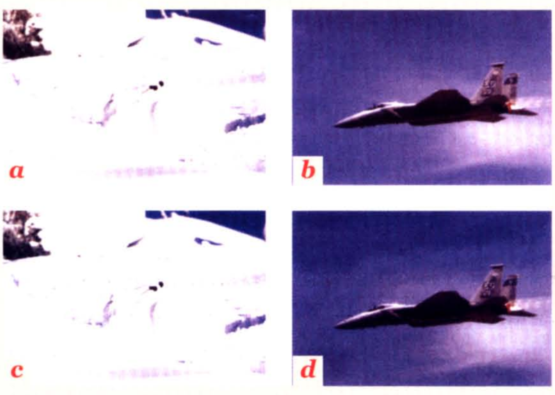
Une idée toute bête pour cacher un message sur un site web est d'utiliser une caractéristique du langage de description des pages web HTML : quel que soit le nombre d'espaces utilisés dans un texte, il en écrit un seul à l'écran. Notons ici chaque espace avec le symbole ∅ pour qu'il soit visible. Nous pouvons transmettre le message « VIVE TANGENTE » dans le texte « La stéganographie consiste à dissimuler un texte caché dans un message anodin » en écrivant dans le code source HTML :

La∅∅∅∅∅∅∅∅∅∅∅∅∅∅∅∅∅∅∅∅∅∅
stéganographie∅∅∅∅∅∅∅∅∅consiste
∅∅∅∅∅∅∅∅∅∅∅∅∅∅∅∅∅∅∅∅∅∅à
∅∅∅∅∅d i s s i m u l e r ∅∅∅
∅∅∅∅∅∅∅∅∅∅∅∅∅∅∅∅∅un∅texte
∅∅∅∅∅∅∅∅∅∅∅∅∅∅caché∅∅∅∅∅
∅∅dans∅∅∅∅∅un∅∅∅∅∅∅∅
∅∅∅∅∅∅∅m e s s a g e ∅∅∅∅
∅∅∅∅∅∅∅∅∅∅∅∅∅∅∅∅ano-
din∅∅∅∅∅

Pour cela, nous avons simplement codé chaque lettre par son numéro d'ordre dans l'alphabet latin. Ce nombre est le nombre d'espaces laissés entre les mots du message anodin. Sur la page web, rien ne peut être remarqué sauf si on réclame son code source. Un petit travail de transcription et le message caché apparaît.

Message dans un dessin

Dans le code RVB, chaque pixel d'une image est constitué de trois octets : un octet pour la composante rouge, un pour la composante verte et un pour la composante bleue. C'est pour cela que l'on parle de RVB (Rouge Vert Bleu). À partir de ces trois octets, on peut donc composer 256^3 c'est-à-dire 16 777 216 couleurs différentes, largement plus que ne peut distinguer l'œil humain. Or l'image n'est que le stockage dans un fichier de tous les pixels RVB composant l'image finale. Y cacher une information est facile, l'astuce est de retirer un bit à chacun des octets RVB. L'image est dégradée mais de façon invisible. Ainsi, en transformant ces bits, on récupère le huitième de la taille de l'image pour cacher un document, quel qu'il soit. Selon la



Le lapin au F15

La dissimulation d'images à l'intérieur d'autres images est devenue un processus facile dont les programmes sont disponibles sur Internet. Par exemple, on cherche à dissimuler l'image d'un avion F15 (*b*) dans l'image d'un lapin (*a*). En remplaçant les bits de poids faible (ceux qui modifient le moins l'image) de chaque pixel de l'image du lapin par les bits de poids fort de l'avion (ceux qui contiennent le plus d'information sur l'image de l'avion), on obtient une image (*c*) où est camouflé un avion. Ce dernier peut être extrait de l'image support : on retrouve l'image de l'avion (*d*) très légèrement dégradée par rapport à sa version initiale, mais tout à fait exploitable.

forme sous laquelle l'information est transmise, différentes méthodes sont ainsi possibles pour dissimuler une information secrète (voir l'encadré *Le lapin au F15*).

On peut également faire passer un message par le son. De faibles variations, imperceptibles pour l'oreille, dans les basses fréquences ou ce que l'on appelle le bruit de fond peuvent contenir une grande quantité d'information. Un grésillement infime peut cacher des secrets. Évidemment, ce bruit doit de préférence être transmis de façon numérique sans quoi les vraies pertes de transmission pourraient effacer entièrement le message caché.

Certains partenaires arrivent ainsi à établir des communications secrètes en établissant un protocole personnel au-dessus d'autres protocoles anodins.

Détection de la stéganographie

Peut-on détecter ou empêcher l'usage de la stéganographie ? L'observateur peut comparer les propriétés statistiques de la communication qu'il soupçonne et les comparer avec celles d'une communication ne contenant pas de messages cachés. De trop grandes différences peuvent être l'indice d'une communication secrète ou n'être qu'une simple anomalie statistique. Une fois découvert, on peut essayer de retrouver le message caché en utilisant des techniques de cassage des codes cryptographiques. Pour interdire toute communication cachée il faut pouvoir intercepter et transformer ou interdire toutes les communications (car elle peuvent potentiellement servir de transport). Un *firewall* possède les propriétés appropriées pour ce genre de contrôle absolu des communications. Celui qui veut empêcher une communication cachée se doit néanmoins de laisser passer le message clair tout en détruisant le message caché. Il peut aussi tout simplement détruire tout message suspect (voire tout message). Sinon, il est obligé de modifier le message tout en lui conservant son sens ou aspect original.

En conclusion, on peut dire que la stéganographie est un sujet encore peu étudié et faiblement médiatisé en tant que tel mais qui va certainement connaître ses heures de gloire dans un futur proche. Les liens entre la stéganographie et la télématique sont évidents et son usage sur l'Internet semble promis à un bel avenir.

M. R.

D'autres procédés techniques

par Alain Zalmanski

Parmi les procédés classiques (ou imaginables), on note également :

- Les messages cachés dans des objets ou média divers insoupçonnables, tel un clou creux enfoncé dans une planche, le dos d'un timbre affranchissant une lettre... ;
- Les perforations minuscules de caractères sélectionnés dans un texte ;
- L'utilisation de la ponctuation ou l'insertion manuelle de commentaires (ou de dessins) dans un texte imprimé ;
- Les micropoints : une page de texte est photographiée et réduite à un point de moins d'un millimètre de diamètre, qui sera ensuite déposé sur le point d'un i ou d'une ponctuation dans une lettre, un livre, une partition musicale ou un document quelconque. Cette technique a été largement utilisée par les services de renseignements et les réseaux de résistance durant la seconde guerre mondiale ;
- L'illustration d'un document anodin à l'aide d'un code graphique comme les hiéroglyphes, l'alphabet des hommes dansants de Conan Doyle ou même le linéaire B, tous déjà réputés inviolables ;
- L'utilisation de signes diacritiques ou de ponctuation.

Les argots

L'argot, ou plutôt les argots, sont des langages permettant à des classes d'âge, des catégories sociales, professionnelles ou à des familles de se distinguer. La fonction de l'argot est de cacher à certains ce que l'on veut communiquer à d'autres. Un article sec à capter.

Être chébran, c'est garin.

Une forme de stéganographie est relative à la langue parlée. Une fois encore, il s'agit de cacher aux non-initiés des propos émis au sein d'une identité sociale, professionnelle, étudiante, religieuse, ethnique ou même familiale. D'où les jargons de toutes sortes, les codes et les langages forgés.

Le verlan

Le *verlan* est un argot à clef, c'est-à-dire qu'il repose sur un système d'encodage fixe, autodescriptif (*verlan,* l'envers). Certains mots d'argot l'utilisaient depuis fort longtemps (comme *balpeau* pour peau de balle). Un mot est décomposé en syllabes qui sont prononcées à l'envers de l'ordre normal. Les monosyllabes – femme (*meuf*), chaud (*auch*), fou (*ouf*), mec (*keum*) sont traités en renversant la place de la ou des consonnes. Parfois l'usage conduit à une élision finale comme pour flic (*keufli* a donné *keuf*) ou arabe (*beurab* a donné *beur*). Laisse tomber devient *laisse béton* et il est de bon ton d'être *chébran* si l'on ne veut pas *avoir la jeura d'être taxé de garin*.

L'hermétisme s'émoussant au fil du temps, de nouvelles techniques plus subtiles sont apparues dans les cités, avec le *veul*. Il s'agit d'une nouvelle déformation du verlan, comme ça devenant par exemple *çacomme* en verlan et *asmeuk* en veul. De plus on mélange les lexiques des diverses cultures cohabitant dans les cités et parfois des jargons, dont le javanais. On aboutit à une langue à clefs multiples, émaillée de mots d'origines variées, argotiques ou non, qui n'est comprise que par les jeunes d'une zone géographique, parfois restreinte, ce qui leur permet de discuter entre eux sans être compris, ni des adultes – parents, enseignants, policiers – ni de bandes d'autre cités ou quartiers.

T'es ouf, y a pas que le cavu dans la vie. T'as vu ses yeuves ? Ils sont telmors. (Tu es fou, il n'y a pas que l'amour dans la vie. Tu as vu ses parents ? Ils sont très fréquentables.)

Le largonji

Le *largonji* est la traduction autodescriptive du mot jargon. La première lettre d'un mot, si elle est une consonne, est remplacée par un *l*, l'initiale se déplace en finale et sert de point de départ à un suffixe qui part du nom de

Quelques argots historiques

Le *largonjem* a été utilisé au bagne de Brest en 1821. Il utilise l'encodage du largonji mais avec un seul suffixe invariable : -em. Ainsi bon devient *lonbem*, boucher *louchébem*. Le mot *louchébem* (langage des bouchers) a été le plus connu car il a été employé par les bouchers de La Villette. Il remonterait à 1850. Ces argots à suffixe unique sont plus pauvres que le largonji.

Le *cadogan*, ou *teutgue*, est attesté en 1896. Sa formation consiste à placer l'élément -dg- après chaque voyelle et à redoubler cette voyelle :

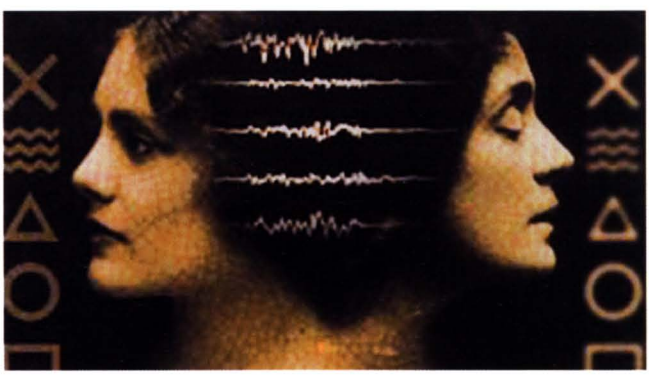
Jedgue veudgue lidguire tandganjandgantedgue. (Je veux lire **Tangente**.)

Le javanais insère l'élément -av-, -va- ou -aj- après chaque consonne. Ainsi bonjour devient *bonvonjouvour*. Le javanais a été utilisé dès 1857 ; il est surtout connu par la tradition scolaire et quelques auteurs comme Serge Gainsbourg ou Frédéric Dard. Exemples :
La gravosse avavavé peverduvu sava chagatte. (La grosse avait perdu sa chatte.)

la lettre (*ji* pour *j*, *bé* pour *b*). Soit : **l + base + initiale**. Le nom *largonji* a été employé dès 1881 par Richepin. Ce code servira à d'autres argots à clefs en changeant le suffixe : *-ic*, *-iche*, *-uche*, *-oc*, *-muche*…
En loucedé le lorgnebé lui liraté son larfeuil du lacsé. (En douce, le borgne lui tira son portefeuille du sac.)

Langages convenus

Il s'agit d'une forme particulière de la stéganographie orale, dans laquelle chaque mot, chaque forme syntaxique, chaque intonation de voix peuvent être codifiés. Le procédé n'est pas déchiffrable par les non-initiés. L'argot de l'imprimerie a lancé le fameux « vingt-deux » qui s'applique maintenant à la police et qui signifie « Attention, l'autorité ». Ce chiffre résulte de l'addition du rang alphabétique des lettres du mot « chef ». Les typographes se prévenaient entre eux de l'arrivée du contremaître en criant « vingt-deux ».



Les messages dits personnels, pleins d'humour et souvent incongrus, émis par la BBC à destination de la Résistance n'étaient compris que de la ou des seule(s) personne(s) habilitée(s) à les recevoir :
« Les gladiateurs circulent dans le sang, je répète, les gladiateurs circulent dans le sang » pouvait annoncer un parachutage d'armes ou demander une intervention contre un convoi allemand et le début du poème de Verlaine *« Les sanglots longs des violons de l'automne bercent mon cœur d'une lan-*

gueur monotone » prononcé à la radio de Londres ne donna de renseignements que sur la date du débarquement et à un seul réseau de résistance.
Début juin 1944 la radio de Londres donna le vrai signal avec les phrases : *« Ouvrez l'œil et le bon »* ; *« Tout le monde sur le pont »* ; *« Messieurs, faites vos jeux »* ; *« Le gendarme dort d'un œil »* ; *« Les carottes sont cuites »* ; *« Les dés sont sur le tapis »* ; *« Les enfants s'ennuient le dimanche »* : autant de messages pour prévenir les résistants d'actions à entreprendre.

Le monde du spectacle et de l'illusion a également utilisé cette forme de langage pour monter des numéros époustouflants de télépathie. Mir et Miroska se rendirent célèbres dans les années 1960 en faisant deviner par « transmission de pensée » le numéro de carte d'identité ou la couleur de la cravate d'un spectateur. Miroska avait les yeux bandés sur scène et Mir, parmi le public, ne faisait apparemment que poser une question précise en brodant constamment sur une phrase : *« Miroska vous êtes avec moi. »* Cela suffisait à Miroska pour tout deviner : nombres, couleurs, patronymes ou métiers, et ce aussi rapidement que

l'énoncé de la question l'avait été. La clef n'a toujours pas été dévoilée et le numéro reste inégalé.

A. Z.

Bibliographie

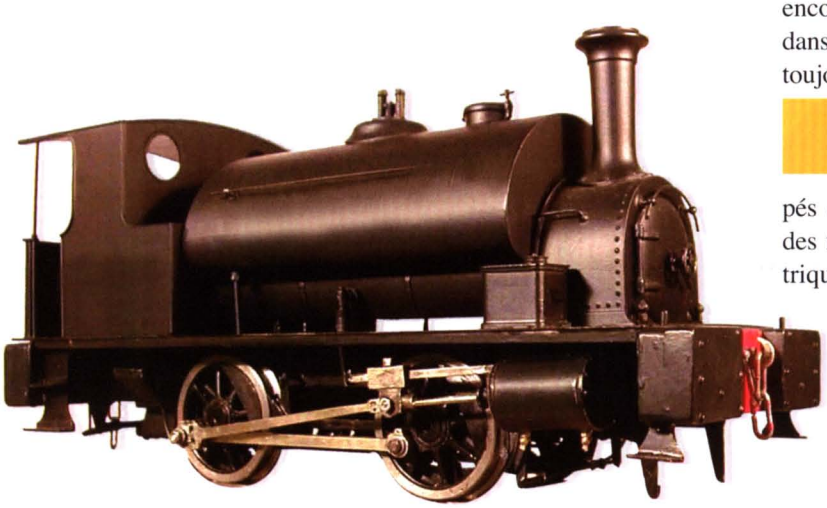
- Paolo Albani, *Dictionnaire des langues imaginaires*, Les Belles Lettres (2001).
- Philippe Pierre-Adolphe, *Tchache de banlieue*, Mille et une nuits (1998).
- Luc Etienne, *L'art du contrepet*, Pauvert (1957).
- Jacques Antel, *Ceux que la Muse habite*, Mille et une nuits (2005).
- Joël Martin, *Le Contrepet*, Que Sais-je ? (2005).

par Alain Zalmanski

Sémaphores

Le sémaphore est un système de signalisation optique qu'utilisaient les premiers télégraphes. Aujourd'hui, il reste en vogue sur les voies ferrées chez les scouts.

Tout est bon pour que l'homme puisse envoyer des messages, si possible codés pour ne pas être interceptés par des oreilles ou des yeux malveillants. Nos ancêtres communiquaient déjà à l'aide de signaux sonores primitifs à faible portée ou de signaux optiques relayés ou non, comme la fumée chère aux Indiens ou la couleur de la voile hissée par les navires. Le phare d'Alexandrie projetait, deux siècles avant notre ère, ses feux jusqu'à 55 kilomètres au large, indiquant aux bateaux la direction de la ville des Ptolémée.



C'est le Français Claude Chappe qui inventa en 1793, un système de signalisation optique moderne qui permettait d'envoyer des messages. Ces premiers télégraphes étaient constitués de hautes tours avec à leurs sommets deux bras dont les positions relatives représentaient des lettres ou des chiffres. Ce codage se transmettait à vue de tour en tour.

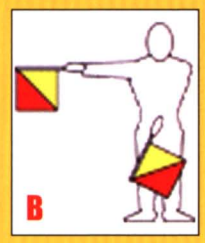
Avec des fanions

Ce système appelé sémaphore – le signe (*sêma*) qui porte (*phoros*) – s'utilise encore aujourd'hui avec un drapeau dans chaque main. Le sémaphore est toujours utilisé sur les voies ferrées, dans les aéroports et de façon plus ludique chez les scouts.

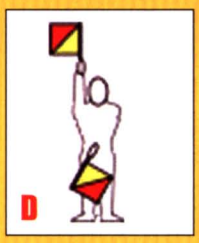
De même, avant d'être équipés d'une radio, les navires utilisaient des fanions portant des formes géométriques de différentes couleurs. Ces fanions étaient hissés le long et entre les mâts et servaient à envoyer des messages, permanents ou ponctuels.

A.Z.

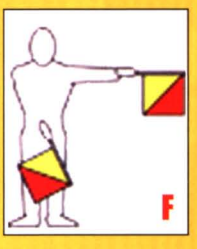

A

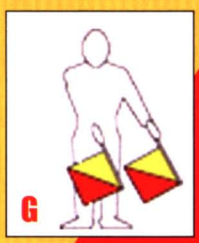

B

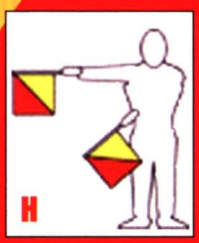

C


D

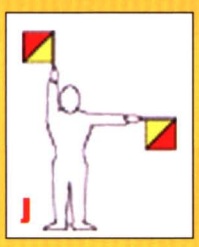

E


F


G


H

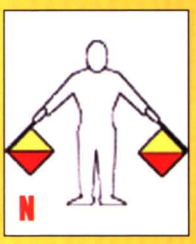

I


J

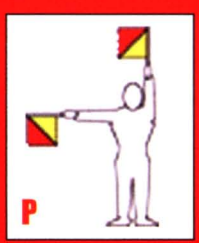

K


L


M


N


O


P

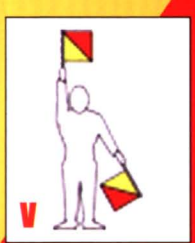

Q


R

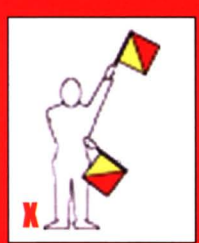

S


T

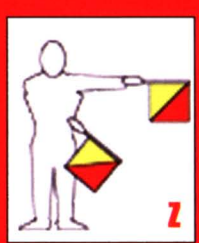

U


V


W


X

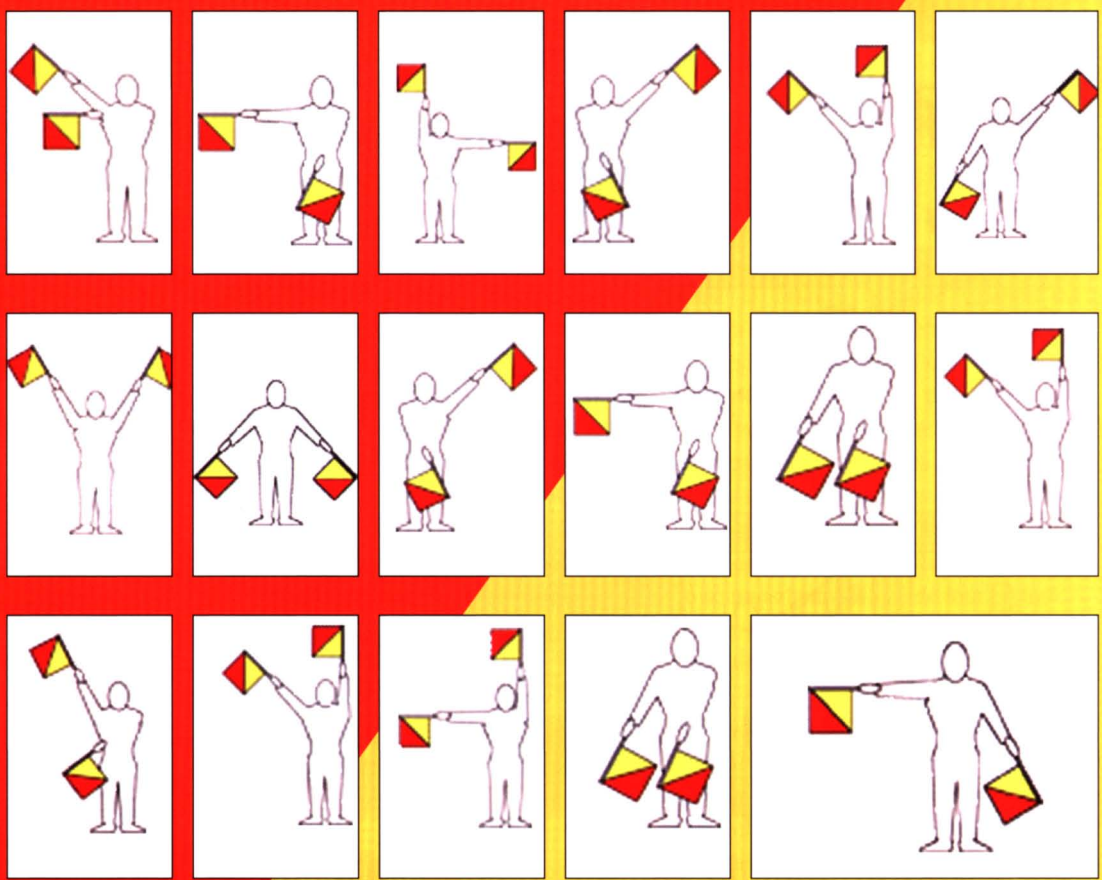

Y


Z

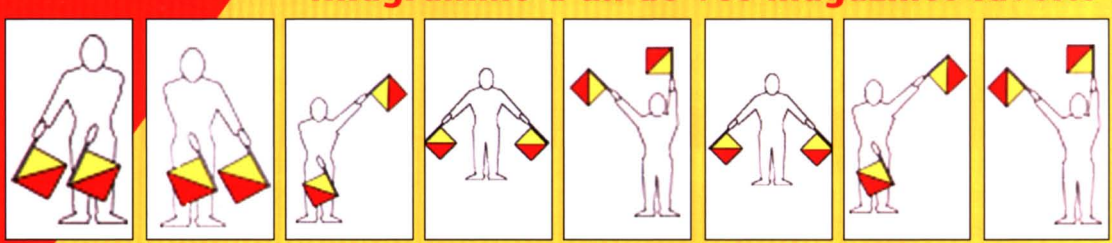
par Alain Zalmanski

Codage double

Pensée philosophique et son auteur



Anagramme d'un de vos magazines favoris



Réponses

• **Objet lu ne bâtit pas**, qui est une pensée et un anagramme de Jean-Baptiste Botulle.

• **Agent net**, anagramme de… *Tangente*.

Le temps de l'artisanat

Au temps de César, changer A en D, B en E, C en F, etc. dans un message suffit pour le rendre incompréhensible à un éventuel intercepteur. Il faudra un millénaire pour qu'un savant arabe, Al-Kindi, trouve une méthode pour décrypter les chiffres obtenus ainsi par simple substitution alphabétique : la méthode des fréquences. Elle consiste à comparer les fréquences des lettres du message avec les fréquences usuelles dans la langue utilisée. La méthode n'est perfectionnée qu'à la Renaissance, par un diplomate : Blaise de Vigenère, en modifiant la substitution en fonction d'une clef. Ce code n'est brisé qu'au XIX^{e} siècle, par Babbage. Il ne survit aujourd'hui dans la réalité qu'à la condition que la clef soit aussi longue que le message, pour le téléphone rouge et la cryptographie quantique.

par Michel Rousselet

César et ses prédécesseurs

Dans l'Antiquité, les rives de la Méditerranée ont connu de nombreuses guerres. La cryptographie est née de la nécessité de conserver le secret des communications entre les chefs militaires et leurs armées, ainsi qu'entre les États qui avaient noué des alliances.

Contrairement à la stéganographie qui s'efforce de dissimuler l'existence d'un message, la cryptographie (du grec *graphein* et *cryptos* : « écriture cachée ») cherche à en dissimuler le contenu.

La scytale de Sparte

Si on en croit ce qui est écrit par l'historien grec Plutarque (47–120 après J.-C.), la cryptographie serait née à Sparte au v^{e} siècle avant notre ère. Pour transmettre des messages confidentiels entre les magistrats de la cité et les généraux en campagne, les Spartiates utilisaient deux bâtons de même longueur et de même diamètre, les *scytales*. L'un des bâtons restait à Sparte tandis que l'autre était emporté par le général. Pour crypter un message, on enroulait en spires jointives un mince bandeau de papyrus (ou de parchemin) sur l'un des bâtons puis on y écrivait le message à protéger. Une fois déroulée, la bande qui ne présentait plus qu'un texte d'apparence incohérente pouvait être acheminée vers son destinataire. Pour lire le message, celui-ci devait enrouler la bande autour du second bâton afin de reformer le texte en clair.

Pour illustrer ces dires, Plutarque rapporte qu'en 404 avant Jésus-Christ, le général spartiate Lysandre vit arriver, venant de Perse, un messager qui lui tendit sa ceinture. Lysandre l'enroula autour de sa scytale et, déchiffrant le message qui lui était envoyé, apprit que les Perses s'apprêtaient à l'attaquer.

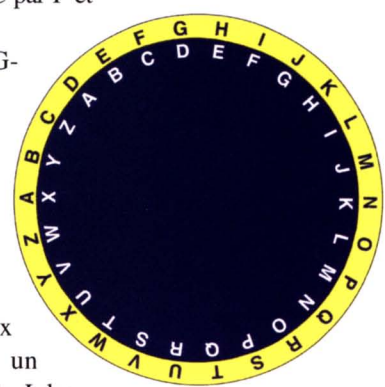
Une scytale

Les méthodes d'Enée le tacticien

Énée le Tacticien était un des généraux de la Ligue arcadienne qui, au IV^e^ siècle avant Jésus-Christ, regroupait un certain nombre de cités grecques. Il a écrit plusieurs ouvrages sur l'art militaire dont un seul nous est parvenu. Il s'agit d'un traité intitulé *Poliorcétiques* qui concerne principalement la défense de la cité mais aussi l'art de crypter des messages. Énée nous apprend par exemple à remplacer les voyelles du texte à cacher par des points : un pour *alpha*, deux pour *epsilon* et ainsi de suite jusqu'à sept pour *omega*. Les consonnes, quant à elles, restent inchangées. Avec ce procédé, le message « DENIS EST HONNÊTE » devient « D..N...S ..ST H....NN..T. ».
Il explique comment camoufler un texte en insérant des lettres quelconques entre celles du texte à protéger. Des marques presque invisibles, des trous minuscules en général, doivent être faites pour signaler au destinataire du message les lettres à utiliser. Au lieu de trous, utilisons par exemple une couleur pour transformer le message « DENIS EST HONNÊTE » en « DARESTFNZAQIS EKLSZT HPTONKKKNÊDTE ». Dans le message transformé, seules les lettres en rouge sont à prendre en considération.
Une troisième méthode consiste à percer vingt-quatre trous, un trou par lettre, dans un osselet ou dans un morceau de bois. Le chiffrement consiste alors à passer un fil à travers les trous qui représentent les lettres du message à envoyer. Pour déchiffrer le message, il faut naturellement connaître l'ordre des trous en partant de celui qui indique l'alpha.

Le chiffre de Jules César

En 58 avant Jésus-Christ, Jules César se lançait à la conquête de la Gaule. Pour communiquer avec ses généraux, il imagina deux procédés de chiffrement. Le premier consistait à remplacer les lettres latines du message à crypter par des lettres grecques. Le second procédé est expliqué par Suétone dans son ouvrage *Les vies des douze César* écrit en 121 après Jésus-Christ. La technique en est particulièrement simple puisqu'il suffit de procéder à une permutation circulaire des lettres de l'alphabet en remplaçant chaque lettre par celle qui est située trois rangs plus loin : A est remplacé par D, B par E, C par F et ainsi de suite.
Ainsi le message « LONGTEMPS JE ME SUIS COUCHÉ DE BONNE HEURE » devient « ORQJWHPV MH PH VXLV FRXFKH GH ERQQH KHXUH ».
La figure ci-contre montre un appareil formé de deux disques servant à crypter un message selon la méthode de Jules César.



Le disque bleu est mobile, on peut donc modifier le décalage qui ici vaut 3.

Aujourd'hui, le chiffre de Jules César n'est plus utilisé que par les écoliers car les messages cryptés avec cette méthode sont très faciles à décrypter.

M. R.

par Michel Rousselet

Affinité et codages

Le codage de Jules César peut être généralisé à des codages par transformations affines. Les propriétés relativement simples de ces codages constituent une bonne approche de la cryptographie moderne.

Numérotons de 0 à 25 et dans l'ordre alphabétique les 26 lettres de l'alphabet français : A porte le numéro 0, B le numéro 1, C le numéro 2 et ainsi de suite. Soient a et b deux entiers naturels et n le rang d'une lettre L quelconque de l'alphabet. Appelons n' le reste de la division euclidienne de l'entier $an + b$ par 26. Ceci peut s'interpréter dans l'anneau $\mathbb{Z} / 26\mathbb{Z}$ (voir l'article *L'arithmétique de la cryptographie*) : on effectue les calculs comme d'ordinaire puis on les remplace par leurs restes dans la division par 26. L'égalité des restes s'écrit avec le symbole ≡. C'est ainsi qu'on écrit $26 \equiv 0$ ou $34 \equiv 8$, $n' \equiv an + b$, etc. Le nombre n' représente le rang d'une certaine lettre L' de l'alphabet puisque $0 \leq n' \leq 25$. En associant L' à L, on réalise ainsi un codage des lettres de l'alphabet par la transformation affine définie par les entiers a et b. Par définition du nombre n', on peut se contenter de choisir les entiers a et b dans l'intervalle [0, 25].

Lettre en clair	n	n'	Lettre codée
A	0	3	D
B	1	8	I
C	2	13	N
D	3	18	S
E	4	23	X
F	5	2	C
G	6	7	H
H	7	12	M
I	8	17	R
J	9	22	W
K	10	1	B
L	11	6	G
M	12	11	L
N	13	16	Q
O	14	21	V
P	15	0	A
Q	16	5	F
R	17	10	K
S	18	15	P
T	19	20	U
U	20	25	Z
V	21	4	E
W	22	9	J
X	23	14	O
Y	24	19	T
Z	25	24	Y

Codage des lettres pour $a = 5$ et $b = 3$

Ce codage par transformation affine des lettres de l'alphabet permet de coder un texte complet en le transformant lettre par lettre. Ainsi, avec a = 5 et b = 3, la phrase Premier essai de codage par transformation affine devient AKXLRXK XPPDR SX NVSDHX ADK UKDQPCVKL-DURVQ DCCRQX.

Comment choisir les nombres *a* et *b* ?

Toutes les valeurs entières de l'intervalle [0, 25] ne permettent pas de réaliser des codages acceptables. Pour qu'un codage puisse remplir sa fonction, il faut que l'application de [0, 25] dans [0, 25] qui à n associe n' soit bijective (voir l'article *Bijectivité et cryptographie*). Ceci conduit à écarter certaines valeurs comme $a = 0$ ou encore $a = 13$ puisque, si $a = 0$, toutes les lettres de l'alphabet sont codées de la même façon ($n' = b$), si $a = 13$ les lettres de l'alphabet ayant des rangs pairs sont codées de la même façon (pour $n = 2k$, $n' = 26k + b \equiv b$).
D'une façon plus générale, nous démontrons plus loin que les seules valeurs de a convenables sont les nombres entiers de l'intervalle [0, 25] premiers avec 26 c'est-à-dire 1, 3, 5, 7, 9, 11, 15, 17, 19, 21, 23 et 25. Cela fait 12 valeurs possibles pour a. Le nombre b quant à lui peut être choisi librement parmi les 26 valeurs entières de l'intervalle [0, 25] (voir l'article *L'arithmétique de la cryptographie*). Il en résulte qu'il y a 12×26 soit 312 transformations affines utilisables pour coder un texte. Si on écarte la transformation identique caractérisée par $a = 1$ et $b = 0$ qui ne présente guère d'intérêt pour sauvegarder la confidentialité d'un message, il nous reste finalement 311 possibilités de codage par transformation affine.

Les codages affines de textes longs sont faciles à décoder.

Décodage d'un texte codé par une transformation affine

Considérons un codage utilisant une transformation affine φ définie par deux nombres a et b. Cette transformation étant nécessairement bijective, elle admet une transformation bijective

réciproque φ' qui doit permettre de réaliser le décodage d'un texte codé avec elle. Cherchons φ' sous forme d'une autre application affine définie par deux nombres a' et b'. L'application composée $\varphi o \varphi'$ est définie par :

$$\varphi o \varphi' (n) \equiv a (a'n + b') + b.$$

Ces deux transformations sont inverses l'une de l'autre si :

$$(aa' - 1)n + ab' + b \equiv 0$$

pour tout n ce qui équivaut à :

$$aa' \equiv 1 \text{ et } ab' + b \equiv 0.$$

Si la première égalité est satisfaite, en multipliant la seconde par a', on obtient : $b' \equiv -a' b$. Ainsi, φ et φ' sont inverses l'une de l'autre si et seulement si :

$$aa' \equiv 1 \text{ et } b' \equiv -a' b.$$

La première égalité équivaut à l'existence d'un entier k tel que : $aa' + 26k = 1$ ce qui correspond à la relation de Bezout et impose effectivement que a soit premier avec 26. De plus, a' peut être déterminé grâce à l'algorithme d'Euclide (voir l'article *Arithmétique de la cryptographie*).

Reprenons l'exemple précédent ($a = 5$ et $b = 3$) et cherchons les nombres a' et b' définissant la transformation affine réciproque de φ. Dans ce cas, nous disposons d'une solution évidente à la relation de Bezout : $5 (-5) + 26 (1) = 1$ donc $a' \equiv -5$ convient d'où : $b' = -(-5) \times 3 = 15$. Nous en déduisons le tableau :

Lettre en clair	n	n'	Lettre codée
	A	0	15 P
B	1	10	K
C	2	5	F
D	3	0	A
E	4	21	V
F	5	16	Q
G	6	11	L
H	7	6	G
I	8	1	B
J	9	22	W
K	10	17	R
L	11	12	M
M	12	7	H
N	13	2	C
O	14	23	X
P	15	18	S
Q	16	13	N
R	17	8	I
S	18	3	D
T	19	24	Y
U	20	19	T
V	21	14	O
W	22	9	J
X	23	4	E
Y	24	25	Z
Z	25	20	U

Décodage des lettres pour $a = 5$ et $b = 3$, $a' = -5$ et $b' = 15$

Bien sûr, on aurait pu également constituer ce tableau en lisant le précédent à l'envers c'est-à-dire de droite à gauche. De même, si un texte est codé avec $a = 7$ et $b = 3$, il se décode avec $a' = 15$ et $b' = 7$.

Encore le décodage

Considérons le texte codé suivant « GXSLN YJYBY NXNQ SLNNXX ». On peut décoder ce texte à partir des deux renseignements suivants :

1) le texte initial a été codé à l'aide d'une transformation affine du type de celles qui ont été décrites ;

2) dans ce codage, E est devenu X et S est devenu N.

Soient a et b les entiers qui définissent la transformation affine ayant servi au codage du texte. Dans l'alphabet en clair, le rang de E est 4, celui de S est 18, celui de X est 23 et celui de N est 13. On peut donc écrire le système suivant :

$$\begin{cases} 4a + b \equiv 23 \\ 18a + b \equiv 13 . \end{cases}$$

Un tel système est qualifié de système diophantien, du nom du mathématicien grec Diophante qui, le premier, a étudié ce genre de problème arithmétique. Soustrayons membre à membre les deux équations, nous obtenons $14\ a \equiv -10$ donc $14\ a \equiv 16$ puisque $-10 \equiv 16$. Cette équation peut s'écrire encore $14\ a = 26\ q + 16$ où q est un entier naturel. En divisant par 2, on obtient $7\ a = 13\ q + 8$ dont une solution immédiate est $a = 3$ et $q = 1$. On a alors $b \equiv 23 - 4 \times 3 \equiv 11$.

D'une façon générale, on peut décoder un texte qui a été codé à l'aide d'une transformation affine φ en essayant de déterminer par calcul les paramètres a et b qui définissent φ. Si le texte à analyser est suffisamment long, on peut faire des hypothèses sur ce que représentent les deux lettres les plus fréquentes du texte à décoder. On sait qu'en français, les deux lettres les plus fréquentes sont le E et le A.

M. R.

par Michel Rousselet

Les fréquences d'Al Kindi

Les érudits et les savants de l'empire arabe sont les inventeurs de la cryptanalyse. C'est une méthode permettant de décrypter les messages chiffrés par substitution mono-alphabétique sans qu'on connaisse la clef du codage ni même le type exact du codage.

En matière de cryptographie, on définit une substitution mono-alphabétique en indiquant de quelle façon remplacer chaque lettre de l'alphabet par une autre. Pour qu'une telle substitution puisse servir au cryptage d'un texte, il faut respecter les deux conditions suivantes :

1) deux lettres différentes sont codées de façons différentes,

2) la même lettre est toujours codée de la même façon.

De façon générale, une substitution mono-alphabétique est une permutation quelconque des lettres de l'alphabet. Par exemple :

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

Un exemple de substitution mono-alphabétique

Le chiffre de Jules César, qui consiste à décaler les lettres de l'alphabet de trois rangs est aussi un exemple de substitution mono-alphabétique :

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

Le chiffre de Jules César

S'il existe seulement 26 permutations circulaires de notre alphabet, il existe en revanche 26 ! permutations soit 403 291 461 126 605 635 584 000 000.

Le rôle d'un savant arabe

Dix siècles après les débuts de la cryptographie, on ne connaissait encore aucune procédure générale applicable au décodage d'un texte crypté à l'aide d'une substitution mono-alphabétique. Le premier traité exposant une telle procédure est publié au IX^{e} siècle après Jésus-Christ par le savant arabe Abu Yusuf al-Kindi sous le titre *Manuscrit sur le déchiffrement des messages cryp-*

tographiques.

Habitué à analyser les textes saints de l'Islam pour en vérifier l'authenticité, Al-Kindi observe que les lettres et les syllabes ont, en arabe, des fréquences qui diffèrent. Ainsi, les lettres *a* et *l* sont les plus courantes tandis que la lettre *j* apparaît dix fois moins souvent. En conséquence, pour décrypter un texte chiffré, Al-Kindi propose de calculer les fréquences des lettres qu'on trouve dans ce texte afin de les comparer aux fréquences constatées dans la langue qui a servi à l'écrire. En Europe, le premier livre à décrire les méthodes de la cryptographie fut écrit au XIII^{e} siècle par Roger Bacon. Sa *Lettre sur les œuvres d'art et sur les nullités de la magie* proposait sept méthodes pour crypter des messages. Il est fort probable que les méthodes d'analyse développées par al-Kindi se soient répandues en Europe à partir du XV^{e}siècle.

La fréquence des lettres en français

En français comme en arabe, les lettres n'ont pas la même fréquence d'apparition. Le tableau suivant montre, exprimées en pourcentages, les fréquences moyennes des lettres utilisées dans les textes écrits en français.

A	B	C	D	E	F
9,42	1,02	2,64	3,39	15,87	0,95
G	H	I	J	K	L
1,04	0,77	8,41	0,89	0,00	5,34
M	N	O	P	Q	R
3,24	7,15	5,14	2,86	1,06	6,46
S	T	U	V	W	X
7,90	7,26	6,24	2,15	0,00	0,30
Y	Z				
0,24	0,32				

Tableau de fréquences des lettres en français

Dans un texte écrit en français, il y a presque toujours beaucoup plus de E que de W ! Il y a donc de fortes chances pour que, dans un texte chiffré, la lettre qui apparaît le plus fréquemment représente un E. Les lettres les moins fréquentes représentent probablement W ou K ou X, etc.

Un exemple de décryptage

Voici un texte crypté par substitution monoalphabétique :

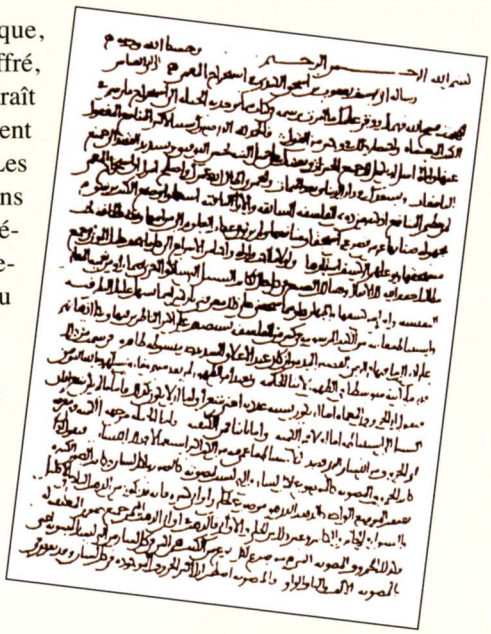
TIXIV KYWXEZ PINIYRI-HMVM-GLPIX IWX RI PI 13 JIZVMIV 1805 E HYVIR, YRI ZMPPI H'EPPIQE-KRI WMXYII E QM-GLIQMR IRXVI EEGLIR IX GSPSKRI. WSR TIVI C IXEMX VIGIZIYV HIW TSWXIW. HMVMGLPIX IWX YR IPIZI FVMPPERX, UYM EGLIZI WIW IXYHIW WIGSRHEMVIW E 16 ERW.

Pour le décrypter, nous déterminons d'abord la fréquence de chaque lettre dans le texte crypté.

A	B	C	D	E	F
0	0	0,53	0	6,84	0,53
G	H	I	J	K	L
4,21	3,68	23,16	0,53	0,00	1,58
M	N	O	P	Q	R
6,84	0,53	0	6,32	1,56	7,37
S	T	U	V	W	X
2,63	1,58	0,53	6,32	7,37	7,37
Y	Z				
4,74	3,16				

Fréquences des lettres dans le texte crypté

Bien entendu, ceci est plus facile à effectuer avec un petit programme informatique. La lettre la plus fréquente dans le texte (I) doit être un E



Une page du manuscrit sur le déchiffrement des messages cryptographiques

et la lettre isolée la plus fréquente (E) doit être un A. Ensuite, nous pouvons nous douter que les autres lettres relativement fréquentes dans le texte (MPRVWX) correspondent aux lettres ILNRSTU. Pour les distinguer, nous analysons les mots de deux lettres : RI, PI, QM, IX et YR. A priori, RI ne peut être que NE, SE ou TE. La répétition RI PI fait plutôt pencher la balance vers NE donc la lettre R correspond donc sans doute à N. Nous continuons le même type de raisonnement pour aboutir à un premier tableau d'hypothèses :

E	I	M	P	R	V
A	E	I	L	N	R
W	X				
S	T				

Premières hypothèses de correspondances

Ces premières hypothèses nous donnent le texte :

xETER xxSTAx LExExNE-xIRIxxLET EST NE LE 13 xExRIER 1805 A xxREN, xNE xILLE x'ALLExAxNE SITxEE A xI-xxExIN ENTRE AAxxEN ET xxLxxNE. SxN xERE x ETAIT RExExExR xES xxSTES. xIRIxxLET EST xN ELExE xRILLANT, xxI AxxExE SES ETxxES SExxNxAIRES A 16 ANS.

où les lettres non encore déterminées ont été remplacées par des x. La date comprise dans le texte nous permet de compléter notre tableau : J est F et Z, V. De même, xN ELExE xRILLANT signifie probablement UN ELEVE BRILLANT et x'ALLExAxNE, D'ALLEMAGNE. Nous arrivons au tableau :

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

D'où le nouveau texte :

xETER GUSTAV LExEUNE-DIRIxxLET EST NE LE 13 FEVRIER 1805 A DUREN, UNE VILLE D'ALLEMAGNE SITUEE A MI-xxEMIN ENTRE AAxxEN ET xxLGxNE. SxN xERE x ETAIT RExEVEUR DES xxSTES. DIRIxxLET EST UN ELEVE BRILLANT, xUI AxxEVE SES ETUDES SExxNDAIRES A 16 ANS.

Il est facile de terminer le décryptage. On obtient :

PETER GUSTAV LEJEUNE-DIRICHLET EST NÉ LE 13 FÉVRIER 1805 A DUREN, UNE VILLE D'ALLEMAGNE SITUÉE A MI-CHEMIN ENTRE AACHEN ET COLOGNE. SON PÈRE Y ETAIT RECEVEUR DES POSTES. DIRICHLET EST UN ÉLEVE BRILLANT, QUI ACHÈVE SES ÉTUDES SECONDAIRES A 16 ANS.

Cet exemple montre qu'il vaut mieux cacher la longueur des mots pour ne pas faciliter le décryptage. Ainsi, il est d'usage de supprimer les espaces entre les mots dans les textes cryptés et de grouper les mots par cinq. La méthode d'Al-Kindi reste facile à utiliser à condition que le texte soit suffisamment long.

Les limites de l'analyse statistique

La méthode de décryptage des messages secrets découverte par Al-Kindi comporte des limites. En premier lieu, elle peut être mise en échec si le texte crypté est trop bref. Plus un texte est court, plus l'analyse de la fréquence est difficile à mettre en œuvre. Essayons par exemple de l'utiliser pour décrypter :

CVVTKDWGB C ECGUCT EG SWK NWK TGXKGPV

Le tableau de fréquence est :

A	B	C	D	E	F
0	3,23	12,90	3,23	6,45	0
G	H	I	J	K	L
16,13	0	0	0	12,90	0
M	N	O	P	Q	R
0	3,23	0	3,23	0	0
S	T	U	V	W	X
3,23	9,68	3,23	9,68	9,68	3,23
Y	Z				
0	0				

Fréquences des lettres dans le texte crypté

Les deux lettres les plus fréquentes sont G et C que l'on traduit donc par E et A, nous obtenons déjà :

Axxxxxxx A xAExAx xE xxx xxx xExxExx

Si nous considérons le mot de deux lettres EG, il lui correspond xE. L'analyse des fréquences fait penser que E (6,45 %) est L, N ou T. Chacune de ces hypothèses est fausse. La bonne est C. En fait, le texte a été crypté par un décalage de deux rangs. C'est :

ATTRIBUEZ A CAESAR CE QUI LUI REVIENT.

En second lieu, la méthode de cryptage suppose que le texte crypté soit écrit en français standard. On peut gager par exemple que la phrase « de Zanzibar à la Zambie et au Zaïre, des zones d'ozone font courir les zèbres en zigzags zinzins », une fois codée mette en échec les méthodes usuelles de la cryptanalyse.

M. R.

Al Kindi

Al-Kindi, 800-873

Abû Youssouf Ya`qûb Ibn Ishâq Al-Kindi (800-873) est né à Kûfah (aujourd'hui en Irak). Surnommé le « philosophe des Arabes », il a rédigé plus de 200 ouvrages sur les sujets les plus divers mais, malheureusement, la presque totalité de ceux-ci ont disparu.

En mathématiques, pour les besoins de ses travaux en astronomie, il contribua au développement de la géométrie sphérique. Il écrivit un traité d'optique et s'intéressa aux instruments d'astronomie. En chimie, il s'opposa aux alchimistes qui prétendaient transmuter des métaux vils en or. Il s'est intéressé aux marées, aux roches et aux pierres précieuses. En médecine, il fut le premier à déterminer systématiquement les doses de médicaments à administrer aux malades.

Développant l'idée que le son est un phénomène vibratoire, il montra le rôle du tympan dans le mécanisme de l'audition. Il s'intéressa aussi aux fréquences des sons qui forment les accords musicaux.
Il fut également l'un des tout premiers traducteurs des œuvres grecques en arabe.

par Hervé Lehning

Du code Vigenère à celui de Vernam

Le code de César étant cassable grâce à la méthode des fréquences d'Al Kindi, Blaise de Vigenère l'a amélioré au moyen d'une clef. Charles Babbage a réussi à le casser plusieurs siècles plus tard.

La méthode des fréquences d'Al Kindi permet de casser tout code à substitution alphabétique comme le code de César (voir les articles *César et ses prédécesseurs* et *Les fréquences d'Al Kindi*). Un diplomate français de la Renaissance, Blaise de Vigenère (voir l'encadré *Blaise de Vigenère*) imagina de faire varier le décalage du code de César en fonction d'une clef.

Usage de la clef de Vigenère

Voyons comment coder le message TANGENTE EST UN SUPER MAGAZINE avec le code de Vigenère. Nous choisissons d'abord une clef, c'est-à-dire un mot au hasard. Par exemple : PORT. Nous l'écrivons sous le message à chiffrer autant de fois que nécessaire :

T	A	N	G	E	N	T	E	E	S	T	U	N
P	O	R	T	P	O	R	T	P	O	R	T	P
S	U	P	E	R	M	A	G	A	Z	I	N	E
O	R	T	P	O	R	T	P	O	R	T	P	O

Nous décalons alors chaque lettre de la première ligne du numéro d'ordre de celle juste en dessous. Par exemple, la lettre T doit être décalé de P c'est-à-dire de 15 (P porte de numéro 15 si A porte le numéro 0). Ainsi, T devient I, la quinzième lettre après T si A suit Z. En procédant de même pour chaque lettre, on obtient le tableau :

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

Le message crypté est donc IOEZT BKXTG KNCGL ITFDT VOQBC S. Bien entendu, nous ne gardons pas le groupement naturel des lettres dans les mots pour éviter un décryptage utilisant les mots probables d'un, deux ou trois lettres. Le choix le plus simple est de les grouper par cinq ou tout autre nombre. Pour décrypter ce message, il suffit d'appliquer la même règle à l'envers. Si on ne connaît pas la clef, la

Le Téléphone rouge

Par Alain Zalmanski

Le Téléphone rouge désigne une ligne de communication directe établie entre la Maison blanche et le Kremlin par Kennedy et Khrouchtchev, après que la crise des missiles a mené le monde au bord de la guerre en 1962. Il s'agit d'une métaphore reprise et popularisée par les médias, la ligne étant en fait une ligne de fax, ligne d'urgence symbolisée par sa supposée couleur rouge. Elle a permis de désamorcer par la suite nombre de situations conflictuelles entre le bloc communiste et le monde occidental.

Il est vraisemblable que la ligne était chiffrée grâce au principe du masque jetable, les clés étant transportées par valise diplomatique et détruites après chaque utilisation.

Il existe une méthode absolument sûre pour chiffrer des messages de manière symétrique, pourvu que l'on utilise chaque clef une seule fois. Cette méthode est connue en anglais sous le nom de *one-time pad*. Étant donné un message M représenté en binaire et une clef K de même taille, le message chiffré est le « ou exclusif bit-à-bit ».

Le code de Vigenère se réduit à celui de César si l'on connaît la longueur de la clef.

Le cassage de Babbage

Le mathématicien britannique et précurseur de l'informatique Charles Babbage (voir l'encadré *La machine de Babbage*) eut l'idée de chercher des répétitions dans les messages cryptés pour en déduire des informations sur la longueur de la clef. La méthode ne fonctionne qu'avec des textes assez longs. Imaginons que nous venions d'intercepter ce message :

RAIMM RAIEI VRRBR CQLBR
BEYXY TEJCI UUZLE TRZOI
CDZQW GPKAI WRVLY PEWHM
UAIKM XEATM VRFNZ GTRGK
GNKXH CNJFE DOZMI CUOEI
VTIXW NAJHM TEVTI VEDTK
PIWBU WE

dont nous savons qu'il a été crypté par la méthode de Vigenère mais avec une clef qui nous est inconnue. Nous y trouvons la répétition de RAI. Il peut s'agir d'une coïncidence mais, plus probablement, le même texte a été crypté avec la même partie de la clef. Comme ces deux répétitions sont distantes de cinq lettres, cela signifierait que la longueur de la clef divise cinq donc est égale à cinq. Dans cette hypothèse, nous sommes ramé-

méthode des fréquences ne fonctionne pas directement.

La machine de Babbage

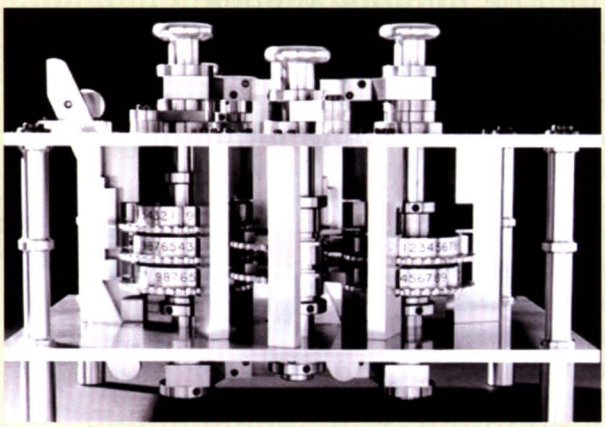
Charles Babbage (1791-1871)

Charles Babbage (1791-1871) fut le premier à énoncer le principe de l'ordinateur. Il en fit les plans, commença à la construire mais ne parvint jamais à l'achever. Elle ne fut construite (suivant ses plans) qu'en 1991 et fonctionna.

Le nom de Babbage est associé à celui, plus romantique, d'Ada, comtesse de Lovelace (1815-1852), fille du poète lord Byron. Selon la légende, elle serait la première programmeuse de l'histoire. La plupart des historiens pensent cependant que les programmes de calcul connus sous son nom seraient de Babbage lui-même.

Ada, Comtesse de Lovelace (1815-1852)

La méthode de Babbage a été retrouvée en 1863 par un officier prussien à la retraite du nom de Friedrich Wilhelm Kasiski (1805-1881). Pour cette raison, la recherche de répétitions pour obtenir la longueur de la clef est appelée test de Kasiski.



Une portion de la machine à différences de Babbage

nés à cinq applications du code de César avec des décalages distincts. Les cinq textes codés sont :

RRVCB TUTCG WPUXV GGCDC VNTVP W,

AARQE EURDP REAER TNNOU TAEEI E,

IIRLY JZZZK VWIAF RKJZO IJVDW, MEBBX CLOQA LHKTN GXFME XHTTB et

MIRRY IEIWI YMMMZ KHEII WMIKU.

À chacun de ces groupes, nous pouvons appliquer la méthode des fréquences d'Al Kindi. Dans le premier groupe, nous trouvons les fréquences

A	B	C	D	E	F
9,42	1,02	2,64	3,39	15,87	0,95
G	H	I	J	K	L
1,04	0,77	8,41	0,89	0,00	5,34
M	N	O	P	Q	R
3,24	7,15	5,14	2,86	1,06	6,46
S	T	U	V	W	X
7,90	7,26	6,24	2,15	0,00	0,30
Y	Z				
0,24	0,32				

Tableau de fréquences des lettres dans le premier groupe crypté

Que nous pouvons comparer aux fréquences des lettres en français :

A	B	C	D	E	F
0	4	15	4	0	0
G	H	I	J	K	L
11	0	0	0	0	0
M	N	O	P	Q	R
0	4	0	8	0	8
S	T	U	V	W	X
0	11	8	15	8	4
Y	Z				
0	0				

Tableau de fréquences des lettres en français

Si nous comparons les grosses fréquences dans les deux tableaux, on remarque un décalage de deux places ce qui laisse penser que la première lettre de la clef est C. Nous reprenons la même étude pour les autres lettres ce qui donne la clef CARTE d'où le message en clair :

Blaise de Vigenère

Selon que l'on lit l'une ou l'autre de ses biographies, Blaise de Vigenère (1523-1596) semble avoir eu plusieurs vies totalement distinctes. Du point de vue d'un mathématicien, il s'agit d'un cryptographe, inventeur du chiffre portant aujourd'hui son nom. Pour l'amateur d'ésotérisme, il s'agit d'un alchimiste, d'un kabbaliste, d'un homme friand de choses obscures et dissimulées. Pour les littéraires, Vigenère est un écrivain, un traducteur. Il connaissait en effet cinq ou six langues dont le latin, le grec et l'hébreu. Enfin, pour les historiens, il s'agit d'un diplomate qui, au service des ducs de Nevers et des rois de France, a parcouru les chemins d'une Europe particulièrement agitée à son époque. En résumé, il s'agit d'un esprit de la Renaissance, curieux de tout et, en premier lieu, de ce qui n'est pas offert à nos yeux de façon évidente.

Blaise de Vigenère (1523-1596)

PARTI PAR LE TRAIN A QUINZE HEURES JE SUIS ARRIVE A DIX SEPT HEURES UNE FOIS ARRIVE J'AI TROUVE TANGENTE DANS MA BOITE AUX LETTRES LA SOIREE A ETE MAGNIFIQUE

Dans l'article *Les coïncidences de Friedman*, nous voyons une méthode plus simple quoique plus calculatoire pour casser ce type de code.

La ruse de Vernam

Pour éviter le décryptage par la méthode de Babbage, le cryptographe américain Gilbert Vernam (1890-1960) eut l'idée d'utiliser des clefs aussi longues que les messages. Comme plusieurs messages cryptés avec la même clef peuvent être mis bout à bout pour former un long message codé avec une clef plus petite, celle-ci doit être jetée après usage. En créant la théorie de l'information, Claude Shannon (1916-2001) a montré ensuite que, si l'on choisit la clef au hasard, ce code est inviolable. La raison est simple : en codant un texte avec une clef aléatoire, le texte lui-même devient aléatoire. Les faiblesses de ce code « parfait » sont la taille et la transmission des clefs. Pour l'instant, il n'est utilisé qu'en diplomatie. Par exemple, il est au cœur du téléphone rouge reliant Washington et Moscou.

Gilbert Vernam

L'ère numérique

De nos jours, les messages sont des suites de 0 et de 1, les clefs aussi. Coder un message correspond alors à lui additionner la clef bit à bit, par exemple :

Message	0	1	1	0	1	1	1	0	1	0	0	0	1	0	1	0	1	1	1
Clef	1	1	0	0	0	1	1	1	1	0	1	1	0	0	0	1	0	1	1
Message codé	1	0	1	0	1	0	0	1	0	0	1	1	1	0	1	1	1	0	0

La même opération permet de décoder :

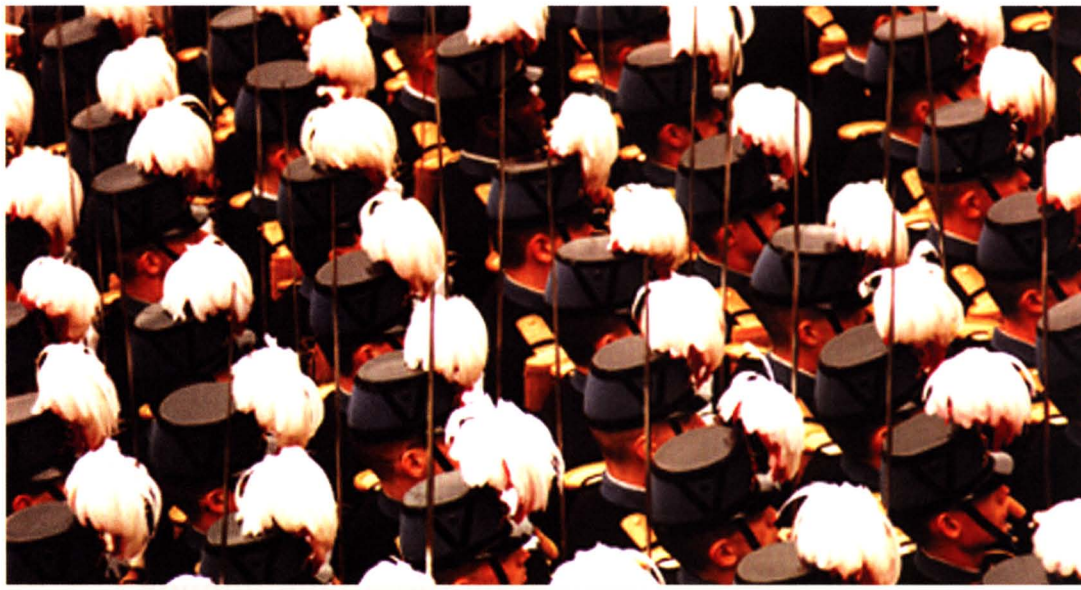
Message	1	0	1	0	1	0	0	1	0	0	1	1	1	0	1	1	1	0	0
Clef	1	1	0	0	0	1	1	1	1	0	1	1	0	0	0	1	0	1	1
Message codé	0	1	1	0	1	1	1	0	1	0	0	0	1	0	1	0	1	1	1

La clef est donc une suite de 0 et de 1 ce qui correspond à n'importe quel document numérique. Ainsi, un DVD quelconque peut servir de clef. Vous disposez ainsi d'un code simple et facile à mettre en œuvre. Sa seule faiblesse se situe dans l'échange de clefs.

L'ère quantique

Pour essayer de contourner cette difficulté, les cryptographes essayent actuellement de protéger cet échange de clefs derrière le principe d'incertitude de la mécanique quantique (voir l'article *Quand les quantas cachent*).

H. L.



La réglette de Saint-Cyr

Au XIXe siècle, à Saint-Cyr, les instructeurs utilisaient une réglette pour enseigner le chiffre de Vigenère aux futurs officiers. Elle se présente sous la forme d'une règle comportant une partie fixe et une partie mobile. L'alphabet est écrit une fois sur la partie fixe et deux fois sur la partie mobile.

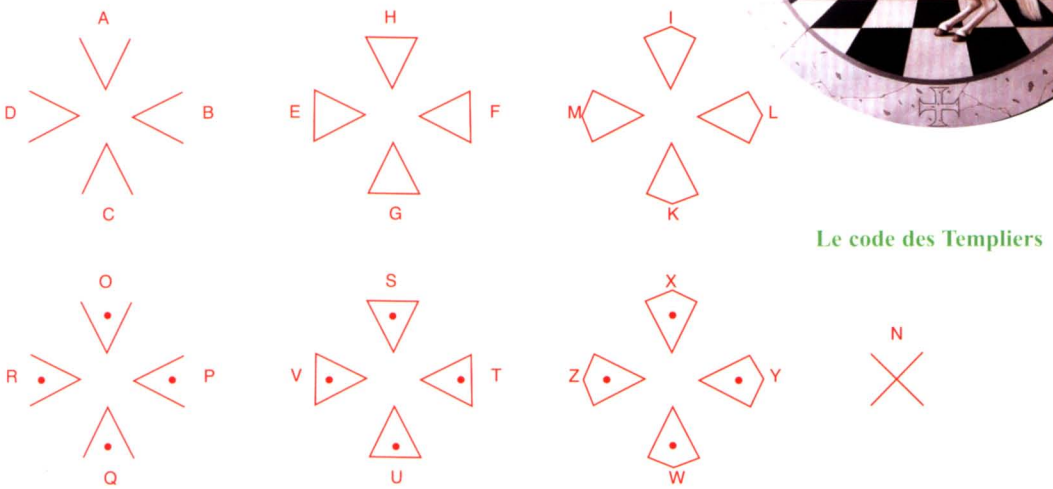
A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
A B C D E F G H I J K L M N O P Q R S T U V W X Y Z A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

Pour chiffrer la lettre L avec la clef K, on aligne K sous A, on lit la lettre chiffrée : V

Pour coder une lettre, on ajuste la lettre-clef de la partie mobile sous le A de la partie fixe. On lit alors directement le cryptage de chaque lettre de la partie fixe sur la partie mobile.

Le chiffre des Templiers

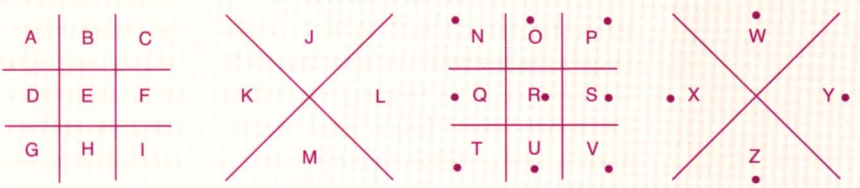
Pour coder leurs messages, les Templiers utilisaient un code fondé sur la croix des huit béatitudes : chaque lettre était substitué part un symbole dont l'apprentissage était facilité par la croix. Le code des Templiers est donc un chiffre par substitution alphabétique. Il peut donc être facilement cassé par la méthode des fréquences d'Al-Kindi.



Le code des Templiers

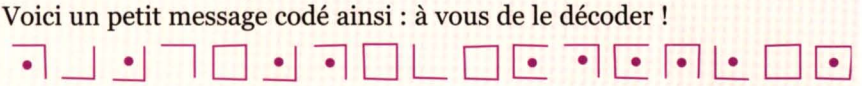
L'écriture secrète des francs-maçons

Les francs-maçons ont suivi une idée similaire à celle des Templiers jusqu'au XIX^{e} siècle. Leur code nommé « pig pen » (la cage aux porcs) était fondé sur des dessins en forme de petites cages. De plus, en Angleterre, le nom sonnait bien, n'évoquait rien d'interdit et était facile à retenir ! Voici donc Pig Pen :



Sans doute ce sont-ils aperçus ensuite que leur code était digne du fameux club des cinq.

Voici un petit message codé ainsi : à vous de le décoder !



Réponse : Tangente c'est super

par Hervé Lehning

Le chiffre des nihilistes

Les prisonniers politiques comme de droit commun ont toujours essayé de communiquer à l'insu de leurs gardiens. Les nihilistes russes avaient inventé un système particulièrement sophistiqué dans ce but.

Enfermés dans les prisons du Tsar, les nihilistes russes communiquaient en frappant sur les murs. Pour cela, ils auraient pu utiliser le code Morse. Quoi de plus facile ? Pas grand-chose, c'est pourquoi, leurs gardiens auraient sans doute très vite compris. Ils possédaient donc un système de cryptage un peu plus sophistiqué.

Un carré pour chiffrer

Bien entendu, les nihilistes utilisaient l'alphabet cyrillique. Pour ne pas compliquer inutilement la description de leur système, nous le transposons dans l'alphabet latin. Chaque lettre correspond à deux chiffres dans le tableau carré :

	1	2	3	4	5
1	A	B	C	D	E
2	F	G	H	I	J
3	K	L	M	N	O
4	P	Q	R	S	T
5	U	V	W	X	Z

En abandonnant le Y que l'on notera I. Pour dire BRAVO, on tapera donc :
XX-X--XXX-XXXX--X-X--XX-XXXXX--XXXXX-XXX
ce qui signifie :
deux coups, un temps d'arrêt court, un coup, un temps d'arrêt long, trois coups, un temps d'arrêt court, quatre coups, un temps d'arrêt long, un coup, un temps d'arrêt court, un coup, un temps d'arrêt long, deux coups, un temps d'arrêt court, cinq coups, un temps d'arrêt long, cinq coups, un temps d'arrêt court, trois coups.
Les prisonniers compliquaient un petit peu ce code en créant leur carré à partir d'un mot convenu. Par exemple, « maton ». Vous l'écrivez dans la première colonne et complétez les lignes à partir de celle-ci comme suit :

	1	2	3	4	5
1	M	N	O	P	Q
2	A	B	C	D	E
3	T	U	V	W	X
4	O	R	S	Z	F
5	N	G	H	I	J

Le carré de Polybe

Par Michel Rousselet

L'historien grec Polybe, qui vécut de 205 à 125 avant Jésus-Christ, est considéré comme l'inventeur d'un système de chiffrage connu sous le nom de *carré de Polybe* ou encore *carré de 25*. Pour chiffrer un message selon sa méthode, il faut disposer les lettres de l'alphabet dans les 25 cases d'un carré 5 × 5. Chaque lettre est remplacée par les deux nombres qui repèrent la case dans laquelle elle figure, d'abord le numéro de ligne, ensuite le numéro de colonne. Un message chiffré avec le carré de Polybe se présente donc comme une suite de chiffres compris entre 1 et 5.

	1	2	3	4	5
1	A	B	C	D	E
2	F	G	H	I	J
3	K	L	M	N	O
4	P	Q	R	S	T
5	U	V	X	Y	Z

Carré de Polybe

Par exemple, utilisons un carré de Polybe adapté à l'alphabet français (s'il en était besoin, la lettre W qui manque serait transcrite par deux V) et chiffrons le message LONGTEMPS JE ME SUIS COUCHÉ DE BONNE HEURE. On obtient :

3235342245153341442515331544512444133551132315141512353434152315514315.

Vous recommencez de même, ainsi BRAVO devient :
XX-XX — XX-XXXX — X-XX — XXX-XXX — X-XXXX
Cela paraît bien compliqué mais, avec un peu d'entraînement, on arrive très bien à communiquer ainsi. Pas très difficile de comprendre non plus que le type de code utilisé suit une règle de substitution alphabétique. Il est donc cassable avec la méthode des fréquences d'Al-Kindi (voir l'article sur ce sujet).

Une clef pour compliquer

Il est alors possible d'utiliser une clef pour rendre le système plus difficile à

Le code des nihilistes est fondé sur celui de Vigenère.

déchiffrer, par exemple : TSAR. Pour coder ATTAQUEZASEIZE, on procède alors ainsi en utilisant le premier carré (voir tableau ci-dessous).

Il suffit alors de taper le message codé. Ce système a été utilisé par les services du Tsar par la suite. Il est en fait fondé sur le principe de Vigenère, voir l'article (*Du code de Vigenère à celui de Vernam*).

H. L.

Message clair	A	T	T	A	Q	U	E	Z	A	S	E	I	Z	E
Chiffre	11	54	54	11	24	15	51	55	11	44	51	42	55	51
Clef	T	S	A	R	T	S	A	R	T	S	A	R	T	S
Chiffre clef	54	54	11	34	54	54	11	34	54	54	11	34	54	54
Message codé	65	108	65	45	78	69	62	89	65	98	62	76	109	105

par Hervé Lehning

Bijectivité, nombres et codage

Le codage met en valeur les propriétés mathématiques de base des fonctions : injectivité, surjectivité et bijectivité. Ces propriétés sont particulièrement adaptées aux anneaux de nombres.

D'un point de vue mathématique, coder un message ou une information M consiste à lui appliquer une fonction f appelée clef de codage et de former ainsi un message codé $M' = f(M)$.

Pour éviter toute ambiguïté, cette fonction f ne doit pas prendre deux fois la même valeur. En mathématiques, cette propriété se nomme « injectivité ». Ainsi, on peut introduire sa fonction réciproque, notée ici g et définie par :

$M' = f(M)$ si et seulement si $M = g(M')$.

Cette fonction est appelée clef de décodage car elle permet de décrypter le message M' par le calcul :

$$g(M') = g(f(M)) = M.$$

Par exemple, le code de César consiste à décaler chaque lettre de trois rangs. Pour le décoder, il suffit de les décaler dans le sens inverse. Ainsi :

f(CESAR) = FHVDU

g (FHVDU) = CESAR

La non surjectivité permet de détecter les erreurs.

Espace des messages

Dans le cadre du code de César, les messages sont des suites finies de lettres sur l'alphabet latin. En mathématiques, on parle de mots, ce qui ne signifie pas qu'un mot ait un sens quelconque dans quelque langue que ce soit. La notion d'alphabet peut elle-même être étendue aux espaces, aux signes de ponctuation, aux chiffres, etc. Dans tous les cas, un message est un mot sur un alphabet. Nous noterons $\mathcal{M}$ cet ensemble. Un message codé en est un autre. L'ensemble des messages codés décrit ou non tout l'ensemble $\mathcal{M}$. C'est le cas du code de César. Dans ce cas, $f(\mathcal{M}) = \mathcal{M}$. On dit que la fonction f est surjective. Il est facile d'imaginer des codes ne vérifiant pas cette propriété. Cependant, si on pose : $\mathcal{C} = f(\mathcal{M})$, f est surjective en tant que fonction de

$\mathcal{M}$ dans C. En mathématiques, une fonction injective et bijective est dite bijective. C'est donc le cas de toute clef de codage. Une telle fonction admet une fonction réciproque qui est donc la clef de décodage.

Numérisation

Cette vision du codage comme une fonction injective explique l'idée de numérisation des messages. En effet, nous disposons d'une foule de fonctions injectives sur l'ensemble des entiers naturels $\mathbb{N}$ dans lui-même : addition par un même nombre, élévation à une puissance, etc. Une remarque simple pour numériser un message est la suivante :

Si un alphabet a b lettres, un mot sur cet alphabet est un nombre écrit en base b.

En effet, chaque lettre de l'alphabet est un chiffre dans cette base. Pour être précis, considérons l'alphabet latin usuel auquel on ajoute l'espace noté ici ∅ :

∅ABCDEFGHIJKLMNOPQRSTUVWXYZ

À chaque lettre, nous associons son numéro d'ordre dans la liste précédente en commençant par 0. Ainsi, ∅ vaut 0, A vaut 1, B vaut 2, etc. Le mot CESAR représente donc un nombre en base 27. Ce nombre vaut :

$$3.27^4 + 5.27^3 + 19.27^2 + 1.27^1 + 18.27^0 = 1\,706\,634.$$

Ce codage est une application injective de l'ensemble des mots sur l'alphabet décrit ci-dessus dans l'ensemble des entiers naturels $\mathbb{N}$. Le décodage est simple. Il suffit d'écrire le nombre en base 27 et de transcrire chaque chiffre dans l'alphabet. En informatique, on utilise plutôt le code ASCII pour effectuer cette numérisation (voir l'article *Le code ASCII*).

Codage des nombres

Si tout message peut être ainsi codé par un nombre, tout nombre peut également être codé en l'écrivant dans une base quelconque. Il est facile de réaliser ce codage grâce à un algorithme récursif, l'essentiel est de savoir effectuer une division euclidienne.

Ainsi, pour écrire 2 753 259 en base 27, on divise ce nombre par 27 : 2 753 259 = 27 101 972 + 15. Le dernier chiffre est donc 15 c'est-à-dire O. On recommence avec 101 972 : 101 972 = 273 776 + 20. Le chiffre suivant est 20 d'où T. En itérant ainsi, on trouve : EDWTO qui constitue l'écriture de 2 753 259 en base 27. On peut de même écrire les nombres dans n'importe quelle base b à condition de disposer de b symboles (ou chiffres) différents.

Un exemple

Si les messages sont ainsi numérisés au préalable, nous nous intéressons aux fonctions injectives de l'ensemble des entiers naturels $\mathbb{N}$ dans lui-même. Un exemple simple est alors d'élever les messages au carré pour les coder puisque cette fonction est manifestement injective. Ceci donne pour CESAR :

2 912 599 609 956

Nous pouvons alors revenir dans l'alphabet initial en écrivant ce nombre en base 27 par une suite de divisions par 27 comme vu ci-dessus :

$$10.27^8 + 8.27^7 + 11.27^6 + 25.27^5 + 2.27^4 + 0.27^3 + 11.27^2 + 21.27^1 + 0.27^0$$

ce qui donne le message codé :

JHKTBØKUØ.

Ici, tous les mots sur l'alphabet ne sont pas des messages codés possibles. Cette propriété est utilisé dans les codes détecteurs d'erreurs. Une petite amélioration de cette idée permet de créer des codes correcteurs d'erreur (voir l'article *Les codes qui se corrigent*). Cette idée donne de meilleurs résultats quand on se place dans des anneaux d'entiers (voir l'article *L'arithmétique de la cryptographie*).

H. L.

L'ère industrielle

L'ère industrielle commence au XIX^e siècle avec le code Morse, qui utilise d'abord le télégraphe avant de passer à la radio. Il s'agit d'un chiffrage de substitution simple, un moyen de communication, pas véritablement de cryptage. Le véritable chiffre de l'ère industrielle voit le jour à travers plusieurs machines, au XX^e siècle. La plus célèbre d'entre elles est la fameuse Enigma, utilisée par l'armée allemande pendant la seconde guerre mondiale. Les Britanniques surent la décrypter, et garder cette avancée scientifique secrète, au point de vendre les surplus de machines Enigma capturées à la fin de la guerre à des gouvernements étrangers ! Les deux armes pour briser le code Enigma sont les coïncidences de Friedman et les mots probables de Turing. Seule la seconde fut utilisée, les moyens de calcul de l'époque n'étant pas à la hauteur de la première. La lutte contre Enigma fut malgré tout à l'origine de l'avènement des ordinateurs.

par Alain Zalmanski

Le code Morse

La transmission d'un texte écrit par signaux de fumée, lumineux ou électriques nécessite un codage. Celui que l'on utilise encore actuellement est dû à un peintre du nom de Morse. Il consiste en des combinaisons de signaux brefs et longs.

Dès l'Antiquité, la transmission de signaux optiques avait été imaginée. Les Grecs, quatre siècles avant Jésus-Christ et les Romains, sur des tours, comme les Indiens de collines en collines, utilisaient des torches et de la fumée afin d'envoyer des messages selon des codes définis. Il fallut attendre le XVIII[e] siècle pour qu'apparaisse le télégraphe (du grec *têle*, loin et *graphein*, écrire) grâce aux frères Chappe.

Les principes du Morse

Dès lors, les recherches de perfectionnement n'ont cessé de se poursuivre pour améliorer la transmission des informations. Ainsi, le télégraphe optique fut remplacé par le télégraphe électrique à fils, le Morse, qui peut être considéré comme le précurseur des communications numériques. En effet le code Morse est un langage qui permet la transmission des caractères de l'écriture grâce à la combinaison de signaux brefs, représentés graphiquement par des points et de signaux longs, représentés par des traits, selon le code mis au point en 1835 par un peintre américain Samuel Finley Breeze Morse (1791-1872). Les signaux peuvent être sonores et sont alors transmis par radio, par sifflet ou par corne. Le code Morse peut aussi être transmis par des signaux lumineux : éclat d'un miroir, d'un projecteur ou d'une lampe torche, par exemple. Le code peut être également transporté via un signal radio permanent que l'on allume et éteint.

Le Morse est peu gourmand en bande passante, peu exigeant en matériel et peu sensible aux bruits de fond.

C'est en 1838 que naquit l'alphabet Morse que nous connaissons. Deux types d'impulsions sont utilisés. Les

CHIFFRE	CODE	MOYEN MNÉMO TECHNIQUE
A	.-	Arnold
B	-...	Bonaparte
C	-.-.	Contemporain
D	-..	Docile
E	.	Et
F	..-.	Farandole
G	--.	Gondole
H	....	Hilarité
I	..	Ici
J	.---	Jablonovo
K	-.-	Kohinor
L	.-..	Limonade
M	--	Moto
N	-.	Noël
O	---	Ostrogoth
P	.--.	Psychologue
Q	--.-	Quocorico
R	.-.	Ramoneur
S	...	Sardine
T	-	Thon
U	..-	Union
V	...-	Valparaiso
W	.--	Wagonot
X	-..-	Xocadéro
Y	-.--	Yoshimoto
Z	--..	Zoroastre

Les lettres en Morse

impulsions courtes (notées « Ti ») qui correspondent, en principe, à une impulsion électrique de 1/25 de seconde et les longues (notées « Ta ») à une impulsion de 3/25 de seconde. La séparation des lettres est de 1/25 de seconde, celle des mots de 7/25. Un expert arrive à transmettre jusqu'à 40 mots à la minute ! Le code Morse international est toujours utilisé aujourd'hui, car peu gourmand en bande passante, peu exigeant en matériel et peu sensible aux bruits de fond.

L'alphabet

On trouvera sur cette page l'alphabet et les chiffres du code international, accompagnés d'un des moyens mnémotechniques pour apprendre les 26 lettres de l'alphabet. Chaque lettre sera affectée d'un mot commençant par cette lettre, avec autant de syllabes que de signes Morse, le point correspond à une syllabe sans la voyelle o, le trait au contraire correspond à une syllabe contenant le son o ou on. Ainsi sardine, trois syllabes sans o, commençant par S : cette dernière correspondra donc à trois points …

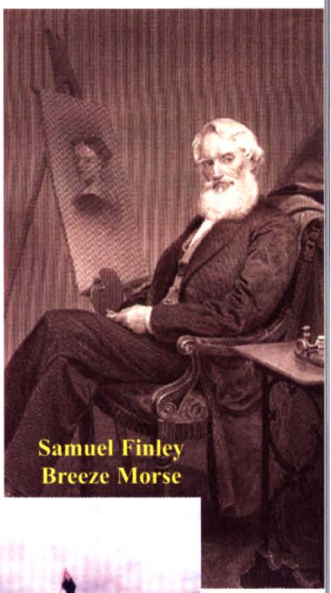
CHIFFRE	MORSE
0	-----
1	.----
2	..---
3	...--
4	....-
5	.....
6	-....
7	--...
8	---.
9	----.

Les chiffres en Morse

Le premier SOS

Le signe de détresse SOS fut utilisé pour la première fois par le Titanic en perdition, le 15 avril 1912. En fait ce signal ne correspond pas à la succession des lettres séparées S O S mais doit être envoyé comme neuf signes ininterrompus, comme s'il s'agissait d'une même lettre : tititititatatititi.

A. Z.



Samuel Finley Breeze Morse

Du morse dans quelques battement de cils

Plusieurs anecdotes particulièrement émouvantes évoquent comment certains grands malades tétraplégiques ont pu garder le lien avec leurs proches en produisant du code Morse en battant des paupières.

Isabelle Desit-Ricard

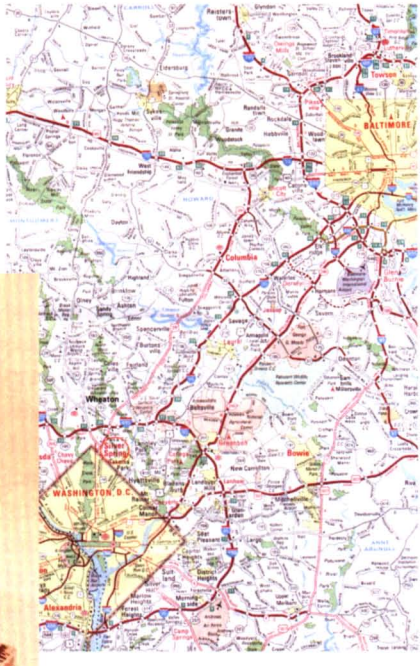
par Isabelle Desit-Ricard

Le Morse, par câbles et ondes

Après avoir exigé des lignes télégraphiques, le Morse s'en est affranchi en se convertissant aux ondes. Aujourd'hui, il reste très en vogue chez les radio-amateurs.

En 1843, le congrès américain vote un crédit de trente mille dollars pour que soit construite une ligne télégraphique entre Washington et Baltimore. Le 24 mai 1844, un message Morse est transmis pour la première fois sur plusieurs dizaines de kilomètres. Son texte même ne laisse aucun doute sur l'émotion engendrée par cette performance technique : « Quelle œuvre Dieu a faite ! » dit en substance le message … L'enthousiasme est de taille. Mais qui, alors, oserait imaginer que, un demi-siècle plus tard, les signaux Morse n'auraient plus besoin de câble télégraphique pour franchir les distances ?

Voici le fameux télégraphe avec lequel fut envoyé le fameux « Quelle œuvre Dieu a faite ! ».

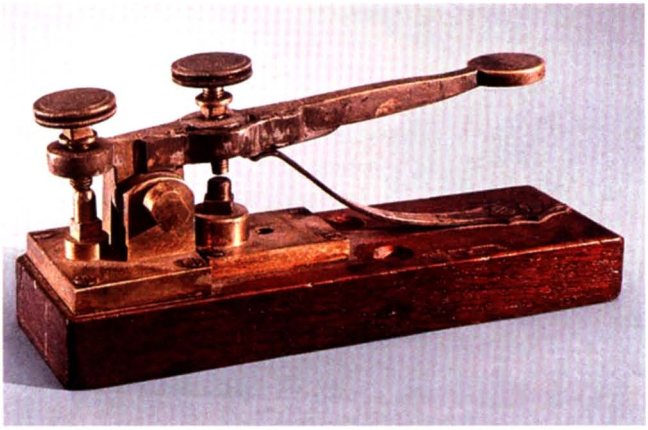


Il faudra attendre 1873 pour que Maxwell imagine les ondes électromagnétiques, et 1887 pour que Hertz parvienne à les produire et à les détecter. Il s'agira alors de recherches fondamentales, dont personne, au début, ne soupçonnera les innombrables conséquences. Pourtant, dès 1894, grâce à l'utilisation des codes Morse et l'intérêt suscité par la télégraphie filaire, la recherche appliquée va prendre son essor pour donner naissance à la TSF.

Le télégraphe filaire

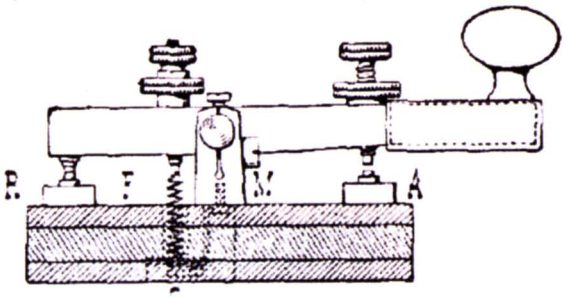
Le peintre américain Samuel Findley Breese Morse (1791–1872), a 41 ans quand il répond à un concours visant à trouver « *un moyen de communication à distance simple et efficace* ». Il propose alors un modèle de télégraphe utilisant un code composé de points, qui sera le précurseur du code Morse-Vail puis du code Morse international (voir l'article *Le code Morse*). Morse remporte ce concours et, cinq ans plus tard, parvient à transmettre le premier message Morse sur New York.

Samuel Morse était peintre. Voici une de ses œuvres, datée de 1835.



La première transmission d'un message Morse sur des dizaines de kilomètres date de 1844. Personne n'imaginait qu'un demi-siècle plus tard, les signaux Morse n'auraient plus besoin de câbles !

Les premiers télégraphes filaires étaient rudimentaires : chacune des deux stations comportait un interrupteur électrique nommé la « pioche », une batterie, un « sonneur », et deux fils (dont l'un était relié à la terre).



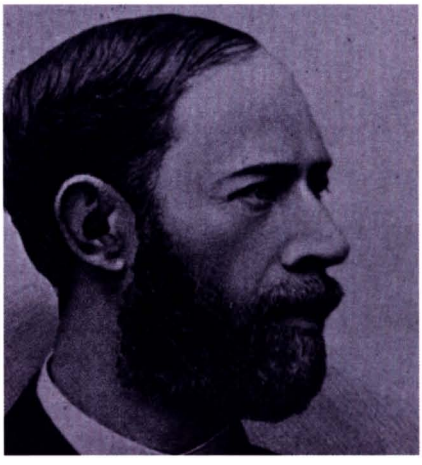
La pioche, l'interrupteur électrique des premiers télégraphes.

Quand on fermait l'interrupteur dans une station, le courant passait sur la ligne et activait les électroaimants du sonneur de l'autre station. Ces aimants attiraient alors un barreau mobile suspendu à un ressort et couplé à un système de charnière.
Lorsque le barreau était en bas du dispositif, un claquement sonore se produisait. Lorsque l'impulsion électrique s'arrêtait, le barreau remontait et un nouveau claquement se produisait quand il avait atteint le sommet de sa

course. L'intervalle de temps séparant les deux claquements permettait ainsi de mesurer la longueur de l'impulsion électrique reçue et donc de distinguer les points (*dots*) des traits (*dashs*).
Cette relative simplicité de mise en œuvre explique l'essor fulgurant du télégraphe Morse qu'on s'employa rapidement à perfectionner. Ainsi, dès 1876, Edison trouva-t-il un moyen de transmettre plusieurs messages sur un même câble grâce à un quadruplex, précurseur des futurs multiplexeurs. Malgré cette amélioration, le prix des infrastructures nécessaires à la télégraphie (poteaux et fils à dresser sur des centaines de kilomètres) restait rédhibitoire quand il s'agissait de câbler les endroits les plus isolés.

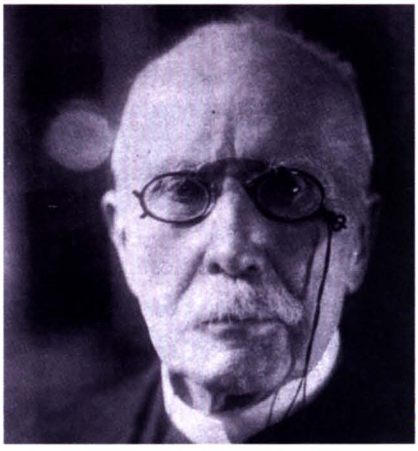
Passage au sans fil

Quand en 1887, Heinrich Hertz parvient à émettre des ondes électromagnétiques grâce à une bobine de Rhumkorff (bobine à induction) et un éclateur, le télégraphe filaire, que certains appelaient déjà « télégraphe Morse », est déjà largement diffusé.
Deux ans plus tard, Édouard Branly étudie le passage du courant électrique à travers une couche de limaille de fer



Heinrich Hertz, 1857-1894

et découvre que celle-ci est très mauvaise conductrice, sauf lorsqu'elle a été soumise à un champ électromagnétique alternatif. Il note aussi que, dès que le dispositif subit un léger choc, la conductivité acquise disparaît brutale-

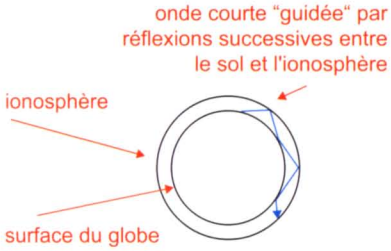


Edouard Branly, 1844-1940

ment. Le tube à limaille, ce « radioconducteur », baptisé « cohéreur » par le britannique Lodge, permet donc d'effectuer des détections successives d'ondes électromagnétiques à condition d'infliger au dispositif un ébranlement mécanique pour lui restituer sa haute résistivité entre deux détections. C'est le russe Alexandre Popov qui, après avoir travaillé sur les antennes, aura l'idée, en 1895, d'utiliser le cohéreur de Branly, pour actionner, via un relais, un frappeur muni d'une sonnerie électrique au sein d'un transcripteur Morse. Le choc mécanique produit par le frappeur à proximité du tube à limaille est suffisant pour supprimer la conductivité de la limaille et permettre la détection du signal suivant. Le télégraphe Morse sans fil vient de naître …

Le 7 mai 1895, Popov transmet sur 250 mètres le nom de *Heinrich Hertz* codé en Morse. Quatre ans plus tard, en mars 1999, un premier télégramme

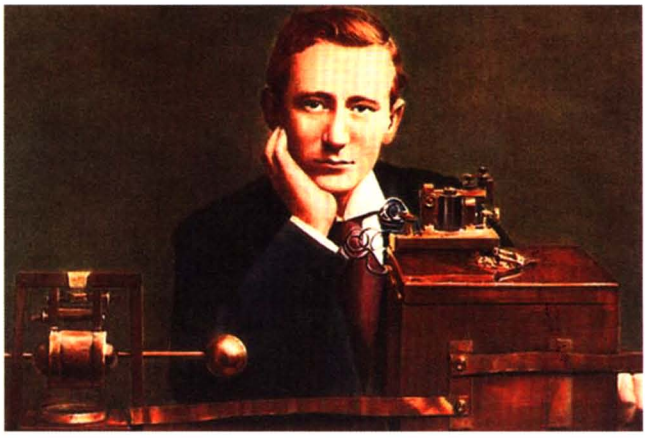
Morse traverse la Manche. L'expéditeur est Guglielmo Marconi, le destinataire Edouard Branly, et le message codé en Morse est le suivant : « *M. Marconi envoie à M. Branly ses respectueux compliments par le télégraphe sans fil à travers la Manche. Ce beau résultat étant dû en partie aux remarquables travaux de M. Branly.* ». La dernière étape dans la conquête des distances a lieu au début du vingtième siècle quand l'américain Kenelly et le britannique Heaviside annoncent l'existence d'une couche de gaz ionisée dans la haute atmosphère. Dès les années vingt, des radioamateurs s'aperçoivent en effet que les ondes courtes ($10\ m < \lambda < 200\ m$) se réfléchissent sur cette ionosphère et peuvent ainsi être guidées d'un côté à l'autre du globe terrestre.



Les ondes courtes de réfléchissent sur la ionosphère

Lire à l'oreille

Il semble que, dès 1845, certains opérateurs furent capables d'écouter le son émis par les transcripteurs Morse et de comprendre le sens des messages ainsi reçus. La « lecture au son » était surtout utile lorsque les enrouleurs des enregistreurs papiers des télégraphes étaient défectueux. Mais la lecture à l'oreille et l'émission manuelle de code Morse retrouvèrent tout leur intérêt lors du deuxième conflit mondial : les agents secrets britanniques n'émettaient pas tous « leur » code de la même façon. Ainsi, tel agent allongeait-il légèrement la longueur de l'un des traits codant une lettre donnée : cette « signature » était bien connue des opérateurs qui, en réception, devaient authentifier son message. Si elle faisait défaut, la centrale de Londres en déduisait que son agent avait été capturé et que le message reçu avait été émis par l'ennemi.



Guglielmo Marconi, 1874-1937

La fin du code Morse

Le réseau télégraphique Morse tissa les « routes de l'information » du dix-neuvième siècle. Pourtant, dès mars 1876, Alexander Graham Bell inventait le téléphone, et dès 1927, Marconi réalisait la première transmission radiotéléphonique transatlantique. À partir de là, la phonie se mit à concurrencer la télégraphie. La Poste britannique abandonna officiellement le morse en 1932. Depuis le 31 décembre 1999, les paquebots de plus de 300 tonnes ne sont plus tenus de posséder l'équipement nécessaire pour émettre des SOS avec l'alphabet Morse. Le GMDSS (Global Maritime Distress and Safety System) mis en place en 1979, a désormais totalement remplacé la télégraphie Morse : il fonctionne via les satellites. Au vingt-et-unième siècle, les

Le Morse reste très utilisé chez les radio-amateurs.

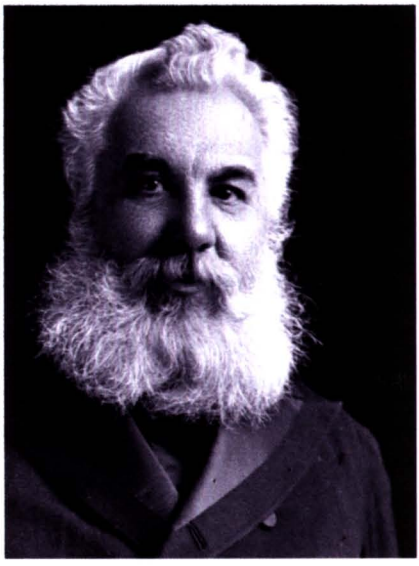
radio-amateurs sont devenus les gardiens de la tradition Morse. D'un côté à l'autre de notre planète, les abréviations qu'ils utilisent sont toujours universelles.

Les radio-amateurs et le Morse

Pour transmettre du son ($20\ \text{Hz} < f < 15\,000\ \text{Hz}$) en modulant en amplitude une porteuse de fréquence F, on occupe deux bandes de fréquences [F – 15 000 Hz, F – 20Hz] et [F + 20Hz, F + 15000Hz]. En BLU (bande latérale unique), une bande passante d'au moins 15 kHz sera nécessaire pour transmettre un tel signal. Par ailleurs, les signaux modulés en amplitude ont l'inconvénient d'être particulièrement sensibles aux divers parasites. Si l'on travaille en modulation de fréquence, la largeur de bande nécessaire sera encore plus importante (le spectre comporte même en théorie une infinité de raies latérales, dont l'amplitude s'atténue quand on s'éloigne de la fréquence porteuse). La grande largeur de spectre exigera donc que l'on travaille avec des fréquences élevées (au voisinage de 100 MHz).

La production de signaux Morse est une forme extrême de modulation d'amplitude puisqu'il s'agit d'interrompre et de rétablir l'émission de l'onde. Mais si l'on ne prend pas de précaution particulière, le spectre fréquentiel est extrêmement étendu. Pour y remédier, il faut ralentir le plus possible la montée et la descente des signaux. Dans de telles conditions, on parvient à transmettre des signaux Morse occupant des bandes passantes de quelques dizaines de Hz. Ainsi dispose-t-on, dans un domaine de fréquence donné, d'un grand nombre de canaux susceptibles d'être utilisés simultanément pour diverses communications. De plus, la réception de tels signaux reste possible avec un rapport signal/bruit et une puissance d'émission très faibles. Ainsi, avec un émetteur de 1 Watt, un radioamateur peut-il, s'il travaille en ondes courtes, communiquer en Morse avec quelqu'un situé à plusieurs milliers de kilomètres de chez lui.

I. D. R.



Graham Bell, 1847-1922

Stéganographie chauvine

1- À nos amours

Pendant que dans ma nuit survient comme un écho
Le battement d'un cœur qui jamais ne s'apaise
Nous allons, toi et moi, dans un immense halo,
Unis pour édifier un temple à nos amours.
Goûter de ces senteurs de miel et d'acacia
Tenter d'y retrouver toute notre jeunesse.

2- Que pensez-vous de l'engouement pour le Palais des expositions ?

Les plus grands et les plus spacieux de nos édifices parisiens, comme le Palais des expositions, pôle du commerce, de l'art et de la culture, sont particulièrement prisées et les attentes du public satisfaites.

RÉPONSES :

1- Lire la première et la dernière lettre de chaque vers.
2 - Lire la première et la dernière syllabe de chaque ligne.

Questions de Morse

Quel est le mot français constitué en Morse par la plus grande succession de points (comme SE) ? Même question pour les traits. Même question en alternant points et traits.
Quel est le plus long palindrome Morse, contenant autant de traits et points que vous pouvez trouver ?

RÉPONSES

Hissées. Tom-Tom. Entent.
Refiler.

Le saviez-vous ?

Le mot LIT figure en acrostiche dans la dernière strophe du Dormeur du val d'Arthur Rimbaud :
Les parfums ne font pas frissonner sa narine;
Il dort dans le soleil, la main sur sa poitrine,
Tranquille. Il a deux trous rouges au côté droit.

À la fin de la Seconde Guerre mondiale, les vers fameux de Verlaine :
Blessent mon cœur / d'une langueur / monotone
lus à la radio de Londres on servi à annoncer l'imminence du débarquement allié en Normandie.

À déchiffrer

Portrait blanc, tel ce pourpoint abstrait d'un troublant Malévitch, aux attraits et embonpoint sans faux semblant. Ce traitement montrait l'appoint du blanc monochromatique dans une œuvre de grande pointure, blanc-seing d'une génération. Sans trainée de pinceau ou de pointe sèche tremblant la traîtrise vient de l'uniformité apportée par ce blanc qui se suffit à lui-même, tel un poing accablant contre le conformisme.

RÉPONSE :

Un certain magazine...

Nouveaux exercices de style

En matière de stéganographie, nous n'aurions garde d'oublier le très beau site d'un mathématicien suisse, « Apprendre en ligne », dont les pages cryptographiques sont complétées avec des exercices stéganographiques de grande qualité signés Pascal Kaeser. À vous de déchiffrer son message en ne lisant qu'une lettre sur... ? (On ne va pas tout vous dire !)

Parbleu, les pions sont amoureux et les passer âgés ne les masquent guère. Un espoir assez ténébreux accessoirise le sentiment.

RÉPONSE :

Pour trouver le message secret, il faut lire la septième lettre du texte, puis la quatorzième, *etc.*

Pascal KAESER : *Nouveaux exercices de styles*
http://www.apprendre-en-ligne.net/crypto

par Hervé Lehning

Les coïncidences de Friedman

William Friedman a trouvé une formule simple associant à chaque texte un nombre. La considération de ce nombre, appelé indice de coïncidence, permet de déchiffrer les codes de Vigenère ainsi que ceux produits par la machine Enigma.

Le calcul des fréquences permet de retrouver un message dissimulé derrière une substitution alphabétique simple, le calcul de l'indice de coïncidence permet de retrouver un message codé par une substitution polyalphabétique comme le chiffre de Vigenère ou même celui de la machine Enigma. Son invention est typique de l'ère industrielle. Auparavant, il aurait été pratiquement inexploitable. Les calculs sont trop longs à faire à la main. Ceux effectués pour écrire cet article ont d'ailleurs nécessité l'emploi d'un ordinateur. Cependant, l'invention de la mécanographie les a rendus possibles dès le début du XX^{e} siècle.

L'indice de Friedman n'est calculable qu'avec les moyens techniques du XX^{e} siècle.

L'indice de coïncidence

L'idée de Friedman est d'introduire un indice invariable par permutation des lettres. Sa valeur n'est donc pas affectée par une substitution alphabétique simple, telle celles engendrées par le code de César. Il permet ainsi de retrouver facilement la longueur d'une clef de Vigenère. L'idée la plus simple est d'additionner les carrés des fréquences des lettres. En fait, on aboutit à un meilleur résultat en modifiant légèrement cette idée. On compte l'occurrence de chaque lettre dans le message : n_A pour A, n_B pour B, …, n_Z pour Z. On en fait la somme n puis la somme des produits

Friedman ou l'amour de la cryptanalyse

Après des études de génétiques, William Friedman s'intéressa soudainement à la cryptographie quand il rencontra Elizebeth (sic) Smith. Cette jeune femme, qu'il épousa ensuite, effectuait des recherches de cryptanalyse dans l'institution où il travaillait. Il eut ensuite une carrière de cryptologue dans l'armée américaine.

William Friedman (1891-1969) et son épouse Elizebeth Smith

$n_A(n_A-1), n_B(n_B-1), \ldots, n_Z(n_Z-1)$. On divise ensuite le résultat par $n(n-1)$. Cet indice est appelé indice de coïncidence pour une raison détaillée en encadré *Les coïncidences de l'indice*. Il est relié au précédent par une formule simple n'impliquant que la longueur n du message (voir l'encadré). De plus, pour n suffisamment grand, ils sont quasiment égaux.

À partir des fréquences usuelles, on peut ainsi déterminer l'indice de coïncidence moyen d'un texte français, éventuellement codé par un chiffre tel celui de César !

En faisant la somme des carrés des fréquences des lettres données dans l'article *Les fréquences d'Al Kindi*, nous obtenons 0,0746 qui est donc l'indice de coïncidence moyen des textes écrits en français. Dans le cas d'un message où les lettres seraient choisies au hasard, nous obtenons une fréquence moyenne de 1/26. L'indice est donc égal à 26 fois $(1/26)^2$ c'est-à-dire à 1/26, soit 0,038. Il s'agit de l'indice de coïncidence moyen d'un texte aléatoire.

Cas d'un message codé

Dans le cas du message :

HONES WMBCY CMGDN EHGZM YDDTI KWQDZ XZGUE YXSRZ NUGOL QPKJK YOWWM FYAUH PTFSO DBQOU DGNCC OZEVY IFGAK YX

Les occurrences des lettres et leurs carrés sont donnés par le tableau :

LETTRE	A	B	C	D	E	F	G	H	I	J
n	2	2	4	6	4	3	6	3	2	1
$n(n-1)$	2	2	12	30	12	6	30	6	2	0
LETTRE	K	L	M	N	O	P	Q	R	S	T
n	4	1	4	4	6	2	3	1	3	2
$n(n-1)$	12	0	12	12	30	2	6	0	6	2
LETTRE	U	V	W	X	Y	Z				
n	4	1	4	3	7	5				
$n(n-1)$	12	0	12	6	42	20				

La somme des occurrences est égale à 87, celle des $n(n-1)$, 276. L'indice de coïncidence de ce texte est donc égal à 276 divisé par $87 \times 86 = 7\,482$ soit 0,037. Cet indice est très proche de ce que l'on obtient avec une distribution aléatoire des lettres. Si nous utilisons la somme des carrés des fréquences, nous obtenons 0,048 ce qui est loin de l'indice normal mais plus éloigné de l'indice d'une distribution aléatoire.

Les coïncidences et Vigenère

Reprenons le message intercepté dans l'article *Du code Vigenère à celui de Vernam* :

RAIMM RAIEI VRRBR CQLBR BEYXY TEJCI UUZLE TRZOI CDZQW GPKAI WRVLY PEWHM UAIKM XEATM VRFNZ GTRGK GNKXH CNJFE DOZMI CUOEIV TIXWN AJHMT EVTIV EDTKP IWBUW E

Pour pouvoir le décrypter par la méthode des fréquences d'Al Kindi, il

Les coïncidences de l'indice

Considérons un texte T de longueur n. Tirons deux lettres au hasard dans ce texte, quelle est la probabilité qu'elles soient identiques ?

Si le nombre de A est égal à n_A, le nombre de couples de deux A est égal à $\frac{n_A(n_A-1)}{2}$, celui de B, $\frac{n_B(n_B-1)}{2}$, etc.

En faisant la somme de tous ces nombres, on trouve le nombre de couples formés de deux lettres identiques. Le nombre de couples quelconques dans le texte est égal à $\frac{n(n-1)}{2}$ donc la probabilité pour que deux lettres d'un texte coïncident vaut :

$$I_c = \frac{n_A(n_A-1) + n_B(n_B-1) + \ldots + n_Z(n_Z-1)}{n(n-1)}.$$

Si nous notons $f_A, f_B, \ldots, f_Z$ les fréquences des lettres dans le texte, $n_A = n f_A$, etc., donc :

$$I_c = \frac{nf_A(nf_A-1) + \ldots}{n(n-1)} = \frac{n}{n-1}(f_A^2 + \ldots) - \frac{1}{n-1}(f_A + \ldots).$$

Nous en déduisons que : $I_c = \frac{n}{n-1} S - \frac{1}{n-1}$ où S est la somme des carrés des fréquences des lettres. Dans les applications, nous pouvons donc employer S au lieu de l'indice de coïncidences.

suffit de trouver la longueur de la clef utilisée. Elle n'est pas égale à 1 car l'indice de coïncidence du texte est de 0,0434. Pour tester si la clef est de longueur 2, prenons le texte obtenu en ne retenant que les lettres de deux en deux, c'est-à-dire :

RIMAE VRRQB BYYEC UZERO CZWPA WVYEH UIMET VFZTG GKHNF DZIUE VIWAH TVIET PWUE

Son indice de coïncidence est égal à 0,0461 ce qui exclut que la clef soit de longueur 2. Nous continuons ainsi en testant toutes les longueurs possibles. Pour 3, nous trouvons 0,0421, pour 4, 0,0383 et pour 5, 0,0677 ce qui rend hautement probable que la clef soit de longueur égale à 5. La méthode des fréquences permet alors de retrouver le message comme vu dans l'article sur le code de Vigenère.

Décryptage d'Enigma

Le message codé donné ci-dessus l'a été au moyen d'une machine Enigma à trois rotors (voir l'article *Les rouages d'Enigma*). Pour le décoder, il suffit de trouver dans quelle position se trouvent les rotors et le tableau de connections. Pour cela, on pourrait essayer toutes les possibilités au moyen d'un ordinateur et ne retenir que celles fournissant des fréquences compatibles avec la langue utilisée. En théorie, cela peut fonctionner. En pratique, le nombre de possibilités est trop grand du fait du tableau de connections. La faille d'Enigma se situe pourtant là. Ce tableau ne permute que douze lettres deux à deux. Autrement dit, quatorze sont inchangées ! Si nous trouvons la bonne disposition des rotors, un certain nombre de lettres sont à leur place : toutes celles qui n'ont pas rencontré une des lettres modifiées du tableau de connections. Si le message est suffisamment long, cela se retrouve dans les calculs statistiques le concernant.

Position des rotors

Nous utilisons six ordinateurs, chacun correspondant à un ordre des trois rotors. Chacun génère les positions des rotors l'une après l'autre. On code le message avec cette position sans utiliser le tableau de connections. Dans chacun de ces cas, on calcule l'indice de coïncidence. Par exemple, pour l'ordinateur simulant une Enigma avec les rotors dans l'ordre I, II, III, chacun étant sur la lettre A, on trouve :

IIOWB AOIZG WNHFO NWUFY EOFWX TOSNH BWQIZ FGJFV FENLG LVMLB IFDTI JQVIU WISEF ONIAE MXCYB TYXWE SWVLI JK

On calcule alors l'indice de coïncidence. On trouve 0,047. On fait de même avec les 17 576 positions des rotors possibles et on garde le record. On compare les records des six simulateurs, ce qui fournit l'indice 0,0509 pour les rotors dans l'ordre III, I, II réglés aux lettres C, Z et D. Le nouveau texte est :

CJSSL ZENPX NECTE FIMKE TLNZE NTNEQ IENTD EROBQ UQCWA EEREN IMMLR TCBKY FIHTP CJCIQ NFLWB IWSSE CDAEH DE

Tableau de connections

Si nous avons bien trouvé la bonne position des rotors dans le message qui précède, un certain nombre de lettres sont bien placées. Pour disposer le premier câblage, nous avons 325 possibilités que nous essayons toutes. Nous trouvons que la connexion GZ optimise l'indice de coïncidence avec le texte :

CJSSL GENPX NERTE FISDE TLNGE NTNEV ITETD EROBU UQEWA EEREN IMMLR TCBKY FIHTP CJCIQ NSLWB ILSSE CDEEH DE

dont l'indice est égal à 0,063. En recommençant, nous trouvons la connexion HO, le texte :

MESSL GENPX NERTE FCSDE TLNGE NTNEV ITETD ERHBU UQECA EEREN IGMLR TCBKP FIOTF CJCGQ NSLWB RLSSE CDEEO DE

et un indice de coïncidence égal à 0,068.

Nous changeons maintenant de tactique car la plupart des lettres importantes sont en place. Le début du texte est assez clair : MESSLGE signifie probablement « message ». Ceci invite à essayer la connexion AL. On obtient alors effectivement :

MESSA GENPX NERTE FCSDE TANGE NTNEV ITETD ERHBU UQECA EEREN IGMAR TCDKP FIOTF CJCGQ NSAWB RASSE CDERO DE

En continuant ainsi, nous trouvons successivement les connexions CR, BQ puis FU. Cher lecteur, vous avez donc tous les éléments pour décoder ce message.

H. L.

Indice de coïncidence dans diverses langues

Remarquez que le calcul se fait de la même façon quel que soit l'alphabet utilisé.

Allemand	0,0762
Anglais	0,0667
Arabe	0,0758
Espagnol	0,0770
Français	0,0746
Grec	0,0691
Hébreu	0,0768
Hollandais	0,0798
Italien	0,0738
Portugais	0,0745
Russe	0,0529

par Hervé Lehning

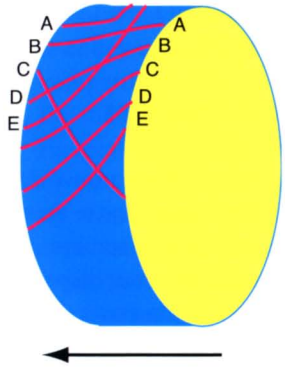
Les rouages d'Enigma

Pendant la Seconde Guerre mondiale, les Allemands disposaient d'une redoutable machine à chiffrer connue sous le nom d'Enigma. Cette machine était composée d'un brouilleur et d'un tableau de connexions. Voyons comment elle fonctionnait.

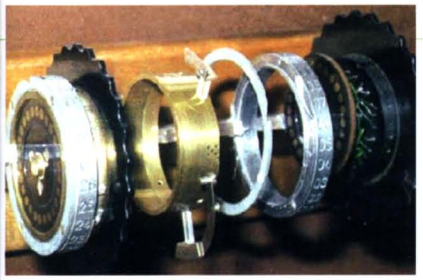
La machine Enigma utilisée par les armées allemandes lors de la Seconde Guerre mondiale permet un chiffrement automatisé suivant un code de Vigenère (voir l'article *Du code de Vigenère à celui de Vernam*). Chaque fois que l'on frappe l'une des vingt-six lettres de son clavier, la lettre chiffrée correspondante s'éclaire sur un tableau lumineux. Pour ce faire, Enigma utilise essentiellement deux composants : le brouilleur composé de plusieurs rotors associés à un réflecteur et le tableau de connexions.

Les rotors

À lui seul, un rotor permet de réaliser une substitution alphabétique, c'est-à-dire une permutation des vingt-six lettres de l'alphabet au moyen de câbles électriques. La machine Enigma classique en dispose de trois notés I, II et III. Pour l'améliorer, les machines de la marine de guerre allemande en auront quatre puis cinq. Les principes restant les mêmes, dans cet article nous décrivons l'Enigma à trois rotors.

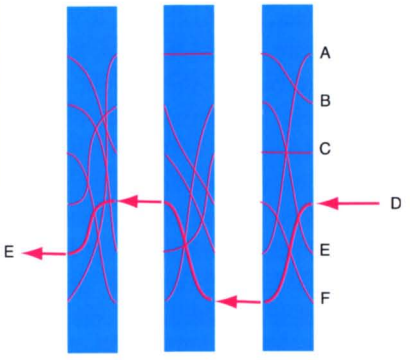


Chaque rotor d'Enigma possède vingt-six crans représentant les lettres de A à Z. Des connections électriques, notées en rouge sur la figure, joignent les crans de droite à gauche, ce qui permet d'échanger les lettres de l'alphabet. Par exemple, ci-dessus, le rotor III est représenté partiellement : A devient B et B, D, *etc.*



Un rotor d'Enigma, entier à gauche et démonté à droite. Les fils réalisant les connections sont visibles.

Pour simplifier, nous figurons chaque rotor sous forme plane avec seulement six crans (de A à F). Plusieurs rotors sont disposés l'un contre l'autre, de gauche à droite, ce qui permet de composer les permutations comme le montre la figure suivante :



La suite de rotors ci-dessus transforme D en E.

Si elle restait fixe, cette disposition ne permettrait qu'une seule permutation par texte. Pour les rotors de la figure ci-dessus, cette permutation correspond au tableau :

A	B	C	D	E	F
A	F	B	E	C	D

ce qui signifie que A est changé en A, B en F, C en B, D en E, E en C et F en D. Un tel codage est très facile à décrypter par la méthode des fréquences d'Al Kindi (voir l'article *Les fréquences d'Al Kindi*). En fait, si les rotors se présentent sous la forme de petits cylindres, c'est que chacun peut tourner autour de son axe. Quand on commence à chiffrer un message, les rotors sont choisis parmi un lot donné (I, II et III, voir l'encadré *Les connexions du brouilleur*) puis placés dans un certain ordre et une certaine position. Ces données forment une partie de la clef. On frappe alors la première lettre. Le premier rotor tourne d'un cran, la permutation des lettres est ainsi modifiée. Dans notre exemple, elle devient :

A	B	C	D	E	F
E	C	A	F	B	D.

Quand on frappe la deuxième lettre, il tourne à nouveau. Quand il a tourné de vingt-six crans, le deuxième rotor tourne d'un cran et ainsi de suite.

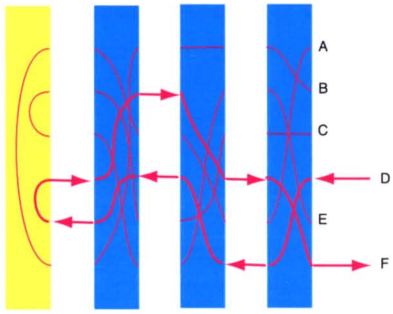
Des calculs de combinaisons

Pour enficher la première fiche du tableau de connexions, nous devons choisir une première lettre parmi 26 et une seconde parmi 25. Cela fait *a priori* 26×25 possibilités mais, comme chacune se retrouve deux fois, on obtient 325 façons de placer la première fiche. De même, nous obtenons 276 façons de placer la deuxième fiche, 231, 190, 153 et 120 façons de placer les suivantes. En procédant ainsi, nous obtenons un nombre de montages égal au produit de ces six nombres. Cependant, l'ordre des fiches n'intervenant pas, chaque montage se retrouve autant de fois qu'il y a de façons de classer six nombres. Pour cela, on choisit le premier parmi 6, le second parmi 5, *etc.* Cela fait donc $6 \times 5 \times 4 \times 3 \times 5 = 720$ fois. En tout, nous obtenons donc 100 391 791 500 façons de monter le tableau de connexions.

La puissance d'Enigma réside dans le nombre de clefs possibles.

Le réflecteur

À l'extrémité des rotors, se trouve un dispositif permettant de réfléchir le signal vers l'entrée du premier rotor :

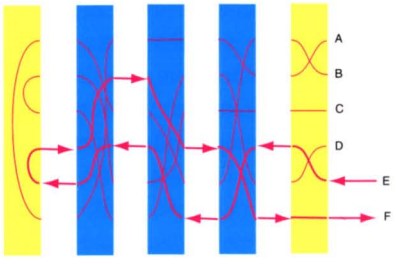


Le brouilleur d'une Enigma : le réflecteur permet de réfléchir le signal. Le décodage se fait ainsi avec la même machine dans la même disposition.

Il correspond à une nouvelle permutation donnée dans l'encadré *Les connexions du brouilleur*. L'ensemble constitué par les rotors et le réflecteur est appelé le brouilleur.

Le tableau de connections

Pour compliquer un peu le système de chiffrement d'Enigma, un tableau de connexions est placé au début du dispositif précédent. Au moyen de six fiches, il permet d'échanger douze lettres. Voici donc le dispositif complet :



Le fonctionnement d'Enigma : si on frappe E, F s'allume et *vice-versa*.

Chiffrement avec Enigma

Pour coder un message, il faut partager une clef secrète avec son destinataire afin de configurer l'Enigma de décodage comme celle de codage. Dans la pratique, l'armée allemande disposait d'une clef par jour. Ce système de clefs devait être tenu secret sinon tous les messages pouvaient être décodés avec n'importe quelle Enigma. Cette clef comporte :

1 la position des six fiches du tableau de connections,
2 le choix et l'ordre des rotors,
3 la position initiale de chaque rotor, cette position étant repérée par une lettre de l'alphabet.

Le nombre de clefs possibles est facile à calculer. Nous obtenons 100 391 791 500 façons de monter le tableau de connexions (voir l'encadré *Des calculs de combinaisons*). Si on dispose de trois rotors seulement, on obtient six façons de les ordonner. Enfin, chacun a vingt-six façons d'être disposé. En tout, les rotors fournissent 6×26^3 soit en tout 105 456 possibilités. Finalement, nous obtenons 10 586 916 764 424 000 soit plus de 10^{16} possibilités. Ce calcul montre l'utilité du tableau de connexions. Sans lui, le nombre de clefs possibles serait ridiculement bas. La difficulté de décrypter Enigma tient au tableau de connexions, pas au brouilleur ! Dans l'article *Les coïncidences de Friedman*, nous voyons que sa faiblesse se situe également là, car le tableau de connexions n'opère pas sur toutes les lettres.

Exemple de chiffrement

En n'utilisant que le brouilleur, c'est-à-dire sans utiliser le tableau de connexions, en positionnant les rotors de gauche à droite : I, II, III chacun sur la lettre A, le message :
TANGENTE C'EST L'AVENTURE MATHEMATIQUE
devient :
ODOAR EHYZC JEETL NMIBV YZQWZ ZLHPS HLR.

Les connexions du brouilleur

Par commodité, nous représentons les rotors et le réflecteur à plat, la partie droite en haut. Les connexions sont alors données par les tableaux suivants :

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

Rotor I

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

Rotor II

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

Rotor III

Dans chacun de ces tableaux, les crans notés de façon identique sont connectés ensemble. Par exemple, si le rotor I est positionné sur la lettre A, la lettre A est transformée en E, B en K, *etc.* Si nous positionnons le rotor sur la lettre B, nous devons tout décaler d'un cran mais les connexions restent les mêmes puisqu'il s'agit d'un câblage physique. La lettre A est à l'ancienne place de B qui était transformée en K ce qui donne J par effet de la rotation effectuée. De même, B est transformée en L, *etc.*

De façon plus simple, le réflecteur correspond à la permutation fixe suivante :

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

Réflecteur

Ces connexions étaient connues des Britanniques grâce aux études d'un mathématicien polonais Marian Rejewski puis la capture de deux sous marins allemands, le U-110 le 9 mai 1941 et le U-559 le 30 octobre 1942. Ces captures furent des exploits britanniques même si Hollywood les attribue implicitement aux Américains dans le film U-571. Quant à lui, le U-571 réussit l'exploit de survivre jusqu'à la fin de la guerre.

Tous les signes de ponctuation sont éliminés ou remplacés par des groupes de lettres peu courantes comme XX ou XY et les lettres sont groupées par cinq pour ne pas permettre un décryptage utilisant les mots probables d'une, deux ou trois lettres. Tout ceci est ensuite transmis en morse via les ondes radio, bien sûr. Voyons comment s'effectue le codage de la première lettre. Nous frappons T, le rotor III tourne d'un cran. Ainsi, T se transforme en J qui, avec le rotor II devient B puis avec le rotor I, K. Le réflecteur transforme K en N. Le rotor I opère maintenant à l'envers, N devient K, K devient D avec le II et D, O avec III. À la lettre suivante, le codage change puisque le rotor III tourne à nouveau d'un cran.

Une Enigma à trois rotors.

H. L.

par Hervé Lehning

Les mots probables de Turing

Les moyens de calcul à l'époque de Turing ne permettaient pas l'utilisation de l'indice de coïncidence pour décrypter Enigma. Pour cela, il commença par deviner la présence de certaines phrases dans le texte.

Pour casser les messages d'Enigma, Turing utilisa une méthode fondée sur les mots probables d'un texte. Cela peut être « divisions », « quartier général » ou tout simplement « bulletin météo » voire ce bulletin même (voir l'encadré *L'erreur de procédure des messages météo*). Tout d'abord, il les localise par une méthode que nous voyons plus loin. Il dispose alors d'un texte et de son chiffrement, par exemple :
DUHAU TQUAR TIERG ENERA L et
ANOTC FLNLF GOPEN SUTQP M.
Il utilise alors ce renseignement pour trouver la disposition des rotors et du tableau de connexions de l'Enigma valable pour le jour et le réseau donné puisque ces données changent chaque jour, à minuit (voir *Les rouages d'Enigma*). Dans ce qui suit, nous utilisons une Enigma à trois rotors, les principes restent valables pour les Enigma à quatre ou cinq rotors.

Utilisation des paires

Dressons le tableau de correspondance des lettres dans notre premier exemple (voir tableau ci-dessous).

Certaines lettres se correspondent plusieurs fois, ici U et N, deux fois dans un sens et une fois dans l'autre, ce qui revient au même car Enigma code de façon symétrique. Comme toutes les lettres ne sont pas échangées par le tableau de connexions, il est possible de supposer que U et N ne le sont pas. Nous configurons six ordinateurs, un

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
D	U	H	A	U	T	Q	U	A	R	T	I	E	R	G	E	N	E	R	A	L
A	N	O	T	C	F	L	N	L	F	G	O	P	E	N	S	U	T	Q	P	M

Correspondance des lettres à chaque étape du chiffrement. Les paires ont été notées en jaune.

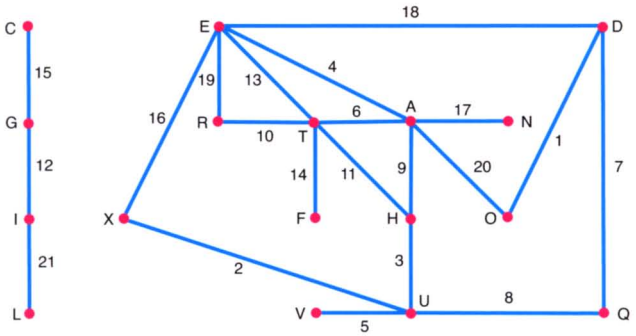
Les stéréotypes des codeurs sont les alliés des décodeurs.

par ordre des rotors comme dans l'article *Les coïncidences de Friedman*. Pour chacun, on essaye toutes les configurations des rotors pour ne retenir que celles échangeant U et N en deuxième, huitième et dix-septième positions. Nous trouvons ainsi que les rotors sont en position III, I, II et réglés sur les lettres A, D et Y. Bien sûr, cela ne marche pas toujours. Premièrement, il est possible que l'on ne trouve aucune paire. Il est possible aussi qu'elles donnent un grand nombre de possibilités de disposition des rotors.
Cependant, si on trouve plusieurs paires dans plusieurs messages, il est probable que l'on trouve ainsi la disposition des rotors.

Utilisation des cycles

Prenons un autre cas :
DUHAU TQUAR TIERG ENERA L
est codé en OXUEV ADQHT HGTFC XADEO I.
Nous obtenons ainsi la correspondance (voir tableau ci-dessous).

Aucune paire ici, ce tableau ne révèle pas grand-chose sur la correspondance trouvée, il est beaucoup plus intéressant de la représenter sous la forme d'un graphe en reliant les lettres par des arcs que l'on marque par le numéro de l'étape du chiffrement :



Graphe représentant les correspondances des lettres entre le message et son code.

Ce graphe est composé de deux parties. Celle de gauche n'a guère d'intérêt. En revanche, celle de droite avec ses nombreux cycles est pleine d'enseignements. Imaginons pour commencer que A soit inchangé. On essaye toutes les dispositions des rotors telle que A soit changé en une lettre en sixième position, elle-même changée en une autre lettre en onzième qui est changée en A en neuvième. Ici, nous ne trouvons aucune solution. Nous recommençons en échangeant A avec les autres lettres de l'alphabet. En tenant compte des autres cycles, nous trouvons que les rotors sont en position III, I, II et réglés sur B, D et Q. La technique donne simultanément le tableau de connexions. Ici par exemple, A est échangé avec V, E avec T, H avec Y, J avec F, L avec P et D avec W.

Position du mot

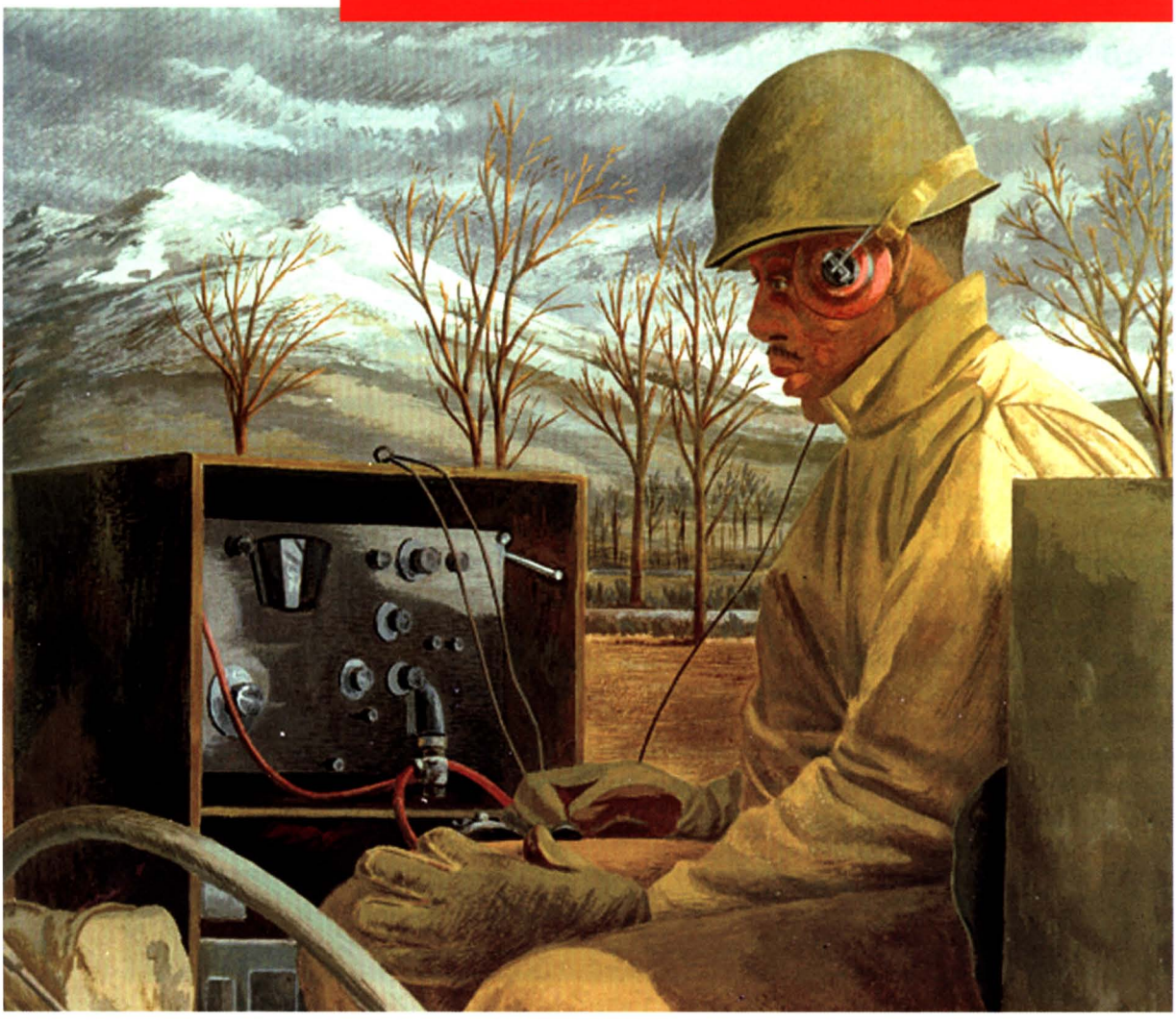
Pour utiliser ce qui précède, il faut encore trouver la position du mot. Une particularité d'Enigma rend la tâche

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
D	U	H	A	U	T	Q	U	A	R	T	I	E	R	G	E	N	E	R	A	L
O	X	U	E	V	A	D	Q	H	T	H	G	T	F	C	X	A	D	E	O	I

Correspondance des lettres à chaque étape du chiffrement.

L'erreur de procédure des messages météo

La météo n'étant pas vraiment un secret militaire, la marine allemande envoyait à son quartier général les messages météo cryptés mais avec une machine Enigma à trois rotors. Chaque jour, cette Enigma était réglée comme celles à quatre rotors, mais avec le quatrième rotor en moins !
Les Britanniques savaient déchiffrer ces messages de manière aisée. Ils avaient donc la position des trois premiers rotors, il restait juste 26 possibilités pour le dernier ! De plus, les Allemands offraient ainsi des mots probables gigantesques à Turing et son équipe ! Les Allemands ne s'aperçurent jamais de cette erreur de procédure monumentale ! Pourtant, ils se doutaient que les Britanniques savaient déchiffrer l'Enigma à trois rotors puisqu'ils étaient passés à quatre.



possible : à cause du réflecteur, aucune lettre n'est codée par elle-même. On fait donc défiler les mots probables comme « panzer » le long du texte et on cherche un cas sans coïncidence. Il est possible alors que nous ayons trouvé une occurrence de notre mot. Bien entendu, cela fonctionne mieux avec des mots longs que des mots courts. L'armée allemande en fournissait beaucoup, en particulier dans les messages météo (voir l'encadré sur les messages météo). Par exemple, considérons le message : TJNWU TBIMQ WMYIC CGNDD DUIXW MXNZN NRTP. Nous y cherchons le mot probable : BULLETIN METEO NORD ATLANTIQUE. Pour cela, nous les mettons en correspondance :

Turing et les bombes

Pour utiliser la méthode du mot probable, on doit pouvoir faire fonctionner un grand nombre de machines Enigma ensemble et automatiquement. Ceci était réalisé par des machines reproduisant un grand nombre d'Enigma travaillant ensemble. Elles portaient le nom de bombes, deux cent furent construites pendant la guerre. Pour les faire fonctionner, il fallait les câbler selon les paires ou les cycles trouvés. Ensuite, elles ne donnaient pas directement la position des rotors utilisée mais des positions impossibles. Il fallait ensuite en déduire manuellement la position correcte.

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

Les trois coïncidences de deux lettres en jaune rendent cette position impossible. Nous continuons ainsi en décalant le mot probable :

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

Une coïncidence rend à nouveau cette position impossible. Nous continuons ainsi jusqu'à obtenir :

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

Aucune coïncidence cette fois-ci. Nous pouvons donc essayer nos méthodes sur la correspondance :

T	B	I	M	Q	W	M	Y	I	C	C	G	N	D	D	D	U	I	X	W	M	X	N	Z	N	N	R
B	U	L	L	E	T	I	N	M	E	T	E	O	N	O	R	D	A	T	L	A	N	T	I	Q	U	E

Nous y trouvons une paire ce qui permet d'appliquer la méthode du premier paragraphe.

H. L.

L'importance d'Enigma dans la guerre

Le décryptage d'Enigma par Alan Turing et son équipe a influé considérablement sur le déroulement de la guerre, en particulier dans la bataille sous marine de l'Atlantique nord. La tactique des sous-marins était simple. Ils quadrillaient l'océan. Le premier qui repérait un convoi prévenait les autres et ils prévoyaient une attaque coordonnée. Cette méthode n'était véritablement efficace que si le secret était gardé. C'est pourquoi la marine de guerre allemande fut équipée d'Enigma possédant quatre puis cinq rotors et un tableau de connections de dix fiches à partir de février 1942. Elle était donc bien supérieure à l'Enigma classique à trois rotors et six fiches (voir l'article *Les rouages d'Enigma*).
Pendant un an, l'équipe de Turing fut incapable de déchiffrer cette nouvelle Enigma. Les conséquences se voient dans les statistiques de 1941 à 1943 concernant le tonnage coulé par les sous marins allemands ainsi que le nombre de sous marins détruits :

	jan.	fév.	mars	avril	mai	juin	juil.	août	sept.	oct.	nov.	déc.
Tonnage coulé*	21	39	41	43	58	61	22	23	53	32	13	26
Sous-marins coulés	0	0	5	2	1	4	4	2	2	2	5	10
1942	jan.	fév.	mars	avril	mai	juin	juil.	août	sept.	oct.	nov.	déc.
Tonnage coulé*	62	85	95	74	125	144	96	108	98	94	119	60
Sous-marins coulés	3	3	6	3	4	3	11	10	10	16	12	6
1943	jan.	fév.	mars	avril	mai	juin	juil.	août	sept.	oct.	nov.	déc.
Tonnage coulé*	37	63	108	58	50	20	46	16	20	20	14	13
Sous-marins coulés	7	18	15	16	41	17	38	24	11	28	18	10

* En milliers

De février 1942 à avril 1943, les britanniques furent incapables de décrypter les messages entre les sous marins. Les tonnages coulés furent importants. Ensuite, ils diminuent tandis que le nombre de sous marins coulés augmente. Bien sûr, le décryptage d'Enigma ne fut pas le seul facteur. L'introduction du sonar en fut un autre mais l'influence du déchiffrement d'Enigma reste primordiale.

Le génie britannique et l'amateurisme d'UNIX

Alan Turing fut un génie. Grâce à lui, les secrets allemands de la Seconde Guerre Mondiale furent souvent dévoilés en temps utile. Le génie britannique fut de tenir cette découverte secrète longtemps après la guerre. Au contraire, ils firent croire jusqu'en 1973 que la machine était indéchiffrable. Ils purent ainsi revendre les machines Enigma prises aux Allemands en 1945 à des gouvernements et des entreprises étrangères. L'espionnage britannique ultérieur fut ainsi subventionné par les espionnés eux-mêmes !

De façon similaire, mais sans doute involontaire, l'algorithme de chiffrement d'Enigma a été intégré dans la distribution du système d'exploitation UNIX (commande crypt). De nombreux laboratoires civils et militaires l'ont ainsi utilisé pour protéger leurs communications. Les espions intéressés se sont gardés de le signaler.

L'ère informatique

L'avènement de l'informatique est à l'origine d'un nouveau visage de la cryptographie. Il est issu de la lutte sans fin entre, d'un côté, les concepteurs d'ordinateurs ou de systèmes informatisés, et de l'autre, les pirates, espions et autres usurpateurs. Les mathématiques sont omniprésentes dans le codage informatique. En particulier, de nombreux systèmes de codage se basent sur la difficulté de factoriser les grands nombres.

par Abdelkader Necer

Le code DES

Le Data Encryption System (DES) fut adopté comme standard par l'administration américaine avant d'être remplacé par l'Advanced Encryption System (AES) en 2002, quand des attaques ont montré sa vulnérabilité.

En mai 1973, le National Bureau of Standards (NBS) américain lança un appel d'offres pour un système de chiffrement. Comme IBM disposait déjà d'un algorithme de chiffrement symétrique nommé Lucifer (voir l'encadré *Lucifer et Horst Feistel*), elle le proposa au concours. Il fut retenu et modifié pour devenir le code DES. Au passage, sa clef passa de 122 bits à 56 bits. Il fut à nouveau modifié plusieurs fois avant d'être abandonné car trop vulnérable.

Description générale

Comme toujours en informatique, les messages sont des suites de bits, c'est-à-dire de 0 et de 1. Pour être codés, ils sont tronçonnés en mots de 64 bits. Le code DES opère ensuite sur ces mots de 64 bits de la façon suivante.
Premièrement, le mot subit une permutation initiale (voir l'encadré *Les permutations de DES*).

Deuxièmement, le mot obtenu M est alors scindé en sa partie gauche G_0 et sa partie droite D_0 ($M = G_0\, D_0$), chacune de 32 bits, pour être transformé en $M_1 = G_1\, D_1$ où :

$G_1 = D_0$ et $D_1 = G_0 + f(D_0, K_0)$.

K_0 désigne une partie de la clef utilisée à ce niveau (clef intermédiaire de 48 bits, voir plus loin), f, une fonction décrite ci-dessous et +, l'addition bit à bit (voir l'article *Du code Vigenère à celui de Vernam*).

Troisièmement, on itère ce procédé pour obtenir $M_2 = G_2\, D_2$, à partir de $M_1 = G_1\, D_1$ et ainsi de suite seize fois jusqu'à obtenir $M_{16} = G_{16}\, D_{16}$.

Quatrièmement, à la fin, on inverse les moitiés gauche et droite et on applique la permutation inverse de la permutation initiale.

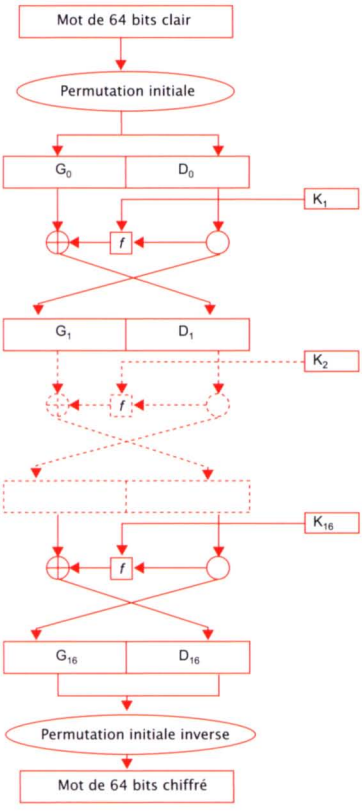


Schéma général du code DES

La fonction *f*

La fonction f opère sur deux arguments. L'argument de gauche, R, possède 32 bits ; il est expansé en un mot de 48 bits en plaçant le trente-deuxième bit en première position, le premier en seconde, et ainsi de suite selon la table :

32	1	2	3	4	5
4	5	6	7	8	9
8	9	10	11	12	13
12	13	14	15	16	17
16	17	18	19	20	21
20	21	22	23	24	25
24	25	26	27	28	29
28	29	30	31	32	1

On additionne ensuite ce mot avec l'argument de droite désigné par K (addition bit à bit). On subdivise le mot obtenu en huit mots de 6 bits. Chacun est transformé en un mot de 4 bits en utilisant les boîtes de substitutions décrites ci-dessous. Les mots ainsi obtenus sont « concaténés » pour former un mot de 32 bits auquel on applique une permutation en plaçant le seizième bit en première position, le septième en seconde et ainsi de suite selon la table :

15	7	20	21
29	12	28	17
1	15	23	26
5	18	31	10
2	8	24	14
32	27	3	9
19	13	30	6
22	11	4	25

On dispose de huit boîtes de substitution différentes. En entrée, chacune prend un mot de 6 bits pour fournir en sortie un mot de 4 bits. Les boîtes sont représentées par des tableaux à deux lignes et seize colonnes. Voici la première en guise d'exemple :

S_1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	14	4	13	1	2	15	11	8	3	10	6	12	5	9	0	7
1	0	15	7	4	14	2	13	1	10	6	12	11	9	5	3	8
2	4	1	14	8	13	6	2	11	15	12	9	7	3	10	5	0
3	15	12	8	2	4	9	1	7	5	11	3	14	10	0	6	13

Pour les autres, voir l'encadré *Les boîtes de substitution*. Les premiers et derniers bits du mot en entrée déterminent une ligne du tableau, les autres bits déterminent une colonne. La valeur numérique (écrite en base 2) trouvée à cet endroit (au croisement de la ligne et de la colonne) indique la valeur des quatre bits de sortie. Par exemple, si le mot à l'entrée de la première boîte est 010101. Le premier bit et le dernier bit forment le mot 01 qui indique le numéro de la ligne de S_1 soit ici la ligne 1. Les 4 bits restants, soit 1010, fournissent le numéro de la colonne soit ici 10 (en base 2 : $10 = 1 \times 2^3 + 0 \times 2^2 + 1 \times 2^1 + 0 \times 2^0$). À l'intersection de la ligne 1 et la colonne 10, on trouve le nombre 12 dont la représentation binaire est 1100. En sortie, on aura donc 1100.

Lucifer et Horst Feistel

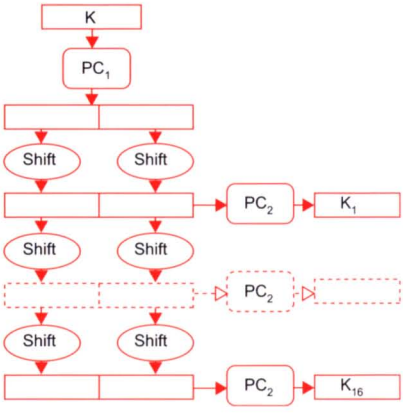
À titre anecdotique, l'appellation *Lucifer* proviendrait de *Demon* obtenu en tronquant *Demonstration*, le nom du système sur lequel travaillait Feistel. Les systèmes d'exploitation de l'époque n'acceptaient pas des noms aussi longs, d'où *Demon* qui fut transformé en *Lucifer*.

Horst Feistel

Horst Feistel (1915-1990) fut l'un des premiers cryptographes universitaires. Il émigre aux États-Unis en 1934 et est incorporé au sein du centre US Air Force Cambridge Research Center dans la section des appareils « identification des amis ou ennemis » en 1944. Il travaille dans ce domaine jusqu'en 1950 et rejoint plus tard le MIT et IBM où il reçoit une récompense pour ses travaux en cryptologie.

La diversification de la clef

La clef secrète K est une clef de 56 bits plus 8 bits de parité ou de contrôle. Ces bits de contrôle occupent les positions dont les numéros sont des multiples de 8 (8, 16, 24, etc.). Ainsi, si les sept premiers bits de la clef secrète sont 0011010, le huitième bit sera 1 pour qu'il ait un nombre pair de 1. Le principe de diversification de la clef est schématisé par la figure ci-dessous :



Diversification de la clef

Les permutations de DES

La permutation initiale consiste à placer le cinquante-huitième bit en première position, le cinquantième, en seconde et ainsi de suite selon la table :

58	50	42	34	26	18	10	2
60	52	44	36	28	20	12	4
62	54	46	38	30	22	14	6
64	56	48	40	32	24	16	8
57	49	41	33	25	17	9	1
59	51	43	35	27	19	11	3
61	53	45	37	29	21	13	5
63	55	47	39	31	23	15	7

La permutation initiale

La permutation inverse est alors donnée par le tableau :

40	8	48	16	56	24	64	32
39	7	47	15	55	23	63	31
38	6	46	14	54	22	62	30
37	5	45	13	53	21	61	29
36	4	44	12	52	20	60	28
35	3	43	11	51	19	59	27
34	2	42	10	50	18	58	26
33	1	41	9	49	17	57	25

L'inverse de la permutation initiale

Ces permutations n'ont aucun rôle dans la sécurité de l'algorithme.

La permutation PC_1 est quant à elle définie par le tableau ci-dessous, où les 56 bits sont numérotés de 1 à 64 en évitant les multiples de 8 :

57	49	41	33	25	17	9
1	58	50	42	34	26	18
10	2	59	51	43	35	27
19	11	3	60	52	44	36
63	55	47	39	31	23	15
7	62	54	46	38	30	22
14	6	61	53	45	37	29
21	13	5	28	20	12	4

La permutation PC_1

De même, la règle d'extraction PC_2 est définie par :

14	17	11	24	1	5
3	28	15	6	21	10
23	19	12	4	26	8
16	7	27	20	13	2
41	52	31	37	47	55
30	40	51	45	33	48
44	49	39	56	34	53
46	42	50	36	29	32

La règle d'extraction PC_2

À la fin des années 1990, DES n'était plus considéré comme un algorithme assez sûr.

On applique une permutation PC_1 à la clef K (voir l'encadré *Les permutations de DES*). Puis à chacune des seize étapes du schéma général de l'algorithme, à chaque moitié du mot de 56 bits obtenu on fait subir un décalage à gauche d'un cran aux étapes 1, 2, 9, 16 et deux crans aux autres étapes. À chacune des ces étapes, on obtient une clef partielle de 48 bits en appliquant la règle PC_2.

Les boîtes de substitution

S_1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	14	4	13	1	2	15	11	8	3	10	6	12	5	9	0	7
1	0	15	7	4	14	2	13	1	10	6	12	11	9	5	3	8
2	4	1	14	8	13	6	2	11	15	12	9	7	3	10	5	0
3	15	12	8	2	4	9	1	7	5	11	3	14	10	0	6	13

S_2	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	15	1	8	14	6	11	3	4	9	7	2	13	12	0	5	10
1	3	13	4	7	15	2	8	14	12	0	1	10	6	9	11	5
2	0	14	7	11	10	4	13	1	5	8	12	6	9	3	2	15
3	13	8	10	1	3	15	4	2	11	6	7	12	0	5	14	9

S_3	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	10	0	9	14	6	3	15	5	1	13	12	7	11	4	2	8
1	13	7	0	9	3	4	6	10	2	8	5	14	12	11	15	1
2	13	6	4	9	8	15	3	0	11	1	2	12	5	10	14	7
3	1	10	13	0	6	9	8	7	4	15	14	3	11	5	2	12

S_4	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	7	13	14	3	0	6	9	10	1	2	8	5	11	12	4	15
1	13	8	11	5	6	15	0	3	4	7	2	12	1	10	14	9
2	10	6	9	0	12	11	7	13	15	1	3	14	5	2	8	4
3	3	15	0	6	10	1	13	8	9	4	5	11	12	7	2	14

S_5	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	2	12	4	1	7	10	11	6	8	5	3	15	13	0	14	9
1	14	11	2	12	4	7	13	1	5	0	15	10	3	9	8	6
2	4	2	1	11	10	13	7	8	15	9	12	5	6	3	0	14
3	11	8	12	7	1	14	2	13	6	15	0	9	10	4	5	3

S_6	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	12	1	10	15	9	2	6	8	0	13	3	4	14	7	5	11
1	10	15	4	2	7	12	9	5	6	1	13	14	0	11	3	8
2	9	14	15	5	2	8	12	3	7	0	4	10	1	13	11	6
3	4	3	2	12	9	5	15	10	11	14	1	7	6	0	8	13

S_7	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	4	11	2	14	15	0	8	13	3	12	9	7	5	10	6	1
1	13	0	11	7	4	9	1	10	14	3	5	12	2	15	8	6
2	1	4	11	13	12	3	7	14	10	15	6	8	0	5	9	2
3	6	11	13	8	1	4	10	7	9	5	0	15	14	2	3	12

S_8	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	13	2	8	4	6	15	11	1	10	9	3	14	5	0	12	7
1	1	15	13	8	10	3	7	4	12	5	6	11	0	14	9	2
2	7	11	4	1	9	12	14	2	0	6	10	13	15	3	5	8
3	2	1	14	7	4	10	8	13	15	12	9	0	3	5	6	11

Contre les attaques : le triple DES

En janvier 1977, les laboratoires RSA ont lancé un défi consistant à décrypter un message chiffré par DES par recherche exhaustive de la clef. Des milliers d'ordinateurs ont travaillé de concert pour aboutir à la découverte de la bonne clef le 17 juin 1997, après avoir exploré environ un quart de l'ensemble des clefs. Depuis, on a découvert d'autres attaques plus rapides.

Pour pallier la faiblesse du DES, des chercheurs du projet DES chez IBM sous la houlette de Walter Tuchman ont développé un algorithme de chiffrement qui enchaîne trois applications successives de l'algorithme DES sur le même bloc de données de 64 bits, avec deux ou trois clefs DES différentes. La version la plus sûre utilise un chiffrement, suivi d'un déchiffrement pour se conclure à nouveau par un chiffrement. Le triple DES est généralement utilisé avec seulement deux clefs différentes. Si on désigne par k_1 et k_2 deux clefs DES, le triple DES peut s'écrire formellement :

$$\text{Triple-DES}_{k_1,k_2} = \text{DES}_{k_1} \circ \text{DES}_{k_{2-1}} \circ \text{DES}_{k_1}$$

Une clef triple DES est donc composée de deux clefs DES et fait 112 bits ce qui met le triple DES hors de portée d'une attaque exhaustive. Cependant cet algorithme, assez simple à mettre en œuvre, est lent. Il a laissé la place à AES, dont les principes restent analogues.

A. N.

Faiblesse du DES

L'algorithme de chiffrement de DES est entièrement connu, seule la clef de 56 bits est inconnue. Malgré l'énormité du nombre de possibilités (2^{56}), on peut imaginer d'essayer toutes les clefs possibles l'une après l'autre et examiner le texte obtenu à chaque fois pour voir si la clef est la bonne. On parle alors d'attaque par force brute. Si le message est un texte français traduit en code ASCII, l'indice de coïncidence et les calculs de fréquences permettent de reconnaître de façon automatique si la clef est la bonne. En moyenne, il faut 2^{55} essais pour décrypter le message puisque la découverte de la clef peut se produire à tout moment entre le premier et le dernier essai (le numéro 2^{56}). Ce nombre est énorme mais, de nos jours, un simple ordinateur de bureau peut analyser plusieurs millions de clefs par seconde. Seul, il ne pourra pas casser un message, il lui faudrait plusieurs centaines d'années mais des milliers d'ordinateurs mis en réseau peuvent le faire en quelques semaines, voire quelques jours. Une machine dédiée a été fabriquée uniquement pour montrer que l'algorithme DES n'offre plus la sécurité requise. À elle seule, cette machine de moins de 200 000 € peut décrypter un message en quatre jours. Pour cette raison, DES a été abandonné au début des années 2000. Il reste utilisé pour certaines applications, comme le cryptage de chaînes de télévision. La clef doit donc impérativement être changée régulièrement.

Hervé Lehning

par Hervé Lehning

Le code RSA

À l'heure actuelle, la méthode RSA est considérée comme la méthode de cryptage à clef publique la plus sûre. Vérité scientifique ou acte de foi ? L'inviolabilité (en pratique tout au moins) des codes RSA tient à la difficulté de factoriser les très grands nombres.

Une méthode de cryptographie à clef publique répond aux critères suivants. Premièrement, une clef de codage est divulguée publiquement de sorte que toute personne puisse coder un message M pour obtenir un message codé P(M). Deuxièmement, la clef de décodage S permettant de retrouver M à partir de P(M), c'est-à-dire telle que S[P(M)] = M, est gardée secrète. Bien entendu, cela suppose que la connaissance de P ne donne pas S trop aisément. La difficulté de la factorisation en nombres premiers fonde la méthode de cryptographie à clef publique la plus répandue de nos jours. Nommée RSA, du nom de ses inventeurs (Rivest, Shamir et Adleman, 1978), elle repose essentiellement sur les deux résultats d'arithmétique classique suivants (voir l'article *L'arithmétique de la cryptographie* et l'encadré *Preuve de RSA*) :

– Si p et q sont deux nombres premiers et k un nombre entier tels que :
$k-1$ soit divisible par $(p-1)\ (q-1)$
alors $x^k - x$ est divisible par pq.

– Les éléments inversibles de $\mathbb{Z} / m\,\mathbb{Z}$ sont les nombres premiers avec m.

La découverte de nouvelles méthodes de factorisation pourrait être tenue secrète dans le but de décoder les messages d'autrui.

Codage RSA

La méthode RSA consiste à choisir deux nombres premiers p et q puis à calculer leur produit $n = pq$ et $m = (p-1)\ (q-1)$. On choisit ensuite un nombre α inversible dans $\mathbb{Z} / m\,\mathbb{Z}$. On crypte alors tout nombre de $\mathbb{Z} / m\,\mathbb{Z}$ en $P(x) = x^{\alpha}$ (dans $\mathbb{Z} / m\,\mathbb{Z}$). Pour crypter un nombre supérieur à n, il suffit de l'écrire en base n puis de crypter chacun de ses chiffres. Pour crypter un texte, on le transforme d'abord en nombre. Nous avons ainsi défini la clef publique $P(x)$.

Par exemple, si $p = 5$ et $q = 11$ alors $n = 55$ et $m = 40$. Le nombre $\alpha = 3$ est premier avec 40 donc il convient. La clef publique P est complète. Voyons comment elle fonctionne sur un exemple simple. Si nous voulons crypter « *Tangente*, l'aventure mathématique », nous commençons par transformer cette phrase en nombre. Pour cela, nous pouvons utiliser le code ASCII (voir l'article *Le code ASCII*). Nous obtenons :

01010100011000010110111001100111
01100101011011100111010001100101
00101100001000000110110000100111
01100001011101100110010101101110
01110100011101010111001001100101
01101101011000010111010001101000
11101001011011010110000101110100
01101001011100010111010101100101

En base 10, ce nombre binaire devient :

38166425425932513455594326643974
44525031398608744883197404061685
39554882090341.

Cette étape est en fait inutile car le but est de l'écrire en base 55, d'où la suite d'éléments de $\mathbb{Z} / 55\,\mathbb{Z}$:

01 00 40 07 00 18 28 47 13 24 12 46
15 06 47 52 52 51 51 12 05 25 35 27
24 42 50 29 30 17 43 36 26 46 35 19
54 49 38 14 18 49 50 08 26

Nous cryptons maintenant chacun de ces nombres en l'élevant à la puissance 3 dans $\mathbb{Z} / 55\,\mathbb{Z}$ c'est-à-dire que pour chaque opération, on ne garde que le reste de son résultat dans la division par 55. Voici les calculs dans le cas du chiffre des unités (26) :

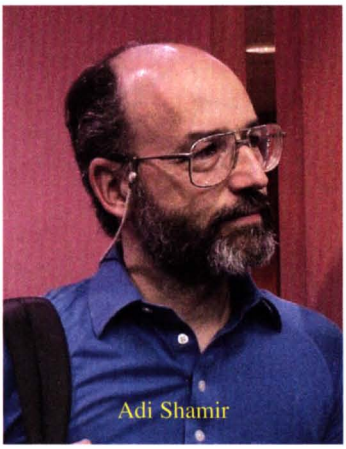
Preuve de RSA

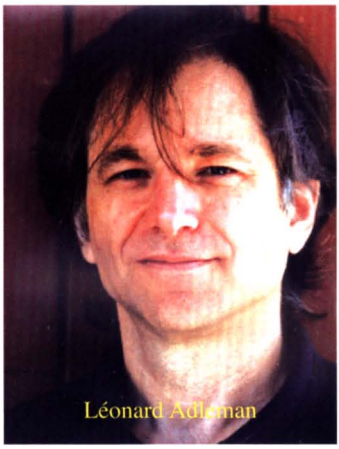
La propriété à la base de la méthode RSA dérive directement du petit théoréme de Fermat (voir l'article *L'arithmétique de la cryptographie*).
Soit k un nombre entier tel que $k - 1$ soit divisible par $p - 1$. Il existe un entier K tel que $k = 1 + K(p - 1)$. Soit x un nombre entier :

$$x^k - x = x[\,(x^{p-1})^K - 1\,].$$

D'après le petit théorème de Fermat, cette expression est nulle dans $\mathbb{Z} / p\,\mathbb{Z}$ ce qui signifie que $x^k - x$ est divisible par p.
Si $k - 1$ est divisible par $(p - 1)(q - 1)$ alors, d'après ce qui précède, pour tout x, $x^k - x$ est divisible par p et q donc par le produit pq puisque p et q sont premiers. On en déduit que, pour tout x, $x^k - x$ est divisible par pq.


Ron Rivest


Adi Shamir


Léonard Adleman

$26^2 = 26 \times 26 = 676 = 12 \times 55 + 16$ donc $26^2 = 16$ dans $\mathbb{Z} / 55\,\mathbb{Z}$,
$26^3 = 26^2 \times 26 = 16 \times 26 = 416 = 7 \times 55 + 31$ donc $26^3 = 31$ dans $\mathbb{Z} / 55\,\mathbb{Z}$.
Nous cryptons ainsi chacun des éléments de $\mathbb{Z} / 55\,\mathbb{Z}$. pour obtenir la nouvelle suite :

01 00 35 13 00 02 07 38 52 19 23 41
20 51 38 28 28 46 46 23 15 05 30 48
19 03 40 24 50 18 32 16 31 41 30 39
54 04 37 49 02 04 40 17 31

qui représente le nombre binaire :

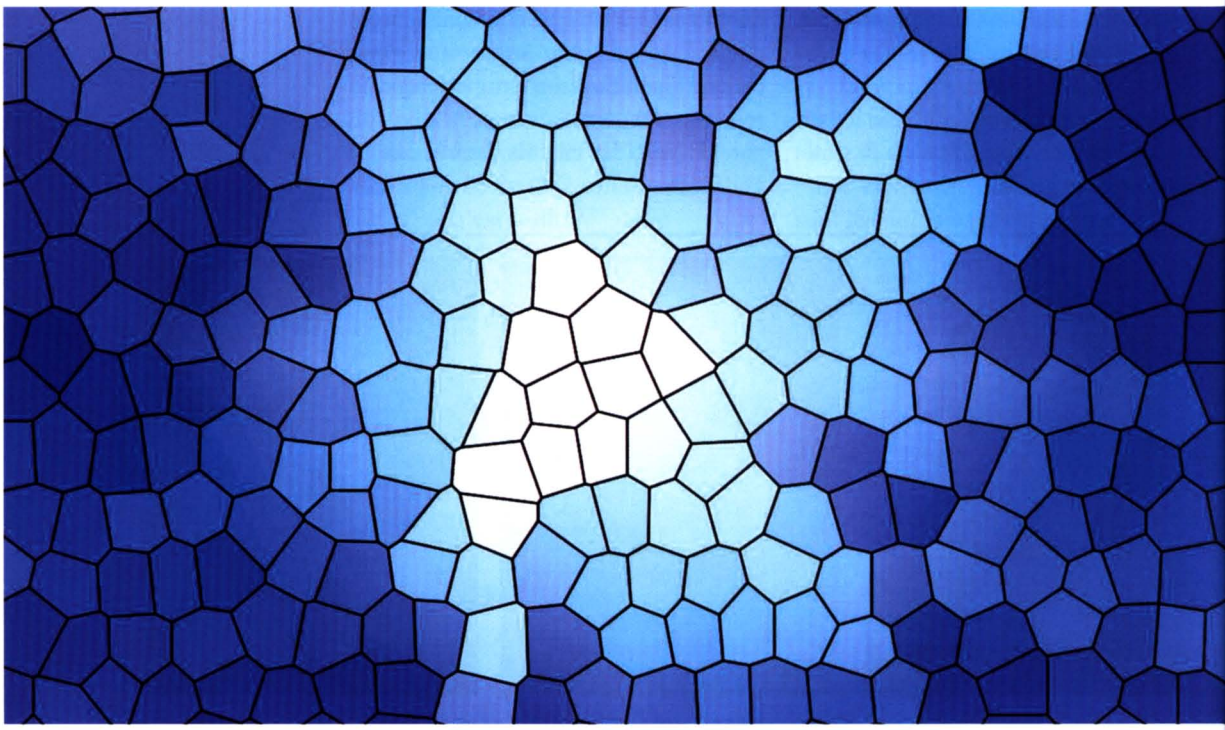
10100011000001100101011110011101
00010011100000011100101111010000
10011010111100010011111101100001
11010000011001111101111000010011
00101011100011111001000111111011
11000001010010111101000111100101
10011100000001010100000001000100
11100010111100011010011000111100

qui est donc le message codé.

Décodage RSA

Pour décrypter, il « suffit » de calculer l'inverse β de α dans $\mathbb{Z} / m\,\mathbb{Z}$. Cette propriété assure que $\alpha\beta - 1$ est divisible par $m = (p - 1)\ (q - 1)$ donc, d'après la propriété citée dans le préambule, $x^{\alpha\beta} - x$ est divisible par $pq = n$. La clef secrète est identique à la clef publique si l'on change α en β.

Voyons comment elle fonctionne dans le cas de notre exemple. Tout d'abord, il s'agit d'inverser 3 dans $\mathbb{Z} / 40\,\mathbb{Z}$. Pour cela, on utilise la relation de Bezout (voir l'article *L'arithmétique de la cryptographie*), ici :
$3 \times 27 - 40 \times 2 = 1$ donc $\beta = 27$ convient.
La clef secrète consiste à reprendre les opérations précédentes en remplaçant 3 par 27.
Nous partons donc du nombre binaire ci-dessus et l'écrivons en base 55, nous retrouvons la liste d'éléments de $\mathbb{Z} / 55\,\mathbb{Z}$:



01 00 35 13 00 02 07 38 52 19 23 41 20 51 38 28 28 46 46 23 15 05 30 48 19 03 40 24 50 18 32 16 31 41 30 39 54 04 37 49 02 04 40 17 31

Voyons comment on élève à la puissance 27 sans trop de calculs sur l'exemple du chiffre des unités :
$31^2 = 31 \times 31 = 961 = 17 \times 55 + 26$
donc $31^2 = 26$ dans $\mathbb{Z}/55\mathbb{Z}$,
$31^4 = 26^2 = 26 \times 26 = 676 = 12 \times 55 + 16$
donc $31^4 = 16$ dans $\mathbb{Z}/55\mathbb{Z}$,
$31^8 = 16^2 = 16 \times 16 = 256 = 4 \times 55 + 36$
donc $31^8 = 36$ dans $\mathbb{Z}/55\mathbb{Z}$,
$31^{16} = 36^2 = 36 \times 36 = 1296 = 23 \times 55 + 36$
donc $31^{16} = 31$ dans $\mathbb{Z}/55\mathbb{Z}$,
donc : $31^{27} = 31^{12} = 31^8 \times 31^4 = 36 \times 16 = 576 = 10 \times 55 + 26$ donc $31^{27} = 26$ dans $\mathbb{Z}/55\mathbb{Z}$.

Bien sûr, nous retrouvons la liste :

01 00 40 07 00 18 28 47 13 24 12 46 15 06 47 52 52 51 51 12 05 25 35 27 24 42 50 29 30 17 43 36 26 46 35 19 54 49 38 14 18 49 50 08 26

et donc notre message.

Inviolabilité de RSA

Dans la pratique, les nombres $n = p\,q$ utilisés ont plusieurs centaines de chiffres. L'inviolabilité de la méthode RSA vient de la difficulté de factoriser n pour en déduire p et q puis m et β. Casser ce code est une question de factorisation (voir l'encadré *Comment casser RSA* ainsi que les articles *L'anniversaire des briseurs de codes* et *La carte qui dit « oui »*). Cette idée d'inviolabilité vient de l'expérience, ce n'est ni un fait démontré, ni même un fait démontrable. La découverte de nouvelles méthodes de factorisation pourrait d'ailleurs être tenue secrète dans le but de décoder les messages d'autrui.

H. L.

Comment casser RSA

Prenons un RSA de clef publique :
(49 808 911, 5 685 669).
Pour retrouver la clef secrète, il faut d'abord factoriser le nombre 49 808 911. Ceci est facile avec système de calcul symbolique comme *Maple* ou *Mathematica*.

En guise d'exemple, voici les instructions en *Maple* :

```
with(numtheory); factorset(49808911);
```

ce qui donne $p = 17\,669$ et $q = 2\,819$ donc $m = 49\,788\,424$.

Pour trouver l'inverse de α, on utilise la formule de Bezout (voir l'article *L'arithmétique de la cryptographie*) : puisque α et m sont premiers entre eux, il existe u et v tels que $\alpha\,u + m\,v = 1$, l'inverse est donc u.
Ces coefficients de Bezout sont donnés par la fonction *igcdex* de *Maple* :

```
igcdex(5 685 669,49 788 424,'u','v'); u;
```

ce qui donne : 843 589. La clef secrète est donc (49 808 911, 843 589).

Les procédures de codage et de décodage sont symétriques, elles utilisent la procédure puissance :

```
Mult:=proc(b, c, n)irem(b*c, n);
end;
Puissance:=proc(x, a, n)local y, z;
if a=0 then 1 else
y:=Puissance(x, iquo(a, 2), n);
z:=Mult(y, y, n);
if irem(a, 2)=0 then z else
Mult(z, x, n) fi fi;
end;
```

On code avec :

```
Puissance(x, 5 685 669,49 808 911);
```

et on décode avec :

```
Puissance(x, 843 589,49 808 911);
```

Il reste bien sûr à transformer les textes en nombres et à écrire les nombres en base 49 808 911 pour coder chaque chiffre.

par J.-G. Dumas et D. Trystram

RSA : les faiblesses d'un code mythique

Le code RSA, qui protège en particulier les cartes bancaires, a longtemps passé pour l'exemple même du code inviolable. Il est pourtant aujourd'hui en danger, à cause du progrès constant en matière de factorisation d'entiers.

La cryptographie a pour objet le développement de méthodes de codage de l'information de sorte que le décodage soit difficile à quiconque ne possédant pas la clef adéquate. Avec l'extension considérable d'Internet depuis les années quatre-vingt-dix, ce problème est devenu crucial. En effet, l'information transitant sur le réseau est visible et potentiellement accessible à tous, il faut donc la coder pour cacher son sens afin de garantir une certaine confidentialité. Les systèmes de cryptage à clef publique permettent à des émetteurs désirant envoyer des messages confidentiels de coder leurs messages avec une clef publique connue. Seul le destinataire pourra les décoder.

Casser un système de cryptage, c'est trouver une méthode de décodage sans posséder la clef. Nous allons montrer comment ceci peut être obtenu à partir de la théorie des nombres (en déterminant les facteurs premiers d'un grand nombre). S'il est facile de multiplier deux nombres premiers, même très grands, il est par contre plus difficile d'en déterminer les facteurs premiers. Cependant, vu les progrès constants en algorithmique et en architecture des systèmes informatiques, les limites imposées pour la taille des clefs de cryptage seront bientôt dépassées.

Systèmes à clefs publiques

Coder un message x, c'est produire un nouveau message y à l'aide d'une fonction de codage f, construite de sorte que y soit très différent de x, mais aussi telle que le décodage soit possible à l'aide d'une autre fonction g, différente de f. f et g sont les clefs du système de codage. Pour la confidentialité, il est essentiel que g ne soit connue que de celui (ou ceux) à qui le message est destiné (c'est la clef privée), f est la clef publique, elle est connue de tous. RSA (voir l'article *Le code RSA*) est le système de codage à

clef publique le plus connu. C'est aussi le plus utilisé, par exemple lors de transactions sécurisées sur Internet (pour la confidentialité du courrier ou l'authentification des utilisateurs).

Pour rendre le système RSA utilisable, il faut tout d'abord générer des clefs, puis, coder le flux d'information que l'on désire envoyer, enfin, il faut que le destinataire puisse le décoder :

- Pour générer des clefs, il faut tout d'abord être capable de produire des nombres premiers très grands (de plusieurs milliers de chiffres), c'est-à-dire de tester rapidement si un nombre donné est premier ou non. Un tel calcul est extrêmement rapide. En outre, il est possible de prendre des nombres premiers déjà découverts (par exemple sur la page des nombres premiers : http://www.utm.edu/research/primes) et les combiner deux par deux pour obtenir directement le modulo. Dans RSA, la clef publique est composée de deux nombres premiers p et q : $m = pq$.

- À partir des clefs, il faut coder l'information à envoyer. Pour cela, il faut calculer des puissances modulo m. Pour que le codage reste praticable, c'est-à-dire pour que le calcul des puissances modulo des nombres de plusieurs milliers de chiffres ne soit pas très coûteux, on ne peut actuellement guère dépasser des nombres de plusieurs centaines de chiffres pour les clefs. Il nous reste à vérifier que cet ordre de grandeur est suffisant pour garantir une bonne fiabilité du codage.
- Celui qui possède la clef privée peut alors décoder le message.

Illustrons cette méthode par un petit exemple en RSA. D'après les calculs de l'encadré *Le code RSA*, pour le modulo : $55 = 5 \times 11$, on peut choisir les puissances 3 et 27. Autrement dit, on code en élévant à la puissance 3 (modulo 55) et on décode en élevant à la puissance 27 (modulo 55). Supposons qu'un émetteur veuille envoyer un message confidentiel au destinataire. Par exemple, le mot « bonjour ». Il faut tout d'abord qu'il se procure la clef publique fermée du couple (3, 55), puis qu'il code son message (par exemple en utilisant la correspondance entre les lettres du mot b-o-n-j-o-u-r et leur numéro dans l'alphabet 02-15-14-10-15-21-18). Le codage du mot s'obtient en calculant les cubes de ces nombres successifs modulo 55 :

$2^3 \equiv 8$, $15^3 \equiv 20$, $14^3 \equiv 49$, $10^3 \equiv 10$, $15^3 \equiv 20$, $21^3 \equiv 21$ et $18^3 \equiv 2$.

Le code RSA est protégé par la difficulté de factoriser les grands nombres.

Longtemps limitée à 320 bits la taille des clefs de cryptage des cartes bancaires est maintenant de 768 bits... mais tout porte à croire que l'on saura bientôt les craquer.

Une fois reçu le message 08204910202102, le destinataire le décode facilement en calculant les puissances $27^{\text{ièmes}}$. On peut le vérifier sur le premier caractère : $8^{27} \equiv 2$ [55] qui permet bien d'obtenir le « b ». Le lecteur courageux pourra le vérifier pour tous les autres caractères !

Fiabilité de RSA

Une fois des clefs construites, il faut établir si notre système de cryptage est fiable, c'est-à-dire si la clef privée u peut rester secrète. Pour l'obtenir, il faut factoriser m, pour trouver p et q (voir l'article *Le code RSA*). Si pour notre petit exemple précédent ($m = 55$) le problème n'est pas ardu, il en est tout autrement dès que m devient grand (plusieurs centaines de chiffres) !

Le décodage de RSA revient au problème de la factorisation d'entiers. Avant toute factorisation, la première difficulté est de savoir reconnaître les nombres premiers, c'est-à-dire d'avoir un algorithme qui permette de décider si un nombre donné est premier ou non. Nous disposons d'algorithmes efficaces en pratique : les méthodes de cribles classiques (qui consistent à essayer de diviser le nombre par tous les nombres plus petits) sont chères et surtout, très gourmandes en place mémoire. On préfèrera des méthodes probablistes comme le test de Rabin-Miller (ici, on cherche à déterminer si un nombre est premier avec une probabilité d'erreur arbitrairement faible).
La factorisation d'entiers est un problème qui peut s'exprimer de manière relativement simple mais qui n'a pas, jusqu'à présent, de solution vraiment efficace. De nombreux algorithmes très différents existent selon les tailles des nombres à factoriser (l'algorithme Pollard rho pour des petits nombres, les courbes elliptiques pour les nombres de quelques dizaines de chiffres et le crible quadratique, actuellement recordman du monde).

Le champion du monde

L'algorithme le plus rapide actuellement sur les ordinateurs classiques pour la factorisation des codes RSA est l'algorithme NFS « *Number Field Sieve* » ou crible de corps de nombre, dérivé du crible quadratique. Le principe général consiste à déterminer des couples de nombres tels que leurs carrés soient congrus modulo m : $x^2 \equiv y^2$ [m]. Dans ce cas, le produit $(x - y)(x + y)$ est un multiple de m et, avec de la chance, l'un de ces deux nombres permet donc de décomposer m. Toute la difficulté de l'algorithme consiste à trouver de tels entiers x et y ! L'idée est d'essayer de trouver des x et y tels que leurs carrés soient proches de m, de sorte que le résultat modulo m soit petit. Si ce résultat est petit, alors il est facile de trouver les petits nombres premiers qui le divisent. Ensuite, il ne reste plus qu'à trouver des combinaisons idoines de ces petits nombres.
Pour préciser la méthode, cherchons à factoriser l'entier $m = 7\,429$. Pour cela, on choisit un entier proche de sa racine carrée, 86 par exemple, et décomposons en facteurs premiers les carrés modulo m des entiers proches de 86 :

$$79^2 = -\,2^2 3^3 11^1 \ [7\,429] ;$$
$$80^2 = -\,3^1 7^3 \ [7\,429] ;$$
$$81^2 = -\,2^2 7^1 31^1 \ [7\,429] ;$$

$82^2 = -\ 3^1 5^1 47^1$ [7 429] ;
$83^2 = -\ 2^2 3^3 5^1$ [7 429] ;
$84^2 = -\ 373$ [7 429] ;
$85^2 = -\ 2^2 3^1 17^1$ [7 429] ;
$86^2 = -\ 3^1 11^1$ [7 429] ;
$87^2 = 2^2 5^1 7^1$ [7 429] ;
$88^2 = 3^2 5^1 7^1$ [7 429] ;
$89^2 = 2^2 3^1 41^1$ [7 429] ;
$90^2 = 11^1 61^1$ [7 429] ;

Ainsi, on trouve deux couples dont le produit donne de petits facteurs avec seulement des puissances paires :
$(79 \times 86)^2 = (2 \times 32 \times 11)^2$ [7 429] ;
$(87 \times 88)^2 = (2 \times 3 \times 5 \times 7)^2$ [7 429].

On pose alors $x = 87 \times 88$ et $y = 2 \times 3 \times 5 \times 7$. D'après ce qui précède, le produit des deux nombres $87 \times 88 \pm 2 \times 3 \times 5 \times 7$ est un multiple de 7 429. On considère alors les *pgcd* de ces nombres et de 7 429. On trouve 17 et 437. Bien entendu, ce sont des diviseurs de 7 429. On a factorisé $7429 = 17 \times 437$.
Si l'idée de base est relativement simple, la programmation est délicate mais très performante. On met les exposants des facteurs premiers dans un tableau et on cherche des combinaisons de lignes qui ne donnent que des valeurs paires. Les calculs des différentes lignes sont indépendants et peuvent se faire en parallèle sur Internet. La dernière phase de la résolution doit cependant se faire sur une seule machine à grosse mémoire.

Faiblesse des clefs actuelles

Les méthodes actuelles de décryptage de clefs de grandes tailles nécessitent de résoudre un problème de factorisation d'entiers. Aujourd'hui, les meilleures méthodes sont basées sur le crible quadratique (NFS) et utilisent des centaines d'ordinateurs tournant parfois pendant plusieurs semaines. La taille des clefs décryptées récemment augmente rapidement (174 chiffres en décembre 2003, 176 au début 2005, et 200 en mai 2005 !), ce qui correspond à des clefs RSA d'environ 665 bits.

Les tailles des clefs de cryptage des cartes bancaires a longtemps été limitée à 320 bits, jusqu'à ce qu'un informaticien (Serge Humpich) arrive à fabriquer de faux codes en 1998. Les autorités ont alors étendu la taille à 768 bits (soit, environ 232 chiffres). Vus les progrès constants sur la factorisation d'entiers, le décryptage sera bientôt à la portée de tous même si les clefs RSA ne sont pas choisies n'importe comment de façon à être plus dures à craquer que des nombres quelconques ! Il faudra sans doute bientôt penser à reculer encore cette limite qui au rythme actuel devrait être atteinte facilement dans quelques années.

J.-G. D. & D. T.

	Exposant de –1	Exposant de 2	Exposant de 3	Exposant de 5	Exposant de 7	Exposant de 11	
79^2	1	2	3	0	0	1	...
80^2	1	0	1	0	3	0	...
...							
86^2	1	0	1	0	0	1	...
87^2	0	2	0	1	1	0	...
88^2	0	0	2	1	1	0	...
...							

Tableau des exposants

par David Delaunay

La carte qui dit « oui »

La carte bleue est protégée par un code appelé code RSA. Pour le casser, il suffit de savoir factoriser un grand nombre. Le progrès mathématique est-il délictueux ?

La progression des moyens informatiques et des outils d'analyse mathématique entraîne une course poursuite entre la cryptologie (la science de la cryptographie) et la cryptanalyse (l'art de briser un cryptosystème). Gare à ceux qui sont à la traîne ! Ce fût le cas du groupement interbancaire (GIE) qui certifie les cartes bleues. Bien qu'averti par des spécialistes de l'obsolescence de leur système face aux techniques modernes, le GIE a utilisé durant plus de 15 ans un système initialement prévu pour ne durer que 5 années. À la fin des années 1990, Serge Humpich a découvert la faille. Avant de voir celle-ci, voyons comment fonctionnent les cartes bleues.

Dès que ce nombre à 232 chiffres sera factorisé, les « yescards » réapparaîtront.

Le principe de la carte bleue

En 1983, le GIE s'appuie sur la cryptographie RSA (le code RSA doit son nom aux mathématiciens Rivest, Shamir et Adleman, qui l'inventèrent en 1977 au MIT) pour former un cryptosystème à clef privée, clef publique basé sur un nombre $n = pq$ de 96 chiffres décimaux (soit 320 bits). Ce cryptosystème permet de former une signature authentifiant la carte bleue. Plus précisément, lorsqu'un usager fait une demande de carte bancaire, sa banque produit un certain nombre d'informations qu'elle fournit au GIE. A partir de celles-ci, le GIE forme un numéro d'identifiant I correspondant au numéro à 16 chiffres visible sur la carte. Parallèlement, à l'aide de sa clef secrète, le GIE forme un numéro d'authentification J. Ces différentes données ainsi que quelques autres sont emmagasinées dans les quelques Ko de mémoire EPROM de la carte à puce qui peut dès lors être transmise au client.

Est-il licite de factoriser ce nombre ?

1550880802783769298423921500751307878471020215206711102793111990113875394553459999757605304671735856091597555389797408938173344043674704780986390069906679096728933081405044935969514508676239942493440750589270015739962374529363251827

Ce nombre de 232 chiffres est utilisé par le GIE pour sécuriser les cartes bancaires. Il est dit qu'un groupe d'étudiants serait parvenu à le factoriser mais aurait préféré garder secret le résultat obtenu, par respect pour leur passion. Cela semble douteux et à ce jour aucune factorisation n'a encore été diffusée. Cependant certains sites peu scrupuleux proposent de communiquer cette factorisation moyennant paiement téléphonique. En tout cas dès que ce nombre sera factorisé, il est probable qu'un grand nombre de « yescards » feront leur réapparition !

Une fois chez un commerçant agréé, le client insère sa carte dans un terminal et commence alors une phase d'authentification (ce message apparaît souvent sur l'écran du terminal). Lors de celle-ci, les numéros I et J sont lus et le terminal contrôle, par le biais de la clef publique du GIE, que ces numéros se correspondent. Si le montant de la transaction est faible et selon la nature de la carte, la phase d'authentification peut en rester là, sinon le terminal contacte un serveur général pour voir s'il n'y a pas d'interdit bancaire, d'opposition ou pour savoir si le compte est suffisamment alimenté. Le client saisit ensuite son code personnel.

À quoi sert le code PIN ?

À chaque carte bancaire est associé un code PIN (Personnel Identifer Number) de quatre chiffres. Celui-ci sert à identifier le porteur de la carte bleue. Lorsque le client tape son code PIN sur le terminal, celui-ci est transmis à la puce de la carte qui vérifie son exactitude. Si la puce reçoit successivement trois codes incorrects, elle se fige et la carte devient inutilisable. Si le code reçu est correct, la puce mémorise le montant et la date de la transaction, elle vérifie que la somme


Serge Humpich dans son atelier

Le premier qui dit la vérité ...

« *Le premier qui dit la vérité, il doit être exécuté* » dit la chanson de Guy Béart. Peut-on appliquer ce proverbe à Serge Humpich ? Nous ne rentrerons pas dans ce débat. Voici simplement le listing Maple du travail qui lui valut dix mois de prison avec sursis :

```
facteur1 : = convert(`c31f7084b75c502caa4d19eb137482aa4cd57aab`,decimal,hex) ;
facteur2: = convert(`14fdeda70ce801d9a43289fb8b2e3b447fa4e08ed`,decimal,hex) ;
produit : = facteur1*facteur2 ;
exposant_public : = 3;
modulo_div_eucl : = (facteur1 - 1)*(facteur2 - 1);
exposant_prive : = expand((1 + 2*modulo_div_eucl)/3);
```

Le nombre utilisé comme clef publique par le GIE était :
213598703592091008239502270499962879705109534182641740644252416500858395774644508840500943086599 9, il s'agit bien du produit des deux facteurs donnés ci-dessus.
Selon les attendus du tribunal, la décomposition de la clef publique en facteurs premiers constitue un délit de contrefaçon de cartes bancaires. Amusant, non ? A quand l'interdiction de l'enseignement de l'arithmétique ?
Si Serge Humpich était coupable selon le tribunal, il est clair que le GIE était fautif d'utiliser une clef dont les mathématiciens savaient qu'elle était factorisable. Dans quelques années, la clef actuelle du GIE le deviendra. Les cartes bancaires seront falsifiables ! Quelle sera sa réaction ? Nous l'attendons avec intérêt. Dans tous les cas, rappelons à cette vénérable institution qu'il est futile de vouloir lutter contre les progrès de la science sur le terrain judiciaire. La papauté l'a fait avant elle et a gagné, et pourtant elle tourne ! N'est-ce pas monsieur Galilée ?

des montants des transactions écoulées dans la semaine n'excède pas un certain plafond et donne son accord au terminal en produisant un numéro de transaction (qui apparaîtra sur la facturette). Le client peut alors repartir avec sa marchandise et le commerçant transmettra dans la soirée à sa banque l'ensemble des transactions réalisées dans la journée.

Serge Humpich et les « yescards »

S'interrogeant sur la fiabilité des cartes bancaires, Serge Humpich est parvenu à comprendre les mécanismes d'authentification d'une carte bleue. En 1997, il factorise le nombre n de 96 chiffres défini par le GIE notamment à l'aide du logiciel de calcul formel Maple. Il lui est alors facile, à partir d'un numéro d'identifiant farfelu I de calculer le numéro d'identification J correspondant. Il conçoit alors de fausses cartes bleues dont la puce répond « oui » à n'importe quel code PIN : ce sont les fameuses « yescards ».

Serge Humpich contacte le GIE pour négocier sa découverte d'une faiblesse dans le protocole d'authentification des cartes bancaires. Le GIE lui demande alors de prouver ses dires et Humpich achète 10 carnets de tickets de métro auprès d'un distributeur de la RATP à l'aide de 10 numéros de cartes bancaires inexistant de la forme :

xxxx xx98 7654 321x, xxxx xx09 8765 432x, xxxx xx10 9876 543x, etc.

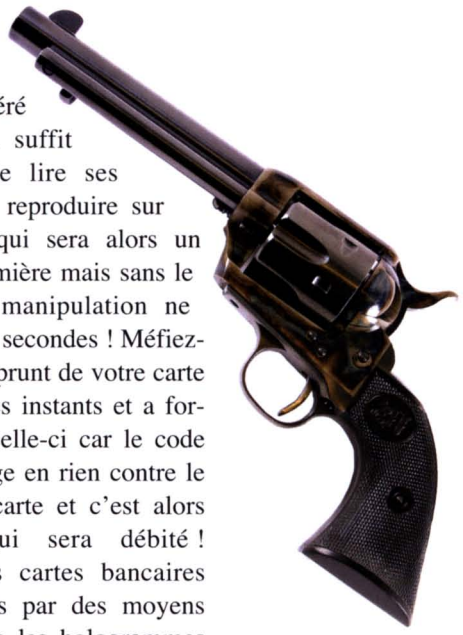
Le GIE semble alors prêt à négocier mais mène parallèlement une enquête lui permettant de remonter jusqu'à l'auteur de ces pratiques « frauduleuses ». Il ordonne une perquisition du domicile de Serge Humpich. Sa découverte est alors diffusée publiquement sur le Net (voir l'encadré *Le premier qui dit la vérité…*) en juin 1999. Serge Humpich est condamné à dix mois de prison avec sursis en février 2000.

Pour combler la faille découverte par Humpich, le GIE est passé progressivement à un système de cryptographie basé sur un nombre $n = pq$ d'environ 230 chiffres ce qui lui a coûté l'adaptation des terminaux de paiement. Cela sera suffisant jusqu'à ce que l'on parvienne à factoriser ce nombre sachant que l'on est déjà récemment parvenu à factoriser un nombre de 200 chiffres décimaux. Malheureusement cela ne comble pas toutes les failles du système de paiement par carte bleue.

Le logiciel geZeroLee

En 2001, plusieurs institutions françaises sont averties de l'imminence d'une attaque informatique contre les systèmes de gestion des paiements par cartes bancaires. Quelques jours plus tard, un pirate informatique répondant au nom de code de « geoli » diffuse sur Internet le logiciel geZeroLee. Ce logiciel invite à suivre « le lapin blanc » et permet, à titre didactique, de fabriquer de fausses cartes bancaires ! Ce logiciel peut être encore trouvé sur le Net à condition de faire preuve d'un peu de courage.

La faille exploitée par les pirates consiste simplement à copier une carte existante. En effet la carte bancaire communique au terminal les numéros d'identification I et J avant même que le propriétaire de la carte ait inséré son code PIN. Il suffit alors au pirate de lire ses numéros et de les reproduire sur une fausse carte qui sera alors un « clone » de la première mais sans le code PIN. Cette manipulation ne prend que quelques secondes ! Méfiez-vous donc d'un emprunt de votre carte bleue pour quelques instants et a fortiori d'un vol de celle-ci car le code PIN ne vous protège en rien contre le clonage de votre carte et c'est alors votre compte qui sera débité ! Heureusement, les cartes bancaires sont aussi protégés par des moyens mécaniques comme les hologrammes ou le relief de la carte qui compliquent sérieusement le travail des pirates.



Page d'accueil d'un logiciel de clonage de carte bancaire

De nos jours une vingtaine d'ingénieurs et d'anciens pirates de cartes bancaires travaillent ensemble à rechercher et à combler les failles de ce moyen de paiement qui reste l'un des plus sûrs.

D. D.

par David Delaunay

Les codes qui se corrigent

Comment éviter les erreurs dans la transmission des informations ? Dans les années 50, Richard Hamming a proposé une réponse à ce problème. Elle est toujours d'actualité.

La transmission d'informations numériques par ligne téléphonique, fibre optique, etc. se fait par l'envoi successif de bits égaux à 0 ou 1. Cette transmission peut être parasitée ce qui conduit à une inversion de certains bits (le taux d'erreur est de l'ordre de 10^{-5} sur une ligne téléphonique et de 10^{-11} en fibre optique) Sans contrôle, l'information est alors irrémédiablement altérée. Voyons différents schémas transformant un message donné en un message « contrôlé » permettant de détecter la présence d'erreurs, voire de corriger celles-ci.

Bit de parité

On sectionne l'information à transmettre en paquets de n bits, correspondant aux messages. À un message donné on forme le message contrôlé en ajoutant un $n+1^{\text{ième}}$ bit de sorte qu'il y ait en tout un nombre de pair de 1. Ce $n+1^{\text{ième}}$ bit est appelé bit de parité. Concrètement : 0 1 0 1 1 1 0 devient 0 1 0 1 1 1 0 | 0 et 1 0 1 1 1 1 0 devient 1 0 1 1 1 1 0 | 1. L'émetteur transmet alors le message contrôlé puis le récepteur vérifie la parité du message reçu. Si lors de la transmission une erreur survient, la parité du message reçu est incorrecte et le récepteur décèle l'existence d'une erreur sans pour autant être capable de la localiser. Il demande alors une nouvelle émission du message. Si lors de la transmission deux erreurs surviennent, la parité du message reçu est correcte, les erreurs ne sont pas décelées.

Tableau de parité

L'information à transmettre est ici découpée en paquets de n^2 bits que l'on organise sous la forme d'un tableau à n lignes et n colonnes. Ce tableau constitue notre message. Pour le contrôler, on y ajoute une $n+1^{\text{ième}}$ ligne et une $n+1^{\text{ième}}$ colonne de sorte que chaque ligne et chaque colonne présentent un nombre pair de 1. Ainsi les n premiers

coefficients de la dernière ligne (respectivement colonne) contrôlent la parité de la colonne (respectivement ligne) qui lui correspond. Le dernier coefficient, quant à lui, contrôle à la fois la parité de la $n+1^{\text{ième}}$ ligne et de la $n+1^{\text{ième}}$ colonne (en fait, il contrôle même la parité du tableau initial).
Concrètement :

$$\begin{matrix}0&0&1\\1&1&1\\0&1&1\end{matrix} \text{ devient } \begin{array}{ccc|c}0&0&1&0\\1&1&1&1\\0&1&1&0\\\hline 0&0&1&1\end{array}.$$

Si lors de la transmission du message ainsi contrôlé une erreur survient, le récepteur détecte une erreur de parité en ligne et une erreur de parité en colonne. Le bit erroné est alors repéré et il suffit de l'inverser pour le corriger. Notons que ce schéma fonctionne même si c'est l'un des bits de parité qui est erroné. Si deux erreurs de transmission sont commises, il y aura, selon leur positionnement, deux ou quatre erreurs de parité. L'existence d'erreurs est détectée, mais on ne pourra pas les corriger. Si trois erreurs sont commises, on peut détecter encore leur présence par la parité globale. Si quatre erreurs sont commises et si celles-ci sont malicieusement positionnées sur deux mêmes lignes et deux mêmes colonnes on ne les détecte pas.

Codage de Hamming

Le message contrôlé comporte ici 2^n bits. Concrètement nous allons prendre $n = 3$, même si une valeur supérieure paraît nécessaire pour percevoir l'efficacité de la méthode. Nous allons donc transmettre un octet dont les bits seront numérotés de 0 à 7 en allant de droite à gauche. Le bit 0 va contrôler la parité de l'octet. Les bits 1, 2 et 4 (correspondant aux puissances de 2) seront les bits de contrôle de notre message. Les bits 3, 5, 6 et 7 (les autres) seront les bits contenant l'information transmise. On décompose leur numéro en sommes de puissances de 2 :
$3 = 2 + 1$, $5 = 4 + 1$, $6 = 4 + 2$ et $7 = 4 + 2 + 1$.
Ensuite, on détermine les bits de contrôle 1, 2 et 4 de sorte que chacun complète la parité de l'ensemble des bits d'information où ils apparaissent dans la décomposition en puissance de 2. Concrètement, si on veut transmettre 1 1 0 1, on forme d'abord :

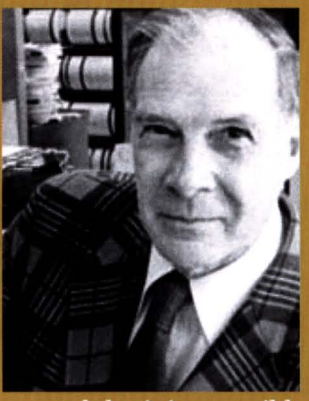
7	6	5	4	3	2	1	0
1	1	0		1			

Puis on détermine le bit 1 de la manière suivante : 1 apparaît dans la décomposition des nombres 3, 5 et 7, il va donc contrôler la parité de ces trois bits et ici valoir 0. De même le bit 2 contrôle 3, 6 et 7, il vaut donc ici 1, etc.
Finalement le message contrôlé est :

7	6	5	4	3	2	1	0
1	1	0	0	1	1	0	0

Richard Wesley Hamming

Mathématicien américain, Hamming (1915-1998) travailla au projet Manhattan pendant la Seconde Guerre Mondiale (construction des premières bombes atomiques). Ses premiers résultats sur les codes correcteurs d'erreur datent de 1950. La citation suivante montre la clarté de son choix des mathématiques appliquées :
« Mathematics is an interesting intellectual sport but it should not be allowed to stand in the way of obtaining sensible information about physical processes ».



Un message erroné est détecté car impair.

L'erreur 404

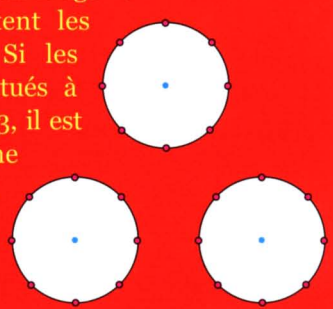
La plus célèbre des erreurs de communication sur l'Internet porte le nom d'une voiture tout aussi célèbre.
Ceux qui ne la connaissent pas n'ont sans doute jamais surfé sur le web. Voici le message poétique qui l'accompagne le plus souvent :
Not Found
The requested URL / o was not found on this server. Mais vous pouvez en trouver toutes sortes d'autres. Des sites lui sont dédiés sur Internet (à rechercher avec un moteur de recherche).

La distance de Hamming

par Hervé Lehning

La notion de distance ne concerne pas uniquement les points de la géométrie ordinaire. Par exemple, dans l'espace des paquets de n bits, on appelle distance de Hamming entre deux paquets le nombre de chiffres à modifier pour passer de l'un à l'autre. Cette notion de distance éclaire les idées de détection et de correction des erreurs de transmission exposées dans cet article. Si deux paquets « légaux » sont toujours à une distance au moins égale à 2, une erreur unique est toujours détectable, s'ils sont à une distance au moins égale à 3, elle est non seulement détectable mais aussi corrigeable.

Sur cette figure, les paquets de bits « légaux » sont en bleus, « les illégaux en rouge ».
Les cercles représentent les boules de rayon 1. Si les paquets bleus sont situés à des distances égales à 3, il est facile de corriger une erreur de transmission car à chaque paquet rouge, il correspond un seul paquet bleu.



Si lors de la transmission, une erreur est commise, on en détecte l'existence par l'imparité du message transmis, on contrôle ensuite la parité des bits 1, 2 et 4. Par exemple si l'erreur est commise sur le bit 5, les bits 1 et 4 sont incorrects. Or $1 + 4 = 5$. C'est donc le bit 5 qui est faux. Si l'erreur est commise sur le bit 2, seule la parité de celui-ci est incorrecte. Si l'erreur est commise sur le bit 0, aucun bit de contrôle n'est incorrect. A chaque fois la somme des numéros des bits de contrôles incorrects donne le bit erroné. Il ne reste alors plus qu'à l'inverser. Si deux erreurs sont commises, la parité du message reçu est correcte mais assurément 1 ou plusieurs bits de contrôle ne le sont pas. L'existence d'erreur est détectée. Notons qu'une erreur sur les bits 1 et 2 ou sur les bits 5 et 6 conduisent aux mêmes erreurs de parité. On ne saura donc pas corriger.

Efficacité

Étudions le rapport entre les tailles de l'information de contrôle et de l'information transmise (inversion). Pour le tableau de parité, nous avons : $\frac{2n+1}{(n+1)^2}$ et pour le codage de Hamming : $\frac{n+1}{2^n}$ (les bits de contrôle correspondent aux puissances de 2 strictement inférieure à 2^n : il y en a n, on y ajoute aussi le bit de parité). Pour une transmission de 128 bits, le tableau de parité nécessite 31 bits de contrôle contre 9 au codage de Hamming.

D. D.

Le code ASCII

Le code ASCII de base utilise sept bits, il permet donc ce coder $2^7 = 128$ caractères différents. De 0 à 31, il s'agit de caractères de contrôle c'est-à-dire signifiant des ordres destinés à l'imprimante comme des retours à la ligne ou des sauts de page. Le caractère 127 est la touche de suppression. Sinon, nous trouvons :

Table des caractères ASCII

33	!	47	/	61	=	75	K	89	Y	103	g	117	u
34	"	48	0	62	>	76	L	90	Z	104	h	118	v
35	#	49	1	63	?	77	M	91	[	105	i	119	w
36	$	50	2	64	@	78	N	92	\	106	j	120	x
37	%	51	3	65	A	79	O	93	]	107	k	121	y
38	&	52	4	66	B	80	P	94	^	108	l	122	z
39	'	53	5	67	C	81	Q	95	_	109	m	123	{
40	(	54	6	68	D	82	R	96	`	110	n	124	|
41	)	55	7	69	E	83	S	97	a	111	o	125	}
42	*	56	8	70	F	84	T	98	b	112	p	126	~
43	+	57	9	71	G	85	U	99	c	113	q		
44	,	58	:	72	H	86	V	100	d	114	r		
45	-	59	;	73	I	87	W	101	e	115	s		
46	.	60	<	74	J	88	X	102	f	116	t		

Cette table ne contient aucun caractère accentué car le code ASCII a été mis au point pour coder l'anglais. Une table étendue sur huit bits les contient. Malheureusement, aucun standard n'existe à ce niveau. Cela explique que vous receviez parfois des messages bizarres où tous les caractères accentués sont remplacés par d'autres plus fantaisistes.

par Hervé Lehning

Les Zips codent

David Huffman a donné son nom à une méthode de compression des données sans perte de qualité. Fondée sur les différences de fréquences d'apparition des lettres dans un texte, elle peut également servir en cryptographie.

Les lettres fréquentes sont codées par des mots courts.

Le codage de Huffman est une méthode de compression des données, utilisée en particulier pour la transmission de messages par télécopie, minitel et Internet. Sur vos ordinateurs, les noms des programmes correspondants se terminent en général par zip. Il utilise le modèle suivant. Dans chaque message, les symboles sont indépendants et apparaissent en toute position avec une fréquence connue indépendante de la position. Chaque symbole est codé en une suite de 0 et de 1 de telle façon que les symboles de forte probabilité aient un code plus court que ceux de faible probabilité. Cette méthode de codage peut également être utilisée en cryptographie.

Un exemple

Pour simplifier, supposons que nous ayons à transmettre un texte composé seulement avec les lettres a, b, c, d et e. Les fréquences d'apparition de ces lettres dans le texte sont données par le tableau suivant :

Lettre	a	b	c	d	e
Fréquence	5 %	50 %	20 %	10 %	15 %

Fréquences d'apparition des symboles dans un texte

En code ASCII (voir l'article *Le code ASCII*), chaque caractère a la taille d'un octet c'est-à-dire huit bits (0 ou 1). Un texte d'un million de caractères utilise donc un méga-octets s'il est codé en ASCII. Dans notre cas, nous n'avons que cinq caractères différents, nous pouvons donc les coder chacun sur trois bits selon la table de codage : $a : 000, b : 001, c : 010, d : 011$ et $e : 100$. Le codage du texte proposé aura alors trois millions de bits au lieu de huit millions, soit un gain de plus de 60% ! Le décodage est simple en lisant la table de codage à l'envers en découpant le texte codé en groupes de trois bits. Une idée plus subtile consiste à utiliser des codes de longueurs variables. Voici une table de codage adaptée :
$a : 1101, b : 0, c : 101, d : 111$ et $e : 100$.

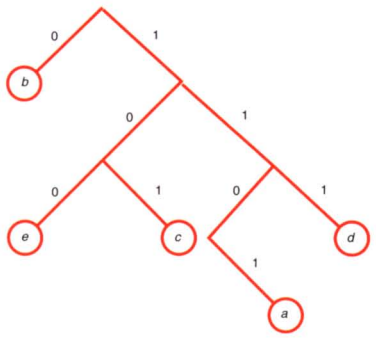
Pour 100 caractères à coder, nous avons donc en moyenne :
$5 \times 4 + 50 \times 1 + 20 \times 3 + 10 \times 3 + 15 \times 3$
soit 205 bits. Pour un million de caractères, nous obtenons donc 2 050 000 bits. Nous obtenons ainsi un gain supplémentaire d'environ 30 %.

La règle des préfixes

La méthode de décodage précédente n'est plus applicable car les codes sont à présent de longueur variable. Le texte codé est une liste de bits. Lors du décodage, ils sont lus et accumulés l'un après l'autre jusqu'à obtenir un code valide.
Supposons, par exemple, qu'avec la table de codage précédente, un mot ait été codé en 11111010100. Pour le décoder, nous lisons le premier bit (1). Il ne constitue pas un code. Nous lisons le bit suivant (1 encore) ce qui donne 11. Ce n'est pas un code non plus. Nous continuons donc et obtenons 111 qui est le code de *d*. Nous notons *d* et remettons l'accumulateur à zéro. Nous reprenons 1, 11, 110 ne sont pas des codes, 1101 en est un, celui de *a*. Et ainsi de suite. Aucune ambiguïté n'est possible car aucun code n'est le préfixe d'un autre. Nous obtenons finalement le mot *dabe*.

Arbre de Huffman

Une méthode simple pour décoder le message est de noter la règle de codage sous forme arborescente (voir ci-dessous). Au départ, nous commençons à la racine de l'arbre. Celle-ci se situe en haut de l'arbre comme en généalogie. Chaque bit correspond à une descente sur une branche suivant la règle suivante : 0 implique une descente à gauche, 1 une descente à droite. Chaque bit nous fait donc passer d'un nœud (le premier étant la racine) à un autre. Après lecture d'un certain nombre de bits, nous parvenons à une feuille étiquetée par une lettre (voir la figure *Arbre de Huffman*). Nous l'écrivons et recommençons à la racine de l'arbre avec le bit suivant.

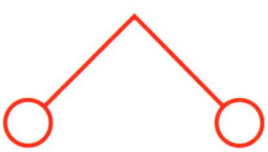


Arbre de Huffman du codage de l'exemple du texte. Le décodage se fait en le parcourant de la racine jusqu'aux feuilles.

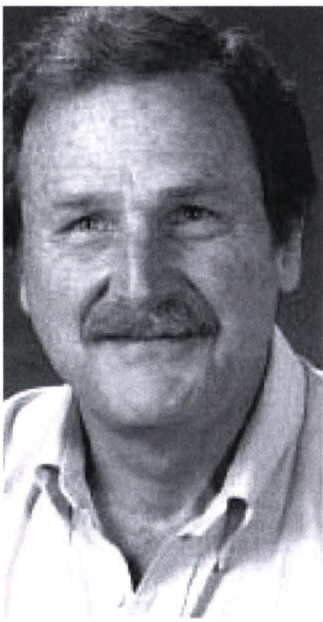
Cette notation arborescente fournit une programmation simple du décodage. Nous allons voir de plus qu'elle est à l'origine d'une idée de création automatique de la table de codage.

Construction de l'arbre

Huffman a proposé un algorithme pour construire un arbre de Huffman associé à un texte. Nous commençons par calculer la fréquence d'apparition de chaque caractère du texte à compresser. Nous prenons alors les deux lettres les moins fréquentes, c'est-à-dire *a* et *d* dans notre exemple et nous formons un arbre les joignant de la façon suivante :

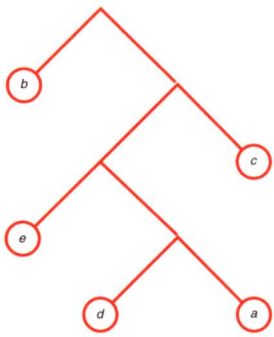


Dans cet arbre, chacune des feuilles est étiqueté par *a* et *d*. La racine est étiqueté par $a \mid d$. C'est très logiquement que nous lui attribuons la somme des fréquences de *a* et de *d* soit 15 %. Nous



David Huffman (1925-1999) a découvert les codes qui portent son nom en 1952 alors qu'il était étudiant au M.I.T.

recommençons en remplaçant a et d dans la liste des caractères par $a \mid d$. Nous adjoignons donc e et remplaçons $a \mid d$ et e par $a \mid d \mid e$ de fréquence 30 %. Notre nouvel arbre comprend une feuille de plus. À la fin de ce procédé, nous obtenons l'arbre de Huffman représenté ci-dessous.



Arbre de Huffman construit avec l'algorithme de Huffman.

La table de codage correspondant à cet arbre est la suivante :
a : 1011, b : 0, c : 11, d : 1010 et e : 100.
À présent, pour 100 caractères à coder, nous avons en moyenne :
$5 \times 4 + 50 \times 1 + 2 \times 3 + 10 \times 4 + 15 \times 3$
soit 161 bits. Le gain est maintenant d'environ 45 % au lieu de 30 %. En fait, ce phénomène est général. En d'autres termes, le codage de Huffman est optimal.

La gloutonnerie ne paye pas

Un voleur dévalise un magasin. Il veut optimiser la valeur de son vol mais il a une contrainte. Il est trop frêle pour dérober plus de 50 kg de marchandise. Il trouve trois articles. Le premier pèse 10 kg et vaut 60 F, le second 20 kg et vaut 100 F et le troisième 30 kg et vaut 120 F. Une stratégie gloutonne consiste à classer les articles par valeur au kg et donc à prendre le premier puis le second et à s'arrêter à cause de la contrainte de poids. Le vol du glouton sera donc de 160 F. Une solution plus raisonnée est de prendre les deux derniers articles ce qui fait 220 F. Mais un homme aussi raisonnable se ferait-il voleur ?

Un algorithme glouton

L'algorithme de Huffman applique une stratégie classique d'optimisation. L'idée sous-jacente à cette stratégie est que l'on peut arriver à une solution optimale en effectuant un choix optimal à chaque étape. Un tel choix est qualifié de « glouton » car il évoque un goinfre se jetant toujours sur le plus gros morceau sans réfléchir à la suite. Un algorithme glouton se comporte ainsi à chaque étape. Sous certaines conditions, il parvient à une solution optimale mais ce n'est pas toujours le cas (voir l'encadré *La gloutonnerie ne paye pas*). Dans notre cas particulier de codage à longueurs variables (suivant la règle des préfixes), les codes les plus longs doivent être attribués aux lettres les moins fréquentes. Ainsi, nous commençons la construction de l'arbre par ces feuilles en reliant les lettres les moins fréquentes. Étant les plus basses sur l'arbre, elles auront donc les codes les plus longs. Plus précisément, le résultat suivant montre que la construction d'un arbre optimal peut commencer par le choix glouton consistant à fusionner les deux lettres de fréquences minimales (voir l'encadré *Le premier choix qui compte*) : Soit T un texte où chaque caractère a a une fréquence $f(a)$. Supposons que x et y soient les deux caractères les moins fréquents de T. Il existe un codage préfixe optimal tel que les caractères x et y aient des codes de même longueur et ne différent que par le dernier bit. L'étape suivante revient à remplacer x et y dans T par $z = \{x, y\}$ de fréquence :

$$f(z) = f(x) + f(y)$$

et à coder le nouveau texte obtenu T' suivant le même principe. Les longueurs des deux textes codés sont les mêmes puisque celles des codes de x et de y aussi. L'algorithme de Huffman apporte donc une solution optimale au problème posé.

Le premier choix qui compte

Le nombre de codages préfixes étant fini, l'un d'entre eux est optimal (la longueur du texte codé correspondant est minimale). Soit A l'arbre correspondant. On considère a et b les deux feuilles sœurs de profondeur maximale dans A. On peut supposer :

$$f(a) \leq f(b) \text{ et } f(x) \leq f(y)$$

puisqu'il ne s'agit que d'une question de notation. Comme x et y sont les lettres les moins fréquentes, on a :

$$f(x) \leq f(a) \text{ et } f(y) \leq f(b)$$

On permute alors les positions de a et de x puis de b et de y dans l'arbre A. On obtient un nouvel arbre B. La longueur du codage du texte correspondant à B est inférieure à celle correspondant à A puisque a et b sont plus fréquentes que x et y. La longueur du code correspondant à A étant minimale, les deux longueurs sont donc égales. L'arbre B est donc optimal. Dans ce codage, les codes de x et de y ont même longueur et ne diffèrent que par le dernier bit.

L'algorithme de Huffman applique une stratégie classique d'optimisation.

Utilisation effective

Le codage de Huffman a de plus l'avantage de donner une programmation simple. Son utilisation comme méthode de compression exige de transmettre en début de message codé l'arbre de codage ce qui réduit son intérêt dans le cas de textes courts. D'autre part, l'élimination de cet arbre dans le message peut transformer cette méthode de compression en méthode de cryptographie. Il reste à utiliser une méthode sécurisée de transfert de l'arbre.

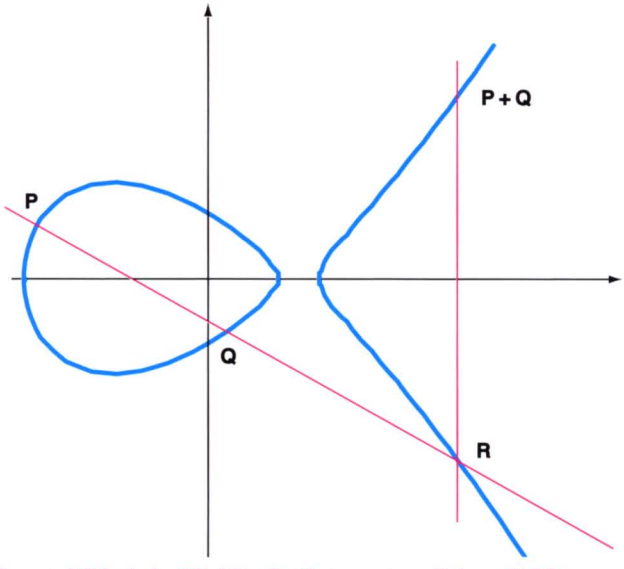
H. L.

par Hervé Lehning

Crypter avec une courbe

Les courbes elliptiques permettent de créer une méthode de cryptographie à clef publique analogue à la méthode RSA. L'idée est qu'un texte peut être transformé en une suite de points d'une courbe.

Une courbe elliptique est une courbe d'équation : $y^2 = x^3 + ax + b$ où a et b sont des nombres réels. Si $a < 0$ et $4a^3 + 27b^2 < 0$, son allure est donnée ci-dessous.



On note PQ la droite PQ si P ≠ Q et la tangente en P sinon. Si PQ recoupe la courbe en un point R, on note P + Q le point où la verticale en R la recoupe. Sinon, on pose P + Q = 0. De même, on pose : P + 0 = P.

On lui adjoint un point fictif dit *point à l'infini* que l'on note ici 0. Une construction géométrique permet alors de définir une loi interne sur une telle courbe comme le montre la figure ci-contre. Des calculs algébriques sont nécessaires pour montrer que cette loi munit bien la courbe elliptique d'une loi de groupe.

La généralité des calculs

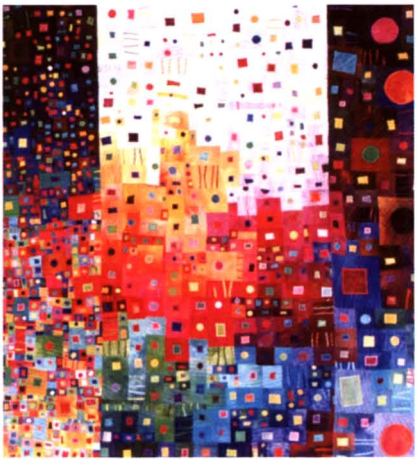
Plus précisément, si P et Q sont distincts et ont pour coordonnées (x_1, y_1) et (x_2, y_2) tels que $x_2 \neq x_1$, les coordonnées (x_3, y_3) de P + Q sont :

$$x_3 = \left(\frac{y_2 - y_1}{x_2 - x_1}\right)^2 - x_1 - x_2$$

$$y_3 = -\frac{y_2 - y_1}{x_2 - x_1} x_3 + \frac{x_1 y_2 - x_2 y_1}{x_2 - x_1}$$

En tenant compte de $y^2 = x^3 + ax + b$, on trouve :

$$\frac{y_2 - y_1}{x_2 - x_1} = \frac{x_2^2 + x_2 x_1 + x_1^2 + a}{y_2 + y_1}.$$



On en déduit une formule valable même si P = Q. Cette formule a de plus le mérite de garder un sens dans un corps quelconque. Nous avons simplement besoin de conserver l'hypothèse $4a^3 + 27b^2 \neq 0$ et d'utiliser un corps de caractéristique différente de 2 et de 3 (voir le glossaire dans l'article *L'arithmétique de la cryptographie*).

Cryptage

Les courbes elliptiques sur des corps $\mathbb{Z} / N\mathbb{Z}$ permettent de définir des méthodes de cryptographie à clef publique analogues au système RSA. L'idée de départ est qu'un texte peut être transformé en une suite de points de la courbe. Cela revient à écrire dans un alphabet ayant autant de signes que la courbe a de points. Notons que le problème sous-jacent n'a rien de simple mais, théoriquement, la question est de transformer un point M de la courbe.

La clef secrète est constituée d'un point P de la courbe et d'un entier k. On calcule ensuite P' = k P. La clef publique est alors le couple de points (P, P'). Pour crypter un point M, le chiffreur choisit un entier l et transmet le couple (U, V) défini par :

U = l P et V = M + l P '.

La connaissance de k suffit pour retrouver M : M = V – k U.

Il est facile de voir qu'elle est même nécessaire.

Logarithme discret

Pour retrouver k connaissant P et P ', il suffit de savoir résoudre l'équation :

P ' = k P.

Ce nombre k est appelé logarithme discret ce qui n'est guère intuitif si on utilise la notation additive ci-dessus. Avec une notation multiplicative de la loi du groupe, cela devient plus habituel puisque l'équation s'écrit alors :

P' = P^k.

Le cryptage avec une courbe elliptique revient à écrire dans un alphabet ayant autant de signes que la courbe a de points.

Dans le corps des réels, k correspondrait alors au logarithme de base P de P' d'où son nom dans le cadre d'un groupe fini.

À l'heure actuelle, ce problème est considéré comme très difficile. On estime qu'une clef de 200 bits (valeur de N) pour les courbes elliptiques est plus sûre qu'une clef de 1 024 bits pour la méthode de RSA. Comme les calculs sur les courbes elliptiques ne sont pas compliqués à réaliser, c'est un gros avantage pour les cartes à puces où on dispose de peu de puissance, et où la taille de la clef influe beaucoup sur les performances.

Les inconvénients sont de deux ordres. D'une part, la théorie des fonctions elliptiques est complexe et relativement récente. Il n'est pas exclu que l'on puisse contourner le problème du logarithme discret. D'autre part, la technologie de cryptographie par courbe elliptique a fait l'objet du dépôt de nombreux brevets à travers le monde. Cela pourrait rendre son utilisation coûteuse !

H. L.

par Hervé Lehning

L'arithmétique de la cryptographie

La cryptographie moderne utilise des résultats arithmétiques anciens, comme la division euclidienne, les théorèmes de Bezout et de Fermat.

En arithmétique, si vous lisez « nombre entier », comprenez « nombre entier relatif », c'est-à-dire avec un signe (+ ou –), comme 0, 1, 2, 3, etc., bien entendu, mais aussi – 1, – 2, – 3, etc. Leur ensemble est noté $\mathbb{Z}$. L'intérêt d'utiliser ces nombres plutôt que 1, 2, 3, etc., tient dans la structure d'anneau commutatif de $\mathbb{Z}$ muni des deux opérations usuelles (voir le glossaire). En bref, on peut soustraire les nombres sans problème.

Division euclidienne

Combien de paquets de 23 dragées peut-on confectionner si on en dispose de 156 ? Ce problème classique de l'école élémentaire peut être résolu en confectionnant les paquets par la pensée. Inutile de passer chez le confiseur ! Au fur et à mesure que nous remplissons nos sacs virtuels, nous utilisons 23, 46, 69, 92, 115, 138, 161 dragées. Nous nous arrêtons donc au

Euclide d'Alexandrie

On ne connaît presque rien de la vie d'Euclide (325—265 environ avant Jésus-Christ). Certains historiens en induisent qu'Euclide est un nom d'auteur choisi par un groupe de mathématiciens, une sorte de Bourbaki antique en quelque sorte. Pourquoi pas ? Cependant, ce type d'attitude correspond davantage à la période contemporaine qu'à l'antiquité. En guise de collaboration, l'époque préfèrait l'esclavage. Euclide serait alors le nom du maître. Ces diverses hypothèses sont affaire d'historiens car l'essentiel n'est pas là. L'œuvre qui porte le nom d'Euclide, *Les Éléments*, a traversé l'histoire. Elle est représentative des mathématiques grecques de l'époque, plus particulièrement de celles de l'école d'Alexandrie.

Euclide par Justus van Gent (1430–1480)

sixième paquet puisque nous ne pouvons remplir le septième. Notre stock imaginaire est épuisé ! Nous pouvons donc confectionner 6 sachets de dragées et il nous en reste $156 - 138 = 18$. Ce résultat s'écrit : $156 = 6 \times 23 + 18$; 6 est appelé quotient de 156 par 23 et 18, reste.

En suivant cette approche, nous démontrons de façon générale que, deux entiers a et b étant donnés (b strictement positif), il existe deux entiers q et r tels que :

$$a = b\,q + r \text{ et } 0 \leq r < b.$$

Ce résultat naturel dès que l'on veut partager a objets en parts égales à b a des conséquences importantes en arithmétique. Pour les analyser, nous commençons par étudier les questions de divisibilité.

Un nombre entier a est dit divisible par un nombre entier b si son reste dans la division par b est nul. *A priori*, cette définition n'a de sens que si b est strictement positif. On peut lui en donner un dans le cas où b est strictement négatif : a est divisible par b s'il est divisible par $-b$. De façon générale, a est donc divisible par b s'il existe un nombre entier q tel que $a = b\,q$.

Diviseurs d'un nombre

Chaque nombre entier a ainsi plusieurs diviseurs (au moins 1 et lui-même). Deux nombres entiers ont toujours des diviseurs en commun (1 au moins). L'ensemble des diviseurs communs à deux nombres entiers a et b a un plus grand élément appelé le *plus grand commun diviseur* (PGCD) de a et b. S'il est égal à 1, on dit que a et b sont premiers entre eux. Pour déterminer le PGCD de deux nombres, le plus simple est d'effectuer une suite de divisions euclidiennes. Voyons comment procéder sur l'exemple de 168 et 45. Nous divisons d'abord 168 par 45 ce qui donne : $168 = 3 \times 45 + 33$, donc tout diviseur commun à 168 et 45 divise aussi 33 (car on peut écrire : $33 = 168 - 3 \times 45$). Réciproquement, un diviseur commun à 45 et 33 divise $168 = 3 \times 45 + 33$. Nous en déduisons que les diviseurs communs à 168 et 45 sont les mêmes que les diviseurs communs à 45 et 33. Le PGCD de 168 et 45 est donc celui de 45 et 33, le reste de la division de 168 par 45.

Il est facile de voir que ce principe est général. Le raisonnement est donc applicable à 45 et 33 : $45 = 1 \times 33 + 12$ donc le PGCD cherché est celui de 33 et 12 ; $33 = 2 \times 12 + 9$ donc le PGCD cherché est celui de 12 et 9 ; $12 = 1 \times 9 + 3$ donc le PGCD cherché est celui de 9 et 3. Comme 3 divise 9, le PGCD de 168 et 45 est égal à 3.

L'algorithme d'Euclide que nous venons d'appliquer donne toujours le PGCD de deux nombres en un nombre fini d'étapes car la suite des restes des divisions effectuées dans ce cadre est strictement décroissante. Ainsi, l'un d'entre eux est nul ce qui implique que le précédent est le PGCD cherché.

Théorème de Bezout

Reprenons l'exemple précédent en écrivant les restes des divisions en partant de la dernière :

$$3 = 12 - 1 \times 9,$$
$$9 = 33 - 2 \times 12,$$
$$12 = 45 - 1 \times 33,$$
$$33 = 168 - 3 \times 45,$$

d'où nous déduisons, en écrivant tout en fonction de 168 et 45 :

$$3 = (45 - 1 \times (168 - 3 \times 45)) - 1 \times ((168 - 3 \times 45) - 2 \times (45 - 1 \times (168 - 3 \times 45)))$$

c'est-à-dire :

$$3 = 15 \times 15 - 4 \times 168.$$

Une conséquence de Bezout

Du théorème de Bezout, on déduit que les éléments inversibles de $\mathbb{Z}/N\mathbb{Z}$ sont les nombres premiers avec N. En effet, si x est premier avec N alors il existe deux entiers u et v tels que $ux + vN = 1$. Le reste de u dans la division euclidienne par N est un élément x' de $\mathbb{Z}/N\mathbb{Z}$ et $xx' = 1$. Tout nombre premier avec N est donc inversible dans $\mathbb{Z}/N\mathbb{Z}$. Réciproquement, si x est inversible dans $\mathbb{Z}/N\mathbb{Z}$ alors il existe x' tel que $xx' = 1$ ce qui signifie que $xx' - 1$ est divisible par N. Un diviseur commun de x et de N divise donc 1, x est donc premier avec N d'où le résultat.

Le raisonnement précédent est général, le résultat aussi. Il s'agit du théorème de Bezout :

Si a et b sont deux nombres entiers et d leur PGCD, il existe deux entiers u et v tels que $a \times u + b \times v = d$.

La façon la plus simple de trouver u et v est l'algorithme d'Euclide décrit précédemment. Parfois, il est cependant possible de le trouver directement comme dans le cas de 33 et 25 :

$$4 \times 25 - 3 \times 33 = 1.$$

Nombres premiers

Un nombre strictement supérieur à 1 est dit premier s'il n'est divisible que par 1 et lui-même. Il est facile d'établir le début de la liste des nombres premiers par la méthode du crible d'Ératosthène. Nous écrivons tous les nombres jusqu'à 100 par exemple, puis nous supprimons de cette liste tous les multiples de 2. Dans le tableau ci-dessous, nous les avons écrits en rouge pour simplifier. Le premier nombre strictement supérieur à 2 encore en noir est premier, c'est 3. Nous en supprimons tous les multiples en les écrivant en bleu. Nous continuons avec 5 puis 7. Tous les nombres en noir sont premiers et réciproquement.

	2	3	4	5	6	7	8	9	10
11	12	13	14	15	16	17	18	19	20
21	22	23	24	25	26	27	28	29	30
31	32	33	34	35	36	37	38	39	40
41	42	43	44	45	46	47	48	49	50
51	52	53	54	55	56	57	58	59	60
61	62	63	64	65	66	67	68	69	70
71	72	73	74	75	76	77	78	79	80
81	82	83	84	85	86	87	88	89	90
91	92	93	94	95	96	97	98	99	100

Crible d'Ératosthène pour déterminer les nombres premiers

Cela peut sembler très simple. Le problème se complique quand on veut déterminer si un très grand nombre est premier. Par exemple, pour déterminer avec cette méthode si le nombre 421 699 est premier, nous devons dresser un tableau de 421 699 nombres et éliminer les multiples de tous les nombres premiers inférieurs à 649. Cela se complique encore si le nombre a plusieurs dizaines voire plusieurs centaines de chiffres.

Factorisation

Tout nombre entier se factorise de manière unique en un produit de nombres premiers. Pour réaliser une telle factorisation, il suffit de diviser le nombre en question successivement par les nombres premiers inférieurs à sa racine carrée. Par exemple, pour factoriser 168, on commence par le diviser par 2 ce qui donne : $168 = 2 \times 84$ et on recommence avec 84 : $84 = 2 \times 42$, $42 = 2 \times 21$. Comme 21 n'est pas divisible par 2, on essaye 3 : $21 = 3 \times 7$. Comme 7 est premier, nous avons achevé la factorisation de 168 :

$$168 = 23 \times 3 \times 7.$$

Ce procédé est applicable à tous les nombres. On en déduit facilement une preuve du théorème :

Tout nombre entier se factorise de manière unique en un produit de nombres premiers.

Factoriser semble donc très simple mais quand les nombres sont grands, les calculs deviennent inextricables. Certaines méthodes modernes de cryptographie sont fondées sur cette difficulté (voir l'article *Le code RSA*). Leur inviolabilité ne tenant qu'à la difficulté de la factorisation, toute nouvelle méthode les met en péril au point que certains juges peu au fait de ce que sont les mathématiques ont assimilé « factoriser un nombre » et « frauder » si celui-ci est la clef d'une méthode de cryptographie (voir l'article *La carte qui dit « oui »*).

Les anneaux d'entiers

Donnons-nous un nombre N, 1 123 257 par exemple, et notons $\mathbb{Z}/N\mathbb{Z}$ l'ensemble des nombres entiers compris entre 0 et N – 1 muni des deux opérations suivantes. Si deux nombres x et y de $\mathbb{Z}/N\mathbb{Z}$ sont donnés, nous définissons leur somme et leur produit dans $\mathbb{Z}/N\mathbb{Z}$ en remplaçant le résultat usuel par son reste dans la division par N. Ainsi :

$$4\,635 \times 78\,934 = 365\,859\,090$$
$$= 325 \times 1\,123\,257 + 800\,565$$

donc dans $\mathbb{Z}/N\mathbb{Z}$:

$$4\,635 \times 78\,934 = 800\,565,$$

ce que l'on note souvent :

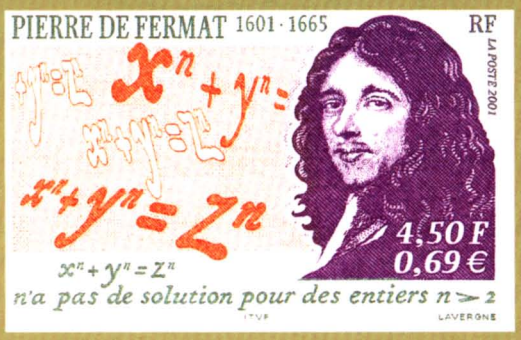
$$4\,635 \times 78\,934 \equiv 800\,565$$
$$\text{ou } 984\,635 \times 78\,934 \equiv 800\,565$$
$$\text{(modulo } 1\,123\,257)$$

pour éviter toute confusion.

L'ensemble $\mathbb{Z}/N\mathbb{Z}$ muni de ces deux opérations est appelé *anneau des entiers modulo* N (voir le glossaire) car les deux opérations ont les propriétés habituelles de l'addition et de la multiplication. Dans un tel anneau, les éléments inversibles sont les nombres premiers avec N. Il s'agit d'une conséquence du théorème de Bezout (voir l'encadré *Une conséquence de Bezout*). L'anneau $\mathbb{Z}/N\mathbb{Z}$ est un corps si tout élément non nul est inversible (voir le glossaire) donc $\mathbb{Z}/N\mathbb{Z}$ est un corps si et seulement si N est premier.

Le petit théorème de Fermat

Fermat est surtout connu pour ses recherches en arithmétique, en particulier pour le théorème cité sur le timbre ci-dessous. Cependant, il a également contribué à la naissance de la géométrie analytique, du calcul infinitésimal et de celui des probabilités.



Voici la preuve de son « petit théorème » :
Soient N un nombre premier et x un élément non nul de $\mathbb{Z}/N\mathbb{Z}$. Si deux des restes dans la division par N des nombres x, $2x$, ..., $(N-1)x$ étaient égaux, leur différence serait divisible par N ce qui conduit à une absurdité puisque x est premier avec N.
L'ensemble de ces restes coïncide donc avec l'ensemble des nombres 1, 2, ..., N – 1. Dans $\mathbb{Z}/N\mathbb{Z}$, la différence des deux produits est donc nulle. Autrement dit, $[x^{N-1} - 1]\,(N-1)!$ est divisible par N. Comme $(N-1)!$ est premier avec N, on en déduit que $x^{N-1} - 1$ est divisible par N d'où $x^N = x$ ce qui est encore vrai pour $x = 0$ donc pour tout élément x de $\mathbb{Z}/N\mathbb{Z}$.

Nous en déduisons que : si N est premier, $x^N = x$ pour tout élément x de $\mathbb{Z} / N\mathbb{Z}$. C'est le petit théorème de Fermat (nous en proposons une preuve directe, c'est-à-dire n'utilisant pas le fait que $\mathbb{Z} / N\mathbb{Z}$ est un corps, dans l'encadré *Le petit théorème de Fermat*). Ceci se traduit par $x^N \equiv x$ (mod N) pour tout entier x.
Ce théorème est à la base de la méthode de cryptographie RSA (voir l'article *Le code RSA*).

H. L.

GLOSSAIRE

Anneau : Ensemble A muni de deux lois notées, en général, + et ×. A est un groupe commutatif pour la loi +, la loi × est associative, distributive par rapport à la loi + et possède un élément neutre noté 1. Si la loi × est commutative, on dit que l'anneau A est commutatif. L'ensemble des nombres entiers relatifs $\mathbb{Z}$ et les ensembles des restes modulo N ($\mathbb{Z} / N\mathbb{Z}$) sont des anneaux commutatifs.

Associative : Une loi * sur un ensemble est dite associative si, pour tout (x, y, z), $x * (y * z) = (x * y) * z$. Pour une telle loi, l'usage des parenthèses est donc inutile.

Commutative : Une loi * sur un ensemble est dite commutative si, pour tout (x, y), $x * y = y * x$.

Corps : Anneau K dans lequel tout élément non nul est inversible pour la multiplication. Si la multiplication est commutative, on dit que K est commutatif. Exemple : l'ensemble des nombres rationnels est un corps, $\mathbb{Z} / N\mathbb{Z}$, est un corps si et seulement si N est premier.
Remarque : dans certains ouvrages, « corps » signifie « corps commutatif ».

Distributive : La loi * est distributive par rapport à la loi + c'est-à-dire que, pour tout (x, y, z), $x * (y + z) = x * y + x * z$. Dans les anneaux de nombres usuels, la multiplication est distributive par rapport à l'addition.

Groupe : Ensemble muni d'une loi associative, admettant un élément neutre et dans lequel tout élément admet un inverse. Les groupes ont été introduits par Évariste Galois (1832) à l'occasion de ses recherches sur la résolubilité des équations. Depuis, ils ont envahi l'ensemble des mathématiques.

Inverse : Une loi * munie d'un élément neutre e étant donnée sur un ensemble, un élément x' est dit inverse de x pour * si $x * x' = x' * x = e$. Un élément admettant un inverse est dit inversible. Les éléments inversibles de $\mathbb{Z} / N\mathbb{Z}$ sont les nombres premiers avec N. Il s'agit d'une conséquence du théorème de Bezout.

Neutre : Un élément e d'un ensemble est dit neutre pour la loi * si, pour tout x, $e * x = x * e = x$. S'il existe, un tel élément est unique. Dans les ensmbles de nombres usuels, 0 est neutre pour l'addition, 1 pour la multiplication.

Le code mystérieux de Marie-Antoinette

Voici un résultat de nature historique qui devrait clore un débat de plus de deux siècles sur la nature de l'affection que Marie-Antoinette portait au comte de Fersen : le décryptement de ses lettres est aujourd'hui complet… même si certains passages ont été remplacés par de petits points dans l'édition parue chez Paleo en 2004, comme s'ils étaient indécryptables. Ces parties ont été rétablies récemment par Jacques Patarin et Valérie Nachef, deux cryptologues, alors à l'université de Versailles-Saint-Quentin-en-Yvelines, en utilisant la clef utilisée pour le reste de la lettre. Il est étonnant que, de nos jours, un éditeur traite encore cette question comme un secret d'État ! Marie-Antoinette utilisait un chiffre connu depuis la Renaissance, celui de Vigenère, encore solide à son époque, mais en commettant une erreur de procédure majeure. Pour économiser son temps sans doute, elle ne chiffrait qu'une lettre sur deux, ce qui rendait son chiffre vulnérable à la méthode du mot probable.

La méthode du mot probable est sans doute la meilleure méthode classique de cryptanalyse, même si elle demande beaucoup d'imagination et d'intuition. Voyons comment elle fonctionne sur un exemple. Sachant que le message CEVOO EFTEU GLYSS AYMER ZEENZ OVNIT KE a été codé en ne chiffrant qu'une lettre sur deux et contient probablement le mot « Marie-Antoinette », nous faisons défiler ce terme le long du message jusqu'à trouver une coïncidence d'une lettre sur deux entre les deux textes. Nous la découvrons à la fin.

M E R Z E E N Z O V N I T K E
M A R I E A N T O I N E T T E

Découverte du mot probable.

Cela signifie que AIATIET est chiffré en EZEZVIK. Si nous savons de plus qu'on a affaire à un chiffre de Vigenère, c'est-à-dire à un décalage variable, nous comptons les décalages nécessaires pour cela, nous obtenons alors les lettres de la clef : EREGNER.

Clair	A I A T I E T
Chiffré	E Z E Z V I K
Décalage	4 17 4 6 13 4 17
Lettre	E R E G N E R

Détermination de la clef.

Celle-ci est donc probablement « REGNE » puisque les clefs de l'époque avaient un sens afin d'être retenues sans avoir besoin de les noter. Le message est alors facile à décrypter, il signifie : *« Le roi est au plus mal. Marie-Antoinette ».* Nous avons sans doute encore beaucoup à apprendre sur le côté historique des codes secrets.

Marie-Antoinete de Habsbourg (1755–1793). Portrait d'Élisabeth Vigée-Lebrun.

par J. -G. Dumas et D. Trystran

Les anniversaires des briseurs de codes

Les factorisations efficaces des nombres entiers font partie des armes des briseurs de codes. Une idée est d'utiliser le paradoxe des anniversaires.

Parmi 50 personnes, deux ont probablement le même anniversaire.

Dans une classe, quelle est la probabilité qu'au moins deux personnes aient le même anniversaire ? Si la classe compte k personnes, il semble logique que ce nombre soit égal à $k / 365$ environ, ce qui fait 1/16 pour une classe de 23 élèves. Cette intuition est totalement fausse ! Pour une classe de 23 élèves, la probabilité que deux élèves aient le même anniversaire est de 1/2, pour 53 ou plus, elle est de 99% !

Origine d'un paradoxe

D'où vient ce paradoxe ? L'explication tient dans une analyse plus fine des probabilités. Considérons l'événement contraire c'est-à-dire celui où toutes les personnes sont nées des jours différents. Pour déterminer une classe où tous les anniversaires sont différents, nous pouvons choisir l'anniversaire du premier ce qui fait 365 possibilités (366 les années bissextiles). Pour le second, nous n'avons plus que 364 possibilités, pour le troisième 363 et ainsi de suite. En ne tenant compte que des anniversaires, nous pouvons donc former $365 \times 364 \times \ldots \times (365 - k + 1)$ classes de k personnes ayant tous leurs anniversaires distincts. Comme il est possible de fabriquer 365^k classes distinctes, la probabilité que toutes les personnes soient nées des jours différents est donc égale à :

$$\frac{365 \times 364 \times \ldots \times (365 - k + 1)}{365^k}$$

donc la probabilité qu'au moins deux personnes aient le même anniversaire est :

$$1 - \frac{365 \times 364 \times \ldots \times (365 - k + 1)}{365^k}.$$

Pour $k = 23$ et $k = 53$, nous trouvons bien les probabilités annoncées.

Plus généralement, prenons n possibilités et k tirages aléatoires indépendants successifs. Le même calcul montre qu'il suffit que k soit égal à environ racine de $\sqrt{n}$ pour avoir une chance sur deux que deux tirages soient identiques (voir l'encadré *Un calcul asymptotique* pour le détail).

Ed Kienholz, *The Birthday*, 1964

Méthode de Monte-Carlo

Soit n un nombre non premier, p son plus petit facteur et q le facteur complémentaire ($n = pq$). Une méthode pour factoriser n est de tirer au hasard des nombres entre 0 et $n - 1$. Parmi ces nombres, q sont multiples de p ($0, p, 2p, \ldots, (q-1)p$), on a donc une chance sur p de tomber sur un multiple de p et donc de trouver p en calculant le plus grand commun diviseur de ce nombre et de n. En moyenne, cet algorithme nécessite donc p tirages pour trouver p. Comme le plus petit facteur premier de n est plus petit que sa racine carrée, en moyenne, cette méthode nécessite un nombre d'étapes de l'ordre de la racine carrée de n pour déterminer le plus petit facteur premier de n. Cette méthode est dite de Monte-Carlo car, de même que la fortune des casinos, elle est fondée sur l'exploitation du hasard.

Un calcul asymptotique

De même que dans le calcul des anniversaires, la probabilité que deux tirages de k nombres parmi n soient identiques est égale à 1 moins :

$$\frac{n\,(n-1)\ldots(n-k+1)}{n^k} = \frac{n\,!}{n^k\,(n-k)\,!}.$$

Cette quantité peut être simplifiée grâce à la formule de Stirling selon laquelle :

$$n\,! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n.$$

Pour $k = \sqrt{n}$, nous obtenons après implifications :

$\left(1-\frac{1}{\sqrt{n}}\right)^{\sqrt{n}-n-\frac{1}{2}} e^{-\sqrt{n}}$ dont le logarithme est égal à : $\left(\sqrt{n}-n-\frac{1}{2}\right)\ln\left(1-\frac{1}{\sqrt{n}}\right)-\sqrt{n}$.

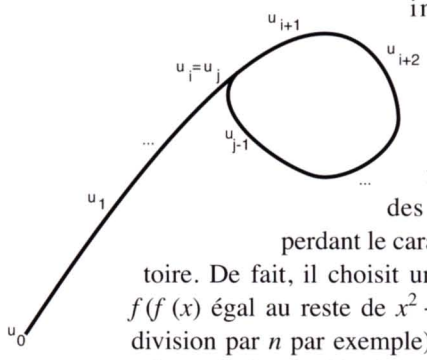
En utilisant un développement limité du logarithme, on en déduit que sa limite est égale à $-\frac{1}{2}$ donc celle de $\frac{n\,!}{n^k\,(n-k)}$ à $e^{\frac{1}{2}}$. Pour n assez grand, la probabilité que deux tirages soient identiques est donc égale à $1 - e^{\frac{1}{2}}$ qui est inférieur à $\frac{1}{2}$.

Amélioration par les anniversaires

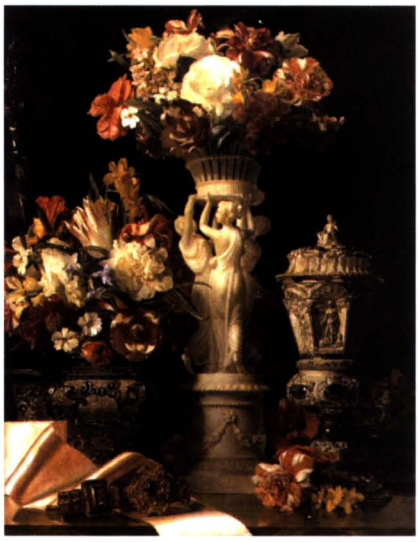
Le paradoxe des anniversaires permet d'améliorer cette méthode de Monte-Carlo. Tirons toujours des nombres au hasard entre 0 et $n-1$, mais en les conservant et en les comparant deux à deux. En racine de p tirages (en moyenne), on obtient deux nombres u et v ayant le même « anniversaire » c'est-à-dire tels que $u-v$ soit un multiple de p ! Il suffit de considérer le plus grand commun diviseur de n et de $u-v$ pour factoriser n. En racine quatrième de n tirages en moyenne, on a donc réussi à factoriser n. Le seul problème restant est le stockage en mémoire de tous ces u. Comme ce sont des très grands nombres, c'est impossible.

Algorithme de Pollard

L'idée Pollard est de ne stocker que certains d'entre eux, et de faire en sorte que, si un « anniversaire » se produit, il se répète à intervalles réguliers. Pour cela, il modifie la règle de formation des u en en perdant le caractère aléatoire. De fait, il choisit une fonction f ($f(x)$ égal au reste de x^2+1 dans la division par n par exemple) et un premier u de façon aléatoire, soit u_0. Les autres sont donnés en appliquant f successivement : $u_1 = f(u_0)$, $u_2 = f(u_1)$, etc. Si f est bien choisie, cette suite se comporte comme une suite aléatoire. Comme les u sont des nombres entre 0 et $n-1$, ils ne peuvent être tous distincts. Cette répétition forme alors un cycle et est représentée par un schéma en forme de ρ d'où son nom.



Le ρ de Pollard : il existe i et j tels que $u_i = u_j$. La règle de formation de la suite implique que les termes se reproduisent ensuite : $u_{i+1} = u_{j+1}$, etc.



F. Waldmüller, *The Birthday Table*, 1840

En pratique, cet algorithme factorise en quelques secondes les nombres d'une vingtaine de chiffres (les facteurs de 12 ou 13 chiffres nécessitent environ dix millions d'itérations !), mais il devient très rapidement inutilisable pour des facteurs plus grands.

Casser le code RSA

Le code RSA (voir l'article *Le code RSA*) est le système de codage à clef publique le plus connu. C'est aussi le plus utilisé, par exemple lors de transactions sécurisées sur Internet (pour la confidentialité du courrier ou l'authentification des utilisateurs). La clef publique correspond à un nombre $n = pq$. Pour le casser, il suffit de connaître p et q d'où l'intérêt de savoir factoriser de grands nombres. Le ρ de Pollard suffit pour factoriser des nombres de quelques dizaines de chiffres mais est insuffisant pour casser le code utilisé par la carte bleue mais il peut être utilisé pour des recherches de collisions dans les fonctions de hachage (voir l'article *Signature et hachage*).

J.-G.D. & D.T.

Les protocoles cryptographiques

Toutes sortes d'instruments présents dans notre vie quotidienne comme les terminaux de carte bleue, les cartes SIM de téléphones portables, les formulaires de paiement sur Internet, font appel à des protocoles informatiques de sécurité. Ceux-ci sont supposés garantir l'identité des interlocuteurs tout au long de l'échange, ainsi que la non interception, modification ou dégradation des messages. Dans ce domaine, des techniques nouvelles apparaissent, comme par exemple, celle basée sur la mécanique quantique.

par Véronique Cortier

Ces protocoles qui nous protègent

Terminaux de cartes bancaires, achats sur Internet, télévision cryptée, téléphonie mobile... la cryptographie est souvent utilisée à travers des protocoles dont la sécurité ne dépend pas uniquement de la méthode de chiffrement employée.

Avec le développement des télécommunications, de plus en plus d'informations circulent sur des réseaux accessibles à tous. Prenons l'exemple d'Internet. Il est possible d'y effectuer des achats en ligne en fournissant le numéro de sa carte de paiement, d'y consulter ses comptes bancaires ou de déclarer ses impôts. Ces informations circulent à travers de nombreux serveurs et routeurs avant d'arriver à leurs destinataires. Elles peuvent donc être lues, interceptées ou modifiées par des personnes mal intentionnées qui auraient détourné un des ces serveurs. Si les communications sensibles n'étaient pas protégées, un pirate pourrait facilement lire les numéros de cartes bancaires ou les codes secrets qui circulent sur Internet. La sécurisation des communications se fait à l'aide des protocoles cryptographiques. Il s'agit de petits programmes informatiques chargés de mettre en place des échanges de messages sécurisés. Pour un paiement en ligne par exemple, le navigateur passe du mode *http* au mode *https* et une icône représentant un cadenas fermé apparaît. Le mode *https* est une combinaison du mode *http* et d'un protocole nommé SSL qui assure trois notions de sécurité distinctes :

Inutile de forcer une porte blindée quand la fenêtre est ouverte !

1) la confidentialité : il est impossible d'espionner les données ;
2) l'intégrité : il est impossible de modifier le contenu des données ;
3) et l'authentification : le protocole s'assure de l'identité du destinataire, pour empêcher que l'utilisateur ne divulgue des données confidentielles à une fausse banque par exemple.

Utilité des protocoles

Les protocoles cryptographiques ne sont pas uniquement utilisés sur des ordinateurs connectés à Internet. Ils interviennent également dès que des transactions se font sur un réseau. Par exemple, ils sont utilisés depuis longtemps dans les décodeurs de chaînes de télévision. Il s'agit d'une part de décoder les données transmises par une chaîne cryptée mais également de mettre à jour l'abonnement : chaque mois, l'utilisateur peut s'abonner à de nouvelles chaînes ou au contraire résilier l'abonnement à certaines chaînes. À chaque fois, le client ne doit pouvoir déchiffrer que les chaînes auxquelles il a droit.

De la même manière, les protocoles cryptographiques sont au cœur de la téléphonie mobile. À chaque connexion, le téléphone doit s'authentifier auprès de son opérateur, c'est-à-dire prouver qu'il fait partie de ses clients. Chaque téléphone portable contient en effet une carte SIM qui comporte un numéro unique qui l'identifie, mais il ne suffit pas pour authentifier le téléphone. Un individu mal intentionné pourrait programmer une carte SIM de manière à ce qu'elle fournisse le numéro d'une autre carte, appartenant à un client de l'opérateur. L'authentification de la carte SIM est assurée par un protocole qui se déroule de la manière suivante. Chaque carte SIM valide (c'est-à-dire non fabriquée par un utilisateur frauduleux) partage un secret avec son opérateur. Pour qu'une carte prouve son identité, l'opérateur va lui envoyer un défi qui ne peut être résolu qu'à l'aide du secret partagé. C'est précisément un protocole cryptographique qui gère cet échange de messages entre le téléphone et l'opérateur.

Les protocoles cryptographiques sont utilisés de manière similaire dans les terminaux bancaires (lors d'un paiement au restaurant par exemple). Il s'agit à nouveau d'assurer d'une part l'authenticité de la carte, qui est assurée par un échange entre le terminal et la banque et d'autre part l'authenticité du propriétaire qui est assurée par la vérification du code secret.

Ils interviennent également dans des applications plus récentes et moins connues comme les porte-monnaie électroniques ou la sécurisation des données médicales. Pour les porte-monnaie électroniques, il s'agit d'empêcher la création de fausses cartes ou d'éviter qu'un utilisateur puisse créditer sa carte sans contrepartie. Pour les données médicales, de nombreux projets sont à l'étude pour que différents acteurs de santé (hôpitaux, laboratoires, médecins, infirmières) puissent partager des données concernant la santé de leurs patients tout en préservant la confidentialité des données. Un des problèmes réside dans le fait que chaque acteur ne doit avoir accès qu'à une partie seulement des données. Les médecins et infirmières ne devraient avoir accès qu'aux patients dont ils ont la charge ; la secrétaire d'un service hospitalier doit avoir accès à la liste des patients de son service sans connaître leurs pathologies et un épidémiologiste doit avoir accès à la liste des maladies diagnostiquées sans pouvoir connaître le nom des patients impliqués. D'autre part, les droits d'accès des utilisateurs doivent être mis à jour régulièrement en fonction des départs ou arrivées des patients ainsi que celles des soignants.

Un protocole et sa faille

Alice souhaite transmettre une clef secrète Kab à Bob. Pour cela, elle envoie la clef Kab à un serveur de confiance, chiffrée à l'aide d'une clef Kas, ce que nous notons {Kab}Kas. Le message envoyé au serveur prend la forme :

`Alice, Bob, {Kab}Kas`

La clef Kas est partagée avec le serveur qui retrouve donc la clef Kab et la crypte avec Kbs, une clef qu'il partage avec Bob. Celui-ci récupère ainsi la clef Kab.

Ce protocole peut sembler parfait, il comporte cependant une faille car, sans connaître aucune des clefs, un attaquant peut modifier les identités en entête du message. Ainsi, si un individu mal intentionné, traditionnellement appelé Charlie, possède également une clef Kcs partagée avec le serveur, il peut intercepter le message d'Alice destiné à Bob :

`Alice, Bob, {Kab}Kas`

et envoyer le message :

`Alice, Charlie, {Kab}Kas`

à la place. Le serveur croit alors qu'Alice souhaite transmettre le message à Charlie et envoie le message :

`Alice, Charlie, {Kab}Kcs`

à Charlie, qui peut alors découvrir la clef Kab qui devait rester secrète entre Alice et Bob ! Pour cela, Charlie n'a pas eu besoin de décoder le message {Kab}Kas mais a tout simplement utilisé une faille logique du protocole. Aussi, dans le protocole original (Wide Mouthed Frog Protocol), l'identité des agents est à l'intérieur des messages. Les messages contiennent également la date d'émission de la clef pour éviter qu'un individu mal intentionné puisse réutiliser une clef vieille de plusieurs mois. Ce protocole reste cependant sujet à des attaques plus subtiles.

Le chiffrement

Une brique essentielle dans la conception de ces protocoles est la cryptographie et en particulier le chiffrement. Un premier type de chiffrement est le chiffrement symétrique : la même clef est utilisée pour chiffrer et déchiffrer les messages. On pourrait penser que la principale difficulté consiste à trouver un chiffrement sûr. C'est en fait loin d'être suffisant. En effet, supposons que deux personnes, Alice et Bob, souhaitent s'échanger un secret S. Alice peut envoyer à Bob le secret S chiffré avec une clef secrète k à l'aide d'un algorithme de chiffrement symétrique comme le triple DES ou AES. Pour déchiffrer le message, Bob doit avoir la clef. Il faudrait donc pour cela qu'Alice ait préalablement communiqué la clef secrète k à Bob.

On peut imaginer partager à l'avance un certain nombre de clefs avec des personnes ou entités que l'on connaît. Mais lorsqu'on effectue une transaction sur un site marchand pour la première fois, on n'a en général aucune information préalable sur le site en question. Comment peuvent donc faire Alice et Bob pour s'échanger une clef secrète s'ils ne se connaissent pas ? Une première solution consiste à envoyer la

clef en clair puis l'utiliser. Mais bien sûr, des individus mal intentionnés pourraient alors voir la clef, l'utiliser pour déchiffrer le message et découvrir le secret S. Il faut donc envoyer la clef k elle-même chiffrée par une autre clef k' mais on rencontre à nouveau le même problème : comment s'échanger une clef de chiffrement ?

Une possibilité est de passer par un serveur intermédiaire, appelé tiers de confiance, avec lequel chacun des acteurs (Alice et Bob) partage une clef secrète. Un protocole fonctionnant sur ce principe est présenté dans l'encadré *Un protocole et sa faille*. Cette solution ne peut cependant pas toujours être appliquée car elle suppose qu'un serveur partage des secrets avec tous ses utilisateurs potentiels. S'ils sont très nombreux, cela demande des capacités de stockage trop importantes. D'autre part, il faut réellement avoir confiance dans ce serveur et donc s'assurer qu'il ne peut pas lui-même être attaqué.

Pour échanger une clef secrète sans faire appel à un serveur intermédiaire, on a souvent recours à un autre type de chiffrement : le chiffrement asymétrique ou à clefs publiques comme RSA : les clefs de chiffrement et déchiffrement sont distinctes et la clef de chiffrement peut être rendue publique sans compromettre la sécurité du chiffrement. Autrement dit, même en connaissant le message chiffré et la clef de chiffrement, on ne peut rien déduire sur le message. Ainsi, si Alice souhaite échanger une clef secrète avec Bob, elle peut tout simplement lui envoyer la clef chiffrée par la clef publique de Bob. Seul Bob peut alors déchiffrer le message.

Si ce principe est simple et couramment utilisé, il faut encore se prémunir contre deux grandes familles d'attaques. D'une part, comme la clef de Bob est publique, n'importe qui peut l'utiliser pour envoyer des messages à Bob. En particulier, une personne mal intentionnée peut envoyer une clef secrète en prétendant être Alice. Il faut donc prévoir une première phase d'authentification. D'autre part, les clefs publiques sont distribuées au travers d'un annuaire électronique ou encore d'une page Web. Mais rien ne garantit que la clef indiquée dans l'annuaire appartient bien à l'utilisateur correspondant. Si un attaquant réussit à corrompre l'annuaire en indiquant sa propre clef publique à la place de celle de Bob, il pourra déchiffrer des messages destinés à Bob. Les clefs sont donc souvent accompagnées d'un certificat qui prouve l'identité de l'utilisateur associé.

Sécurité des protocoles

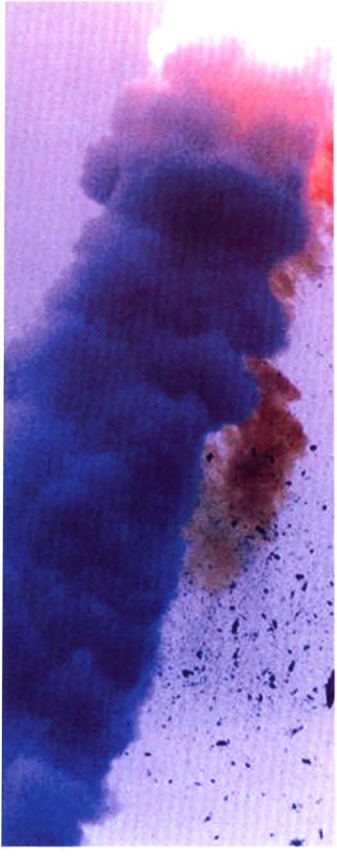
La sécurité des protocoles repose bien sûr sur la fiabilité des algorithmes de chiffrement utilisés mais également sur la fiabilité de la conception du protocole elle-même. En effet, il arrive régulièrement qu'un protocole puisse être attaqué sans casser le chiffrement : inutile de forcer une porte blindée quand la fenêtre est ouverte ! Un exemple d'attaque est expliqué en encadré. L'analyse de la sécurité des protocoles fait donc l'objet de nombreuses recherches en informatique fondamentale à l'heure actuelle.

V .C.

par Véronique Cortier

La vérification des protocoles

Les protocoles cryptographiques peuvent avoir des failles indépendantes de la méthode de chiffrement utilisée. Pour vérifier qu'ils en sont exempts, on a recours à des méthodes de logique mathématique.



Les protocoles cryptographiques sont des programmes utilisés à grande échelle. La moindre faille peut avoir des répercussions très importantes. Il faut donc s'assurer de manière rigoureuse de leur fiabilité. Lorsque l'on conçoit un protocole, on commence en général par vérifier qu'il n'est pas sujet aux attaques déjà existantes. On teste ensuite une série de comportements possibles. Mais si aucune attaque n'est détectée au cours de ce processus, cela ne signifie pas pour autant que le protocole soit infaillible. Des comportements ont pu être oubliés. Pour se convaincre de la fiabilité d'un protocole, il faut donc démontrer que parmi tous les comportements possibles des agents et des attaquants potentiels, aucun ne conduit à une faille. Il est impossible d'énumérer toutes les actions possibles car il y en a une infinité. Pour obtenir une preuve de sécurité, on appelle alors les mathématiques et la logique à la rescousse. Cette démarche se situe plus généralement dans le cadre de la vérification de programmes (voir l'encadré *La vérification de programmes*).

La vérification des protocoles est une question de logique.

Une modélisation logique

Pour analyser rigoureusement les protocoles, la première étape consiste à les modéliser, c'est-à-dire à les formaliser à l'aide d'objets mathématiques que l'on sait manipuler. Ainsi, les messages sont représentés par des objets appelés termes. Supposons par exemple que la suite de bits 011011010 ait été obtenue en chiffrant l'identité d'Alice par la clef publique de Bob. Le message 011011010 sera représenté par le terme [A]pub(B). Pour le chiffrement symétrique, on utilise le symbole { } à la place de [ ]. Pour représenter les échanges de messages effectués lors du protocole, on utilise des formules logiques. La formule atomique R(m) signifie que le message m circule

sur le réseau. Dans ce qui suit, le message m est A,B,{Kab}Kas ce qui signifie que son expéditeur est A, son destinataire B et que le corps du message est {Kab}Kas c'est-à-dire Kab codé avec la clef symétrique Kas.

Ainsi, pour représenter l'action du serveur dans le protocole inspiré du Wide-Mouthed-Frog protocol (voir l'encadré dans l'article *Ces protocoles qui nous protègent*), on utilise la formule :

$$R(A,B,\{Kab\}Kas) \Rightarrow R(A,B,\{Kab\}Kbs).$$

Cela signifie que si un message de la forme A,B,{Kab}Kas est envoyé sur le réseau, le serveur répondra par le message A,B,{Kab}Kbs. On utilise la variable Kab pour rendre compte qu'il s'agit d'une valeur inconnue du serveur : quelle que soit la valeur prise par Kab, le serveur répondra de la même manière.

Modélisation d'une attaque

Il reste enfin à représenter les actions possibles pour les attaquants. On utilise à nouveau des formules logiques.

$$R(x), R(y) \Rightarrow R(\{x\}y)$$

Un attaquant peut chiffrer un message qu'il connaît à l'aide d'une clef qu'il connaît.

$$R(\{x\}y), R(y) \Rightarrow R(x)$$

Si un attaquant intercepte un message de la forme $\{x\}y$ et qu'il possède la clef y de déchiffrement, il peut en déduire le message x en clair.

$$R(x), R(y) \Rightarrow R(x, y)$$

L'attaquant peut concaténer des messages.

Preuve de la sécurité d'un protocole

À partir de ce paquet de formules mathématiques servant d'axiomes, on peut raisonner mathématiquement sur le protocole et démontrer (le cas échéant !) qu'il n'existe pas d'attaque en prouvant par exemple que la formule R(Kab) est fausse, c'est-à-dire la clef Kab ne circulera jamais en clair sur le réseau (même si le protocole est attaqué). Les premières preuves de sécurité furent effectuées à la main. Cependant, elles sont relativement longues et fastidieuses car il y a de nombreux cas à examiner. Aussi, il n'est pas simple pour un être humain de vérifier si elles sont justes. D'autre part, les protocoles sont souvent implémentés sous de nombreuses variantes. Il faut donc refaire la preuve de sécurité pour chacune des variantes. Une autre approche consiste à construire des algorithmes génériques qui permettent de faire la preuve qu'un protocole quelconque vérifie ou non une propriété de sécurité. Plusieurs prototypes de logiciels ont ainsi été développés pour rechercher automatiquement (c'est-à-dire sans intervention humaine) des attaques ou faire la démonstration qu'il n'y a pas de faille. Ces prototypes ont permis de découvrir de nouvelles attaques, parfois sur des protocoles que l'on croyait sûrs depuis une quinzaine d'années.

Pour revenir à notre exemple, une façon de démontrer que la clef Kab sera (ou ne sera pas) connue d'un attaquant potentiel est d'ajouter petit à petit toutes les formules qui sont des conséquences logiques des formules initiales et observer si la formule R(Kab) en fait partie. Ainsi, la formule $R(x), R(y), R(z) \Rightarrow R(x, \{y\}z)$ est une conséquence des autres : si un atta-

La vérification de programmes

Les logiciels sont désormais utilisés dans de très nombreuses applications, de l'ordinateur familial à la machine à café. Si l'on peut accepter que le système d'exploitation de son ordinateur « plante » régulièrement, il est crucial que les logiciels embarqués à bord des voitures, des avions de ligne ou des fusées par exemple soient fiables. Ainsi, l'explosion de la fusée Ariane V lors de son vol inaugural le 4 juin 1996 n'était pas due à un problème de matériel mais à un *bug* logiciel : la vitesse d'un capteur a dépassé la taille allouée et a causé une erreur imprévue qui a conduit à la perte du contrôle du système de guidage. De manière similaire, la perte de la sonde Mars Climate Orbiter le 26 septembre 1999 a été causée par une erreur de conversion entre *miles* et kilomètres !

Lors du développement d'un logiciel, on procède bien sûr à de nombreux tests. C'est la façon la plus naturelle de détecter des erreurs. Cela ne garantit cependant pas que le programme n'a aucune faille car des problèmes peuvent apparaître seulement pour des séquences d'actions que l'on n'aurait pas explorées. Supposons par exemple que l'on veuille montrer que l'entier n (n + 1) est toujours pair, pour n entier naturel. On peut bien sûr vérifier que 1 (1 + 1) = 2, 2 (2 + 1) = 6, 3 (3 + 1) = 12 sont pairs mais cela ne garantit pas que la propriété soit vraie pour n entier quelconque, même si l'on procède à un très grand nombre de tests. Pour cela, il faut démontrer la propriété à l'aide d'un raisonnement mathématique. De la même manière, pour s'assurer qu'un programme est fiable, il faut démontrer qu'il vérifie les propriétés souhaitées pour toutes les exécutions possibles.

L'industrie est donc de plus en plus sensibles au besoin de vérifier leurs logiciels critiques, c'est-à-dire d'obtenir une preuve mathématique que toutes les exécutions possibles du logiciel vérifieront les propriétés demandées. Ainsi, l'équipe de Patrick Cousot, professeur à l'École normale supérieure (ENS) a établi que « *la preuve mathématique qu'un programme comme celui de la commande de vol de l'A380 ne comprend pas de bogues liés à la limite des capacités de l'ordinateur* » (*Le Monde*, N° 18 741, 27 avril 2005, page 18). Ce programme comportait pourtant environ 500 000 lignes de code.

quant observe les messages x et y sur le réseau, s'il connaît également la clef z, il peut envoyer le message x, $\{y\}z$ sur le réseau. L'inconvénient de cette méthode est qu'il y a une infinité de formules logiques, conséquences des formules initiales. Pour s'en sortir, on développe des stratégies pour d'une part n'ajouter qu'un nombre fini de formules et d'autre part, être sûr de trouver la formule R(Kab) si une attaque est possible.

Une faiblesse de la méthode

Une faiblesse de ces approches pour analyser les protocoles réside, de par la modélisation retenue, sur le fait que seules les attaques logiques peuvent être détectées, c'est-à-dire les attaques qui n'exploitent pas des faiblesses potentielles des algorithmes de chiffrement utilisés. Pourtant, il se pourrait que leurs propriétés mathématiques, combinées au protocole, conduisent à des attaques. Aussi, des travaux de recherche récents cherchent à développer l'approche logique en prenant de mieux en mieux compte les propriétés mathématiques des fonctions utilisées. D'autres éléments encore peuvent être source d'attaques. Des erreurs peuvent par exemple provenir de la programmation du protocole ou de son installation par l'utilisateur. D'autre part, le protocole est en général utilisé dans un environnement où d'autres protocoles fonctionnent simultanément. L'interaction entre les protocoles a rarement été étudiée, d'autant que chaque protocole peut être modifié ou supprimé en fonction des besoins des utilisateurs et des mises à jour. Analyser un protocole dans sa globalité reste donc un véritable défi.

V. C.

Divers protocoles

couramment utilisés en informatique

• SSL et TLS

Les protocoles SSL (Secure Socket Layer) ou TLS (Transport Layer Security), successeur de SSL, sont utilisés sur internet, en combinaison avec http, pour former le mode https. Mais ils sont également combinés à d'autres protocoles comme ftp (pour le transfert de fichiers) ou imap (pour le transfert de mails). Ils visent à établir un canal sécurisé après une phase d'authentification : la phase d'authentification s'effectue entre le client et le serveur qui s'échangent ensuite une clef de session à l'aide d'un chiffrement à clef publique (comme RSA). La suite des échanges est chiffrée (chiffrement symétrique) à l'aide de cette clef de session.

• SSH

Le protocole SSH (Secure Shell) permet à un client d'ouvrir une session sur un ordinateur distant en évitant que le mot de passe soit transmis en clair. Il est utilisé par exemple quand un employé se connecte à son entreprise depuis chez lui. La première étape du protocole consiste en une phase d'authentification mutuelle du client et du serveur au cours de laquelle une clef de session est échangée. Le mode de chiffrement utilisé durant cette phase est préalablement négocié entre le serveur et le client. En effet, SSH permet une grande souplesse sur les algorithmes de chiffrement utilisés. La clef de session sert ensuite à établir un canal sécurisé que le client utilise pour transmettre son mot de passe et s'authentifier auprès du serveur.

• KERBEROS

Le protocole Kerberos est un protocole d'authentification et de délivrement de tickets, permettant à des utilisateurs distants d'accéder à des services. Il repose sur du chiffrement à clefs symétriques comme DES et nécessite la présence d'un serveur de confiance (serveur Kerberos). Le serveur maintient une base de données contenant les clefs secrètes de tous les utilisateurs. Le fonctionnement de Kerberos repose sur la notion de « tickets ». Pour accéder à un service, un client doit contacter le serveur qui lui envoie un ticket chiffré avec la clef du client. Ce premier ticket contient une clef de session et un deuxième ticket. Le client envoie alors le deuxième ticket chiffré avec la clef de session au serveur de tickets qui lui donne un troisième ticket pour obtenir l'accès au service demandé.

• CHAP

CHAP (Challenge-Handshake Authentication Protocol) est un protocole d'authentification couramment utilisé lors de la connexion d'un utilisateur à son fournisseur d'accès Internet. L'utilisateur et le fournisseur d'accès partagent un secret comme le mot de passe de l'utilisateur. Le fournisseur procède à l'authentification de son client à l'aide d'un « défi » (challenge) de la manière suivante : il envoie un nombre aléatoire de 16 bits et le client doit répondre par le résultat d'un calcul faisant intervenir ce nombre et son mot de passe. Le fournisseur d'accès vérifie si le résultat est correct. D'autres défis peuvent être envoyés ultérieurement si le fournisseur d'accès souhaite vérifier que la personne connectée est toujours son client.

• EAP

EAP (Extensible Authentication Protocol) est un mécanisme d'authentification fréquemment utilisé sur les réseaux sans fil. Il ne s'agit pas d'un protocole d'authentification à proprement parler mais d'un cadre pour plusieurs méthodes d'authentification possibles. EAP permet de négocier le mode d'authentification choisi entre les participants et fournit également plusieurs fonctions communes aux différentes méthodes d'authentification.

par J.-G. Dumas et J.-L. Roch

Signature électronique et hachage

Un document signé ne doit pouvoir être modifié sans le consentement de son signataire. Pour le garantir, la fabrication des signatures électroniques passe par la confection de résumés électroniques des messages. Pour cela, on utilise des fonctions de hachage.

Par définition, une signature est une apposition validant une pièce comptable ou un acte, ou affirmant l'exactitude, la sincérité ou encore la responsabilité d'un écrit. Jusqu'en 2002, en France, seule l'apposition de son nom manuscrit pouvait faire foi. Dorénavant, il est également possible d'apposer une signature électronique ! Nous allons voir comment de telles signatures sont réalisables et quelles sont les contraintes techniques inhérentes à leurs confections. L'article *Signature et authentification* en montre une application importante.

Intégrité et non-répudiation

La première fonction que doit réaliser une signature est l'intégrité. C'est-à-dire que le contenu d'un document signé ne doit pas pouvoir être modifié sans le consentement de son signataire. En outre, celui-ci ne doit pas pouvoir ensuite renier sa signature : c'est la non-répudiation. Pour cela il est nécessaire d'établir des résumés électroniques servant de preuve de non modification. L'élément technique utilisé est une fonction de hachage, c'est à dire une application H qui transforme une chaîne de caractères, c'est-à-dire une séquence M de bits (0 ou 1), de longueur arbitraire en un résumé $R = H(M)$, c'est-à-dire une séquence de bits de longueur fixe donnée à l'avance n.

Pour garantir l'intégrité, il faut que toute modification du message originel M engendre une modification complète de son résumé. Les fonctions de hachage généralement utilisées sont de plus uniformes, c'est-à-dire qu'elles garantissent l'équiprobabilité des résumés possibles : pour toute séquence de n bits R, la probabilité que H (M) soit égal à R est égale à $1/2^n$. Enfin, au niveau de la sécurité, il faut garantir la difficulté de retrouver un texte initial à partir de son seul résumé. Nous verrons que, dans le cas contraire, cela peut entraîner la fal-

sification de signatures. En particulier les fonctions de hachage sont classées suivant trois niveaux de sécurité donnés ici dans l'ordre du plus résistant au moins résistant :

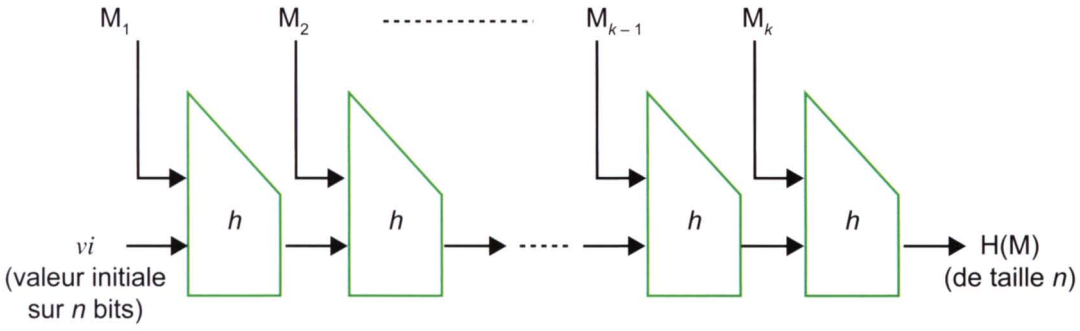
– Résistance à la préimage :
Ayant R, il est difficile (c'est-à-dire extrêmement coûteux) de trouver M tel que H (M) = R.

– Résistance à la seconde préimage :
Ayant M, il est difficile de trouver M' tel que H (M) = H (M').

– Résistance aux collisions :
Il est difficile de trouver M et M' tel que H (M) = H (M').

En pratique, comment réaliser de telles fonctions de hachage ? Partant d'une fonction de hachage restreinte *h* qui fournit un résumé sûr de taille *n* d'un message de taille $2n$, Merkle et Damgård ont proposé une construction très simple d'une fonction de hachage H sûre de taille quelconque. Il suffit d'itérer les résumés fournis par *h* comme décrit dans la figure ci-dessous.

Ainsi, Merkle et Damgård ont montré que si *h* est résistante aux collisions alors l'itération H l'est aussi ! De nombreuses variantes existent alors pour fabriquer des fonctions de hachage. Par exemple en modifiant légèrement une fonction de cryptage, on obtient une boîte *h* qui correspond à nos attentes : en effet un cryptage par bloc (avec votre système préféré) de 128 bits par exemple, combine une clef secrète de taille 128 bits à un bloc de message de 128 bits pour sortir un bloc crypté de 128 bits. Il suffit alors d'introduire la clef secrète en valeur initiale de l'itération ci-dessus puis d'itérer sur tous les blocs du message. On obtient une fonction de hachage très résistante. Cependant, souvent les fonctions de cryptage sont un peu longues à calculer ; aussi des standards de fonctions de hachage sont donc apparus, tels MD5 (128 bits) ou SHA-1 (160 bits), qui proposent des fonctions de cryptages simplifiées à itérer. Malheureusement les nombres de bits de ces dernières ne sont pas suffisants ! Revenons en effet un instant sur la notion de résistance. En pratique, quelle est la difficulté de trouver des collisions ? Par la force



Construction de Merkle-Damgård. Le message M est découpé en blocs de tailles *n* puis on résume succesivement des chaînes de $2n$ bits en partant d'une chaîne arbitraire *vi*.

Algorithme de Yuval

Les entrées de l'algorithme de Yuval sont :
un message légitime M_1 (un contrat par exemple) ; un message frauduleux M_2 (un contrat truqué) ; une fonction de hachage h sur n bits.
Ses sorties sont :
M_1' très proche de M_1 et M_2' très proche de M_2 tels que $h(M_1') = h(M_2')$

Voici les étapes à suivre pour passer des entrées aux sorties :

1. Générer $\sqrt{2^n}$ modifications mineures de M_1, notées M_1'
2. Pour chacune, calculer $h(M_1')$
3. Générer des M_2', modifications mineures de M_2, jusqu'à collision avec un M_1'.

brute, combien de messages doit-on essayer pour trouver une collision ? Réponse : entre 2 et $2^n + 1$! La sécurité viendrait donc du fait que 2^n est trop grand. Ainsi, avec $n = 128$, réaliser 2^{128} opérations reste aujourd'hui encore totalement impossible : il faudrait un milliard d'ordinateurs travaillant plus de mille fois l'âge de l'univers pour simplement compter jusqu'à 2^{128}.

L'attaque de Yuval

Cependant, il est possible d'utiliser le paradoxe des anniversaires (voir l'article *Les anniversaires des briseurs de codes*) pour réduire ce nombre de manière drastique. L'attaque de Yuval (voir l'encadré L'*algorithme de Yuval*) permet de trouver des collisions en seulement racine carrée du temps du nombre précédent. Le paradoxe des anniversaires assure en effet que le temps de calcul de cet algorithme n'est en moyenne que de l'ordre de $\sqrt{2^n}$. Une application simple est alors d'envoyer M_1' et de soutenir plus tard, qu'en fait M_2' avait été envoyé, le résumé étant la preuve de notre bonne foi. Grâce au paradoxe des anniversaires, cet algorithme ne nécessite que de l'ordre de $\sqrt{2^n}$ opérations. Si l'on a choisi MD5 comme fonction de hachage, il suffit alors de l'ordre de 2^{64} opérations pour trouver une collision, ce qui est tout à fait faisable : un ordinateur à un 1GHz permet de faire $10^9 = 2^{30}$ opérations par seconde ; donc, en trois semaines, dix mille ordinateurs peuvent fabriquer de faux contrats. En pratique, en utilisant des astuces supplémentaires, Xiaoyun Wang et Hongbo Yu de l'université de Shandong en Chine ont réussi à encore diviser ce nombre et trouver en 2004 des collisions sur MD5 en quelques heures avec une grappe de PCs. Par

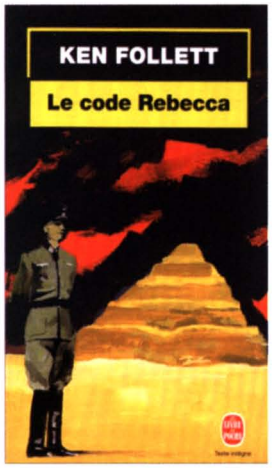
ailleurs, plusieurs astuces techniques ont récemment permis de trouver des collisions non seulement sur des fonctions de 128 bits comme MD5, mais également sur des fonctions supposées plus résistantes, comme SHA-1 (160 bits, base de la sécurité de la X-box par exemple). Ainsi de nouveaux standards sur 256 bits au moins émergent, comme par exemple SHA-256. Pour ces standards, même l'attaque de Yuval ne peut encore rien !

En travaillant trois semaines, dix mille ordinateurs peuvent fabriquer un faux contrat.

J.-G.D & J.-L.R.

Le code Rebecca

Dans son roman *le Code Rebecca*, Ken Follett a imaginé une amélioration du code de Vigenère proche de celui de Vernam.



Guerre, espions et contre-espions

Dans l'éternelle quête du Graal que constituerait le système parfait de cryptage, celui que nul ne parviendrait à percer, Ken Follett, dans *le Code Rebecca*, a presque trouvé la solution infaillible. Presque… L'écrivain britannique renommé pour ses romans d'espionnage a conçu dans cet ouvrage une aventure d'espionnage. Elle se déroule au Caire en 1942, pendant la Seconde Guerre mondiale, au moment où Rommel et les Anglais se livrent une guerre sans merci. Rommel vient de prendre Tobrouk et croit dur comme fer que la conquête de l'Égypte n'est plus qu'une question de jours. Alex Wolf, super espion nazi, va infiltrer les services du contre-espionnage britannique pour leur voler des renseignements, et les envoyer codés à l'aide d'un émetteur à l'état major de l'Afrika Korps.

Coder avec un livre

Voici comment l'auteur a imaginé ce système de codage :

*« Wolf s'approcha du buffet où il dissimulait l'émetteur radio. Il prit le roman anglais (*Rebecca *de Daphné du Maurier) et la feuille de papier sur laquelle était inscrit le chiffre du code. Il l'étudia. On était aujourd'hui le 28 mai. Il fallait ajouter 42 – le chiffre de l'année – à 28 pour arriver au numéro de la page du roman qu'il devait utiliser pour coder son message. Mai était le cinquième mois de l'année, aussi fallait-il supprimer une lettre sur cinq dans la page* [...]

Il décida d'envoyer comme message SUIS ARRIVE, M'INSTALLE, ACCUSEZ RECEPTION [...]

Commençant en haut de la page 70 du livre, il chercha la lettre S. En supprimant une lettre sur cinq, le S était le dixième caractère de la page. Dans son code il serait donc représenté par la dixième lettre de l'alphabet le J. Il lui fallait ensuite un U. Dans le livre la troisième lettre après le S était un U. Le U de suis serait donc représenté par la troisième lettre de l'alphabet, le C. Il y avait des façons particulières pour représenter les lettres rares comme le X, par exemple. [...]

Ce type de code était une variation unique de bloc, la seule forme de code indéchiffrable en théorie comme en pratique. Pour décoder le message, il fallait avoir tout à la fois le livre et la clef. »

Un système inviolable ?

Ken Follet a raison, ce système est indécryptable, mais il y a tout de même une faille d'ordre matériel, le chiffreur et le déchiffreur doivent garder la clef et le livre à portée de main donc susceptibles d'être découverts par le camp adverse. Voilà qui ne fait pas mentir notre ami Edgar Allan Poe dans *le Scarabée d'or* : *« Il est vraiment douteux que l'ingéniosité humaine puisse créer une énigme de ce genre dont l'ingéniosité humaine ne vienne à bout par une application suffisante. »*

Le code Vernam

Le code imaginé par Ken Follet peut être comparé au chiffre de Vernam utilisé, entre autres, par le Téléphone rouge. La faiblesse reste la même : la transmission de la clef (voir l'article *Du code de Vigenère à celui de Vernam*).

par David Delaunay

Signature et authentification

Les méthodes de cryptographie à clefs publiques donnent naissance à des protocoles de signature et d'authentification. Ceux-ci permettent de s'assurer de l'intégrité du message transmis, de l'identité, de la fiabilité de l'expéditeur, etc.

D'un point de vue mathématique, chiffrer un message M consiste à lui appliquer une fonction f appelée clef de chiffrement et de former ainsi un message crypté $M' = f(M)$. Pour éviter toute ambiguïté, cette fonction f ne doit pas prendre deux fois la même valeur (voir l'article *Bijectivité et codage*). Ainsi, on peut introduire sa fonction réciproque, notée ici g et définie par :

$M' = f(M)$ si et seulement si $M = g(M')$.

Cette fonction est appelée clef de déchiffrement car elle permet de décrypter le message M' par le calcul :

$$g(M') = g(f(M)) = M.$$

Pour certains cryptosystèmes, connaître f suffit pour déterminer aisément g, c'est le cas des chiffrements par substitution ou permutation. En revanche, il en existe pour lesquels la connaissance de la clef de chiffrement ne permet pas techniquement de déterminer la clef de déchiffrement. C'est le cas du cryptosystème RSA (voir l'article *Le code RSA*). Pour ce système, le concepteur du code détermine à la fois f et g, mais une personne étrangère à la conception du code ne peut déduire une fonction de l'autre car cela la conduirait à des calculs infaisables en pratique. Cette dissymétrie entre les deux clefs f et g permet au créateur du code de diffuser l'une d'elle, par le biais d'un annuaire par exemple. Cette clef est alors appelée clef publique tandis de l'autre est appelée clef privée.

À l'aide d'une clef publique, quiconque peut chiffrer un message, seul le concepteur du code saura le déchiffrer. À l'inverse, à l'aide de sa clef privée, seul le concepteur du code saura chiffrer un message que quiconque pourra déchiffrer. Ce dernier principe, apparemment sans intérêt, est utilisé dans les protocoles décrits dans cet article. Nous convenons d'appeler Alice la conceptrice du code, f sa clef privée, g sa clef publique et Bob le qui-

dam souhaitant communiquer avec Alice.

Signature électronique

Alice transmet un message à Bob. Ce dernier veut s'assurer qu'elle en est bien l'expéditrice et que son message n'a pas été corrompu. De plus, il souhaite qu'elle ne puisse nier l'avoir envoyé. Pour cela, Alice accompagne son message d'une signature électronique. Celle-ci est réalisée en deux temps :

– Alice forme un résumé $R = h(M)$ de son message par le biais d'une fonction de hachage comme par exemple le MD5 (voir l'encadré *La fonction de hachage MD5*),

– Alice crypte alors le résumé par le biais de sa clef secrète pour former $S = f(R)$, la signature du message transmis.

À la réception du message M accompagné de sa signature S, Bob calcule son résumé R ainsi que $g(S)$ à l'aide de la clef de chiffrement publique d'Alice. Il lui suffit alors de vérifier si $g(S) = R$ pour s'assurer qu'elle est bien l'expéditrice du message. Par ce mécanisme, un intermédiaire mal intentionné ne peut pas se substituer à elle et l'intégrité du message est assurée. Notons que cette signature est ici plus sûre qu'une signature papier puisqu'elle est fonction du document signé.

Authentification par défi

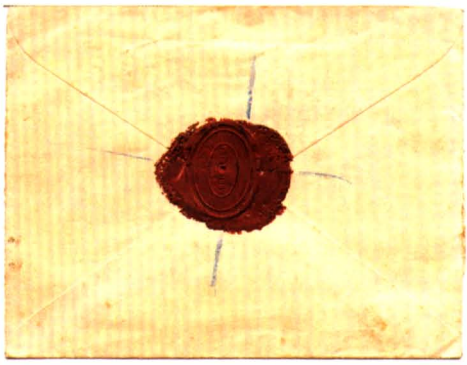
Bob veut transmettre des informations à Alice mais il veut préalablement être certain de communiquer avec elle et non avec un usurpateur. Il souhaite donc l'authentifier. Pour cela il lui lance un défi. Il choisit arbitrairement

La fonction de hachage MD5

Une fonction de hachage forme le résumé d'un texte en remplissant les deux objectifs suivants :
– la moindre modification du message initial entraîne une modification majeure du résumé ;
– il n'est pas possible de former un message dont le résumé soit égal à une expression donnée.

Le résumé MD5 d'un texte est obtenu en commençant par compléter ce texte avant de le sectionné en blocs de 512 bits. Un procédé itératif modifie ensuite une valeur initiale de 128 bits convenue en fonction de chacun des blocs du texte par application de fonctions complexes (et notamment de valeurs prises par la fonction sinus). La valeur finale obtenue est le résumé cherché. Par exemple, « Cette fille est polie » est résumé en « ce0d33c5d48f75451be3587712732b8 » alors que « Cette fille est jolie » est résumé en « 55d1f1b640797594c86e771cf55aa07d ».

un message M et demande à Alice de le chiffrer. Étant en possession de f, elle peut former $M' = f(M)$ et le lui transmettre. À réception, Bob vérifie si $M = g(M')$. Si tel est le cas, il est assuré qu'elle est en possession de la clef privée f. En effet, un usurpateur ne connaissant par f ne peut chiffrer un message aléatoire, il serait automatiquement démasqué. Ce principe d'authentification se retrouve lors d'échanges électroniques où les interlocuteurs ont besoin de s'identifier mutuellement avant d'échanger des informations sensibles.



Un tiers de confiance

Bob souhaite échanger des données sensibles avec Alice qu'il rencontre pour la première fois. Cette situation est celle d'un client entrant dans un site de commerce en ligne. Alice communique à Bob sa clef publique. Il peut ainsi chiffrer ses informations avant de les transmettre à Alice qui est la seule à pouvoir les décrypter. Soupçonneux, il se demande si Alice ne serait pas un truand avide d'informations sensibles. Peut-il faire confiance à cette inconnue ? Pour résoudre ce problème, Alice et Bob vont faire appel à un tiers de confiance, une autorité de certification. L'une des plus connues est *Verisign* mais pour notre explication celle-ci sera appelée Charlène. Alice produit auprès de Charlène différentes informations personnelles ainsi que sa clef de chiffrement publique. Charlène forme alors un certificat regroupant ces informations, un numéro de série, une période de validité ainsi qu'une signature numérique qu'elle aura réalisée à l'aide de sa propre clef privée.

Lors d'une transaction avec Alice, Bob reçoit ce certificat. Il prend connaissance de la clef publique d'Alice et peut vérifier la validité du certificat à l'aide de la clef publique de Charlène. En faisant confiance à Charlène, Bob peut faire confiance à Alice.

Ces différents protocoles sont les bases permettant les échanges d'informations sécurisées sur tout type de réseaux. Bien sûr ces protocoles sont invisibles de l'utilisateur, ce sont les logiciels exploités, comme par exemple le navigateur Internet et les différents serveurs qui les mettent en place.

D. D.

Législation et cryptographie

Pendant longtemps les moyens de transmission et les messages envoyés ont été strictement contrôlés, particulièrement en temps de guerre, conduisant à la censure, aux écoutes téléphoniques et au brouillage radio. Jusqu'à une période récente, même les pigeons voyageurs étaient contrôlés par le ministère des Armées.

Le commerce électronique se développe de plus en plus, notamment sur Internet. Or, pour que les transactions puissent se dérouler en toute sécurité sur des réseaux informatiques, on a besoin d'outils de chiffrement efficaces. D'une façon plus générale, les nécessités de cryptage industriel ou commercial et les nouvelles technologies de communication, ont conduit le législateur, dans un premier temps, à définir les catégories de moyens et de prestations de cryptologie pour lesquelles la procédure de déclaration préalable était nécessaire, puis à substituer à cette demande de déclaration une simple procédure d'information selon le principe du « Tiers de confiance ». C'est le décret n° 99-199 du 17 mars 1999 :

« Sont concernés les matériels ou logiciels offrant un service de confidentialité mis en oeuvre par un algorithme dont la clef est d'une longueur supérieure à 40 bits et inférieure à 128 bits sous certaines conditions. »

En fait l'État n'a plus aucun contrôle sur les échanges et a bien été obligé de passer de l'interdiction complète de chiffrage à une simple information de la part des utilisateurs.

Il est évident que la masse d'informations qui circulent sur Internet et sur les réseaux télévisés rend quasiment impossible leur surveillance par un quelconque organisme. À plus forte raison si les messages sont cryptés ou à plus forte raison stéganographiques. D'où l'inquiétude et l'inefficacité relative des pouvoirs publics face aux nécessités relatives à la sécurité ou aux trafics en tout genre.

Les nombres premiers

Le Haut Comité européen de lutte contre les codages illicites rappelle que la détention de nombres premiers supérieurs à 2^{128} est désormais soumise à autorisation. Il demande donc que tout détenteur d'un nombre strictement supérieur le fasse tester. Un test rapide portant sur le dernier chiffre permet d'éliminer 60 % des suspects. Les 40 % restant doivent impérativement subir un test de Miller-Rabin. Nous rappelons qu'il est aujourd'hui sans douleur.

Bletchley Park

Bletchley Park se situe à 80 km au nord-ouest de Londres, à Milton Keynes. Ce site fut choisi pour abriter l'équipe britannique de décodeurs, 7 000 personnes environ, des mathématiciens, des linguistes et même six cruciverbistes recrutés en 1942 par un concours organisé par le Daily Telegraph. L'éloignement de Londres et des centres industriels la mettait à l'abri des raids aériens allemands. Elle restait cependant proche des grands centres de communication. Bletchley Park est devenu un musée en 1991, on peut y voir des Enigma et des « bombes ».

par David Delaunay

SSL, le vigile d'Internet

Le protocole SSL sécurise les transactions commerciales sur Internet en utilisant des méthodes de cryptographie asymétriques pour créer un canal sécurisé.

Lors d'un achat en ligne ou plus généralement lors d'échanges d'informations sensibles, il est préférable de sécuriser la communication. En effet lorsqu'une information transite sur un réseau, comme Internet, elle passe de serveurs en serveurs jusqu'à atteindre son destinataire. Entre-temps un analyseur de réseaux a pu écouter le trafic. L'information ne peut donc être transmise en clair. De plus, il est possible de mystifier une adresse IP et donc de se faire passer pour quelqu'un que l'on n'est pas. L'identification des interlocuteurs est indispensable. Ces problèmes peuvent être résolus à l'aide de la cryptographie. Nous voyons ici les principes de cryptographie à clefs symétriques et asymétriques et comment ils interviennent dans le protocole SSL, le plus utilisé pour la sécurisation des achats sur le Net.

Clefs symétriques et asymétriques

La grille de Vigenère (voir l'article *Du code Vigenère à celui de Vernam*) est un système de cryptographie dit à clef symétrique car la même clef permet à la fois de chiffrer et de déchiffrer les messages. De nos jours, le codage DES en est une complexification calculatoire mais le principe reste le même : qui sait chiffrer, sait aussi déchiffrer. Les cryptosystèmes asymétriques sont eux d'une autre nature : ce sont deux clefs différentes qui chiffrent et déchiffrent. Seuls les concepteurs du code les connaissent, et il est techniquement impossible de déduire l'une de l'autre. Le codage RSA ou l'algorithme Diffie-Hellmann sont les cryptosystèmes asymétriques les plus connus.

Dans la pratique, les systèmes symétriques et asymétriques sont aussi efficaces l'un que l'autre pour protéger l'information qu'ils chiffrent. Néanmoins, les systèmes symétriques nécessitent l'existence d'un canal sécurisé pour permettre l'échange de la clef de chiffrement. *A contrario*, dans un système à clefs asymétriques il est possible de communiquer la clef de chiffrement à quiconque sans pour autant affaiblir l'information chiffrée. Cette clef est dite publique, la clef de déchiffrement est dite privée. Elle est gardée secrète par le concepteur du code. Malheureusement,

les cryptosystèmes asymétriques sont gourmands en calculs alors que les cryptosystèmes symétriques sont beaucoup plus rapides. L'idée du protocole SSL est de créer, par un système asymétrique, un canal sécurisé permettant l'échange d'une clef symétrique et de poursuivre la communication en chiffrant par le biais de cette clef.

Le protocole SSL

Le protocole SSL (pour *Secure Socket Layer*) a été développé par Netscape Communications Corp. en collaboration avec RSA Data Sécurity Inc. afin de sécuriser les échanges d'information sur les réseaux. Partons d'une situation concrète : Alice veut acheter un CD à la boutique en ligne de Bob. La communication va passer en mode sécurisé, l'adresse du serveur commencera par « https :// » au lieu du classique « http :// » et il apparaît généralement un cadenas fermé en bas du navigateur d'Alice. C'est toute la communication *http* qui sera sécurisée par le système qui va se mettre en place. Mais avant de parler de cette transaction, revenons quelques temps en arrière, au jour où Bob a décidé d'ouvrir un site de vente en ligne. Pour sécuriser ses futures transactions, Bob met au point un système de cryptographie à clefs asymétriques. Il fait ensuite appel à une autorité de certification qui va certifier sa clef publique. L'autorité de certification joue ici le rôle du tiers de confiance qui assure aux futurs consommateurs le sérieux du site de vente en ligne.

Revenons à la transaction d'Alice. Son navigateur contacte le serveur de Bob en lui faisant part de son souhait de passer en mode sécurisé. Il transmet aussi la liste des systèmes de cryptographie symétriques qu'il supporte. En retour le serveur lui envoie la clef publique de Bob et précise le cryptosystème le plus performant avec lequel il est compatible. Le navigateur d'Alice vérifie que la clef publique de Bob est certifiée par au moins une autorité dont il reconnaît la compétence. Si tel est le cas, il génère aléatoirement une clef symétrique qu'il chiffre par le biais de la clef publique de Bob avant de la lui envoyer. Bob reçoit cette clef qu'il déchiffre à l'aide de sa clef privée.
À ce stade, le navigateur d'Alice et le serveur de Bob ont convenu d'un cryptosystème symétrique et se sont échangés une clef par le biais d'un canal sécurisé. Cette clef ne servira qu'à cette transaction. Pour cette raison, on parle de clef de session. Les messages échangés seront ensuite décomposés par blocs, chaque bloc signé numériquement afin d'en assurer l'intégrité et le tout chiffré par la clef de session.

Atouts et faiblesses de SSL

Le protocole SSL est rapide, c'est son premier atout. Ce n'est pas le seul, en voici une liste rapide. Tout d'abord, l'intégralité de la transaction est chif-

Les autorités de certification

Quelles sont les autorités de certification dont votre navigateur reconnaît la compétence ? Pour le trouver, si vous utilisez Internet Explorer, faites :
Menu Outils | Options Internet, onglet Contenu puis bouton Certificats.
Avec Firefox :
Menu Outils | Options, onglet Avancé puis bouton Gérer les certificats
Comment désactiver tous les certificats ? Il suffit de changer la date de l'ordinateur et de passer en l'an 2029 car bon nombre de certificats auront expiré en 2028.

frée par une clef de session échangée via un canal sécurisé. Ensuite, le client est assuré de l'identité du serveur puisque la clef publique est certifiée par un tiers de confiance. Si quelqu'un usurpe l'identité du serveur il ne pourra déchiffrer la clef de session formée car il n'est pas en possession de sa clef privée. De son côté, le serveur est certain de communiquer avec le créateur de la clef de session car il peut vérifier l'intégrité des messages déchiffrés par leur signature.

Il revient essentiellement au client de vérifier l'intégrité du site sur lequel il transmet des informations sensibles.

La principale faiblesse du protocole SSL se situe au niveau de la liste des autorités de certification, il suffit qu'une seule d'entre elles valide la clef publique de Bob pour que celui-ci soit jugé digne de confiance. De plus, le protocole SSL ne prévoit pas de vérification systématique de la non-révocation des certificats. Cela reste donc essentiellement au client de vérifier l'intégrité du site sur lequel il transmet des informations sensibles.

Le *phishing*

Vous recevez un jour un courriel de la « South Trust Bank » qui, suite à un incident technique, vous demande de bien vouloir lui communiquer à nouveau vos identifiants bancaires. Vous êtes surpris car vous n'êtes pas client de cette banque ! Vous êtes en fait victimes de *phishing*, phénomène responsable de nombreux *spams*. Le *mail* que vous venez de recevoir contient un lien menant sur un site pirate très semblable à celui de la « South Trust Bank ». L'objectif du pirate est de récolter les identifiants bancaires de quelques clients de cette banque. En général la connexion vers ces sites n'est pas sécurisée contrairement à ce qu'ils peuvent prétendre. En effet, ici, cela ne servirait pas à grand-chose !

Des faiblesses en dehors du protocole

Le principal souci d'Alice est que son numéro de carte bancaire ne soit pas dévoilé durant la transaction. Le protocole SSL protège la transmission de ce numéro mais ne protège pas celui-ci ni au départ ni à l'arrivée. Si l'ordinateur d'Alice est espionné par un *keylogger* (programme enregistrant les saisies du clavier) alors le numéro de carte bancaire d'Alice peut être piraté. D'autre part, une fois le numéro de carte bancaire parvenu sur le site de Bob, celui-ci doit le transmettre à sa banque et Alice n'a aucun contrôle sur la fiabilité de cette transmission. De plus Bob sera peut-être amené à stocker le numéro de carte bancaire dans ses archives afin de garder une preuve de la transaction. De son côté, le principal souci de Bob est qu'Alice soit la véritable détentrice de la carte bancaire dont elle transmet le numéro. Le protocole SSL ne fournit aucune garantie à Bob de cela. Si le porteur de la carte bancaire conteste le paiement, Bob devra rechercher l'identité du destinataire de la marchandise afin d'en exiger le règlement.

Le protocole SET (pour *Secure Electronic Transaction*) pallie ces faiblesses en faisant intervenir l'autorité bancaire. Le numéro de carte est directement transmis à la banque accompagné d'une signature électronique du client empêchant des contestations ultérieures. L'autorité bancaire transmet alors au site marchand l'autorisation de transaction et Bob est ainsi assuré du paiement de sa marchandise. Pour des raisons économiques et de lobbying le protocole SET ne parvient pas encore à s'imposer.

D. D.

Folies douces

Redoutable carré d'ordre 6

L'idée de déposer un carré magique dans les fondations d'un édifice n'est pas rare. On dit qu'Abraham aurait déposé un carré d'ordre 100 dans les fondations de La Mecque. L'emploi d'autres carrés magiques pour la conservation des bâtiments est attesté par ailleurs, comme dans les pyramides d'Égypte lors de leur construction. La durée d'un édifice sera encore accrue si le carré est d'ordre 6 et écrit à l'encre sur une tuile lorsque Saturne entre dans le signe du Capricorne. L'Arche de la Défense, pas plus que les colonnes de Buren ou le forum des Halles, n'ont certes pas bénéficié de ces précautions élémentaires…

Les carrés d'ordre 6 sont particulièrement efficaces : celui que Moïse établit sur une feuille d'or lui permit de faire émerger le cercueil de Joseph des profondeurs du Nil. Un tel carré, gravé au moment de l'exaltation de Mercure, alors que cette planète est libre des influences néfastes, de la combustion dans la lumière du Soleil et du nadir de Mars, au moment de la conjonction de la Lune et de Jupiter et de l'ascendant de la Vierge et des Gémeaux, verra son porteur sortir en vainqueur de toutes les joutes. Enfin, si on jette une tuile avec ce carré dans un puits, l'eau en deviendra potable.

Secrets de marabouts

À ces propriétés architecturales, il faut ajouter les nombreuses vertus prophylactiques des carrés magiques. Selon la tradition islamique, ils empêchent la peste, les épidémies et autres maladies graves de pénétrer dans une maison qui les contient. Le maître de cette maison sera préservé de la lèpre, de la goutte, de la paralysie faciale, de la colique et de la mort subite. En outre, le carré magique recèle un secret étrange pour la cessation de la migraine et des autres douleurs de la tête. On possède également le témoignage d'un médecin milanais qui explique comment guérir d'une morsure de serpent en avalant trois fois de suite la formule SATOR/AREPO tracée en carré sur une croûte de pain. L'utilisation du même carré sur bandelettes coraniques est encore très utilisée chez les marabouts africains pour guérir les maladies de peaux et l'asthme.

Capable de venir à bout de différentes fièvres, de maux de dents, de la transmission de la rage, de morsure de serpent, d'incendies et même d'erreurs ou malversations comptables, la fonction magique du carré est donc évidente, comme l'atteste la pérennité de son ancestrale valeur prophylactique et apotropaïque.

par Chérif Zananiri

Quand les quanta cachent

Théoriquement, la mécanique quantique fournit une méthode de cryptographie dans laquelle toute interception du message modifie ce message et peut donc être détectée. Elle pourrait également fournir un nouveau type d'ordinateur capable de casser les codes de cartes bleues.

Les codages courants utilisent des opérations mathématiques transformant un fichier en un message codé. Cependant, aucune information, même codée, n'est à l'abri d'espionnage ou de falsification. Le moyen le plus efficace est d'utiliser une clef de codage « jetable » c'est-à-dire que l'on n'utilise qu'une fois. Le problème est alors de trouver une manière inviolable d'échanger cette clef. Charles Bennet, du laboratoire de recherche d'IBM à Yorktown Heights et Gilles Brassard de l'Université de Montréal ont proposé d'utiliser la mécanique quantique pour résoudre la question. Depuis, d'autres laboratoires effectuent leurs propres expériences.

Perturbation par la mesure

En 1900, Max Planck suggère que les ondes électromagnétiques évoluent par paquets ou quanta. On parle dès lors de mécanique quantique. Dans les années trente, Werner Heisenberg découvre que l'observation de ces corpuscules est difficile car toute mesure exige de leur fournir une énergie et cette énergie modifie leurs propriétés. Autrement dit, l'observation modifie l'objet observé ! Ce principe d'incertitude s'accompagne d'une relation mathématique dont la connaissance exacte est inutile pour comprendre les principes généraux qui suivent. L'idée essentielle de la cryptographie quantique est de créer un canal de communication au sein duquel toute interception fausse le message. Ce canal est utilisé pour transmettre la clef de codage qui est aussitôt essayée sur un message convenu d'avance comme « *Tangente* est un super magazine » par exemple. Toute interception de la clef est donc détectée.

Pour réaliser cette idée, on utilise une propriété des photons polarisés. Chaque photon peut être polarisé, c'est-à-dire que l'on peut imposer une direction à son champ électrique. La polarisation est mesurée par un angle qui varie de 0° à 180°. Dans le protocole de cryptographie que nous décrivons, la polarisation prend quatre valeurs : 0°, 45°, 90°, 135°. Pour détecter la polarisation des photons, on utilise un filtre polarisant suivi d'un détecteur de pho-

tons. Si le filtre est orienté à 0°, un photon orienté de même le traverse puis est détecté. Un photon orienté à 90° est stoppé. En revanche, un photon orienté à 45° ou 135° traverse le filtre une fois sur deux c'est-à-dire avec une probabilité de 0,5. Ainsi, on peut donc distinguer entre un photon à 0 ou 90° mais pas entre des photons à 45 ou 135°. De même, un filtre orienté à 45° distingue entre les photons à 45 et 135° mais pas entre ceux à 0 ou 90°.

Envoi de la clef secrète

Comment envoyer la clef secrète grâce à ces photons polarisés ? Tout d'abord, les interlocuteurs (appelons-les Alain et Brigitte) disposent de deux canaux de communication : un canal quantique (une fibre optique par exemple) à travers lequel Alain envoie des photons polarisés à Brigitte et un canal classique (radio par exemple). Par le canal classique, ils décident que les photons polarisés à 0 ou 45° représentent 0, et ceux polarisés à 90° ou 135° représentent 1, par exemple. Alain émet, sur le canal quantique, une suite de photons polarisés au hasard. Brigitte les mesure avec un filtre ou l'autre. Elle note 0 si le photon traverse, 1 sinon. Pour éliminer les cas d'incertitude, elle donne l'orientation de son filtre à la réception (diagonal ou rectiligne). S'il diffère de l'orientation à l'émission, le bit envoyé est incertain donc supprimé. La clef transmise est la suite des bits conservés. Le tableau suivant en montre un exemple :

Émission	0°	45°	90°	45°	0°	135°	90°	0°	45°
Bit envoyé	0	0	1	0	0	1	1	0	0
Filtre réception	45°	45°	0°	45°	0°	0°	0°	45°	0°
Traverse ?	non	oui	non	oui	oui	non	non	non	oui
Bit reçu	1	0	1	0	0	1	1	1	0
Clef	X	0	1	0	0	X	1	X	X

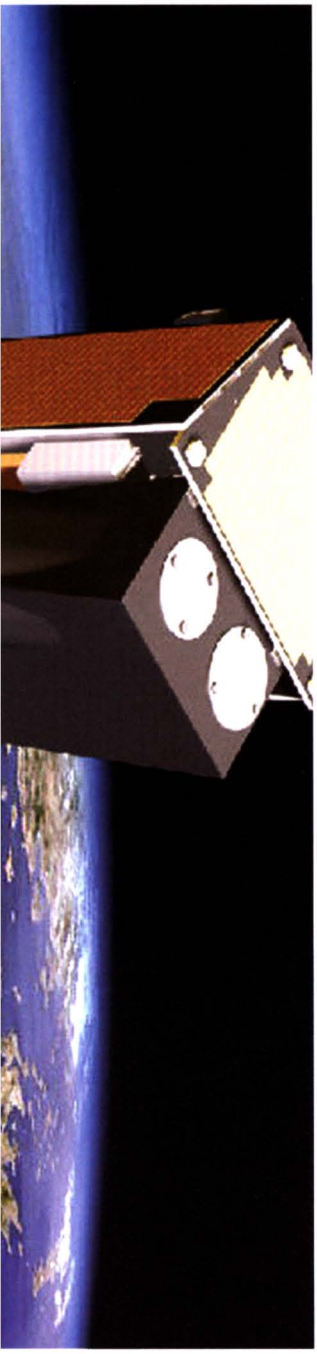
Suite d'émission de photons et réception

Dans cet exemple, la clef transmise est donc : 01001. Si l'émission est interceptée, l'espion doit ré-émettre des photons polarisés vers Brigitte et, du fait de l'incertitude sur sa mesure, il va se tromper une fois sur deux (voir l'encadré *Espionnage d'un*

Espionnage d'un émetteur quantique

Que se passe-t-il si un espion (Charles) surveille la communication entre Alain et Brigitte ? Pour écouter ce que Alain envoie à Brigitte, Charles doit utiliser un filtre polarisant et, tout comme Brigitte, le placer dans une orientation aléatoire. Pour éviter que Brigitte ne se doute d'une écoute, il lui renvoie l'information qu'il a reçue. Cependant, il n'est pas sûr de la polarisation du photon émis par Alain, il connaît seulement la valeur obtenue avec l'orientation de son filtre polarisant. La probabilité qu'il soit orienté correctement est de 50%. Charles renvoie donc les photons avec une polarisation différente de celle choisie par Alain dans la moitié des cas. Il est ainsi très simple pour Alain et Brigitte de savoir qu'ils ont été écoutés : il leur suffit de comparer les résultats obtenus pas Brigitte, sur un canal non sécurisé. S'ils constatent qu'ils ne correspondent pas à ce que Alain a envoyé, alors il leur faut se méfier. Cette comparaison signifie que Brigitte doit annoncer à Alain une partie du message qu'elle a reçu. Pour éviter de révéler des éléments confidentiels lors de cette vérification, Alain insère des morceaux aléatoires dont il note la position dans le message. Ce sont ces morceaux qu'il demande à Brigitte de révéler.

Nicolas Delerue



émetteur quantique). Si la clef est assez longue, nos interlocuteurs s'en rendront compte. Ils pourront la modifier. Une application aussi précise de la logique quantique se heurte à plusieurs obstacles. Il s'avère délicat de manipuler des photons pour leur donner un état précis et lors des échanges, ces états peuvent être altérés.

Et pourtant, ça marche !

Historiquement, le premier succès réel du prototype qui implante les idées ci-dessus a eu lieu le 27 février 1991. Ce jour-là, environ 715 000 impulsions d'intensité moyenne 0,12 photon par impulsion ont été transmises entre deux interlocuteurs. Étant donné que les détecteurs ne sont pas très efficaces, ceci a résulté en une suite de 2 000 bits contenant 79 erreurs, c'est-à-dire qu'environ 4 % des bits ont été mal reçus. Le protocole de réconciliation a néanmoins réussi à découvrir et corriger toutes ces erreurs en ne dévoilant que 550 bits à l'espion. En fonction de l'intensité moyenne des impulsions, de l'efficacité des détecteurs et du nombre d'erreurs de transmission, on peut estimer que l'espion n'avait qu'une infime probabilité d'avoir obtenu plus que 601 des 2 000 bits avant réconciliation, par espionnage du canal quantique. Par conséquent, on a sacrifié 1 172 bits par l'intermédiaire du protocole d'amplification de confidentialité, de telle sorte que la clef secrète finale était de 828 bits, avec une probabilité inférieure à un sur un million que l'espion n'en connaisse ne fût-ce qu'un bit. (Ce nombre de bits sacrifiés provient du calcul $1\,172 = 550 + 601 + 21$, dans lequel le 21 est un paramètre de sécurité imposé par le protocole d'amplification de confidentialité).

L'ordinateur quantique

Une des propriétés les plus prometteuses de l'information quantique est qu'elle peut être simultanément en plusieurs états différents. C'est ce qu'on appelle le principe de superposition. Sous certaines conditions, un électron, par exemple, peut être simultanément sur deux « orbites » différentes du même atome (modèle planétaire ultra-simple de Bohr). On dit alors de cet

Utiliser des photons enchevêtrés

Une autre propriété intéressante (et surprenante) de la physique quantique est que deux photons peuvent être enchevêtrés, c'est-à-dire avoir les même propriétés tout en voyageant dans deux directions différentes. Ces propriétés ne sont pas forcément fixées au moment de l'envoi des photons mais au moment où ils sont mesurés. En 1991, Arthur Ekert a proposé d'utiliser cette propriété en cryptographie quantique. La méthodologie est presque la même que dans le protocole proposé par Bennett et Brassard mais Alice ne fixe la polarisation du photon qu'au dernier moment, lorsque Brigitte a alors déjà reçu le photon, rendant encore plus difficile toute tentative d'espionnage. Cette méthode permet aussi une communication dans les deux sens : Brigitte peut « répondre » à Alain en fixant elle-même la polarisation des photons qui sont alors lus par Alain.

Nicolas Delerue

électron qu'il est dans un état superposé : à la fois ici et là-bas. Toutefois, l'électron ne peut pas être détecté sur les deux orbites en même temps ; dès qu'il est détecté à un endroit plutôt qu'à un autre, l'information cesse d'être quantique et devient « classique ». Ce phénomène de superposition pourrait avoir des conséquences révolutionnaires dans la conception des futurs ordinateurs. Un ordinateur quantique tirerait justement parti de ce principe. En utilisant des particules dans des superpositions de plusieurs états simultanément, un tel ordinateur pourrait parvenir à trouver la réponse à certains problèmes extrêmement complexes en diminuant de beaucoup le temps nécessaire à leur solution. C'est ce qu'on appelle le parallélisme quantique. En fait, l'accélération pourrait être d'autant plus grande que le problème est difficile : la factorisation de très grands nombres apparaît d'ailleurs comme l'exemple le plus spectaculaire, étant donné ses répercussions en cryptographie classique.

Entre aujourd'hui et demain

Où en sommes-nous aujourd'hui, à la fin de l'année 2012 ? Entre progrès indéniables et secrets d'État.

Progrès d'abord : une grande avancée rendue publique par les chercheurs d'IBM qui porte sur la réduction des marges d'erreur sur les données dans les calculs élémentaires. Pour être clair, nous introduisons le concept de *qubit*. Les *quantum-bits* sont des objets quantiques (atomes, ions, *etc.*) qui peuvent stocker dans un état superposé des 1 et des 0. Le problème est que leur durée de vie est très brève, seulement quelques milliardièmes de seconde. Or depuis peu, IBM a créé un qubit tri-dimensionnel à partir de circuits faits de matériaux supraconducteurs (qui conduisent l'électricité sans résistance, donc sans pertes). Quand ils sont refroidis près du zéro absolu, ces circuits se comportent comme des qubits stables jusqu'à 100 microsecondes, soit deux à quatre fois mieux comparativement aux précédents records.

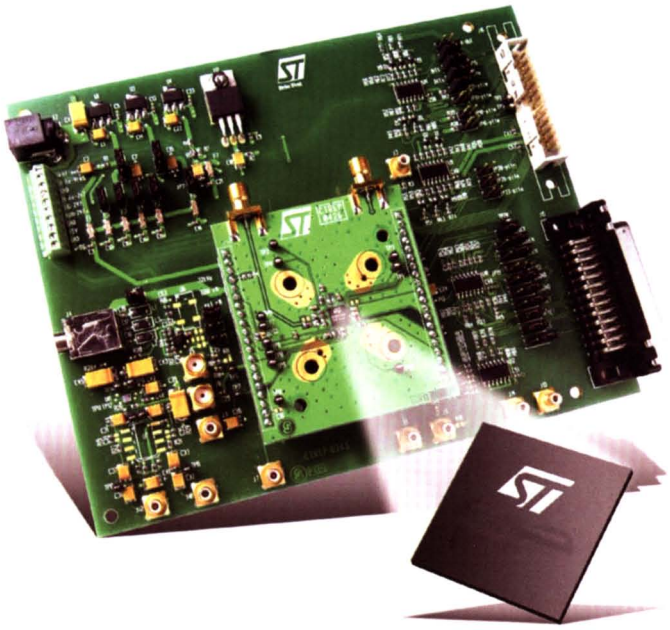
Enfin, l'ajout de qubits dans un pro-

Tenir compte des erreurs

Lorsqu'ils voyagent d'Alain à Brigitte, la plupart des photons conservent leur polarisation. Cependant, il est impossible d'exclure que certains photons perdent leur polarisation initiale en chemin. Alain et Brigitte doivent donc être prêts à admettre que certains morceaux de l'information qu'ils se transmettent soient abîmés lors de la communication. Pour éviter que cela n'affecte la compréhension du message, ils doivent donc utiliser des codes de correction d'erreur en ajoutant une certaine redondance dans le message. Plus il y a de redondance dans le message, plus il sera facile pour Brigitte de corriger les erreurs et donc d'obtenir le message correctement. Un espion peut cependant lui aussi essayer de profiter de ces erreurs (et de cette redondance) en interceptant seulement un nombre limité de photons : si ce nombre est suffisamment petit, Alain et Brigitte ne détecteront rien et penseront que la qualité de la transmission était simplement mauvaise. Si le message inclut de nombreuses redondances, la fraction d'information interceptée par l'espion peut être suffisante pour lui permettre de récupérer certaines informations essentielles.

L'apparition d'erreurs lors de la transmission de photons polarisés est pour l'instant la principale limitation de la cryptographie quantique. Actuellement les communications par cryptographie quantique sont limitées à une centaine de kilomètres avec des photons transmis par fibre optique. Des recherches sont en cours pour permettre d'utiliser des satellites, ce qui permettrait des transferts sur des distances beaucoup plus longues.

Nicolas Delerue



cesseur quantique augmente sa puissance de manière exponentielle puisque chaque qubit ajouté double sa puissance de calcul. Ainsi, un hypothétique ordinateur de « seulement » 100 qubits permet de simuler un cerveau humain, tandis qu'avec 300 qubits on pourrait simuler la totalité de l'univers visible depuis le big-bang ! Il en découle que l'ordinateur quantique est parfait pour résoudre des problèmes combinatoires (tel justement le cassage des mots de passe). Autre question : cet ordinateur existe-t-il vraiment ? Il semblerait que oui.

Une puce fabriquée par la société canadienne D-wave, insérée dans des calculateurs utilisés par Google, aurait récemment impressionné par la rapidité de son action. L'entreprise prétend avoir réussi à intriquer 128 qubits alors que le record actuel (2012) est de 14 qubits.

Vrai ? Intox ? Pas de réponse. Un pays qui dispose de cet outil, le clamerait-il haut et fort ? Pas si sûr ! Que faut-il penser ? Il n'y a pas de loi physique qui s'oppose à la création d'un ordinateur quantique ; le problème est uniquement d'ordre technologique. Juste une affaire, donc, de recherche et de développement, et surtout d'investissement.

C. Z.

Serge Haroche, prix Nobel de physique 2012

Le 9 octobre 2012, Serge Haroche est récompensé par le comité Nobel pour ses travaux sur les photons. Parmi les applications possibles de ses recherches, les ordinateurs quantiques.

En 2007, le chercheur et son équipe avaient réalisé l'exploit expérimental de piéger quelques photons entre deux miroirs et de mesurer leurs propriétés physiques sans les détruire. *« On aime comparer le photon au soldat de Marathon. Il transmet son message, et puis il meurt. Avec notre expérience, il survit ! »* Cette expérience est très délicate (17 ans de mise au point) car il faut que le système étudié soit parfaitement isolé des perturbations extérieures. Le cœur du dispositif expérimental est une cavité optique : deux petits miroirs courbes, hyperrefroidis et éloignés de 3 cm l'un de l'autre. Le but de l'opération est de piéger quelques photons entre ces miroirs, les faisant rebondir quelques milliards de fois afin d'avoir le temps de mesurer leurs interactions avec des atomes spéciaux utilisés comme des sondes. Le temps de piégeage de ces photons est d'un dixième de seconde, temps pendant lequel ils parcourent 30 000 km dans la cavité.

Le mystérieux palimpseste d'Archimède

La sauvegarde des manuscrits de l'Antiquité n'a pu se faire que par des copies permanentes, dont la pérennité doit beaucoup au passage du rouleau de papyrus au codex, empilement de feuillets de parchemin. L'histoire des œuvres d'Archimède est celle, passionnante, de trois de ces codex. Deux d'entre eux, répertoriés A et B, furent traduits en 1269, du grec au latin, pour le Vatican. On perd trace du codex B dès 1311. Encore propriété du pape Nicolas V en 1450, le codex A, copié à la demande de Laurent le Magnifique, atterrit dans la famille Pie et disparait en 1564. Cependant, le corpus des manuscrits et la légende d'Archimède sont sauvés par les traductions de Gérard de Crémone et Guillaume de Moerbeke.

Le codex C, copié à Constantinople au X^{e} siècle, est transformé vers 1229 en livre de prières. Les doubles pages sont détachées, grattées, décapées, poncées, puis coupées en deux et tournées de 90° pour constituer les folios d'un palimpseste deux fois plus petit de 177 pages. Confié au monastère Mar Saba près de Bethléem, on retrouve sa trace vers 1840 au Métochion, bibliothèque de l'église du Saint-Sépulcre de Jérusalem à Constantinople. En 1906, le philologue danois Johan Heiberg étudie le document à partir de photos qu'il a prises, juste avant que le codex ne disparaisse. En 1915, il réédite complètement l'œuvre d'Archimède en combinant les codex A, B et C, publiant pour la première fois la *Méthode* et le *Stomachion*, dont seul le codex C a gardé la trace.

Archimède par Domenico Fetti (1620)

Après maintes péripéties au cours du XX^{e} siècle, qui lui causeront plus de dégâts que tous les siècles précédents, on retrouve le codex C mis aux enchères en 1998 dans un piteux état. Un mystérieux acquéreur confie le document au Walters Art Museum de Baltimore, avec la mission de le restaurer et de nous livrer les ultimes travaux d'un des plus grands penseurs de l'humanité (voir en page 147).

Références :

www.archimedespalimpsest.org

Mathématiques discrètes et combinatoire. Bibliothèque Tangente 39, 2010.

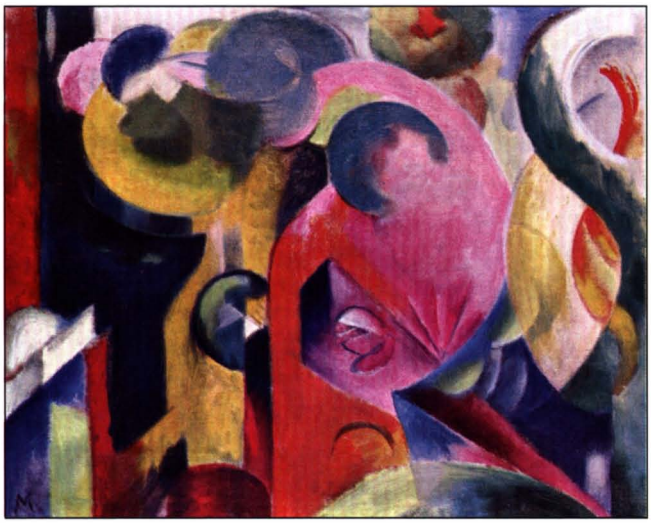
Le codex d'Archimède. Reviel Netz et William Noel, JCLattès, 2008.

Œuvres d'Archimède, en quatre tomes (bilingues), Les Belles Lettres, Paris, 2003.

par Élisabeth Busser

AKS, l'algorithme efficace

AKS, trois lettres miraculeuses pour un algorithme de reconnaissance en temps polynomial des nombres premiers, même très grands, voilà qui va intéresser la cryptographie. Mais qui se cache derrière cette découverte et quels en sont les ressorts mathématiques ?



***Composition III* (1914) de Franz Marc (Karl-Ernst-Osthaus-Museum).**

Les procédés cryptographiques sont grands consommateurs de nombres premiers, de préférence très grands. D'Ératosthène, au troisième siècle avant J.-C. à 2002, année de naissance du fameux algorithme AKS, en passant par l'algorithme de Rabin en 1976 ou celui d'Adleman (le « A » du code RSA) en 1992, les algorithmes de reconnaissance de ces nombres spéciaux ne manquent pas, mais aucun jusqu'en 2002 n'était exécutable en un temps relativement court.

Premier ? Composé ? Historique d'une interrogation

Un nombre premier, on le sait, est un entier n'ayant que deux diviseurs, 1 et lui-même. On connaît depuis Euclide (III^{e} siècle avant J.-C.) un moyen simple de savoir si un nombre n est premier ou non : on le divise par la suite des nombres premiers de 2 à $n - 1$, processus dont on peut énoncer le déroulé sous forme d'un premier algorithme :

- Si la division « tombe juste », énoncer : le nombre n n'est pas premier.
- Si aucune des divisions ne « tombe juste », énoncer : le nombre n est premier.

Cela peut nécessiter jusqu'à $n - 1$ opérations, mais peut-on faire mieux ? Poser cette question c'est s'intéresser

à ce que les mathématiciens nomment la *complexité* de l'algorithme, c'est-à-dire le nombre d'opérations nécessaires pour le mettre en œuvre.

Ératosthène, vers 240 avant J.-C., généralise l'algorithme précédent en construisant le « crible » qui va porter son nom pour permettre de dresser la liste des nombres premiers jusqu'à n. Une fois écrite la liste (L) des entiers de 1 à n :

- Barrer de (L) tous les multiples de 2 (sauf 2). Le premier nombre restant après 2 est 3.
- Barrer de (L) tous les multiples de 3 (sauf 3). Le premier nombre restant après 3 est 5.
- Barrer de (L) tous les multiples de 5 (sauf 5). Le premier nombre restant après 5 est 7. *Etc.*

Inutile de continuer au-delà de $\sqrt{n}$ puisque tous les nombres susceptibles d'être barrés le sont déjà. L'algorithme nécessite au plus $\sqrt{n}$ divisions, chacune étant (voir encadré) en $O((\log_2 n)^2)$: il est donc en $O((\sqrt{n}\log_2 n)^2)$ et il est *déterministe*, c'est-à-dire permet de répondre par oui ou par non à la question « Ce nombre est-il premier ? ».

À défaut, maintenant, de savoir reconnaître si un nombre est premier, au XVII^{e} siècle, on sait au moins discerner qu'il ne l'est pas. Voici comment : on sait – c'est le « petit théorème de Fermat », énoncé par ce dernier en 1640 et démontré par Euler en 1736 – que si n est premier et a premier avec n, a^{n-1} a pour reste 1 dans la division par n ; on dit que a^{n-1} est congru à 1 modulo n, ce qu'on écrit $a^{n-1} \equiv [n]$. En utilisant la contraposée de ce théorème, si pour un certain entier n on trouve un « témoin de non-primalité » ou « certificat » a compris entre 2 et $n-1$ tel que $a^{n-1} \neq 1\ [n]$, alors on est sûr que n n'est pas premier : c'est le *test de Fermat*. Le seul calcul à effectuer étant une élévation à la puissance $n-1$ en arithmétique modulo n, il est en $O(\log_2(n-1)\times(\log_2 n)^2)$, donc en $O((\log_2 n)^3)$.

Dommage que le test soit à sens unique : il existe des entiers qui « passent » ce test

Échelle de complexité

La vitesse de calcul d'un ordinateur est à peu près proportionnelle au nombre des chiffres 0 ou 1 utilisés pour écrire les nombres en numération binaire. On sait par ailleurs que le nombre de chiffres de l'écriture binaire de n est égal à la partie entière du logarithme de base 2 de n ($\log_2 n$) augmentée de 1, donc de l'ordre de $\log_2 n$. Les mathématiciens, à qui il fallait un langage capable de préciser le temps de calcul f en fonction du nombre utilisé n, ont donc cherché, pour chaque algorithme, une fonction simple g à laquelle comparer f. Ils disent par exemple que $f(n)$ est « en $O(g(n))$ » (on dit « en O de $g(n)$ ») pour exprimer que f ne croît pas plus vite que g. C'est en ces termes que nous définirons la complexité des algorithmes des opérations arithmétiques élémentaires entre deux entiers dont le plus grand est n :

Opérations	Complexité
Addition, soustraction	$O(\log_2 n)$
Produit, quotient	$O((\log_2 n)^2)$

Il en est de même pour les opérations portant sur les restes dans la division par n (on écrit, pour traduire que a et b ont même reste dans la division par n : $a \equiv b\ [n]$, qu'on lit « a est congru à b modulo n » et on parle d'arithmétique modulaire) :

Opérations modulo n	Complexité
Addition, soustraction modulo n	$O(\log_2 n)$
Produit modulo n	$O((\log_2 n)^2)$
Exponentiation ($a^c\ [n]$)	$O(\log_2 c \times (\log_2 n)^2)$

Même pour les algorithmes plus sophistiqués, la complexité mesure le nombre d'opérations binaires nécessaires. On dit par exemple qu'un algorithme est « polynomial » s'il s'effectue en temps polynomial, c'est-à-dire en $O(\log_2 n)^k)$ pour un certain entier k.

mais ne sont pas premiers pour autant. 91, par exemple, est tel que $3^{90} \equiv 1$ [91] et pourtant... $91 = 7 \times 13$. On dit que 91 est *pseudo-premier* pour la base 3. Oui mais, direz-vous, il n'est pas pseudo-premier pour la base 2 et le test avec 2 pourra dire qu'il est composé. Le comble, c'est qu'il existe même des nombres n dits « a-pseudo-premiers » pour toute base $a > 1$ premier avec n et inférieur à n : ce sont les *nombres de Carmichael*, dont 561 est le plus petit ! Le test de primalité de Fermat est donc probabiliste. Le mettre en œuvre, c'est :

- choisir un nombre a au hasard entre 2 et $n - 1$;
- si $a^{n-1} \equiv 1$ [n], déclarer n premier ;
- déclarer n composé dans tous les autres cas.

En réalité, si la dernière assertion est certaine, pour la précédente, n est *probablement* premier. On admet que pour le test d'un entier compris entre 2 et 10^{100} le risque d'erreur est de $2{,}8 \times 10^{-8}$ et qu'il descend à $1{,}2 \times 10^{-123}$ si on remplace 10^{100} par $10^{1\,000}$; mais ce test ne sera jamais un test infaillible de primalité, à cause de nombres comme ceux de Carmichael.

C'est dans les années 1970 que deux autres tests vont contourner l'obstacle des nombres de Carmichael : en 1976, le test de Solovay-Strassen, et en 1977, celui de Rabin-Miller, sorte de raffinement du test précédent. Tous deux reposent également sur une interprétation du petit théorème de Fermat. Dans ses conditions d'application, $a^{(n-1)/2}$ est une racine carrée de 1 modulo n qui, si n est premier, vaut soit 1 soit –1 modulo n. Donc, quand on fait ce calcul et qu'on trouve d'autres valeurs que 1 ou –1, c'est que n n'est pas premier. Le test de Rabin-Miller va même plus loin en poursuivant : si en plus $(n - 1)/2$ est encore pair, alors $a^{(n-1)/4} \equiv \pm 1$ [n] ; on continue en calculant $a^{(n-1)/8}$, *etc*. Si à l'une des étapes on passe à une autre valeur que 1 ou – 1, on est sûr que n est composé et a est un certificat de non-primalité de n. Dans le déroulement de l'algorithme, on préfère, au lieu de diviser les exposants par 2, procéder à l'envers en les multipliant par 2, c'est-à-dire effectuer une suite d'élévations au carré. Le processus se décrit donc ainsi :

- écrire $n - 1$ sous la forme $2^k \times m$ avec m impair ;
- choisir un entier aléatoire a, compris entre 1 et $n - 1$;
- calculer $b_0 = a^m$ [n] ;
- si $b_0 \equiv 1$ [n], conclure que n est premier.
- Calculer successivement $b_1 = {b_0}^2$, $b_2 = {b_1}^2, \ldots, b_k = {b_{k-1}}^2 = a^{n-1}$;
- s'il existe $i < k$ tel que $b_i = -1$ et donc pour tout $j > i$, $b_j = 1$, conclure que n est premier.
- Sinon, conclure que n est composé.

Testons par exemple 61 en choisissant $a = 5$: $61 - 1 = 60 = 2^2 \times 15$;
$5^{15} \equiv -1$ [61]. On dit que 61 est fortement pseudo-premier pour la base 5 et on le soupçonnera d'être premier.
Testons maintenant 221 avec $a = 5$:
$221 - 1 = 220 = 2^2 \times 55$;
$5^{55} \equiv 112$ [221] ; $112^2 \equiv 168$ [221].
221 est composé (en effet : $221 = 13 \times 17$).
Les calculs les plus coûteux en temps de ce test sont ceux de a^m [n] et de ses carrés successifs, dont on fait au plus k itérations, ce qui fait de ce test, tout comme d'ailleurs le test de Solovay-Strassen, un algorithme en $O((\log_2 n)^3)$. On progresse ! Cependant, ces deux tests, de complexité « polynomiale », sont encore probabilistes : conclure « n est premier » au test de Rabin-Miller veut, ici aussi, simplement dire qu'il est probablement premier. On le qualifiera de nombre *fortement pseudo-premier pour*

la base a. On peut prouver que si, avec le test de Solovay-Strassen, la probabilité d'erreur sur la primalité de n est ½, l'algorithme de Rabin-Miller annonce qu'un nombre est premier avec une probabilité d'erreur inférieure à ¼ dès que n est supérieur à 9.

L'algorithme AKS : comment ça marche ?

Les trois découvreurs de l'algorithme AKS : Manindra Agrawal, mathématicien et professeur à l'Institut indien de technologie de Kanpur, et ses deux élèves, Neeraj Kayal (à gauche) et Nitin Saxena.

D'autres tests de primalité, reposant sur d'autres critères que le petit théorème de Fermat, ont encore vu le jour jusqu'en 2002, ayant soit le défaut d'être trop lents soit celui de n'être pas déterministes, et c'est précisément cette année-là qu'est né l'algorithme AKS dont nous allons voir qu'il possède beaucoup de qualités. Une seconde version, améliorée, a été présentée en mars 2003.
A comme Agrawal, le professeur, K comme Kayal et S comme Saxena, ses deux élèves de licence à l'Institut indien de technologie de Kanpur, en Inde. Trois lettres pour une idée géniale, celle d'avoir introduit des polynômes dans une variante du petit théorème de Fermat. Leurs recherches vont mener à un test de primalité déterministe dont ils vont démontrer qu'il permet de conclure en temps polynomial.
L'idée de départ de ce test est en quelque sorte une généralisation du petit théorème de Fermat aux polynômes : si a et $n \geq 2$ sont deux entiers premiers entre eux, alors dire que n est premier équivaut à dire que $(X + a)^n$ a même reste que $X^n + a$ dans la division par n, ce que les mathématiciens écrivent
$(X + a)^n \equiv X^n + a\ [n]$.
C'est la congruence (*).

L'identité précédente va donc fournir un critère très simple de primalité, et c'est là son grand mérite : un nombre n étant donné, il suffira de choisir un entier

La base de l'AKS

$(X + a)^n \equiv X^n + a\ [n]$: dit comme cela, c'est simple, mais la démonstration n'est pas très compliquée non plus.

En effet, le coefficient de X^i dans le développement de $(X + a)^n$ est $\binom{n}{i} \times a^{n-i}$ et, vu la forme des $\binom{n}{i}$, on sait que $i \times \binom{n}{i} = n \times \binom{n-1}{i-1}$, donc n divise $i \times \binom{n}{i}$.

- Si n est premier, pour $i = 1, 2, ..., (n-1)$, n est premier avec i. C'est donc qu'il divise $\binom{n}{i}$ d'après le théorème de Gauss (*Si a est premier avec b et qu'il divise le produit bc, alors c'est qu'il divise c*). Ainsi, tous les coefficients hormis ceux des deux extrémités, sont multiples de n, et $(X + a)^n \equiv X^n + a^n\ [n]$, d'où (petit théorème de Fermat... encore lui !) $(X + a)^n \equiv X^n + a\ [n]$, soit la congruence (*).
- Si maintenant n est composé, soit i un de ses facteurs premiers et k le plus grand exposant tel que i^k divise n. Alors i^k ne peut pas diviser $\binom{n}{i}$ et est premier avec a^{n-i}. Aucun des coefficients de X^i n'est donc nul modulo n et on ne peut pas avoir $(X + a)^n \equiv X^n + a\ [n]$.

a premier avec n et de vérifier par la suite si la congruence en question est satisfaite. Le problème est que cette vérification, qui demande d'évaluer n coefficients, prend un temps en $O(n)$, ce qui est bien long ! Il va donc falloir réduire le nombre des coefficients. On peut le faire, disent les auteurs du test, en comparant modulo n les restes de $(X + a)^n$ et de $X^n + a$ dans la division par un certain polynôme $X^r - 1$, r étant un entier convenablement choisi qui dépendra de n. Il s'agirait alors de vérifier si $(X + a)^n$ est congru à $X^n + a$ modulo $(X^r - 1)$ en tant que polynôme, en calculant modulo n sur les entiers pour pouvoir conclure sur la primalité de n. Un petit problème cependant, c'est qu'il existe des nombres composés, comme certains nombres de Carmichael (toujours eux...), qui vérifient (*) pour quelques valeurs de a et de r. C'est le cas pour le troisième, 1 729, égal à $7 \times 13 \times 19$, et qui vérifie (avec $a = 5$ et $r = 3$), en calculant sur les entiers modulo 1 729, que $(X + 5)^{1729} \equiv X^{1729} + 5$ modulo $X^3 - 1$. Alors, 1 729 serait-il premier ? Tout de même pas car (*) n'est plus vérifiée pour $a = 5$ et $r = 5$.

Nos trois chercheurs ont néanmoins prouvé qu'il était quand même possible d'utiliser la caractérisation énoncée en exhibant un r tel que la congruence soit vérifiée pour plusieurs entiers a. Le nombre de ces entiers ainsi que les valeurs de r sont bornés par un polynôme en $\log_2 n$, ce qui fait de l'algorithme qu'on appelle désormais AKS un test déterministe en temps polynomial de la primalité de n. Ouf !

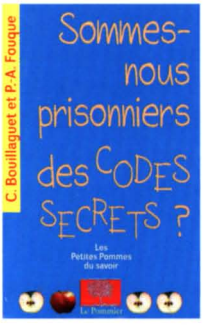
Cet algorithme, pour déterminer si un nombre n donné est premier, se déroulera donc, selon la démonstration de ses auteurs, en trois phases :

- Dans la première, on élimine le cas où n est de la forme a^b avec $b > 1$ (on conclut alors que n est composé).
- Dans la deuxième, en considérant successivement les nombres premiers à partir de 2, on cherche la première valeur de r, de préférence petite par rapport à n, telle que, si k est le plus petit entier tel que $n^k \equiv 1[r]$, k soit strictement supérieur à $(\log_2 n)^2$.
- Dans la troisième, on cherche entre 1 et $2\sqrt{r}\ \log_2 n$ un nombre a entier associé à r qui permettra d'attester que n n'est pas premier. Il faut pour cela tester continuellement, en calculant modulo n, l'équation $(X + a)^n \equiv X^n + a\ [X^r - 1]$ et, dès qu'une valeur de a la vérifie, conclure « n est composé ». Si par contre aucun a de l'intervalle considéré n'a permis de vérifier l'équation, on conclut que n est premier.

En fin de compte, Manindra Agrawal, Neeraj Kayal et Nitin Saxena ont bel et bien prouvé, et c'est là toute la puissance de leur raisonnement, d'abord que leur algorithme renvoie « n est premier » si, et seulement si, il l'est vraiment ; ensuite que si n est supérieur à 3, r n'est pas plus grand que $(\log_2 n)^3$, le polynôme de $\log_2 n$ dont nous avions parlé, et que les calculs s'effectuent en un temps polynomial en $O(\log_2 n)^{21/2 + \varepsilon}$ pour tout $\varepsilon > 0$. Voilà donc un magnifique algorithme – répétons-le – déterministe, capable de dire en un temps polynomial si un entier, même très grand, est premier ou non.

É. B.

Dessine-moi un code secret

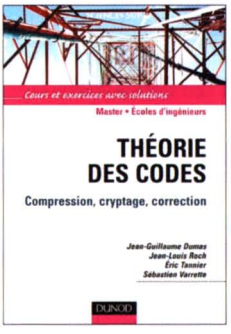


Sommes-nous prisonniers des codes secrets ?
Charles Bouillaguet et Pierre-Alain Fouque, Le Pommier, 64 pages, 2011, 4,90 euros.

S'il y a un domaine de pointe qui se prête à une vulgarisation de qualité auprès du jeune public, c'est bien la cryptographie. À tel point que l'on se demande pourquoi il faut attendre l'année 2011 pour tenir enfin ce minuscule ouvrage entre les mains. *« Des réponses brèves, claires et sérieuses aux questions que vous vous posez sur le monde »* : encore une fois, le credo de la collection « Les petites pommes du savoir » est respecté. Plus qu'une simple introduction aux codes secrets, le livre s'interroge intelligemment sur la question de l'intrusion des codes secrets dans notre quotidien (cartes bleues, DVD, transfert de fichiers sur Internet, communications…). Certains passages seront sans doute difficiles pour le jeune lecteur auquel la collection s'adresse ; un adulte accompagnera avec bénéfice la lecture et la progression de l'enfant. On regrettera par contre qu'il soit nécessaire d'aller sur Internet pour voir la solution des deux énigmes proposées.

É.T.

Les codes : toute la théorie



Théorie des codes. Compression, cryptage, correction.
Jean-Guillaume Dumas, Jean-Louis Roch, Éric Tannier et Sébastien Varrette, Dunod, 350 pages, 2007, 35 euros.

Pour qu'un téléphone portable communique avec une antenne-relais ou que la sonde Cassini transmette les clichés de son aventure à la Nasa, il est nécessaire que les intervenants conviennent de codes pour communiquer. Partant de l'exemple de la communication par fax, les auteurs illustrent la dégradation induite par la numérisation et la nécessité de la compression lorsque l'image numérisée est essentiellement constituée de longues successions de 0 ou de 1. Ils expliquent aussi les conséquences d'un parasitage et la nécessité d'adjoindre une information contrôlant l'intégrité de la transmission. Enfin, si la transmission est sensible, il faudra la crypter ! Au terme de cette introduction, l'ouvrage donne quelques rappels mathématiques qui seront suffisants pour rafraîchir la mémoire du lecteur averti mais ne pourront pas se substituer à de véritables cours d'algèbre pour le lecteur profane.

L'ouvrage présente ensuite les techniques de compression de l'information : algorithme d'Huffman, compression à dictionnaire dynamique, réduction d'entropie… Suit un exposé de cryptologie expliquant le chiffrage symétrique DES et asymétrique RSA, approfondissant les différents protocoles associés tout en exposant quelques outils de cryptanalyse motivant les choix des algorithmes. En fin d'ouvrage sont présentés les codes détecteurs et correcteurs d'erreurs : CRC, BCH… qui nous permettent, entre autres, de lire des CD même fortement rayés.

Notons que la linéarité de l'ouvrage est interrompue par de multiples exercices qui, en guise d'exemples, permettent d'approfondir les notions présentées tout en aiguisant notre curiosité et en nous situant dans un contexte moderne. Ces exercices sont corrigés en fin d'ouvrage.

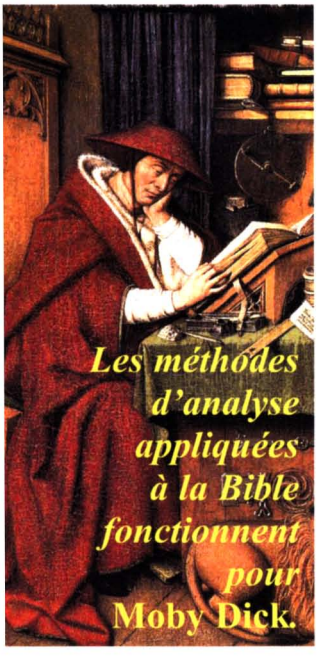
C'est véritablement un livre de référence illustrant les applications techniques de l'algèbre moderne, livre à posséder par toute personne intéressée par le sujet.

D.D.

par Sophie de Vaucorbeil

Le code de la Bible

Dans un livre devenu *best-seller*, un journaliste américain « montrait » qu'il était possible de lire tout l'avenir du monde crypté dans le texte de la Bible. En utilisant un code bien choisi, on trouve en effet que de grands événements des temps modernes sont inscrits dans *Le Livre*... et dans d'autres.



En 1998, un journaliste américain, Michael Drosnin, publiait un livre devenu très vite un *best-seller* dont le titre, *La Bible, le code secret*, et le contenu accrocheur allaient soulever une belle polémique dans le monde laïque juif et chrétien. En effet, dans son livre, le journaliste affirme l'existence d'un code secret annonçant d'une manière irréfutable des événements tels que l'assassinat d'Anouar el Sadate, d'Yitzhak Rabin ou la Shoah. La chose n'est pas nouvelle, cela fait des siècles que des hommes s'échinent à vouloir unir la beauté de l'absolu mathématique et le plus connu des textes religieux du monde occidental.

Structures numériques

Au début du XX^{e} siècle, le mathématicien russe Ivan Panin (1855–1942) a consacré toute son existence aux structures numériques de la Bible. Il a eu l'idée de transformer chaque lettre du texte biblique en un nombre. Par exemple, parmi ses découvertes figurait la récurrence du chiffre 7 (chiffre de la perfection divine) ou de ses multiples. Chaque lettre en hébreu ayant aussi une valeur numérique, il était tout à fait possible d'additionner les lettres d'un mot, d'une phrase ou d'un livre complet et de le réduire à un résultat numérique. Il se dégageait ainsi une harmonie mathématique structurante dans l'ensemble des textes qui composent les cinq premiers livres de l'*Ancien Testament* ! Ces résultats surprenants amenèrent cet athée nihiliste à se convertir au christianisme. Dans son livre, Michael Drosnin avance l'idée que « *la Bible contiendrait à l'intérieur de ses lignes l'histoire en détail de toute l'humanité. Les prophéties y sont bien cachées. Pourquoi est-ce à notre génération qu'il a été donné de connaître cela ? Pourquoi seulement maintenant ? Tout simplement parce que la découverte n'a été rendue possible que par la nouvelle technologie de*

l'ordinateur ». Pour ce faire, il s'appuie sur les travaux du mathématicien Eliyahu Rips qui a repris les travaux du rabbin Weissmandel, cette fois-ci avec l'aide de l'informatique. Celui-ci utilise la méthode des *suites de lettres équidistantes* (ELS). Ce chercheur et son équipe ont mis au point un logiciel qui prend le texte hébreu de la Bible, saute un nombre déterminé de lettres, imprimant par exemple une lettre sur quinze. Ces lettres sont alors arrangées dans une matrice, c'est-à-dire un cadre rectangulaire, dans lequel on pourrait chercher des mots, comme dans les livres de jeux que l'on trouve dans tous les pays. Les éléments peuvent être lus horizontalement, verticalement ou en diagonale.

En utilisant le code secret, Rips découvre dans la Genèse les noms de nombreux sages et grands rabbins du premier millénaire de notre ère. Probabilité qu'un tel événement soit dû au hasard ? Moins d'une chance sur deux milliards selon lui. Sceptique au début, Drosnin dit avoir été convaincu quand Rips lui a montré la prédiction de la guerre en Irak !

La preuve par *Moby Dick*

Mais le « code secret de la Bible », malgré son succès public, ne rencontre pas l'adhésion de tout le monde. Nombre d'hébraïstes et de mathématiciens se penchent sur la question au grand dam de Drosnin qui lance un défi dans le magazine *Newsweek* : *« Si ceux qui me critiquent arrivent à trouver dans* Moby Dick *un message codé annonçant la mort d'un premier ministre, je les croirai. »*

Qu'à cela ne tienne. Brendan Mac Kay, professeur de mathématiques à l'Université nationale d'Australie, relève le gant ! En utilisant la même méthode il découvre dans le texte anglais de Melville pas moins de neuf assassinats d'un premier ministre et, plus fort encore, la mort de Lady Di, de son amant et de son chauffeur !

À l'affirmation de Drosnin que *« toutes les bibles en hébreu actuellement disponibles sont concordantes lettre pour lettre »*, on pourrait répondre, sans risque de se tromper, qu'aucun original de la Bible n'est connu et que Rips et ses collaborateurs, à l'instar d'Ivan Panin, ont utilisé le « codex de Léningrad » qui est une copie postérieure à l'an mil.

Prédire le passé

Pour Jean-Paul Delahaye, professeur et chercheur en informatique au CNRS de Lille, le livre de Drosnin est une vaste escroquerie :

« Le mathématicien Émile Borel (1881–1956) a démontré un théorème remarquable : en prenant au hasard une suite infinie de symboles typographiques, vous y trouverez tous les textes possibles et imaginables. Vous y trouverez donc le livre de Drosnin. Sur un texte de mille lettres avec la méthode des lettres équidistantes, vous pouvez extraire plusieurs millions de mots, et c'est le diable si vous n'en trouvez pas un qui vous évoque votre grand-tante ou les poilus de 14–18. À ce jeu on ne perd jamais ! »

Enfin, on peut s'étonner à juste titre que toutes les « prédictions » ont toutes été faites *a posteriori*, lorsque les événements faisaient déjà partie du passé. On pourrait donc en conclure que derrière l'affaire du code secret se cache l'erreur classique qui consiste à voir de la structure dans le hasard et à en déduire qu'il s'agit là d'une information alors que tout cela ne procède que de l'esthétique du hasard, comme voir dans les nuages ou dans l'écume des vagues des apparitions ou des personnages.

S. V.

par Sophie de Vaucorbeil

La Kabbale

Dans la recherche des codes de la Bible, la Kabbale a plusieurs millénaires d'avance. Les kabbalistes cherchent à connaître, à travers les chiffres, le lien entre le créateur et sa création. Leur analyse du texte sacré utilise combinaisons de mots, permutations de lettres et équivalences numériques.

Il est impossible de traduire ici en quelques lignes la complexité de la Kabbale, cette somme d'écrits, révélatrice de toute une tradition mystique juive à travers les siècles. Ceux qui étudient la Kabbale y vouent en général leur existence entière. Très modestement, en voici une toute petite approche.

Histoire de la Kabbale

La Kabbale, de l'hébreu *kabbalah* qui signifie littéralement « recevoir », est la somme des mystères de la tradition mystique juive (ne pas confondre avec la Kabbale « moderne », voir l'encadré). Ce sont, d'une part les techniques de lecture et de déchiffrage des textes pour en dévoiler et en communiquer les secrets et, d'autre part une approche mystique qui consiste à recevoir la sagesse et la lumière d'en haut, de l'infini. Le kabbaliste voit dans un texte jusqu'à douze niveaux de signification !

La Bible donne $\pi = 3$ et ne se trompe jamais.

En tant que tradition orale, on dit la Kabbale probablement aussi ancienne que la date de rédaction du Pentateuque (recueil des cinq premiers livres de la Bible). La tradition écrite (doctrine de la diaspora) voit le jour en plein Moyen-Âge. Gershom Scholem précise que la Kabbale est née dans le Sud de la France. Elle prospéra en Espagne et se développa très rapidement jusqu'à la publication en 1280 et en 1300 du *Zohar* ou *Livre de la Splendeur* de Moïse de Léon. C'est une compilation de textes en araméen transmis par la tradition orale.

La Kabbale a joué un rôle important dans le bouillonnement humaniste de la Renaissance. On pense à Pic de la Mirandole, très intéressé par la vision d'un monde en perpétuel changement. L'espace de liberté d'interprétation qu'offrait la Kabbale, à une époque où la liberté religieuse était très restreinte, a attiré les esprits libres et curieux au risque de subir l'excommunication voire le bûcher !

Un peu de vocabulaire kabbalistique

Le Tserouf : art combinatoire des lettres. La racine d'un mot hébraïque est formée en général de trois lettres. L'analyse consiste à examiner toutes les combinaisons de ces trois lettres et à y découvrir une logique interne ou externe. D'abord un sens premier donné par la loi de succession des lettres – un degré zéro de signification, étape d'égarement nécessaire. Enfin « l'essence du mot supplie le mot de pouvoir dire autre chose ». Le sens fait alors brèche et le mot est mûr pour une nouvelle combinaison.

Le Notarikon : interprétation de chaque lettre d'un mot comme abréviation d'une phrase entière (principe du sigle ou de l'acronyme). L'apparente simplicité d'un mot peut dévoiler grâce au Notarikon des ressorts subtils. Ce procédé peut aisément s'appliquer à d'autres langues que l'hébreu.

La Temoura : consiste à permuter les lettres d'un mot selon des règles très précises (principe de l'anagramme). C'est un procédé fréquemment utilisé par les kabbalistes.

Le tétragramme : sans aucune voyelle, c'est un vide dans le langage, à partir duquel il prend tout son sens. Imprononçable, il crée une distance infranchissable, qui interdit toute possibilité de prendre Dieu pour un objet : YHWH, (certains prononcent ce mot imprononçable : Yahvé, on pourrait plutôt dire Yahou !) valeur numérique 26, (parce que c'est une valeur unique dans tout l'univers mathématique en ce sens qu'il est le seul entier compris entre un carré et un cube).

La Guematria : (du grec *gematria*, art de mesurer tout ce qui est dans le ciel et la terre d'où la géométrie) est une technique de lecture et d'interprétation. Elle autorise une lecture infinie des mots et des textes en permettant de rapprocher des mots dont la somme des lettres qui les composent est identique. En Hébreu, il n'existe pas de chiffres et chaque lettre de l'alphabet est associée à un nombre. Il y trois sortes de guématria de la simple (addition des valeurs numériques des lettres constituant le mot) à la plus complexe (chaque lettre vaut le nombre donné par la guématria simple mais au carré). Cette façon de coder et de décoder les textes a pour but d'assouplir l'esprit et de véhiculer des informations importantes sous une forme anodine. On peut imaginer qu'à l'époque des persécutions à l'encontre des juifs, ce procédé a dû connaître un plein essor. Ce n'est pas une simple machine à démontrer mais un outil pour relativiser la façon dont on perçoit la signification d'un texte. Elle est directement liée aux particularités de la langue hébraïque et ne peut être appliquée aux autres langues.

Recherche du « code secret » de la Bible

La lecture kabbalistique de l'Écriture biblique est fondée sur le Sod (secret), considéré dans la tradition classique comme le niveau le plus complexe de l'écriture scripturaire. Les kabbalistes étudient et analysent minutieusement la parole créatrice et tous les mots qui animent notre monde. Ils se servent de combinaisons de mots, de permutations de lettres et d'équivalences numériques. Il y a eu des influences réciproques entre les kabbalistes et les pythagoriciens. À noter qu'elles ont donné lieu à la numérologie occidentale, science que l'on ne peut pas qualifier d'exacte à l'heure actuelle.

La Kabbale cherche à accéder à la connaissance de ce qui peut unir Dieu le créateur et sa création. La lumière est un

des mots les plus importants de la Kabbale, c'est la plus haute métaphore de l'infini et du divin. Lumière, vibration et énergie en sont les trois mots clefs. En hébreu, lumière et infini ont la même valeur numérique 207. Pour passer de la lumière de l'infini ou lumière d'en haut à la réception de cette lumière, les intermédiaires sont multiples. La Kabbale est l'étude de ces mondes intermédiaires qui existent entre le monde inférieur et le monde supérieur. Ils sont : les chiffres et l'univers mathématique, la décade des éléments fondamentaux (appelés Sefirots), les noms multiples de Dieu, le tétragramme, etc. L'encadré de la page précédente en donne un aperçu.

Bibliographie

• Gershom Sholem, *La kabbale et sa symbolique*. Payot, 1999 ;

• Charles Mopsick, *La cabale et les cabalistes*. Albin Michel, 1989.

Il y a Kabbale et Kabbale

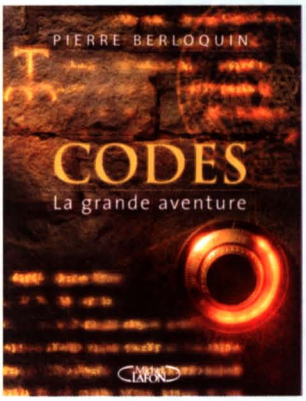
La Kabbale n'a rien à voir avec la Kabbale (fondée par un ancien agent d'assurances) dont se réclament nombre de stars (comme Madonna), une version *New Age*, mercantile et fortement soupçonnée d'être sectaire.

Ah ce nombre Π !

La Bible (livre des Rois) parle d'un chaudron de mer en métal fondu qui avait dix coudées d'un bord à l'autre (diamètre) et était environné d'un fil de trente coudées (périmètre). Ce qui donne pour Π la valeur 3. Mais le rabbin Néhémiah (II^{e} siècle) affirmait que la Bible ne se trompait pas et qu'il s'agissait du diamètre utile, le diamètre intérieur. Après quelques remaniements des valeurs numériques des lettres, cela colle ! Le réalisateur américain Darren Aronofsky nous a concocté, en 1998, *Pi*, un film tout à fait réjouissant sur la Kabbale et le nombre Π. Un jeune mathématicien surdoué (évidemment) et sujet à de terribles migraines croit que les mathématiques sont le langage de la nature. Il voit Π partout, même dans l'analyse des valeurs de Wall Street ! Il tente de trouver la formule du marché de la Bourse à l'aide d'un gigantesque ordinateur. Tout le monde lui court après, une belle analyste financière avec son matériel très performant, et bien sûr un groupe de juifs kabbalistes sacrément intéressés par l'idée de trouver la preuve du vrai nom de Dieu !

S. V.

Les codes, tous les codes

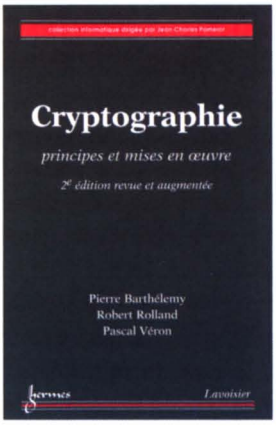


Codes : la grande aventure. **Pierre Berloquin, Michel Lafon, 420 pages, 2010, 19,95 euros.**

Cet ouvrage se distingue des livres sur les codes secrets en ce sens qu'il ne se limite pas à la cryptographie proprement dite. Pour Pierre Berloquin, que les codes passionnent depuis toujours, les aspects symboliques et esthétiques des codes ont autant d'importance que leurs côtés purement techniques. Le premier amateur de codes a été Pythagore, qui considérait que *« tout est nombre »*. Or, nous vivons à une époque où tout est numérisé : images, sons, mais aussi toutes nos données personnelles, nos courriels comme nos données médicales, donc tous les aspects de notre vie passent par un codage. L'auteur évoque aussi les essais de langues universelles ou le manuscrit de Voynich, qu'il considère comme une œuvre d'art. Le lecteur trouvera enfin près de cent cinquante codes à déchiffrer, dont les solutions sont données en fin d'ouvrage. Ce livre, qui a déjà été réédité plusieurs fois aux États-Unis, doit figurer dans la bibliothèque de tout amateur.

M.C.

Mise en œuvre de la cryptographie : les dernières avancées



Cryptographie : principes et mises en œuvre. **Pierre Barthélemy, Robert Rolland et Pascal Véron, Lavoisier, 472 pages, 2012, 95 euros.**

Il est difficile, même pour un amateur très éclairé, de trouver un ouvrage bien écrit qui permette de faire le point sur les dernières avancées en matière de cryptographie. La nouvelle édition de *Cryptographie* est une telle référence pour qui a un niveau de maîtrise de mathématiques. Pour la cryptographie à clé publique comme à clé secrète, de nombreux développements théoriques récents sont présentés, les résultats sont souvent démontrés. Des annexes présentent les bases nécessaires pour suivre les développements, mais il vaut mieux être déjà familier avec l'arithmétique, l'algorithmique et la théorie de la complexité. Les références bibliographiques, nombreuses, ont été actualisées depuis l'édition de 2005. Le point fort de l'ouvrage est l'attention portée aux aspects pratiques : implémentation et mise en œuvre, protocoles, normes et standards, et surtout sécurité des transmissions, sous tous ses aspects.

La signature en aveugle, le vote électronique, l'authentification, le respect de l'anonymat, l'élaboration et la délivrance de certificats sont devenus des problématiques sociétales à la frontière des mathématiques, de l'algorithmique, de l'informatique et de la technologie. Pourtant, bien qu'elles soient amenées à se développer à grande échelle, elles sont rarement soulignées dans les ouvrages. Une place importante leur est réservée ici. Enfin, les auteurs nous alertent longuement sur les nouvelles technologies (puces RFID, Wi-Fi, forums de discussion en ligne, messagerie électronique, réseaux GSM, téléphones portables et *smartphones*, assistants personnels, cartes à puce, réseaux sociaux, sans parler des clés USB…), qui présentent un niveau de sécurité et de confidentialité des échanges au mieux insuffisant, au pire inexistant.

É.T.

par David Delaunay

Les codes-barres décodés

Apparus dans les années 1970, les codes-barres sont devenus indispensables pour la gestion des marchandises, mais aussi pour l'identification du courrier, des bagages, des médicaments et même des patients. Quel est donc leur secret ?

Outre ceux scannés par les hôtesses de caisse, on retrouve les codes-barres sur les palettes de marchandises ou encore sur les courriers que nous expédions. Ces codes-barres permettent d'attribuer une information individuelle ou collective à un produit et ceci au simple coût d'une impression ; un lecteur optique permet ensuite de lire cette information.

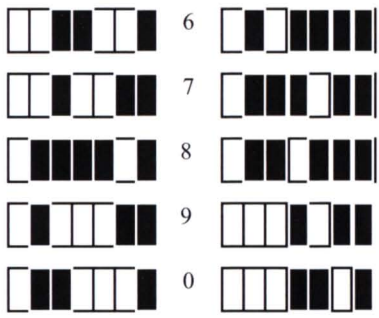
Pour comprendre comment sont définis les codes-barres, nous allons préciser comment est construit le code EAN 13 (*European Article Numbering* à treize chiffres) : c'est celui duquel nous sommes le plus familier car nous le retrouvons sur la plupart des produits manufacturés. Ce sont treize chiffres décimaux qui sont codés par trente barres noires de largeurs variables séparées par des espaces qui sont eux-mêmes de largeur variable. Ces largeurs sont toutes égales à une, deux, trois ou quatre fois une largeur unité que nous appellerons *module*. Ce sont ces largeurs qui déterminent les chiffres codés. Aux extrémités et au milieu du code-barres figurent deux barres noires de largeur égale au module séparées d'un espace de même largeur ; ces barres sont souvent figurées plus longues que les autres comme dans la figure en bas ci-contre. Elles ne contiennent pas d'information relative à la marchandise étiquetée mais permettent au lecteur optique de déterminer la largeur du module définissant le code-barres. En pratique, le lecteur optique peut en effet ne pas être bien parallèle au code-barres, ou celui-ci peut ne pas être imprimé sur une surface plane ; la connaissance du module aux extrémités du code et en son milieu permet alors au lecteur optique d'estimer la valeur du module sur toute la longueur du code-barres.

Codage des chiffres

Entre ces barres déterminant le module sont codés les chiffres associés à notre code-barres. Cependant, ils ne sont pas tous codés de la même façon. En particulier les chiffres de gauche et de droite sont codés différemment : cela permet au lecteur optique de déterminer dans quel sens est lu le code. Il y a exactement trois façons de coder ces chiffres, mais pour chacune le code est formé par l'alternance de deux barres

noires et de deux barres blanches sur une largeur égale à sept modules.

Le premier code, que nous appellerons le *code G*, commence par une barre blanche et se termine par une barre noire. Chaque chiffre y est codé comme suit :



Codage des dix chiffres dans le code G.

Le deuxième code, que nous appellerons le *code D*, commence par une barre noire et est défini en considérant le complémentaire du code précédent. Ainsi, le chiffre 1 est codé de la façon suivante :

Codage du chiffre 1 dans les codes G et D.

Enfin, il existe un troisième code, G', commençant par une barre blanche et finissant par une barre noire (comme le code G), défini en symétrisant les éléments du code D. Le chiffre 1, par exemple, est ainsi codé :

Codage du chiffre 1 dans les codes G' et D.

Bien entendu, il n'y a pas d'ambiguïtés sur ces différents codes et à chaque configuration ne correspondent qu'un chiffre et un codage.

Encodage des treize chiffres

Des treize chiffres que nous devons coder, nous isolons le premier et séparons les chiffres restant en deux blocs de six qui constitueront les portions gauche et droite de notre code-barres. Les chiffres de la portion droite seront simplement transformés en barres par le code D. Pour la portion gauche, c'est plus compliqué : ce sont les codes G et G' qui vont être utilisés. Plus précisément, le premier chiffre que nous avons isolé détermine la position des chiffres qui seront codés par le code G, les autres étant codés par G'. Par exemple, si le premier chiffre du code-barres est égal à 3, il est convenu que les six chiffres de la portion gauche seront codés en suivant la séquence : G G G' G' G' G.

Dans les treize chiffres définissant un code-barres, le dernier est en fait un chiffre de contrôle permettant de vérifier l'intégrité de la lecture du code c'est-à-dire de vérifier qu'il n'y a pas eu (trop) d'erreurs lors du décryptage du code. Ce dernier chiffre est calculé de la façon suivante :
– on somme les chiffres de rang impair, ce qui donne S_1 ;
– on somme les chiffres de rang pair, ce qui donne S_2 ;
– on détermine le chiffre C des unités de $S_1 + 3S_2$.
Le dernier chiffre du code vaut alors $10 - C$.

Voyons ce qui se passe sur un exemple. Considérons la suite des douze chiffres : 311111122222. Calculons d'abord le chiffre de contrôle. On a $S_1 = 10$, $S_2 = 9$, $C = 7$, et donc le chiffre de contrôle est $10 - 7 = 3$. Les treize chiffres que nous allons coder se répartissent ainsi : 3/111111/222223. Isolons le chiffre 3 initial ; les 1 du

bloc de gauche seront codés en G ou G' comme il est dit ci-dessus tandis que les chiffres du bloc de droite seront codés en D. On obtient finalement :

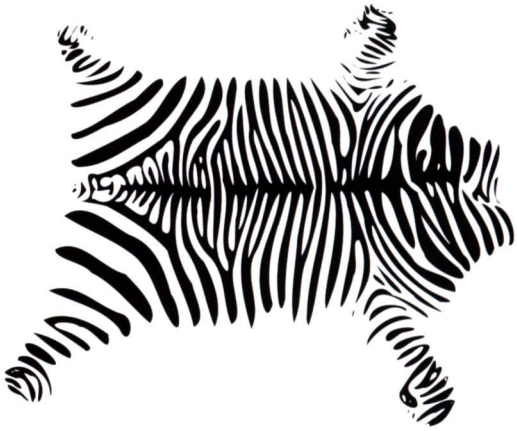
Encodage des treize chiffres 3111111222223.

Ces codes qui nous sont si familiers nous sont désormais (un peu) moins obscurs.

D. D.

Qui peut le plus ?

Sans considérer que son treizième chiffre est un chiffre de contrôle, on peut affirmer qu'un code-barres permet de chiffrer 10^{13} combinaisons. Pour cela, il exploite $7 \times 12 = 84$ modules. Un simple codage binaire permettrait 2^{84} combinaisons. Ce qui représente environ deux mille milliards de fois plus de combinaisons qu'avec les code-barres !



Des codes exhaustifs

Décomposer sept modules en une alternance de quatre barres de longueurs variables revient à déterminer des naturels non nuls x_1, x_2, x_3, x_4 vérifiant :

$$x_1 + x_2 + x_3 + x_4 = 7.$$

(x_1 donne le nombre de modules pour la première barre, x_2 pour la seconde, *etc.*).

Ce problème est celui de la partition d'un entier en une somme. C'est un problème de dénombrement que le mathématicien sait résoudre. Précisément, il existe

$$\binom{p-1}{p-n} = \frac{(p-1)!}{(n-1)!(p-n)!}$$

façons d'écrire un naturel p comme somme de n naturels non nuls, soit $x_1 + \ldots + x_n = p$.

Ici $n = 4$ et $p = 7$: il existe donc exactement vingt écritures possibles de 7 comme somme de quatre naturels non nuls ; on en conclut que les codes G et G' sont exhaustifs, c'est-à-dire qu'ils épuisent toutes les possibilités d'écriture.

Codex C décodé

Heureusement pour nous, et pour la suite de l'histoire, l'effacement du texte du mystérieux palimpseste d'Archimède (voir en page 131) fut incomplet : Heiberg a pu déchiffrer, au prix d'un travail remarquable, 80 % du texte à la simple lumière du jour. Mais les parties cachées par la reliure lui restaient inaccessibles. À ces manques, se rajoutent au XX^{e} siècle l'attaque des moisissures, une restauration catastrophique de la reliure et la destruction de plusieurs pages par ajout d'enluminures pour « valoriser » l'ouvrage.

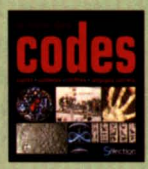
Depuis, les techniques les plus modernes en imagerie, associées à des traitements du signal élaborés, ont été utilisées pour analyser ce document unique. L'étude sous rayonnement ultraviolet met en évidence de nouveaux textes en 2005. En 2006, l'analyse par rayons X avec le synchrotron de l'Université de Stanford permet de mettre en évidence les traces du fer contenu dans l'encre d'origine.

Le codex C nous livre les seules copies connues du *Stomachion*, puzzle combinatoire deux mille ans avant le Tangram, de la *Méthode*, dans lequel Archimède explicite sa technique d'intégration, et la seule copie grecque des *Corps flottants*. En 2007, les analyses montrent que le palimpseste contient également des discours de l'orateur athénien Hypéride, et un commentaire unique d'Alexandre d'Aphrodise sur *les Catégories* d'Aristote.

RÉFÉRENCES BIBLIOGRAPHIQUES

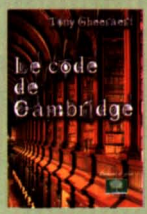
Voici une sélection d'ouvrages qui vous permettront d'en savoir davantage sur la cryptographie et d'approfondir ainsi le sujet :

- *L'Univers des codes secrets – De l'Antiquité à Internet*. Hervé Lehning, Ixelles, 304 pages, 2012, 19,90 euros.



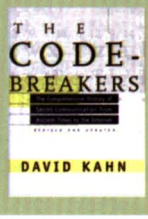
- *Le livre des codes – Signes, symboles, chiffres, langages secrets*. Sélection du Reader's Digest, 280 pages, 2009, 34,95 euros.

- *Le code de Cambridge*. Tony Gheeraert, Le Pommier, 528 pages, 2010, 23,90 euros.



- *Histoire des codes secrets*. Simon Singh, Le Livre De Poche, 504 pages, 2001, 7,10 euros.

- *Comprendre les codes secrets*. Pierre Vigoureux, Ellipses, 302 pages, 2010, 23,75 euros.



- *The Code Breakers – The Comprehensive History of Secret Communication from Ancient Times to the Internet*. David Kahn, Scribner, 1996.

Les codes QR, une autre dimension

Exploités depuis quelques années par les entreprises et depuis peu par les publicitaires, les codes QR sont désormais omniprésents. Ils présentent l'avantage d'utiliser deux dimensions pour le stockage de l'information. Cela permet, avec une surface semblable, de stocker beaucoup plus d'informations qu'avec un code-barres classique.

Trente de Vassily Kandinsky (1937). © Musée National d'Art Moderne, Centre Georges Pompidou.

Les codes QR (*Quick Response*) servent à stocker graphiquement une information sous la forme d'un tableau constitué de cases noires et blanches. Contrairement à un code-barres dans lequel l'information n'est stockée que sur une dimension, les codes QR exploitent les deux dimensions du graphique. L'information y est de plus codée de façon redondante, de sorte que si une partie du code est altérée, celui-ci reste néanmoins compréhensible.

L'information codée se présente sous la forme d'une chaîne de caractères numériques ou alphanumériques. Dans une usine, cette information pourra référencer le produit, son lieu de production, la date de conception… Dans une entreprise de livraison, on pourra noter l'expéditeur, le destinataire, la nature du produit… Ce code QR est destiné à être lu par un scanneur qui décryptera l'information contenue. Le particulier qui rencontre un code QR (dans une publicité ou un lieu publique par exemple) utilisera quant à lui la caméra de son téléphone portable et accèdera à l'information contenue à l'aide d'un logiciel *ad hoc*. Souvent cette information est une adresse Internet, comme « http://tangente/poleditions.com », figurée par le code suivant :

Le logiciel détermine la nature de cette information grâce à la présence dans le

texte du protocole HTTP (*HyperText Transfer Protocol*, soit protocole de transfert hypertexte), et pourra proposer à l'utilisateur l'ouverture de cette page, ce qui lui donnera accès à d'encore plus nombreuses informations. Un code QR peut aussi contenir les informations d'une carte de visite ou, s'il figure sur la plaque d'un immeuble, contenir des informations de géolocalisation permettant de se positionner à l'intérieur d'une grande ville ou d'un circuit touristique.

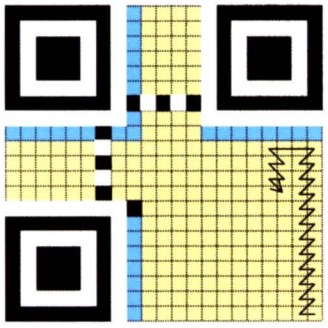
Des damiers et des masques

Les codes QR sont de différentes tailles en fonction de la quantité d'information qu'ils contiennent. Les plus petites grilles sont formées de 21×21 cases et permettent de coder 34 caractères numériques ou 14 alphanumériques ; les plus grandes sont formées de 177×177 cases et permettent de coder plus de 2 000 caractères alphanumériques !

Quelle que soit sa taille, on voit toujours des petits carrés à trois des coins d'un code QR : ceux-ci servent à déterminer la position du code dans l'image scannée, le sens dans lequel celui-ci est présenté, la taille du code et, enfin, la largeur de la grille utilisée. De manière moins apparente, il existe aussi sur une rangée horizontale et une rangée verticale une alternance de cases noires et blanches qui aident le logiciel à déterminer les coordonnées à l'intérieur de la grille (voir figure). Lorsque les grilles deviennent plus grandes, d'autres points de repère apparaissent.

Made in Japan

Les codes QR ont été développés en 1994 par une entreprise japonaise travaillant pour Toyota. Ils constituent une marque commerciale enregistrée par Denso Wave Incorporated. Leur usage étant libre de droit, ils se sont considérablement développés ces dernières années au Japon. D'autres codes bidimensionnels existent comme le *Data Matrix* que l'on rencontre aussi fréquemment et dont le principe de fonctionnement présente des analogies avec les codes QR.



Les cases restantes (jaunes et bleues sur la figure) sont celles qui vont contenir l'information. Les cases bleues servent à en préciser le format (numérique ou alphanumérique), ainsi que le nombre de caractères codés. Les cases jaunes contiennent le message proprement dit, dans sa version binaire, soit une succession de 0 et de 1.

Masque rituel Kwakwaka'wakw.

Une technologie créative

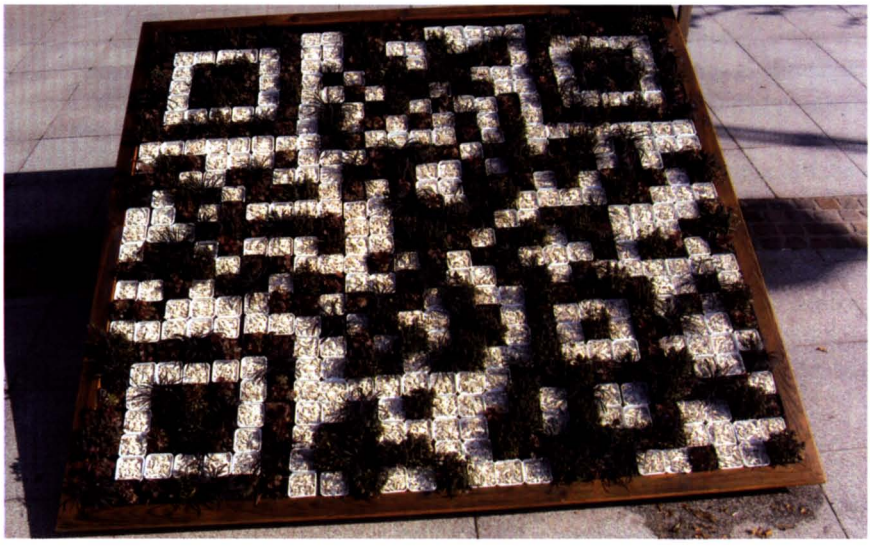
Les codes QR cherchent aujourd'hui à se faire remarquer.

Les codes QR comportent un code correcteur qui permet de les lire même s'ils sont altérés. Cette altération peut être involontaire (rayure, déchirure ou autre) ou... choisie ! On peut ainsi incorporer à un code QR, sans perte d'information à sa lecture, un graphique, le logo d'une entreprise, ou tout autre élément fantaisiste destiné à rendre ce code plus attrayant ou plus original !
Les deux codes suivants renverront bien toujours à la page (fictive) http://tangente/poleditions.com.

Par un algorithme mathématique exploitant calcul polynomial et calcul en congruence, on allonge le message en lui adjoignant une information de contrôle. C'est le codage de Reed–Solomon qui est utilisé (on retrouve celui-ci sur les CD et les DVD) ; il a l'avantage de permettre de reconstruire un message même quand un certain nombre de bits successifs sont altérés. Le message est alors inscrit dans les cases jaunes en suivant le chemin en zigzag de la figure précédente.

Dans un souci d'esthétique et pour éviter une trop grande succession de cases blanches ou de zones noires, on applique plusieurs masques au code produit. Ces masques transforment les cases blanches en des cases noires et inversement. À chacun de ces masques est attribué un indice de qualité esthétique évalué numériquement ; le masque donnant le plus fort indice sera celui finalement retenu. La nature du masque choisi est aussi spécifiée dans les cases bleues, dont toute l'information sera codée avec redondance afin de pouvoir être décryptée même en cas d'altération du code.
Le code est alors prêt à être diffusé puis scanné !

D.D.



© Kirikou

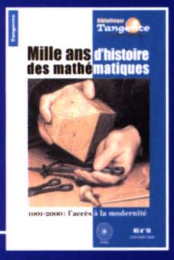
Un code QR végétal (hôtel de ville de Vélizy-Villacoublay).

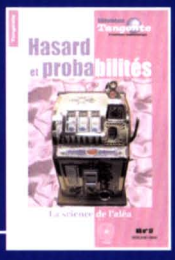
Une nouvelle façon de faire rimer mathématique avec esthétique

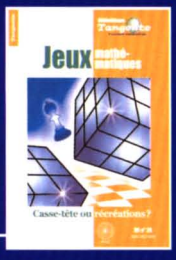
Bibliothèque Tangente

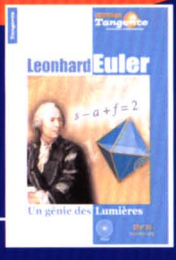
L'aventure mathématique

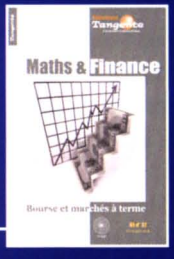
Ces magnifiques ouvrages en couleur de 160 pages feront l'admiration des visiteurs de votre bibliothèque

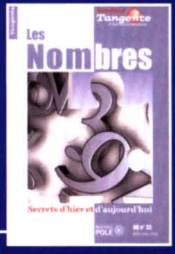


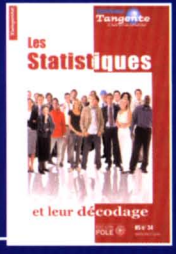


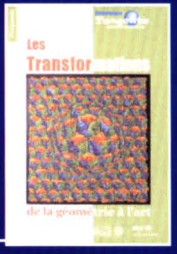


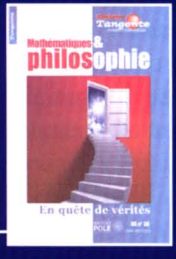


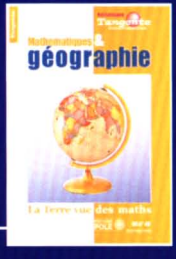


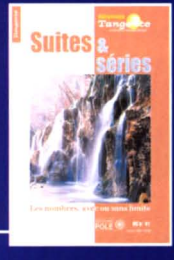


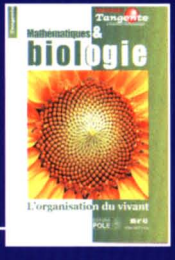


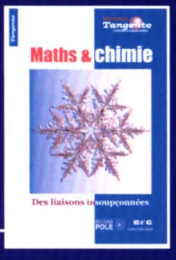


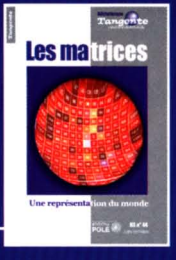












Disponible chez votre libraire ou auprès des Éditions Pole - 80, boulevard Saint-Michel, 75006 Paris

Problèmes à décoder

HS2601 – Les deux messages o

Thomas a reçu un message, mais la machine défectueuse a omis une ou deux barres de chaque lettre de ce mot écrit en script :

Son correspondant, se rendant compte que sa machine « mange » certains traits, renvoie le même message à Thomas, en espérant que ce deuxième message complétera le premier... Hélas, le même mot est transmis ainsi :

Aidez Thomas à reconstituer ce mot.

HS2603 – À la Jules César ✓

Cette citation a été codée à l'aide d'une substitution simple constante (du genre A ⇒ B, B ⇒ C, C ⇒ D, *etc.*, mais peut-être avec un autre décalage !).

OHFULYDLQHVWXQHVRWH-GHYRBDQWHPHUYHLOOHD-QGUHSLHBYHGHPDQGLDUJXHV

HS2604 – Modulo dix ✓✓

Dans ce message codé, chaque lettre a été remplacée par le dernier chiffre de son numéro dans l'alphabet. De plus, on a supprimé les blancs et les signes de ponctuation.

28129045619061925359 45.

Solutions page 156

HS2602 – Le message d'Alice o

Alice n'est pas la seule à être passée au travers du miroir ; ce message aussi !

CEQUIESTAFFIRMÉSANSPREUVEPE
UTÊTRENIÉSANSPREUVE(EUCLIDE).

HS2605 – Code pour mobile ✓✓

Le ou les prénoms et le nom de quatre mathématiciens ont été codés en utilisant les chiffres notés sur les touches d'un téléphone portable (voir le tableau ci-contre).

1	2 ABC	3 DEF
4 GHI	5 JKL	6 MNO
7 PQRS	8 TUV	9 WXYZ

Retrouvez ces quatre mathématiciens.

- **Un Français :** 43674 76462273
- **Un Allemand :** 2275 374337424 42877
- **Un Russe :** 7236889 58684824 82432924333
- **Un Hongrois :** 7285 37367.

HS2606 – À la Collatz ✓✓

Dans le message suivant, on a utilisé un code à substitution, c'est-à-dire que chaque lettre du texte en clair a été remplacée par une autre lettre suivant un certain procédé.

Le procédé utilisé est une substitution variable, inspirée de l'algorithme de Collatz ; on passe de n_k à n_{k+1} de la façon suivante :

- si n_k est pair, alors $n_{k+1} = n_k/2$,
- si n_k est impair, alors $n_{k+1} = 3\,n_k + 1$.

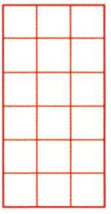
(On rappelle que pour un décalage de $n = 3$, A devient D, B devient E, *etc.*). Ici la valeur de n change à chaque lettre.

À vous de déchiffrer le texte suivant.

Indice : la première valeur est $n_1 = 98$.

IKFADBNCAHEDAZBIRB-MOWLQUWFQBSTPJNFNF.

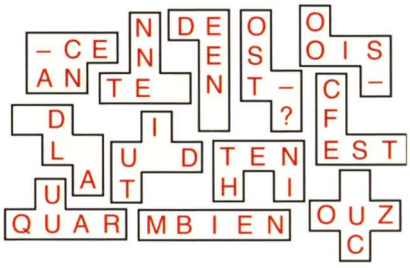
HS2607 – Le rectangle à secrets ✓✓



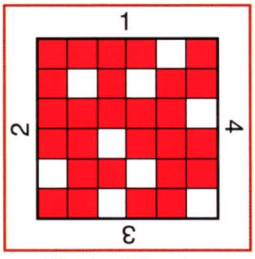
Mathilde et Mathias ont inventé un moyen de communication secret. L'expéditeur écrit le texte dans le rectangle, ligne par ligne, puis le recopie colonne par colonne, en séparant les lettres en trois « mots » de six lettres. Celui qui reçoit le message a vite fait de décoder. Mathilde, pendant le contrôle de mathématiques, a oublié sa calculette. Angoissée, elle adresse à Mathias le message suivant : « S T I U O E E F S A R ? P O Q T Z ».

Quelle doit être la réponse de Mathias (en clair) ?

HS2608 – $12 \times 5 = 6 \times 10$ ✓✓



HS2609 – Une histoire à tourner en rond ✓✓



grille de décodage

E	F	M	N	Q	B	
R	U	D	E	A	R	
I	E	R	E	U	J	
C	E	A	T	A	I	
I	L	E	M	N	P	
P	O	E	R	E	A	

message codé

Alice a envoyé un message codé à Bob. Malheureusement, celui-ci a laissé traîner le message et la grille de décodage. Charles, qui passait par là, a su trouver la signification du message.

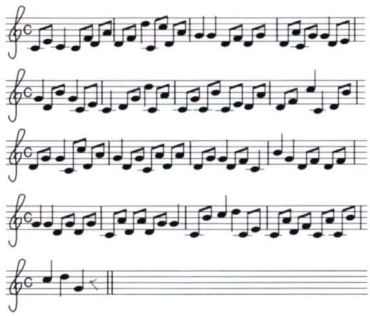
Quelle phrase Alice a-t-elle envoyée ?

HS2610 – Le code des TPP ✓✓

Les enveloppes des lettres destinées à la ville dont le code postal est 0 2 1 0 0 portent la première bande représentée ci-dessous.

Quel est le code postal de la ville pour laquelle la bande est la seconde ?

HS2611 – Cryptocitation musicale ✓✓



Décryptez la citation de Beethoven codée dans la partition ci-dessus.

HS2612 – Code secret ✓✓

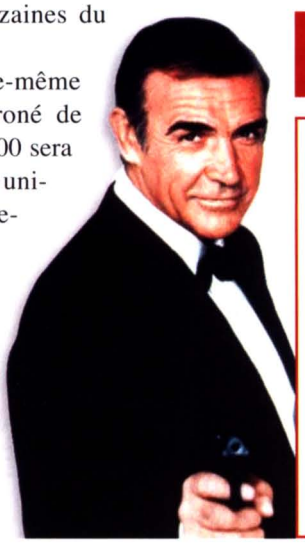
Retrouvez la combinaison du coffre de M. Fochar Jean grâçe aux indications suivantes :

- Une croix indique un chiffre bien placé.
- Un rond indique un chiffre mal placé.

1 9 0 5	o o o
9 0 8 5	x o
8 5 5 0	x
0 8 7 8	x o

HS2613 – Le redresseur de codes ✓✓✓

L'agent 007 est informé du département de sa future mission par deux séries de douze chiffres. Les dix premiers $a_1, a_2, \ldots, a_{10}$ (valeurs de 0 à 9) sont suivis d'une clé $a_{11}a_{12}$.
a_{11} est le reste de la division par 11 de la somme des dix chiffres de gauche.
a_{12} est le reste de la division par 11 de la somme des dix produits $i \times a_i$ d'un chiffre a_i par son numéro i de 1 à 10 ; a_{11} et a_{12} peuvent prendre la valeur 10 notée X.
Par sûreté, chacun des deux codes présente une erreur et une seule sur l'un de ses douze chiffres.
La valeur par laquelle il faut remplacer le chiffre erroné de 258 719 346 07X sera le chiffre des dizaines du département.
La valeur elle-même du chiffre erroné de 049 132 900 000 sera le chiffre des unités du département.
Dans quel département doit officier 007 ?



HS2614– Code secret ✓✓✓

Le coffre-fort de la F.F.J.M. est de la nouvelle génération. Il comprend un clavier électronique n'ayant que quatre touches portant respectivement les chiffres 0, 1, 2 et 9, et un écran de contrôle sur lequel rien n'est affiché. Voici le mode d'emploi permettant l'ouverture de ce coffre :
- La combinaison qui commande l'ouverture doit être tapée en commençant par le chiffre des unités (ou le chiffre le plus à droite), puis le chifre des dizaines, ensuite le chiffre des centaines, *etc*.
- Si l'on tape 0, puis une suite de chiffres X, puis 0 à nouveau, l'écran affiche la suite X (les chiffres de X étant disposés dans l'ordre naturel : ...., centaines, dizaines, unités).
- Si l'on a rentré une suite de chiffres X, et si l'écran affiche une suite Y,
  1) si l'on tape 2, l'écran affichera YY, c'est-à-dire la juxtaposition de deux fois la suite Y ;
  2) si l'on tape 1, l'écran affichera la suite formée en inversant l'ordre des chiffres de la suite Y ;
  3) si l'on tape 9, alors l'écran affichera 0Y, c'est-à-dire le chiffre 0 suivi de la suite Y.

Le coffre ne s'ouvre que si la suite de chiffres affichée à l'écran est identique à la suite qui a été tapée au clavier (dans l'ordre inverse).
Quelle est la plus petite combinaison (le plus petit nombre affiché à l'écran) permettant d'ouvrir le coffre ?

Niveau de difficulté : ❍ très facile ; ✓ facile ; ✓✓ pas facile ; ✓✓✓ difficile ; ✓✓✓✓ très difficile.

Source des problèmes

- Championnat des jeux mathématiques et logiques **(HS2601 ; HS2607 ; HS2608 ; HS2609 ; HS2610)**
- Rallye de l'Université mathématique d'été **(HS2602 ; HS2603 ; HS2604 ; HS2606 ; HS2614)**
- Coupe Euromath des régions **(HS2605)**
- D'après *Jeux & Stratégie* **(HS2611)**
- Tournoi mathématique de Saint-Michel-en-l'Herm **(HS2612)**
- *Y'a pas qu'les maths dans la vie*. Dominique Souder, Aléas, 2002 **(HS2613)**

Solutions

HS2601

CONSTATATION

HS2602

CEQUIESTAFFIRMÊSANSPREUVEPEUTÊTRENIÊSANSPREUVE(EUCLIDE).

HS2603

lécrivainestunesortedevoyantémerveilléandré-pieyredemandiargues

HS2604

LHABITNEFAITPASLEMOINE

HS2605

Henri Poincaré - Carl Friedrich Gauss - Pafnuti Lvovitch Tchebycheff - Paul Erdos.

HS2606

ONASOUVENTBESOINDUN PLUSPETIT-QUESOI.

HS2607

Écrivons le message dans les cases du rectangle, colonne par colonne. Lisons ensuite les lettres ligne par ligne : « SEPTFOISQUATORZE ? ». Mathias doit donc répondre : 98.

S	E	P
T	F	O
I	S	Q
U	A	T
O	R	Z
E	?	

HS2608

La réponse est 70.

C	O	M	B	I	E	N		D	E
F	O	I	S		D	O	U	Z	E
E	S	T	–	I	L		C	O	N
T	E	N	U		D	A	N	S	
H	U	I	T	–	C	E	N	T	–
Q	U	A	R	A	N	T	E		?

HS2609

Quejaimeàfaireapprendrece nombreutile.

HS2610

Le premier code comporte vingt-quatre signes. On peut donc supposer que chaque chiffre du code postal est transcrit à l'aide de six signes. Découpons le code en tranches de six symboles.
On remarque qu'une séquence apparaît trois fois : il s'agit de « point-point-trait-trait-trait-trait ». Il est vraisemblable que cette séquence correspond au chiffre 0 qui lui aussi apparaît trois fois dans 02100 (le code postal de Saint-Quentin, département de l'Aisne, France). Le codage correspondrait alors à 0 0 _ _ 0. On peut alors penser que 02100 est codé en prenant les chiffres dans l'ordre inverse : 0 0 1 2 0. Le chiffre 1 serait codé « point-trait-point-trait-trait-trait » et le chiffre 2 « point-trait-trait-point-trait-trait ».
Selon ces conventions, on lit de gauche à droite 0-0-2-0-1, qui correspond au code postal 10200 (code postal de la ville de Bar-sur-Aube, département de l'Aube, France).

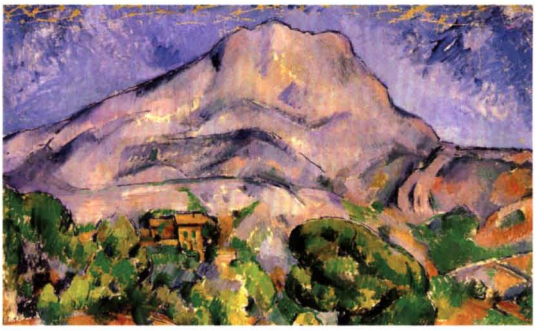
HS2611

La musique est une révélation plus haute que toute sagesse et toute philosophie.

HS2612

9170

HS2613



L'erreur peut être dans la clé ou dans les dix chiffres de gauche. Supposons qu'elle soit dans ceux-ci : il faut soit ajouter 6 soit enlever 5 à l'un des dix chiffres pour obtenir le 1 au lieu du 7. Il faut aussi que dans le remplacement d'un produit $(i \times a_i)$ par un autre, la différence engendrée soit un multiple de 11. Ceci ne peut pas se produire à cause de i qui est inférieur à 11, et la différence de deux coefficients a ne peut valoir 11 (ni 0, ce qui ne serait pas une différence). L'erreur est donc dans la clé : le 7 est faux et la bonne valeur à prendre pour chiffre des dizaines du département est 1.
Code 049 132 900 000 : on calcule la clé qui est 68 et non 00, ce qui fait deux chiffres différents. Comme il ne doit y avoir qu'une erreur, c'est que la clé est juste, et que l'erreur se cache dans les dix premiers chiffres. a_{11} vaut 0 et non 6, le chiffre erroné doit être augmenté de 5 ou diminué de 6. a_{12} vaut 0 et non 8, il faut ajouter 3, ou 14, ou 25, ou 36, ou 47, ou 58… aux produits $i \times a_i$, ou bien leur soustraire 8, ou 19, ou 30, ou 41, ou 52…
En essayant de concilier les deux conditions sur $a_{11}a_{12}$, on ne trouve qu'une possibilité : ajouter 5 au cinquième chiffre, ce qui augmente de 25 la valeur a_{12}. Le vrai cinquième chiffre est donc 8 et non 3. Mais le chiffre des unités du département étant la valeur erronée elle-même, c'est 3. Le département de mission est donc 13.

HS2614

la plus petite combinaison permettant d'ouvrir le coffre est : 1 2 9 1 0 1 2 9 1 0.

Tangente

Publié par Les Éditions POLE
SAS au capital de 42 000 euros

Siège social :
80 bd Saint-Michel – 75006 Paris
Commission paritaire : 1016 K 80883
Dépôt légal à parution

Directeur de Publication et de la Rédaction
Gilles COHEN

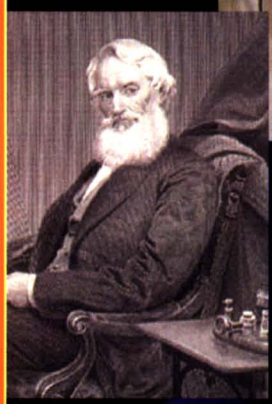
Rédacteur en chef de ce numéro
Hervé LEHNING

Secrétaires de rédaction
Gaël OCTAVIA
Karine BRODSKY (nouvelle édition)

Ont collaboré à ce numéro
Élisabeth BUSSER, Véronique CORTIER, Michel CRITON, David DELAUNAY, Nicolas DELERUE, Isabelle DESIT-RICARD, Jean-Guillaume DUMAS, François LAVALLOU, Abdelkader NECER, Jean-Louis ROCH, Michel ROUSSELET, Édouard THOMAS, Denis TRYSTRAM, Sophie de VAUCORBEIL Alain ZALMANSKI, Chérif ZANANIRI

Maquette & Iconographie
Thibaud DI DOMENICO, Guillaume GAIDOT, Laurence GAUTHIER, Natacha LAUGIER
Photo de couverture : Tobia Revà , *Soglia Celeste*
Autres photos : droits réservés

Abonnements
Tél. : 01 47 07 51 15 – Fax : 01 47 07 88 13
abo@poleditions.com



Abonnez-vous à tangente

l'aventure mathématique

••• Tangente le magazine des mathématiques •••

Pour mieux comprendre le monde : *Tangente*
Le seul magazine au monde sur les mathématiques.
Tous les deux mois depuis 25 ans.

••• Les hors-séries Bibliothèque Tangente •••

Ce sont de magnifiques ouvrages d'en moyenne 160 pages (prix unitaire 19,80€), richement illustrés, approfondissant le sujet du dernier numéro des HS « kiosque » de *Tangente*.
Disponibles
- chez votre libraire
- avec **l'abonnement SUPERPLUS**
- avec **l'abonnement Math++**

à un prix exceptionnel (33% de réduction).

••• Tangente Sup •••

6 numéros par an (ou 4 dont 2 doubles), destinés à ceux qui veulent aller plus loin ou aux étudiants de premier cycle. Dans chaque numéro, un dossier : Surfaces, Groupes, Galois, Prévision…

••• Les hors-séries « kiosque » •••

4 fois par an, un hors-série « kiosque » d'au moins 52 pages, explorent l'actualité des grands dossiers du savoir ou de la culture mathématique.

Les matrices, Théorie des jeux
Disponibles
- chez votre marchand de journaux
- avec **l'abonnement PLUS**
- avec **l'abonnement Math+.**

••• Spécial Logique •••

Nouveau ! Dans la collection Tangente Jeux et Stratégie, un trimestriel contenant près de 200 jeux : tests de logique, grilles à remplir, énigmes mathématiques… Accès numérique gratuit pour les abonnés à la version papier.

••• Tangente Éducation •••

Trimestriel qui traite de thèmes pédagogiques variés : les programmes, les TICE, la formation des enseignants, MathC2+, l'informatique et les sciences du numérique… **Permet l'accès à de nombreuses ressources en ligne.**

---

codif : BIB**26**

Bulletin d'abonnement à retourner à :

Espace Tangente - 80, Bd Saint-Michel - 75006 PARIS

Nom ..... Prénom .....

Établissement .....

Adresse .....

Code Postal ..... Ville .....

Profession ..... E-mail .....

Oui, je m'abonne à	FRANCE MÉTROPOLITAINE		EUROPE	AUTRES
	1 AN	2 ANS	Supplément par an	
TANGENTE	☐ 36 €	☐ 68 €	☐ + 12 €	☐ + 15 €
TANGENTE PLUS	☐ 56 €	☐ 108 €	☐ + 20 €	☐ + 25 €
TANGENTE SUPERPLUS	☐ 88 €	☐ 172 €	☐ + 24 €	☐ + 30 €
TANGENTE SUP	☐ 25 €	☐ 46 €	☐ + 6 €	☐ + 8 €
TANGENTE ÉDUCATION	☐ 12 €	☐ 22 €	☐ + 2 €	☐ + 3 €
SPÉCIAL LOGIQUE	☐ 19,50 €	☐ 37 €	☐ + 8	☐ + 10,50 €
ABONNEMENT MATH+ *	☐ 105 €	☐ 199 €	☐ + 30	☐ + 30 €
ABONNEMENT MATH++ **	☐ 135 €	☐ 260 €	☐ + 32	☐ + 32 €
ABONNEMENT SOUTIEN ***	☐ 155 €	☐ 300 €	☐ + 35 €	☐ + 35 €

* Tous les titres avec les HS « kiosque ». ** Tous les titres avec les HS Bibliothèque. *** Tous les titres avec les deux HS.

Total à payer ☐

Je joins mon paiement par (établissements scolaires, joindre bon de commande administratif) :

❑ Chèque (uniquement **payable en France**)

❑ Carte (à partir de 30 €) numéro : ☐☐☐☐ ☐☐☐☐ ☐☐☐☐ ☐☐☐☐

Date et Signature : crypto : ☐☐☐ Expiration le : ...../.....

Achevé d'imprimer
pour le compte des Éditions POLE
sur les presses de l'imprimerie SPEI à Pulnoy
Imprimé en France
Dépôt légal — Février 2013

Tangente

Cryptographie & codes secrets

- Le temps de l'artisanat
- L'ère industrielle
- L'ère informatique
- Les protocoles cryptographiques

Édition augmentée 2013

L'histoire de la cryptographie est celle d'un combat sans merci entre ceux qui ont quelque chose à cacher et ceux qui aimeraient bien découvrir ce qu'on leur cache. Au temps de César, déjà, les généraux permutaient les lettres des messages envoyés à leurs armées. La méthode s'est perfectionnée pour aboutir, sous l'ère industrielle, à des machines à crypter dont la plus célèbre fut Enigma. Outre le domaine militaire, l'usage d'un code comme le Morse a installé l'idée de communiquer à distance. Aujourd'hui, la cryptographie est omniprésente. Systèmes informatiques, terminaux de cartes bleues, téléphones mobiles sont équipés de protocoles de sécurité que défient les pirates des temps modernes. Sur ce champ de bataille, les armes sont mathématiques et la plus redoutable se nomme factorisation de grands nombres.

9 782848 841403

Prix : 19,80 €

EDITIONS POLE