ARITHMETIQUE ET CODES SECRETS
Un coup d'œil historique

Martine Bühler

Ce conte du lundi doit beaucoup à mes lectures, diverses mais toujours passionnées, sur l'histoire de la cryptographie. Vous trouverez bien sûr à la fin de l'article une bibliographie, mais je tiens à signaler ma dette envers deux ouvrages et un article : les livres de S. Singh et J. Stern et l'article de R. Noirfalise dans *Repères*.

Les premiers systèmes de codes

Jules César a utilisé divers types de codes secrets. Dans *La Guerre des Gaules*[1], il raconte qu'il envoya un message à Cicéron dans lequel les lettres latines étaient remplacées par les lettres grecques correspondantes ; cela suffisait à rendre illisible le message par l'ennemi (pas assez cultivé pour connaître le grec !) mais limpide pour Cicéron. Jules César, conquérant fameux, eut recours à plusieurs codages ; dans *La Vie des douze Césars*[2], Suétone en décrit un autre : on remplace chaque lettre du message par la lettre placée trois rangs après elle dans l'alphabet.

Dans les deux cas, il s'agit d'un codage « monoalphabétique » ou « monograhique ». Chaque lettre est remplacée par un symbole ; dans les deux exemples donnés il s'agissait d'une autre lettre de l'alphabet ou d'une lettre grecque, mais ce pourrait aussi être un symbole n'ayant aucune autre signification ou un dessin.

Mettons-y tout de suite de l'arithmétique, bien que cela soit inutile à ce stade, car, tout ou tard, il faudra bien en venir là !

Chaque lettre de l'alphabet est associé à un nombre comme dans le tableau ci-dessous :

Lettre	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
Nombre associé	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25

Définir un codage monoalphabétique revient à définir une fonction arithmétique de $\{0,1,2,...,23,24,25\}$ dans lui-même ou, pour parler un langage plus moderne de $\mathbb{Z}/26\mathbb{Z}$ dans lui-même. Par exemple, le deuxième procédé de Jules César correspond à l'opération suivante : si x est le nombre associé à une des lettres du message, on définit $f(x) \equiv x+3 \pmod{26}$ avec $0 \le f(x) \le 25$. La lettre cryptée est la lettre associée à f(x) dans le tableau ci-dessus. Les exercices 1 et 2 de l'annexe 1 proposent des codages définis par des fonctions arithmétiques simples.

En fait, on peut définir une permutation des lettres sans congruences, en donnant par exemple un tableau de correspondance entre lettres « en clair » et lettres « cryptées » ; mais encore faut-il être sûr que ce tableau ne tombera pas entre les mains de l'ennemi. Aussi a-t-on souvent utilisé des mots-clefs ou des phrases-clefs. Par exemple, décidons de prendre comme mot-clef MATHEMATIQUES ; ré-écrivons-le sans répétition MATHEIQUS ; nous obtiendrons le tableau de correspondance entre lettres en clair et lettres cryptées de la manière suivante :

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

On a commencé par écrire le mot-clef sans répétition dans le tableau et on a continué à partir de la dernière lettre en suivant l'ordre alphabétique sans répétition. Il suffit donc d'apprendre le mot-clef par cœur et on peut communiquer en sécurité ; on peut même choisir une phrase-clef assez longue. Comme il y a 26 ! permutations possibles des lettres de l'alphabet, ce type de codages paraît difficile à déchiffrer sans le mot-clef ou la fonction arithmétique de codage. Ce mode de communication codée fut donc employé et considéré comme sûr pendant fort longtemps.

Un progrès décisif dans le déchiffrement des messages cryptés fut accompli dans la civilisation arabo-islamique ; les savants arabes se sont intéressés à la linguistique et découvrirent que certaines lettres sont plus utilisées que d'autres. Al Kindi rédigea au IX^{ème} siècle un *Manuscrit sur le déchiffrement des messages cryptographiques* où il utilise cette particularité. Cette méthode porte le nom d'analyse des fréquences. Transposons-la au français : les trois lettres les plus employées en français sont les lettres E, A et I[3] ; en présence

[1] Cesar *De bello Gallico (La Guerre des Gaules)* Traduction L.A. Constans Les Belles Lettres 1937.

[2] Suetone *De Vita Caesarum (Laa Vie des Douze Cesars)* Traduction Henri Ailloud Les Belles Lettres 1961.

[3] En fait, si on consulte des tables de fréquences des lettres en français dans différents ouvrages, on s'aperçoit qu'elles ne concordent pas entièrement ; par exemple, les lettres les plus fréquentes peuvent être E et S au lieu de E et A. Le mieux est d'établir sa propre table à partir de textes du même type que ceux qu'on veut déchiffrer ; en effet, un texte « militaire » ne

d'un texte crypté (dont on a de bonnes raisons de penser qu'il est écrit en français), on compte la fréquence d'apparition de chaque lettre dans le texte crypté. Les trois lettres les plus répandues sont probablement mises pour E, A et I (mais éventuellement dans un ordre différent). On repère également les couples fréquents avec ces trois lettres cryptées : par exemple, si on pense que la lettre H code la lettre E, on repérera dans le texte crypté les lettres qui suivent le plus fréquemment le H et on fera l'hypothèse qu'elles sont mises pour T ou N. Ensuite, on emploie une technique bien connue des amateurs de mots croisés consistant à deviner des mots dont on connaît seulement certaines lettres.

Cette technique permet de venir à bout de tout message crypté grâce à un système monoalphabétique, à condition qu'il soit suffisamment long –une quarantaine de lettres- et composé de mots « normaux ». La méthode fonctionne également si on choisit de remplacer chaque lettre, non pas par une autre lettre, mais par un symbole sans signification particulière (par exemple un rond pour A, un carré pour B, etc.) ; il suffit d'appliquer la méthode des fréquences à ces symboles. Un exemple de ce type de déchiffrement est donné dans *Les hommes dansants*, une nouvelle d'Arthur Conan Doyle où l'on voit Sherlock Holmes déchiffrer des messages où chaque lettre est codée par un « bonhomme » dans une position différente (mais l'analyse est faite en tenant compte de la fréquence des lettres en anglais)[4].

L'Europe à la même époque accuse un retard certain en cryptographie. Le premier livre traitant du sujet est un livre de Bacon au XIII^{ème} siècle . La cryptographie se répand au XIV^{ème} siècle et la cryptanalyse suit au XV^{ème} siècle. On essaie des systèmes dérivés du codage monoalphabétique en le compliquant, mais ces systèmes ne résistent pas à une amélioration de l'analyse des fréquences.

Le système de Vigenère

Afin d'éviter le déchiffrement par analyse statistique, Alberti propose en 1460 d'alterner deux alphabets cryptés par décalage de lettres : par exemple, les lettres ayant un rang pair dans le message en clair seront décalées de 7 et celles de rang impair seront décalées de 11 ; le mot MESSAGE devient donc TPZDHRL. Ainsi la même lettre S est codée différemment. C'est un début de réponse des cryptographes (ceux qui cherchent à rendre les communications illisibles par un supposé espion) aux cryptanalystes (ceux qui cherchent à déchiffrer des messages secrets qui ne leur sont pas *a priori* destinés). Le même type de démarche va mener Vigenère(1523-1596) au système qui porte son nom, qu'il décrit dans son *Traité des chiffres*(1586). Les correspondants choisissent un mot-clef qu'ils gardent secret, par exemple le mot ROSE. Le code consiste à utiliser 4 (nombre de lettres du mot-clef) alphabets translatés : la première lettre du message est translaté de 17 (rang de la lettre R – première lettre du mot-clef – dans l'alphabet « normal »), la deuxième lettre du message est translaté de 14 (rang de O), la troisième de 18 et la quatrième de 4 ; ensuite on recommence le cycle. Par exemple, cryptons les mots CODE SECRET avec le mot-clef ROSE.

Le tableau suivant nous aidera :

présentera pas les mêmes fréquences qu'un roman ou un texte scientifique, car il n'emploie ni le même style, ni les mêmes mots les plus fréquents.

[4] Voir aussi *Le scarabée d'or* d'Edgar Allan Poe.

Mot-clef	R 17	O 14	S 18	E 4	R 17	O 14	S 18	E 4	R 17	O 14
Lettre en « clair »	C	O	D	E	S	E	C	R	E	T
Nombre associé	2	14	3	4	18	4	2	17	4	19
Nombre « translaté »	19	2	21	8	9	18	20	21	21	7
Lettre « cryptée »	T	C	V	I	J	S	U	V	V	H

Ce système permet d'éviter l'analyse de fréquences : la lettre E par exemple est cryptée de trois manières différentes : I, puis S et enfin V. De plus, une même lettre cryptée peut représenter des lettres différentes : V représente D, puis R, puis E. Le système n'est cependant pas utilisé immédiatement car il est jugé trop compliqué à mettre en œuvre ; on lui préfère des codes monoalphabétiques ou dérivés plus simples. Mais au XVIII^{ème} siècle, la cryptanalyse se développe et devient un métier ; les codes monoalphabétiques ne résistent plus au déchiffrement ; les cryptographes se résignent à utiliser le système de Vigenère malgré sa complexité. L'usage des télégraphes au XIX^{ème} siècle accélère la course aux codes « sûrs ». Ce type de codage sera cependant brisé par l'Anglais Babbage au XIX^{ème} siècle. Reprenons l'exemple ci-dessus, avec le mot-clef ROSE : un mot de 3 lettres peut être crypté de 4 manières différentes : par décalages successifs de 17, 14, 18 (correspondant aux rangs des lettres R, O, S dans l'alphabet) si le rang de la première lettre de ce mot dans le message en clair est congru à 1 modulo 4, ou bien par décalages successifs de 14, 18, 4 (correspondant aux rangs des lettres O, S, E) si le rang de la première lettre est congru à 2 modulo 4, ou bien par décalages successifs de 18, 4, 17 (rangs de S, E, R), ou bien par décalages successifs de 4, 17, 14 (rangs de E, R, O). Si ce mot intervient 5 fois dans le message, alors il sera codé 2 fois de la même façon (hé oui ! le fameux principe des tiroirs intervient aussi dans les codes secrets !). Ce qui précède est bien sûr valable pour toute suite de lettres intervenant au moins 5 fois dans le message (5, c'est-à-dire une fois de plus que le nombre de lettres du mot-clef). Venons-en à la méthode de cryptanalyse de Babbage : lorsque le message est suffisamment long, on recherche des séquences de lettres se reproduisant plusieurs fois dans le message crypté. Soit c'est une coïncidence (ce qui est peu probable si on choisit des séquences de 4 lettres ou plus), soit il s'agit d'une même suite de lettres cryptée plusieurs fois de la même façon ; le nombre de lettres séparant deux séquences identiques est alors un multiple du nombre de lettres du mot-clef. Si plusieurs séquences se répètent après des intervalles de n_1, n_2, n_3…, alors P.G.C.D.(n_1,n_2,n_3,…) ou un de ses diviseurs est sans doute le nombre de lettres du mot-clef. Si on a trouvé que ce nombre est d par exemple, on regarde uniquement les lettres du message crypté dont le rang est congru à 1 modulo d et on leur applique la méthode d'analyse des fréquences (qui marche d'autant mieux qu'on connaît la nature de la permutation appliquée aux lettres de l'alphabet : une simple translation), on recommence avec les lettres de rang congru à 2 modulo d, etc. Un exemple de cryptanalyse par cette méthode est proposé dans l'annexe 2.

Babbage n'a pas publié son travail, qui n'est connu que par des papiers non publiés retrouvés au XX^{ème} siècle. Un autre cryptanalyste, Kasiski, est arrivé ultérieurement au même résultat et a publié ses travaux en 1863.

Le cryptage mécanisé

Il faut donc, pour rendre le message indéchiffrable, éviter ces répétitions de séquences de lettres ; les cryptographes trouvent la parade : la clef doit être aussi longue que le message. On peut par exemple choisir les paroles d'une chanson, commode à mémoriser. Mais les cryptanalystes sont gens tenaces et inventifs : ils imaginent la méthode des « mots probables », qu'on retrouvera utilisée de main de maître par les services secrets anglais pendant la deuxième guerre mondiale. L'idée est la suivante : le message contient sûrement des mots courants, par exemple « les », « une », etc. On place ces mots au hasard dans le message crypté et on regarde ce que cela donne pour les lettres de la clef. Si les mots courants sont mal placés, il y a de fortes chances que la suite de lettres obtenue pour la clef soit improbable (par exemple ZGT), alors que, si le mot courant a été bien placé, les lettres du mot-clef semblent cohérentes (par exemple MAT) ; la suite ressemble aux techniques empiriques des cruciverbistes : on essaie de compléter les lettres obtenues pour obtenir des mots qui ont un sens (par exemple, MAT pourraient être le début de MATHEMATIQUES) et on retourne au message codé pour voir ce que cela donne ; si le décodage avec le mot-clef supposé a un sens, on ne s'est probablement pas trompé ; sinon faisons d'autres essais. Le va-et-vient entre la clef et le message codé allié à la technique « cruciverbiste » permet le déchiffrement.

La seule solution pour rendre le message indéchiffrable est de prendre comme clef une suite de lettres choisies au hasard aussi longue que le message (et n'ayant donc aucun sens). Cette fois, la cryptanalyse est impossible : en effet, supposons qu'on dispose d'un message codé de cette manière (mais sans la clef évidemment). Pour tout message en clair ayant le même nombre de lettres, on peut trouver une clef telle que le

cryptage avec cette clef donne le message crypté de départ ; parmi tous ces messages possibles, on ne peut pas savoir lequel est le bon car, la clef ayant été choisie au hasard et n'ayant aucun sens, on ne peut pas s'appuyer sur elle comme précédemment pour vérifier nos hypothèses de déchiffrement. Ce système est donc imparable ! Est-ce à dire qu'il est parfait ?

En fait, ses défauts proviennent de la complexité de sa mise en œuvre. Tout d'abord, il faut une clef différente pour chaque message. Si le cryptanalyste dispose de deux messages différents codés avec la même clef, il peut utiliser sa méthode des « mots probables » : il place au hasard des mots courants dans le premier message, en déduit un morceau de clef possible, puis décrypte une partie du deuxième message avec ce morceau de clef. Si celui-ci a un sens, il essaie de compléter les mots par la technique « cruciverbiste », puis revient au premier message avec une clef complétée et ainsi de suite : le va-et-vient entre les deux messages permet le déchiffrement. Il faut donc établir un grand nombre de longues suites de lettres aléatoires à l'avance. L'expéditeur et le destinataire des messages dispose chacun de deux « carnets de code » identiques ayant des centaines de pages, chaque page comportant une suite aléatoire de centaines de lettres. Le premier message est codé avec la page numéro 1, le deuxième message avec la page numéro 2, etc. On imagine la lourdeur du procédé, le codage et le décodage n'étant par ailleurs pas spécialement rapides ! Et que dire du problème de la distribution des clefs ? Si l'ennemi s'empare d'un carnet de codes (imaginez une armée en campagne où chaque compagnie a son spécialiste chargé de décoder les messages de l'Etat-Major possédant donc le carnet commun), les messages codés seront aussi limpides pour lui que des messages en clair. Ce problème de distribution des clefs est récurrent en cryptographie et nous verrons qu'il ne sera réglé qu'avec le concept de clef publique et l'irruption de l'arithmétique sur le devant de la scène au vingtième siècle.

La complexité du codage est également un obstacle majeur pour des Etats ayant un nombre important de messages à transmettre chaque jour. Dans les années vingt, un inventeur allemand, Scherbius, invente une machine permettant le brouillage automatique des messages : la machine Enigma. Elle dispose d'un clavier ordinaire de machine à écrire sur lequel on tape le message en clair ; chaque lettre est automatiquement décalée pour donner le message crypté. Le décalage change à chaque lettre car la machine possède plusieurs « brouilleurs » électriques connectés les uns aux autres (voir illustration, p.54). Le récepteur et l'émetteur doivent disposer de deux machines identiques et la mettre sur la même position de départ pour communiquer en toute sécurité. Ainsi, les carnets de codes ne servent plus qu'à transmettre un message très court par jour : la position initiale de la machine Enigma. A partir de là, codages et décodages se font automatiquement ; même si l'ennemi s'empare d'une machine, il ne pourra pas décoder les messages, car il ne connaît pas la position initiale (et il y en a de plus en plus :10^{16} au départ et cela augmente dans des proportions astronomiques avec le perfectionnement d'Enigma).

Dans les années trente, l'armée allemande se dote de 30 000 machines Enigma.

Dès 1931, la France se procure les plans de la machine Enigma, mais les services secrets français n'en font aucun usage ; ils les font cependant parvenir aux services de renseignements polonais, qui, se sentant sans doute plus directement menacés par les visées allemandes, vont faire de réels efforts pour déchiffrer Enigma. La mécanisation du système de chiffrement les incite à recruter des scientifiques pour ce travail. Rejewski, mathématicien polonais, analyse les différents aspects du fonctionnement d'Enigma et attaque le système à partir de la répétition du message-clef : le carnet de codes ne contient que la position initiale du jour, servant à réceptionner le message codé ; mais celui-ci n'est pas codé avec la position initiale du carnet de codes, car la multiplication des messages codés de la même façon dans la journée faciliterait la cryptanalyse. Simplement, le message commence par un « message-clef » donnant la position initiale des trois rotors : la suite du message est codé avec cette position initiale, les autres réglages restant ceux du carnet de codes. La position des trois rotors est donnée par une suite de trois lettres, par exemple PFK, et, pour éviter les erreurs à la réception, ce groupe de trois lettres est émis deux fois : le message-clef comporte donc six lettres PFKPFK, codées avec la position initiale donnée par le carnet de codes. Or, cette répétition apporte une faiblesse dans la sûreté du codage ; en effet, les deux groupes de trois lettres PFK sont codés de deux manières différentes car les rotors avancent d'un cran à chaque lettre codée. Ainsi PFKPFK est codée par exemple LTIHVC ; la lettre P est codée successivement par L et H ; les lettres L et H sont donc liées entre elle par le fait que, avec la position initiale choisie, elles codent toutes deux la même lettre , l'une au rang 1 du message et l'autre au rang 4. Si je ne connais pas la position initiale, je sais néanmoins que ces deux lettres sont liées par cette position initiale ; de même pour les lettres T et V, H et C. Rejewski a alors l'idée d'établir un « répertoire » des liaisons : une position initiale étant donnée, quelles sont les liaisons qu'elle induit sur les couples de lettres de l'alphabet ? Les Polonais construisent une version mécanisée de ce répertoire, permettant d'accélérer le décryptage, de la même façon que les machines Enigma permettent d'accélérer le cryptage. Cependant, en 1938, la machine passe à 5 rotors : on commence par choisir 3 rotors parmi les 5 possibles avant de les placer dans leur position initiale ; cela multiplie par 10 le nombre de possibilités et met à mal les méthodes polonaises. En 1939, la Pologne offre aux Alliés deux machines Enigma et les plans des machines à décrypter de Rejewski.

En Angleterre, les services de renseignements lance l'opération «Ultra», transportant à Bletchley Park (Buckinghamshire) le « Government Code and Cypher School », chargé d'intercepter et de décrypter les messages ennemis. On recrute intensément parmi les plus brillants spécialistes : mathématiciens, historiens, linguistes, spécialistes de japonais et d'allemand,…Alan Turing arrive à Bletchley en septembre 1939. Les Anglais travaillent à partir des méthodes élaborées par les Polonais. Mais en 1940, les Allemands suppriment la

répétition du message-clef. Turing opte pour la méthode des « mots probables » ; il recherche des formules courantes (par exemple BULLETIN METEO) pour faire des hypothèses quant au déchiffrement, puis élimine des liaisons impossibles ; il traduit ses raisonnements en un système de relais électriques qui permet d'explorer rapidement les différentes hypothèses et de ne retenir que celles qui sont plausibles. Les Britanniques réussissent ainsi à déchiffrer les messages allemands.

Après la seconde guerre mondiale, le développement de l'informatique impulse de nouvelles recherches en cryptographie. Un message est transformé en une suite de 0 et de 1 et il s'agit de crypter cette suite . Les Etats-Unis adoptent le Data Encryption Standard (DES) à la fin des années soixante-dix. Le DES sépare le message en tranches de 64 bits sur lesquelles on opère des transformations définies par une clef de 64 bits (une suite de 64 termes formée de 0 et de 1 c'est-à-dire un grand nombre écrit en système binaire). Le DES est resté en service jusqu'à nos jours ; le *Monde* annonçait dernièrement qu'il allait être remplacé par un nouveau système inventé par des cryptographes belges (voir article du *Monde* dans l'annexe 5). Cependant, le DES n'échappe pas au problème signalé plus haut de la distribution des clefs.

L'arithmétique au secours de la cryptographie

Avant d'expliquer en quoi consiste le cryptage R.S.A. (acronyme de ses inventeurs Rivest, Shamir, Adleman), devenu célèbre au vingtième siècle, expliquons un système de codage par exponentiation, qui, à ma connaissance, n'a pas été utilisé, mais qui est une bonne introduction aux systèmes actuels. Ce système est très bien expliqué dans l'article de Robert Noirfalise cité en bibliographie.

On choisit un nombre premier p et un nombre entier e tel que e est premier avec p-1. Le couple (p,e) constituera notre clef de codage tenue secrète. Il est hors de question de coder lettre à lettre un message, car ce procédé ne résisterait pas à l'analyse des fréquences. On commence donc par grouper les lettres du message en blocs de m lettres. Un bloc correspond alors à un nombre de la manière suivante : nous avons vu qu'on peut associer à chaque lettre un nombre entre 0 et 25 (qu'on peut considérer comme un nombre à deux chiffres, quitte à rajouter un zéro pour les dizaines éventuellement manquantes : 02 pour la lettre C par exemple). On associe alors à un bloc de lettres le nombre formé en accolant les nombres associés à chaque lettre du bloc : le groupe de lettres SEMA est ainsi associé au nombre 18041200. Comme nous allons travailler modulo p, il est indispensable que les nombres ainsi obtenus soient inférieurs à p, pour éviter que deux blocs de lettres différents ne soient associés à des nombres égaux modulo p, ce qui signifie que p > 2525….25(nombre à 2m chiffres formés de m fois le nombre 25) . La fonction de codage est définie par : $f(x) \equiv x^e \pmod p$ et chaque nombre x associé à un bloc de lettres du message en clair est crypté par le nombre f(x) ainsi obtenu. Pour obtenir la fonction de décodage, on utilise le petit théorème de Fermat. Comme e est premier avec p-1, il existe d tel que : $ed \equiv 1 \pmod{p-1}$; on a alors : $\left[f(x)\right]^d \equiv \left[x^e\right]^d \equiv x^{ed} \pmod{p-1}$ et $ed = 1 + k(p-1)$ (puisqu'on a choisi d tel que $ed \equiv 1 \pmod{p-1}$)donc $x^{ed} = x^{1+k(p-1)} = x^1\left(x^{p-1}\right)^k \equiv x \pmod p$ car $x^{p-1} \equiv 1 \pmod p$. Ainsi la connaissance de e et p permet de calculer d, donc de retrouver x à partir de f(x).

L'exercice de l'annexe 4 explique ce procédé sur un exemple (pris dans l'article de *Repères* déjà cité). L'intérêt de cet exercice est d'être une bonne introduction au système R.S.A. : la théorie en est plus simple. D'autre part, il met en jeu des exponentiations modulo p, ce qui permet d'expliquer aux élèves la méthode d' « exponentiation rapide » .

Ce type de codage résiste à la cryptanalyse, même si on connaît des « mots probables » : on peut imaginer par exemple que le cryptanalyste sache que le nombre y=f(x) du message corresponde à un mot connu (par exemple, il peut savoir que le troisième mot du message est « secret »). Pour une valeur de x, il connaît donc à la fois x et f(x). Peut-il en déduire la clef du message ? En admettant même qu'il connaisse p, cela signifierait qu'il sait résoudre en e l'équation $x^e \equiv y \pmod p$, x , y et p étant connus. Or on ne connaît pas actuellement de méthode permettant d'éviter les essais systématiques pour e, essais dont le coût en termes de temps de calcul sur ordinateurs est élevé.

Ce type de codage est donc efficace, mais ne résout pas le problème de distribution des clés : si on veut communiquer en réseaux, la multiplication des possesseurs de clés augmente les risques de fuite. La réponse à ce problème est apportée par le concept de clé publique, inventé en 1976 par trois chercheurs américains, Diffie, Helmann et Merckle. Jusqu'au vingtième siècle, il a paru évident que le cryptage et le décryptage étaient symétriques : la même clé sert à la fois pour coder et décoder. Or les chercheurs américains pensèrent que ce n'était pas nécessaire : il suffisait de trouver une fonction arithmétique qui ne soit pas inversible ; ainsi, on pourrait rendre publique la fonction de codage, et seul, le destinataire du message pourrait décoder, grâce à une clé secrète permettant d'inverser cette fonction. L'article révolutionnaire de Diffie et Helmann présentait cependant un défaut de taille : ils étaient incapables de donner un exemple d'une telle fonction !

Diffie et Helmann proposèrent malgré tout un concept intéressant : l'échange public de clés secrètes. Nous avons souligné plus haut la difficulté de résoudre le problème de la distribution des clés. Deux personnes désirant communiquer en secret doivent échanger une clé secrète, qui est un grand nombre avec les méthodes modernes de cryptographie (comme le DES par exemple). Voici la procédure imaginée par Diffie et Helmann :

on choisit un grand nombre premier q tel que $\frac{q-1}{2}$ est aussi premier, ce qui rend l'algorithme plus résistant. On choisit également un grand nombre α, de préférence racine primitive modulo q (c'est-à-dire tel que si $0 < n < q-1$ alors $\alpha^n \neq 1 (\text{mod } q)$).Les nombres q et α sont publics. Le personnage A choisit secrètement un grand nombre X_A et le personnage B choisit secrètement un grand nombre X_B ; A calcule $Y_A \equiv \alpha^{X_A} (\text{mod } q)$ et B calcule $Y_B \equiv \alpha^{X_B} (\text{mod } q)$; A transmet Y_A à B et B transmet Y_B à A. La clé secrète commune est alors le nombre $K \equiv Y_B^{X_A} \equiv \left(\alpha^{X_B}\right)^{X_A} \equiv \left(\alpha^{X_A}\right)^{X_B} \equiv Y_A^{X_B} (\text{mod } q)$. La connaissance de q, α, Y_A et Y_B ne permet pas de retrouver les nombres secrets X_A et X_B, donc ne permet pas le calcul de K. La confidentialité repose sur la difficulté de calcul des « logarithmes discrets », c'est-à-dire le calcul de l'exposant d'un nombre x^n connaissant x^n et x modulo q. Mais, si cette méthode permet un échange public de clé (l'échange de Y_A et Y_B pouvant ne pas être secret), elle ne donne pas d'exemple de clé publique de cryptage.

Paradoxalement, c'est en cherchant à montrer l'impossibilité de ce concept de clef publique que Rivest, Shamir et Adleman trouvèrent une fonction convenable en 1978. Le destinataire du message (conventionnellement désigné par le prénom Alice) choisit arbitrairement deux très grands nombres premiers p et q (il existe des programmes d'ordinateurs permettant de le faire) ; il calcule le produit p.q = n et choisit un nombre e premier à (p-1)(q-1). La clef de codage est le couple (n,e) qui peut donc être rendu public. Pour coder un message, on groupe le texte en clair en blocs de m lettres, avec n > 2525…25 (nombre de 2m chiffres). Chaque bloc est associé à un nombre (en accolant les nombres associés à chaque lettre du bloc). Ainsi le message est transformé en une suite de nombres x_1, x_2, x_3... tous strictement inférieurs à n. La fonction de codage est définie par : $f(x) \equiv x^e (\text{mod } n)$. Le message codé est constitué de la suite de nombres $f(x_1), f(x_2), f(x_3)$... Pour déchiffrer le message, il faut déterminer d tel que $ed \equiv 1 (\text{mod}(p-1)(q-1))$ ce qui est possible car on a choisi e premier à (p-1)(q-1) ; seule Alice peut le faire car, si elle a rendu public le couple (n,e), elle a gardé secrètes les valeurs de p et q. Le déchiffrement se fait comme ci-dessus car , si $y = f(x) \equiv x^e (\text{mod } n)$ alors $y^d \equiv x^{ed} (\text{mod } n)$ avec ed = 1 + k(p-1)(q - 1). Donc $x^{ed} \equiv x (\text{mod } p)$ et $x^{ed} \equiv x (\text{mod } q)$ et, comme p et q sont premiers entre eux, $x^{ed} \equiv x (\text{mod } pq)$ c'est-à-dire $x^{ed} \equiv x (\text{mod } n)$. On peut bien sûr utiliser également la forme généralisée par Euler du théorème de Fermat, en remarquant que, dans ce cas, $\varphi(n)=(p-1)(q-1)$, mais il est plus simple avec les élèves de raisonner directement sur p et q, en employant éventuellement le théorème de Gauss : si p divise $x^{ed} - x$ et q également, alors $x^{ed} - x$ = ap = bq ; donc p divise bq et, comme p est premier avec q, p divise b donc $x^{ed} - x$ = b'pq donc n = pq divise $x^{ed} - x$.

Un indiscret ne peut pas se procurer p et q, même connaissant n, car on ne connaît pas actuellement de méthode rapide permettant de factoriser de très grands nombres. La factorisation des grands nombres est un problème qui a occupé les mathématiciens bien avant l'invention de la cryptographie à clef publique et on trouvera dans ce même numéro de *Mnemosyne* un exercice d'arithmétique utilisant une lettre de Fermat à Mersenne où Fermat donne une méthode de factorisation de grands nombres[5] ; mais toutes les méthodes actuellement connues sont impuissantes à factoriser de très grands nombres (plus de 300 chiffres décimaux mais les records tombent malgré tout de temps en temps !). L'hypothèse de la difficulté de ce problème dans l'absolu sert de base à la cryptographie moderne. Cependant, on est actuellement incapable de démontrer cette hypothèse. Rien ne prouve non plus qu'il soit nécessaire de factoriser n pour « casser » le code R.S.A. : en 1998, Boneh et Verkatesan[6] ont montré qu'on peut casser le code sans factoriser n si l'exposant e est trop petit. Et, en admettant qu'un mathématicien travaillant pour des services secrets ait réussi à trouver une méthode efficace de factorisation, il ne l'a sans doute pas publiée !

A défaut de chercher une méthode de cryptanalyse de la méthode R.S.A., le lecteur intéressé pourra s'exercer au décryptage par analyse des fréquences et à la méthode de Babbage pour le système de Vigenère ; vous trouverez en effet dans les annexes 1 et 2 deux textes cryptés inventés par S. Singh qui a lancé un concours de décryptage, et, pour ces deux textes, un exemple de cryptanalyse par les méthodes exposées dans l'article. Les annexes 3 et 4 donnent le texte d'exercices donnés à des élèves de spécialité de Terminale S en 2000 sur ce thème.

[5] Voir également l'étude: *Factorisation de grands nombres*, page 17 dans ce numéro.

[6] D. Boneh et R. Venkatesan *Breaking RSA May Be Easier Than Factoring*, Eurocrypt 1998, LNCS 1403, Springer-Verlag.

Annexe 1 : texte à décrypter par analyse des fréquences

Extrait de *Histoire des codes secrets* de Simon Singh (Lattès 1998)

Cryptanalyse du texte par analyse des fréquences

Dans les six premières lignes, voici le nombre d'apparitions de certaines lettres :
A : 7 B : 14 X : 33 G : 10 J : 13 M : 17 N : 9 T : 14 V : 14 U : 8
Il est raisonnable de penser que X est mis pour e. On considère les mots de deux lettres ; on voit fréquemment XT et XJ ; or les deux mots de deux lettres commençant par e les plus fréquents sont *en* et *et*.Essayons les deux possibilités.
Si X=e, T=t, J=n, le texte commence par : et .en .t.n.tn (les points remplaçant les lettres non encore décryptées). C'est impossible. Que donne l'autre choix ?
Si X=e, T=n, J=t, le texte commence par : en .et .n.t.nt Pourquoi pas ?
Le M très fréquent et visiblement ici mis pour une voyelle est certainement mis pour a. Continuons dans cette voie :

en .et .n.tant

Tout cruciverbiste lit alors : en cet instant
Essayons : X=e T=n J=t M=a A=c B=i R=s
en cet instant, a..a…ent .es ..i.ts .'.ne .ain .'….e et i.s ec.i.i.ent
Remettons les lettres cryptées (en majuscules pour les distinguer des lettres déchiffrées) pour bien voir apparaître les lettres doubles :
aQQaVUVent = apparurent
ecViDiVent=ecrivirent
N'Une = d'une
en cet instant apparurent Ges dCiWts d'une Lain
Et ainsi continue le décryptage maintenant facile : on a reconnu Ges=les, dCiWts=doigts Lain=main.

Annexe 2 : texte à décrypter par la méthode de Babbage

Extrait de *Histoire des codes secrets* de Simon Singh (Lattès 1998)

Cryptanalyse par la méthode de Babbage

Il faut repérer des suites de 4 ou 5 lettres se répétant à l'identique. On remarque DOEOY répétée deux fois avec un écart de 45, NUOC avec un écart de 80. Rappelons que la longueur du mot-clef est un diviseur commun des écarts notés, donc ici de 45 et 80. Le mot-clef a donc probablement 5 lettres. Ceci est confirmé par la répétition de UU avec des écarts de 35 puis 105 lettres. Cela signifie que, si on écrit une lettre sur cinq du message, la partie du message obtenue ainsi est cryptée par un décalage unique à la mode de César. On va donc réécrire le message de façon à faire apparaître cela.

J'ai écrit les cinq premières lettres du message crypté verticalement, puis à côté les cinq suivantes et ainsi de suite. Ainsi la première ligne écrite comporte les lettres 1, 6, 11, 16, etc. du message ; la deuxième ligne comporte les lettres 2, 7, 12,etc. Nous pouvons alors appliquer l'analyse des fréquences à chaque ligne. J'ai commencé par la dernière.

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

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

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

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

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

Dans la dernière ligne, E est manifestement la lettre la plus courante. Faisons l'hypothèse qu'elle est mise pour a ou e ou i.

Si E = a, cela signifie qu'on a chiffré par un décalage de +4 et on déchiffre la dernière ligne par un décalage de –4. Alors U = q. On s'intéresse à cette lettre car les appariements possibles sont rares ; par quoi cet hypothétique q serait-il suivi ? Probablement de la lettre u sauf cas particulier. Or la lettre cryptée suivant un U de la dernière ligne est la lettre cryptée de la colonne suivante à la première ligne (vue la disposition que j'ai adoptée). Examinons différents U de la dernière ligne : ils sont suivis par des lettres différentes : J ou K ou A ; ceci n'est pas possible.

Si E = i alors on déchiffre la dernière ligne par un décalage de +4. Alors M = q. On remarque que M est suivi deux fois par W. Pourquoi pas ? Alors pour la première ligne, on a : W = u. Le déchiffrement de la première ligne se fait par un décalage de –2. Mais alors, à la cinquième ligne, S = w est souvent suivi d'un T (de la première ligne) = r ou de N = l ou de I = g ou de S = q ou de M = k. Tout ceci est impossible.

La seule hypothèse valable pour la dernière ligne est donc E = e. La dernière ligne se déchiffre sans décalage ! La lettre Q de la dernière ligne est plusieurs fois suivie de la lettre M à la première ligne ; donc, on suppose que M = u et que la première ligne se déchiffre par décalage de +8.Essayons ; on écrit le message en déchiffrant deux lettres sur cinq et en écrivant en minuscules les lettres déchiffrées et en majuscules les autres.

sQWOenVJPurUUNusGLMesJINmeUXFquRJBge…

On peut ensuite continuer, soit avec les appariements rares, soit avec les syllabes probables. Par exemple, le premier en est peut-être suivi de t, auquel cas, dans la deuxième ligne, V = t et le déchiffrage se fait par un décalage de –2. Cela donne :

soOWentJPursUNuseLMeshINmesXFquoJBge

Le premier mot serait-il « souvent » ? Alors les lignes 3 et 4 se déchiffre par des décalages de +6 et –1. On obtient :
souventpoursamuserleshommesdequipageprennentdes
souvent pour s’amuser les hommes d’équipage prennent des.

Annexe 3: travaux pour des élèves de terminale scientifique (spécialité maths)
Exercices faits en classe

Systèmes de codages monographiques

Dans ce type de codages, chaque lettre de l'alphabet est transformée par codage en une autre lettre de l'alphabet.

Question : combien y a-t-il de permutations des lettres de l'alphabet ?

Dans la suite, chaque lettre de l'alphabet est associée à un nombre entier compris entre 0 et 25 (à l'aide de son rang dans l'alphabet). Un système de codage monographique est donc défini par une application f de $\{0,1,2,...,25\}$ dans lui-même.

Lettre	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
Nombre associé	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25

EXERCICE 1 : codage par translation ; système de César

Soit x le nombre associé à une lettre de l'alphabet « en clair » et f(x) le nombre associé à la lettre « cryptée ». On considère le codage tel que :

$$f(x) \equiv x+3 \pmod{26} \text{ avec } 0 \le f(x) \le 25 .$$

1°)Coder le mot « CHOIX ».

2°)Décoder le mot « PHVVDJH ».

EXERCICE 2 : codage par transformation affine.

I)Un exemple de codage « affine ».

On considère le codage tel que f est définie par :

$$f(x) \equiv 7x+15 \pmod{26} \text{ avec } 0 \le f(x) \le 25 .$$

1°)Coder le mot « MESSAGE ».

2°)Montrer qu'il existe un unique entier relatif a' tel que :

$$7a' \equiv 1 \pmod{26} \text{ et } 0 < a' < 26.$$

3°) Pour x entier de $\{0,1,2,...,25\}$, on pose : $y = f(x)$.

Montrer : $x \equiv a'.y + b' \pmod{26}$ où b' est un entier relatif à déterminer..

4°) Décoder le message : « GTGRVRCSKTVPCDMR ».

II)Cas général de codage « affine ».

Soient a et b des nombres entiers relatifs. On considère le codage tel que f est définie par : $f(x) \equiv ax + b \pmod{26}$ avec $0 \le f(x) \le 25$. Le codage est utilisable à la condition que deux lettres différentes soient codées différemment.

1°)On suppose : a est premier avec 26.

Montrer qu'alors : $f(x_1) = f(x_2) \Rightarrow x_1 = x_2$.

2°)On suppose : PGCD(a, 26)=d avec d > 1. On a alors : 26 = d.k avec 0 < k < 26.

Montrer que, dans ce cas, $f(k) = f(0)$.

Dans la suite de l'exercice, a est premier à 26.

3°)Fonction de décodage.

a)Montrer qu'il existe a' entier relatif tel que $a.a' \equiv 1 \pmod{26}$.

b)Pour x entier de $\{0,1,2,...,25\}$, on pose : $y = f(x)$.

Montrer : $x \equiv a'.y + b' \pmod{26}$ où b' est un entier relatif à déterminer.

Annexe 4 : travaux pour des élèves de terminale scientifique (spécialité maths)
Devoir à la maison(annexe faite en travaux dirigés en classe)

Dans les deux exercices suivants, on utilisera le « petit théorème de Fermat » :
Soit p un nombre premier et a un entier relatif premier à p , alors : $a^{p-1} \equiv 1 (\mathrm{mod}\, p)$**.**

Comme pour les codages monographiques, chaque lettre de l'alphabet en « clair » est associée à un « équivalent numérique », mais qui a obligatoirement deux chiffres, éventuellement en ajoutant un zéro devant le rang de la lettre.

Lettre	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
Nombre associé	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25

Pour éviter le déchiffrement par analyse statistique, on groupe le message en blocs de m lettres, ce qui donne un nombre de 2m chiffres (en comptant les éventuels zéros en tête de l'écriture). Le codage consiste alors à déterminer une fonction arithmétique f que seul le destinataire du message peut inverser. Les exercices suivants propose deux types de codage actuellement utilisés.

Exercice 1 : codage par exponentiation.

Soit p un nombre premier. Nous allons travailler « modulo p » ; il est donc nécessaire que deux nombres de 2m chiffres obtenus par regroupement en blocs de m lettres d'un message initial soit toujours différents modulo p ; on choisira m le plus grand possible (pour rendre le déchiffrement moins aisé) tel que le plus grand nombre obtenu par regroupement de m lettres soit inférieur strictement à p. Pour m =2 par exemple, le plus grand nombre possible ainsi obtenu est 2525 et on ne pourra faire des blocs de deux lettres que si p>2525.

1°) Compléter le tableau suivant :

Entier premier p choisi	Plus grand entier m convenable
25 <p < 2525	m= 1
2525 < p <	m= 2
	m=
	m=

Remarque : ce tableau est infini.

Pour éviter un décodage facile par analyse des fréquences, il faut donc que p soit un très grand nombre premier.

Dans la suite de l'exercice, e désigne un entier naturel premier à p-1.

Soit x un nombre entier tel que : $0 \le x \le p-1$. On définit la fonction arithmétique f par : $f(x) \equiv x^e (\mathrm{mod}\, p)$ avec $0 \le f(x) \le p-1$. Le codage consiste à remplacer chaque bloc de m lettres par le nombre f(x) correspondant. Le couple (p, e) constitue la clef de codage.

2°)Calcul de x^e modulo p pour e et p donnés.

Nous avons vu en exercice comment calculer x^n modulo p pour n'importe quel entier naturel n. Ce calcul est simple à effectuer lorsque p est petit ; mais, lorsque p est grand, ce qui est toujours le cas dans les problèmes de codage, il faut trouver un moyen de réduire le temps de calcul : comment calculer effectivement, par exemple, $251^{35} (\mathrm{mod}\, 1987)$ avec des opérations qui ne dépassent pas les capacités de la calculatrice et qui ne prennent pas trop de temps ? Une méthode efficace est la suivante :

*on décompose 35 en somme de puissances de 2 : $35 = 2^5 + 2 + 1$.

*on calcule 251^{2^n} modulo 1987 pour n=0 à 5 par élévations au carré successives de la manière suivante :

$$251 \equiv 251 (\mathrm{mod}\, 1987)$$

$$251^2 \equiv 1404 (\mathrm{mod}\, 1987)$$

$$251^{2^2} = \left(251^2\right)^2 \equiv 1404^2 \equiv 112 (\mathrm{mod}\, 1987)$$

$$251^{2^3} = \left(251^{2^2}\right)^2 \equiv 112^2 \equiv 622 (\mathrm{mod} 1987)$$

$$251^{2^4} = \left(251^{2^3}\right)^2 \equiv 622^2 \equiv 1406 (\mathrm{mod} 1987)$$

$$251^{2^5} = \left(251^{2^4}\right)^2 \equiv 1406^2 \equiv 1758 (\mathrm{mod} 1987)$$

*on calcule enfin $251^{35} (\mathrm{mod}\, 1987)$:

$$251^{35} = 251^{2^5} \times 251^2 \times 251 \equiv 1758 \times 1404 \times 251 (\mathrm{mod}\, 1987)$$
$$\equiv 378 \times 251 \equiv 1489 (\mathrm{mod}\, 1987)$$

Nous avons effectué en tout 5 élévations au carré et 2 multiplications modulo 1987 ; nous sommes loin des 34 multiplications nécessaires pour élever à la puissance 35 .

Pour voir si vous avez compris, expliquez comment calculer 304^{29} modulo2633 et faites les calculs intermédiaires avec votre calculatrice. Les calculatrices disposent en général d'une fonction donnant directement le reste de la division euclidienne d'un nombre a par un nombre b(par exemple, dans la TI80, on trouve dans le menu MATH NUM la fonction REMAINDER ; REMAINDER(304^2,2633) vous donne le reste de la division euclidienne de 304^2 par 2633,i.e.261).

3°)Codage d'un message.

Dans cette question, p=2633 et e=29.

a)Comment faudrait-il procéder pour vérifier que p est premier à l'aide de la table de nombres premiers de votre livre de spécialité (donner les explications sans faire les calculs) ?

b)Vérifier que 2632 et 29 sont premiers entre eux.

c)Quelle est la valeur de m correspondant à p ? p est-il suffisamment grand pour éviter un décodage par analyse des fréquences ?

d)Coder le message : « CODE SECRET ».Pour cela :

*Ecrire le message groupé en blocs de m lettres sans tenir compte de l'espace et, sous le message ainsi écrit, noter l'équivalent numérique de chaque bloc.

*Elever chaque nombre correspondant à un bloc à la puissance 29 modulo 2633.

L'élévation à la puissance 29 modulo 2633 est suffisamment fastidieuse pour mériter un traitement informatique. Voir annexe avant de faire les calculs.

4°)Décoder un message. Dans cette question, p=2633 et e=29.

a)Montrer qu'il existe un entier relatif u tel que : $29u \equiv 1(\mathrm{mod}(2632))$. En déduire qu'il existe un unique entier naturel d tel que : $29d \equiv 1(\mathrm{mod}(2632))$ avec $0 \le d < 2632$ et calculer d.

b)En déduire : $f(x)^d \equiv x(\mathrm{mod}\, 2633)$. La fonction de décodage est donc définie par : $g(y) \equiv y^d (\mathrm{mod}\, 2633)$ avec $0 \le g(y) < 2633$.

b)Décoder le message :

0500	1868	0951	0815	2165	0680	1130

Pour cela, il faut élever chaque nombre du tableau à la puissance d modulo 2633. Expliquer la marche à suivre et utiliser Excel pour le faire.

5°)Cas général.

On rappelle que p est un nombre premier quelconque et que e est un nombre entier positif premier avec p-1.

a)Montrer qu'il existe un entier relatif u tel que : $e.u \equiv 1(\mathrm{mod}(p-1))$. En déduire qu'il existe un unique entier naturel d tel que : $e.d \equiv 1(\mathrm{mod}(p-1))$ avec $0 \le d < p-1$.

b)En déduire : $f(x)^d \equiv x(\mathrm{mod}\, p)$. La fonction de décodage est donc définie par : $g(y) \equiv y^d (\mathrm{mod}\, p)$ avec $0 \le g(y) < p$.

Ce type de codage résiste bien à la cryptanalyse si p est suffisamment grand. Cependant, si les messages doivent circuler sur un réseau, un nombre important de personnes doivent connaître les clés de codage, ce qui augmente les risques de fuite. L'exercice 2 présente un moyen d'éliminer ce défaut.

Exercice 2 : codage à clés publiques ; système R.S.A.

La clé de codage est un couple d'entiers naturels (e,n) rendu public.

On a : n = p.q où p et q sont des nombres premiers distincts et très grands. La décomposition de n en facteurs premiers n'est connue que de la personne destinataire du message. Le nombre e est premier avec (p-1)(q-1).

Pour coder le message, on découpe le texte en blocs de m lettres, chaque bloc ayant un équivalent numérique comme dans l'exercice 1. Nous allons travailler modulo n, d'où une condition sur m en fonction de n.

Si x est l'équivalent numérique d'un bloc, la fonction de codage est définie par : $f(x) \equiv x^e \pmod{n}$ avec $0 \leq f(x) < n$.

1°)Montrer qu'il existe un unique entier naturel d tel que : $e.d \equiv 1 \pmod{(p-1)(q-1)}$ avec $0 \leq d < (p-1)(q-1)$.

2°)a)Montrer : $x^{ed} \equiv x \pmod{p}$ et $x^{ed} \equiv x \pmod{q}$.

b)En déduire : $x^{ed} \equiv x \pmod{n}$.

3°)Quelle est la fonction de décodage ?

Annexe au problème(aidant à traiter la question 3°) de l'exercice 1)

1°)Ecrire un algorithme permettant à une calculatrice programmable de calculer N^{29} modulo 2633 pour N quelconque. L'algorithme commence par : $? \rightarrow N$ afin de pouvoir entrer n'importe quelle valeur pour N ; utiliser la fonction REMAINDER pour simplifier l'écriture de l'algorithme.

2°)Voici enfin l'occasion d'apprendre à se servir d'un tableur !

Tous les ordinateurs du lycée disposent d'un tableur[7], capable d'exécuter des calculs comme une calculatrice programmable. Nous allons nous en servir pour calculer 304^{29} modulo 2633. Voici la marche à suivre intelligemment, c'est-à-dire en essayant de comprendre ce qui se passe :

Ouvrir une feuille de calcul ; on voit apparaître un tableau à double entrée dont chaque case s'appelle une « cellule ». Pour sélectionner une cellule, on clique dessus ; pour sélectionner un groupe de cellules, on clique sur la première du groupe et, en maintenant le bouton de la souris enfoncé, on glisse jusqu'à la dernière où on relâche le bouton. Les cellules sélectionnées sont en surbrillance.

Commençons par indiquer sur la feuille de calculs ce que nous calculons :

Sélectionner la cellule A1 puis taper n puis entrée.

Sélectionner la cellule B1 puis taper a^(2^n)mod2633 puis entrée.

Maintenant, nous allons programmer les calculs. Il est nécessaire pour comprendre d'avoir traité la question 2°) de l'exercice 1.

Sélectionner la cellule A2 puis taper 0 puis entrée.

Sélectionner la cellule A3puis taper =A2+1 (pour taper A2 dans cette formule, cliquer sur la cellule A2), puis taper entrée.

Sélectionner les cellules A3 jusqu'à A6, puis cliquer sur Edition et, en maintenant le bouton de la souris enfoncé, glisser jusqu'à recopier, puis dans le sous-menu qui s'ouvre alors, glisser jusqu'à en-bas et relâcher le bouton de la souris. Examiner ce qui se passe.

Sélectionner B2 ; taper 304 puis entrée.

Sélectionner B3 ; taper = ; dans le menu Insertion, choisir Fonction puis math&trigo et MOD ; lire les indications données et mettre dans B3 =MOD(B2^2,2633) puis entrée.

Sélectionner les cellules B3 à B6 et recopier vers le bas.

Avec la même procédure, mettre dans C2 =MOD(B2*B4,2633) et dans C3 =MOD(B5*B6,2633).

Que faut-il mettre dans C6 pour obtenir le résultat final, c'est-à-dire 304^{29} modulo 2633 ?

Si on remplace 304 par 1502, qu'obtient-on ?

[7] Cette annexe a été établie pour des élèves du lycée Flora Tristan(93 – Noisy-le-Grand), alors que nous disposions du tableur Excel. Les instructions peuvent différer légèrement pour un autre tableur(Lotus par exemple).

Annexe 5 : articles du *Monde*.

Un nouvel algorithme de cryptage belge s'impose aux Etats-Unis

Cette victoire improbable a réjoui les spécialistes européens de la cryptographie, cette « science du secret » qui consiste à coder et décoder efficacement des données. Au terme d'une compétition internationale qui a duré trois ans, le département du commerce américain a choisi, lundi 2 octobre, un algorithme de cryptage belge, baptisé Rijndael, pour succéder au DES (Data Encryption Standard).

…

Contrairement au DES, qui avait été développé en grand secret par IBM,l'AES a été choisi au terme d'une procédure transparente particulièrement sévère.

Le Monde octobre 2000

www.cryptonline.com

Crypter n'importe quel document, gratuitement, rapidement et sans formalité

…Le service, baptisé « Cryptonline »est gratuit, automatique, ouvert à tous, et ne prend que quelques secondes….Cryptonline utilise un cryptage à 56 bits, c'est-à-dire de moyenne puissance, faisant appel à des algorithmes standard….Cryptonline n'est donc pas fait pour protéger les secrets d'Etat

Le Monde 18 janvier 2001

Bibliographie

Simon SINGH *Histoire des codes secrets* Ed . J.C.Lattès paris 1999

Jacques STERN *La science du secret* Ed. Odile Jacob Paris 1998

Jean-Paul DELAHAYE *Merveilleux nombres premiers* Ed. Belin-Pour la Science Paris 2000

Robert NOIRFALISE *Arithmétique et cryptographie* in *Repères* n°37 octobre 1999 (pages 41-62)

Martin HELMANN *Les mathématiques de la cryptographie à clef révélée* in *Pour la Science 1979*

Jean-Paul DELAHAYE *La cryptographie RSA vingt ans après* in *Pour la Science* n° 267 Janvier 2000