Théorie de la communication des systèmes de chiffrement

C. E. Shannon

1. Introduction et résumé

Les problèmes de cryptographie et de systèmes de chiffrement sont des applications intéressantes de la théorie de la communication [1]. Cet article développe une théorie des systèmes de chiffrement. L'approche se situe au niveau théorique et vise à compléter le traitement des ouvrages de référence sur la cryptographie [2]. Une étude détaillée des nombreux types standard de codes et de chiffrements, ainsi que des moyens de les déchiffrer, est menée. Nous nous intéresserons davantage à la structure mathématique générale et aux propriétés des systèmes de chiffrement.

Le traitement est limité à certains égards. Premièrement, il existe trois grands types de systèmes de chiffrement : (1) les systèmes de dissimulation, incluant des méthodes telles que l'encre invisible, la dissimulation d'un message dans un texte innocent, ou dans un faux cryptogramme de couverture, ou d'autres méthodes dissimulant l'existence du message à l'ennemi ; (2) les systèmes de confidentialité, par exemple l'inversion de la parole, dans lesquels un équipement spécial est nécessaire pour récupérer le message ; (3) les "véritables" systèmes de chiffrement où la signification du message est dissimulée par un chiffrement, un code, etc., bien que son existence ne soit pas cachée, et où l'ennemi est supposé disposer de l'équipement spécial nécessaire pour intercepter et enregistrer le signal transmis. Nous ne considérerons que le troisième type : les systèmes de dissimulation posent principalement un problème psychologique, tandis que les systèmes de confidentialité posent essentiellement un problème technologique.

Deuxièmement, le traitement dans le présent article se limitera au cas d'informations discrètes où le message à chiffrer est constitué d'une séquence de symboles discrets, chacun choisi dans un ensemble fini. Ces symboles peuvent être des lettres dans une langue, des mots d'une langue, des niveaux d'amplitude d'un signal vocal ou vidéo "quantifié", etc., mais l'accent et la réflexion principaux ont été portés au cas des lettres.

Cet article est divisé en trois parties. Les principaux résultats seront brièvement résumés. La première partie traite de la structure mathématique fondamentale des systèmes secrets. Comme en théorie de la communication, une langue est considérée comme représentée par un processus stochastique produisant une séquence discrète de symboles selon un système de probabilités. Un paramètre D est associé à une langue, appelé redondance de cette langue. D mesure, en quelque sorte, de combien il est possible de réduire la longueur d'un texte dans cette langue sans perdre d'information. À titre d'exemple, comme u suit toujours q dans les mots anglais, le u peut être omis sans perte. Des réductions considérables sont possibles en anglais en raison de la structure statistique de la langue, de la fréquence élevée de certaines lettres ou mots, etc. La redondance est

Le contenu de cet article est paru dans un rapport confidentiel "*A Mathematical Theory of Cryptography*" daté du 1er septembre 1946, qui a maintenant été déclassifié.

Transcription en LaTeX et traduction assistée de Google Translate : Denise Vella-Chemla, juin 2025.

[1] Shannon, C. E., "*A Mathematical Theory of Communication*", Bell System Technical Journal, juillet 1948, p. 379 ; octobre 1948, p. 623.

[2] Voir, par exemple, H. F. Gaines, "*Elementary Cryptanalysis*" ou M. Givierge, "*Cours de Cryptographie*".

d'une importance capitale dans l'étude des systèmes secrets.

Un système secret est défini abstraitement comme un ensemble de transformations d'un espace (l'ensemble des messages possibles) en un second espace (l'ensemble des cryptogrammes possibles). Chaque transformation particulière de l'ensemble correspond à un chiffrement avec une clé particulière. Les transformations sont supposées réversibles (non singulières), de sorte qu'un déchiffrement unique est possible lorsque la clé est connue.

Chaque clé, et donc chaque transformation, est supposée associée à une probabilité a priori : la probabilité de choisir cette clé. De même, chaque message possible est supposé associé à une probabilité a priori, déterminée par le processus stochastique sous-jacent. Ces probabilités pour les différentes clés et messages sont en réalité les probabilités a priori du cryptanalyste adverse pour les choix en question et représentent sa connaissance a priori de la situation.

Pour utiliser le système, une clé est d'abord sélectionnée et envoyée au point de réception. Le choix d'une clé détermine une transformation particulière dans l'ensemble constituant le système. Ensuite, un message est sélectionné et la transformation particulière correspondant à la clé sélectionnée est appliquée à ce message pour produire un cryptogramme. Ce cryptogramme est transmis au point de réception par un canal et peut être intercepté par l'ennemi. [3]." À l'extrémité réceptrice, l'inverse de la transformation particulière est appliqué au cryptogramme pour récupérer le message d'origine.

Si l'ennemi intercepte le cryptogramme, il peut calculer les probabilités a posteriori des différents messages et clés possibles qui auraient pu produire ce cryptogramme. Cet ensemble de probabilités a posteriori constitue sa connaissance de la clé et du message après l'interception. La "connaissance" est ainsi identifiée à un ensemble de propositions auxquelles sont associées des probabilités. Le calcul des probabilités a posteriori constitue le problème généralisé de la cryptanalyse.

À titre d'exemple, dans un chiffrement par substitution simple à clé aléatoire, il y a 26 ! transformations correspondant aux 26 ! différentes manières de substituer les 26 lettres de l'alphabet. Elles sont toutes également probables et ont donc chacune une probabilité a priori de $1/26!$. Appliqué à l'anglais normal, cela revient à dire que la probabilité a priori est de $1/26$. Le cryptanalyste étant supposé ignorer la source du message, si ce n'est qu'il produit du texte anglais, les probabilités a priori des différents messages de N lettres ne sont que leurs fréquences relatives dans un texte anglais normal.

Si l'ennemi intercepte N lettres de cryptogrammes dans ce système, ses probabilités changent. Si N est suffisamment grand (par exemple, 50 lettres), il existe généralement un seul message dont la probabilité a posteriori est proche de l'unité, tandis que tous les autres ont une probabilité totale proche de zéro. Il existe donc une "solution" essentiellement unique au cryptogramme. Pour des messages plus petits (par exemple $N = 15$), il existe généralement de nombreux messages et clés de probabilité comparable, sans qu'aucun ne soit proche de l'unité. Dans ce cas, il existe plusieurs "solutions" au cryptogramme.

[3]Le mot "ennemi", issu des applications militaires, est couramment utilisé dans le travail cryptographique pour désigner toute personne susceptible d'intercepter un cryptogramme.

Si l'on considère un système secret représenté de cette manière, comme un ensemble de transformations d'un ensemble d'éléments en un autre, deux opérations de combinaison naturelles produisent un troisième système à partir de deux systèmes donnés. La première opération de combinaison est appelée opération de produit et correspond au chiffrement du message par le premier système secret R et au chiffrement du cryptogramme résultant par le second système S, les clés pour R et S étant choisies indépendamment. Cette opération totale est un système secret dont les transformations sont constituées de tous les produits (au sens habituel de produits de transformations) de transformations de S par des transformations de R. Les probabilités sont les produits des probabilités des deux transformations.

La deuxième opération de combinaison est l'"addition pondérée".

$$T = pR + qS \qquad\qquad p + q = 1$$

Elle correspond à un choix préliminaire : utiliser le système R ou S avec les probabilités p et q, respectivement. Dans ce cas, R ou S sont utilisés de la manière qu'on a définie initialement.

Il est démontré que les systèmes de chiffrement avec ces deux opérations de combinaison forment essentiellement une "algèbre associative linéaire" à élément unitaire, une variété algébrique largement étudiée par les mathématiciens.

Parmi les nombreux systèmes de chiffrement possibles, il en existe un qui possède de nombreuses propriétés particulières. Ce type est appelé système "pur". Un système est pur si toutes les clés sont équiprobables et si, pour trois transformations T_i, T_j, T_k quelconques de l'ensemble, le produit

$$T_i T_j^{-1} T_k$$

est également une transformation de l'ensemble. Autrement dit, chiffrer, déchiffrer et chiffrer avec trois clés quelconques doivent être équivalents à chiffrer avec une clé donnée.

Avec un chiffrement pur, il est démontré que toutes les clés sont essentiellement équivalentes : elles conduisent toutes au même ensemble de probabilités a posteriori. De plus, lorsqu'un cryptogramme donné est intercepté, il existe un ensemble de messages susceptibles d'avoir produit ce cryptogramme (une "classe résiduelle") et les probabilités a posteriori des messages de cette classe sont proportionnelles aux probabilités a priori. Toutes les informations obtenues par l'ennemi en interceptant le cryptogramme constituent une spécification de la classe résiduelle. De nombreux chiffrements courants sont des systèmes purs, incluant une simple substitution avec clé aléatoire. Dans ce cas, la classe résiduelle comprend tous les messages présentant le même schéma de répétition de lettres que le cryptogramme intercepté.

Deux systèmes R et S sont définis comme "similaires" s'il existe une transformation fixe A ayant un inverse, A^{-1}, telle que

$$R = AS.$$

Si R et S sont similaires, une correspondance biunivoque entre les cryptogrammes obtenus peut être établie, conduisant aux mêmes probabilités a posteriori. Les deux systèmes sont identiques d'un point de vue cryptanalytique.

La deuxième partie de l'article traite du problème du "secret théorique". Dans quelle mesure un système est-il protégé contre la cryptanalyse lorsque l'ennemi dispose d'un temps et d'une main-d' œuvre illimités pour analyser les cryptogrammes interceptés ? Ce problème est étroitement lié aux questions de communication en présence de bruit, et les concepts d'entropie et d'équivoque développés pour ce problème trouvent une application directe dans ce domaine de la cryptographie.

Le "secret parfait" est défini en exigeant d'un système qu'après l'interception d'un cryptogramme par l'ennemi, les probabilités a posteriori de ce cryptogramme représentant divers messages soient identiques aux probabilités a priori de ces mêmes messages avant l'interception. Il est démontré que le secret parfait est possible, mais nécessite, si le nombre de messages est fini, le même nombre de clés possibles. Si le message est considéré comme constamment généré à un "taux" R donné (à définir ultérieurement), la clé doit être générée à un taux identique ou supérieur.

Si un système de chiffrement à clé finie est utilisé et que N lettres du cryptogramme sont interceptées, l'ennemi disposera d'un certain ensemble de messages avec certaines probabilités que ce cryptogramme pourrait représenter. À mesure que N augmente, le champ se rétrécit généralement jusqu'à ce qu'il existe finalement une "solution" unique au cryptogramme ; un message avec une probabilité essentiellement égale à un, tandis que tous les autres sont pratiquement nuls. On définit une quantité $H(N)$, appelée équivoque, qui mesure statistiquement la proximité du cryptogramme moyen de lettres avec une solution unique ; autrement dit, le degré d'incertitude de l'ennemi quant au message original après l'interception d'un cryptogramme de N lettres. On en déduit diverses propriétés de l'équivoque ; par exemple, l'équivoque de la clé n'augmente jamais avec N. Cette équivoque constitue un indice de secret théorique ; il est théorique dans la mesure où il laisse à l'ennemi un temps illimité pour analyser le cryptogramme.

La fonction $H(N)$ pour un certain type idéalisé de chiffrement, appelé chiffrement aléatoire, est déterminée. Avec certaines modifications, cette fonction peut être appliquée à de nombreux cas pratiques. Cela permet de calculer approximativement la quantité de données interceptées nécessaire pour résoudre un système secret. Il ressort de cette analyse qu'avec les langages ordinaires et les types habituels de chiffrements (et non de codes), cette distance d'unicité est approximativement $H(K/D)$. Ici, $H(K)$ est un nombre mesurant la taille de l'espace de clés. Si toutes les clés sont a priori également probables, $H(K)$ est le logarithme du nombre de clés possibles. D est la redondance du langage et mesure la quantité de contrainte statistique imposée par le langage. En substitution simple avec clé aléatoire, $H(K)$ est égal à $\log_{10} 26\,!$, soit environ 20, et D (en chiffres décimaux par lettre) est égal à environ $0,7$ pour l'anglais. L'unicité est donc obtenue à environ 30 lettres.

Il est possible de construire des systèmes de chiffrement à clé finie pour certains langages dans lesquels l'équivoque ne s'approche pas de zéro lorsque $N \to \infty$. Dans ce cas, quelle que soit la quantité de données interceptées, l'adversaire n'obtient pas une solution unique au chiffrement, mais dispose de nombreuses alternatives, toutes de probabilité raisonnable. De tels systèmes sont appelés systèmes idéaux. Il est possible, dans n'importe quel langage, d'approcher ce comportement, c'est-à-dire de faire en sorte que la distance qui sépare $H(N)$ de zéro diminue jusqu'à une valeur arbitrairement grande de N. Cependant, ces systèmes présentent un certain nombre d'inconvénients, tels que la complexité et la sensibilité aux erreurs de transmission du cryptogramme.

La troisième partie de l'article porte sur le "secret pratique". Deux systèmes ayant la même taille de clé peuvent tous deux être résolus de manière unique lorsque N lettres ont été interceptées, mais diffèrent considérablement quant à la quantité de travail nécessaire pour parvenir à cette solution. Une analyse des faiblesses fondamentales des systèmes de chiffrement est effectuée. Cela conduit à des méthodes de construction de systèmes dont la résolution nécessitera un travail considérable. Enfin, une certaine incompatibilité entre les différentes qualités souhaitables des systèmes secrets est abordée.

Partie I
Structure mathématique des systèmes secrets
2. Systèmes secrets

Comme première étape dans l'analyse mathématique de la cryptographie, il est nécessaire d'idéaliser la situation de manière appropriée et de définir de manière mathématiquement acceptable ce que nous entendons par système de chiffrement. Un diagramme "schématique" d'un système de chiffrement général est présenté dans la figure 1. À l'extrémité émettrice, il y a deux sources d'information : une source de message et une source de clé. La source de clé produit une clé particulière parmi celles qui sont possibles dans le système. Cette clé est transmise par un moyen, supposé non interceptable, par exemple par messager, à l'extrémité réceptrice. La source de message produit un message (le "clair") qui est chiffré et le cryptogramme résultant envoyé à l'extrémité réceptrice par un moyen potentiellement interceptable, par exemple la radio. À l'extrémité réceptrice, le cryptogramme et la clé sont combinés dans le déchiffreur pour récupérer le message.

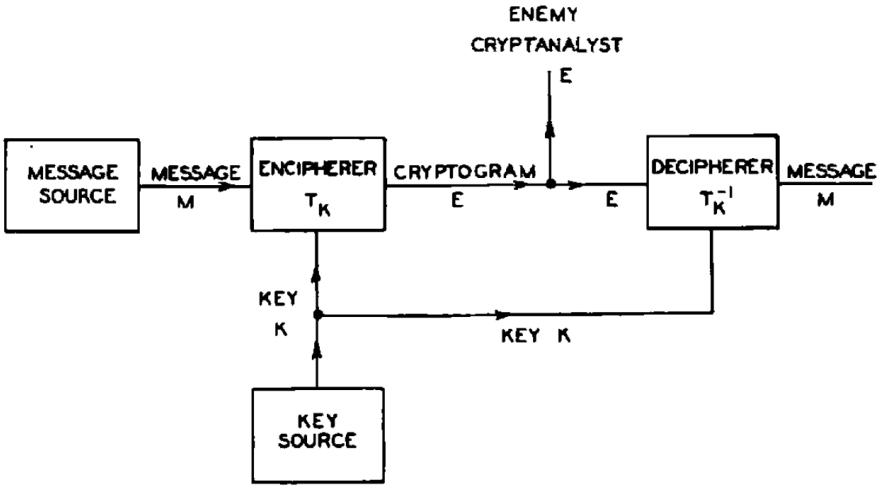

FIG. 1. Schéma d'un système secret général

De toute évidence, le chiffreur effectue une opération fonctionnelle. Si M est le message, K la clé et E le message chiffré, ou cryptogramme, on a

$$E = f(M, K),$$

c'est-à-dire que E est une fonction de M et K. Il est toutefois préférable de considérer cela non pas comme une fonction à deux variables, mais comme une famille d'opérations ou de transformations (à un seul paramètre), et de l'écrire

$$E = T_i M.$$

La transformation T_i appliquée au message M produit le cryptogramme E. L'indice i correspond à la clé utilisée.

Nous supposerons, en général, qu'il n'existe qu'un nombre fini de clés possibles, chacune étant associée à une probabilité p_i. Ainsi, la source de la clé est représentée par un processus statistique qui en choisit une parmi l'ensemble des transformations $T_1, T_2, \ldots, T_m$ avec les probabilités respectives $p_1, p_2, \ldots, p_m$. De même, nous supposerons généralement un nombre fini de messages possibles $M_1, M_2, \ldots, M_n$ avec les probabilités a priori associées $q_1, q_2, \ldots, q_n$. Les messages possibles, par exemple, pourraient être les séquences possibles de lettres anglaises de longueur N, et les probabilités associées seraient alors les fréquences relatives d'occurrence de ces séquences dans un texte anglais normal.

À la réception, il doit être possible de récupérer M, connaissant E et K. Ainsi, les transformations T_i de la famille doivent avoir des inverses uniques T_i^{-1} tels que $T_i T_i^{-1} = I$, la transformation Identité. Ainsi :

$$M = T_k^{-1} E.$$

Quoi qu'il en soit, cet inverse doit exister de manière unique pour tout E pouvant être obtenu à partir d'un M de clé i. On en arrive ainsi à la définition suivante : un système de chiffrement est une famille de transformations T_i unique et réversibles d'un ensemble de messages possibles en un ensemble de cryptogrammes, la transformation T_i ayant une probabilité associée p_i. Inversement, tout ensemble d'entités de ce type sera appelé "système de chiffrement". L'ensemble des messages possibles sera appelé, par commodité, "espace des messages" et l'ensemble des cryptogrammes possibles "espace des cryptogrammes".

Deux systèmes de chiffrement seront identiques s'ils sont constitués du même ensemble de transformations T_i, avec les mêmes messages et le même espace des cryptogrammes (portée et domaine) et les mêmes probabilités pour les clés.

Un système de chiffrement peut être visualisé mécaniquement comme une machine dotée d'une ou plusieurs commandes. Une séquence de lettres, le message, est introduite en entrée de la machine, et une seconde série apparaît en sortie. Le réglage des commandes correspond à la touche utilisée. Une méthode statistique doit être prescrite pour choisir la touche parmi toutes les possibles.

Pour rendre le problème mathématiquement traitable, supposons que l'adversaire connaisse le système utilisé. Autrement dit, il connaît la famille de transformations T_i et les probabilités de choisir différentes clés. On pourrait objecter que cette hypothèse est irréaliste, car le cryptanalyste ignore souvent quel système a été utilisé et donc les probabilités en question. Il y a deux réponses à cette objection :

1. La restriction est bien plus faible qu'il n'y paraît à première vue, en raison de notre définition large de ce qui constitue un système secret. Supposons qu'un cryptanalyste intercepte un message et ignore si un chiffrement par substitution-transposition ou de type Vigenère a été utilisé. Il peut considérer le message comme chiffré par un système dont une partie de la clé spécifie lequel de ces types a été utilisé, la partie suivante étant la clé spécifique à ce type. Ces trois possibilités différentes se voient attribuer des probabilités selon les meilleures estimations a priori par le chiffreur des probabilités utilisant les types de chiffrement respectifs.

2. Cette hypothèse est en fait celle couramment utilisée dans les études cryptographiques. C'est une approche pessimiste et donc sûre, mais réaliste à long terme, puisqu'il faut s'attendre à ce que le système de cryptage qu'on utilise finisse par être découvert. Ainsi, même lorsqu'un système entièrement nouveau est conçu, de sorte que l'adversaire ne peut lui attribuer aucune probabilité a priori sans le découvrir effectivement, il faut toujours l'utiliser en s'attendant à ce qu'il finisse par être découvert.

La situation est similaire à celle observée dans la théorie des jeux [4], où l'on suppose que l'adversaire "connaît" la stratégie de jeu utilisée. Dans les deux cas, cette hypothèse permet de délimiter précisément la connaissance de l'adversaire.

Une deuxième objection possible à notre définition des systèmes secrets est qu'elle ne tient pas compte de la pratique courante consistant à insérer des valeurs nulles dans un message et à utiliser plusieurs substituts. Dans de tels cas, il n'existe pas de cryptogramme unique pour un message et une clé donnés, mais le chiffreur peut choisir librement parmi plusieurs cryptogrammes différents. Cette situation pourrait être gérée, mais ne ferait qu'ajouter de la complexité à ce stade, sans modifier substantiellement les résultats de base.

Si les messages sont produits par un processus de Markov pour représenter une source d'information, les probabilités des différents messages sont déterminées par la structure du processus de Markov. Pour l'instant, cependant, nous souhaitons adopter une vision plus générale de la situation et considérer les messages comme un simple ensemble abstrait d'entités avec leurs probabilités associées, pas nécessairement composé d'une séquence de lettres et pas nécessairement produit par un processus de Markov.

Il convient de souligner que, tout au long de cet article, un système de chiffrement désigne non pas une, mais un ensemble de plusieurs transformations. Une fois la clé choisie, une seule de ces transformations est utilisée, ce qui pourrait conduire à définir un système de chiffrement comme une transformation unique sur un langage. L'adversaire, cependant, ne sait pas quelle clé a été choisie et les clés qui "auraient pu être choisies" sont aussi importantes pour lui que la clé réelle. En effet, seule l'existence de ces autres possibilités confère au système son caractère secret. Puisque le secret est notre principal intérêt, nous sommes contraints d'adopter le concept assez élaboré de système de chiffrement défini ci-dessus. Ce type de situation, où les possibilités sont aussi importantes que les réalités, se produit fréquemment dans les jeux de stratégie. Le déroulement d'une partie d'échecs est largement contrôlé par des menaces non mises à exécution. L'existence virtuelle d'intentions non réalisées est quelque peu similaire dans la théorie des jeux.

Il convient de noter qu'une seule opération sur un langage constitue un type dégénéré de système de chiffrement selon notre définition : c'est un système avec une seule clé de probabilité unitaire. Un tel système est dépourvu de secret : le cryptanalyste trouve le message en appliquant l'inverse de cette transformation, la seule du système, au cryptogramme intercepté. Le déchiffreur et le cryptanalyste possèdent alors la même information. En général, la seule différence entre les connaissances du chiffreur et celles du cryptanalyste adverse réside dans le fait que le chiffreur connaît la clé utilisée, tandis que le cryptanalyste ne connaît que les probabilités a priori des différentes

[4] Voir von Neumann et Morgenstern "*The Theory of Games*", Princeton 1947.

clés de l'ensemble. Le processus de déchiffrement consiste à appliquer au cryptogramme l'inverse de la transformation utilisée lors du chiffrement. Le processus de cryptanalyse consiste à tenter de déterminer le message (ou la clé) à partir du cryptogramme et des probabilités a priori des différentes clés et messages.

Il existe un certain nombre de questions épistémologiques difficiles liées à la théorie du secret, ou en fait à toute théorie qui implique des questions de probabilité (en particulier les probabilités a priori, le théorème de Bayes, etc.) lorsqu'elle est appliquée à une situation physique. Traitée de manière abstraite, la théorie des probabilités peut être posée sur une base logique rigoureuse grâce à l'approche moderne de la théorie de la mesure [5] [6]. Cependant, lorsqu'on applique la théorie des probabilités à une situation physique, notamment lorsqu'il s'agit de probabilités "subjectives" et d'expériences non reproductibles, de nombreuses questions de validité logique se posent. Par exemple, dans l'approche du secret présentée ici, les probabilités a priori de diverses clés et messages sont supposées connues du cryptanalyste adverse. Comment peut-on déterminer opérationnellement si ses estimations sont correctes, compte tenu de sa connaissance de la situation ?

On peut construire des situations cryptographiques artificielles de type "urnes et dés" dans lesquelles les probabilités a priori ont une signification précise et univoque, et l'idéalisation utilisée ici est certainement appropriée. Dans d'autres situations imaginables, par exemple une communication qu'on intercepterait entre des envahisseurs martiens, les probabilités a priori seraient probablement si incertaines qu'elles seraient dénuées de signification. La plupart des situations cryptographiques pratiques se situent entre ces limites. Un cryptanalyste pourrait être disposé à classer les messages possibles dans les catégories "raisonnable", "possible mais improbable" et "déraisonnable", mais il est possible qu'il estime qu'une subdivision plus fine soit inutile.

Heureusement, en pratique, seules les erreurs extrêmes dans les probabilités a priori des clés et des messages entraînent des erreurs significatives dans les paramètres importants. Cela est dû au caractère exponentiel du nombre de messages et de cryptogrammes, et aux mesures logarithmiques employées.

3. Représentation des systèmes

Un système de chiffrement tel que défini ci-dessus peut être représenté de différentes manières. Un diagramme linéaire, comme dans les figures 2 et 4, est pratique à des fins d'illustration. Les messages possibles sont représentés par des points à gauche et les cryptogrammes possibles par des points à droite. Si une certaine clé, disons la clé 1, transforme le message M_2 en le cryptogramme E_4, alors M_2 et E_4 sont reliés par une ligne étiquetée 1, etc. De chaque message possible, il doit y avoir exactement une ligne sortante pour chaque clé différente possible. Si la même chose est vraie pour chaque cryptogramme, nous dirons que le système est fermé. Une manière plus courante de décrire un système est d'énoncer l'opération que l'on effectue sur le message pour une clé arbitraire afin d'obtenir le cryptogramme. De même, on définit implicitement les probabilités pour diverses clés en décrivant comment une clé est choisie ou ce que l'on sait des habitudes de l'adversaire en

[5] Voir J. L. Doob, "*Probability as Measure*", Annals of Math. Stat., v. 12, 1941, p. 206–214.

[6] A. Kolmogoroff, "*Grundbegriffe der Wahrscheinlichkeitsrechnung*", Ergebnisse der Mathematic, v. 2, n° 3 (Berlin 1933).

matière de choix de clés. Les probabilités des messages sont déterminées implicitement en fonction de notre connaissance a priori des habitudes linguistiques de l'ennemi, de la situation tactique (qui influencera le contenu probable du message) et de toute information particulière que nous pouvons avoir concernant le cryptogramme.

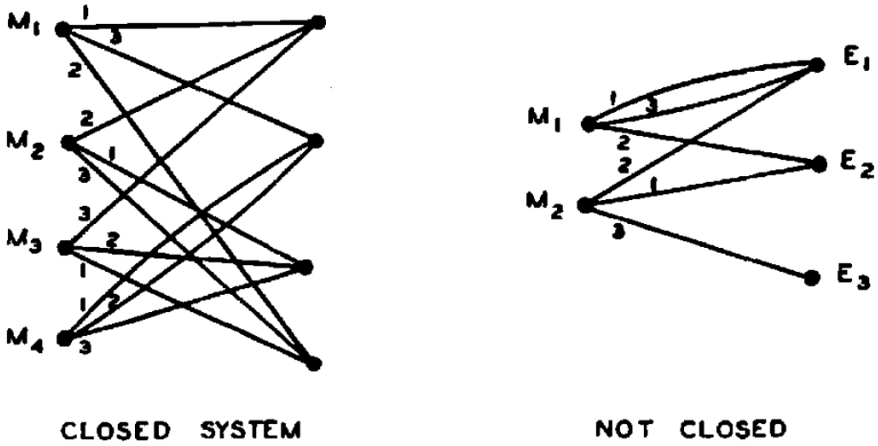

FIG. 2. Illustrations de systèmes simples

4. Quelques exemples de systèmes secrets

Dans cette section, plusieurs exemples de chiffrements seront donnés. Ils seront souvent mentionnés dans la suite de l'article à titre d'illustration.

1. Chiffrement par substitution simple.

Dans ce chiffrement, chaque lettre du message est remplacée par un substitut fixe, généralement une lettre également. Ainsi, le message,

$$M = m_1 m_2 m_3 m_4 \dots$$

où $m_1, m_2, \dots$ sont les lettres successives devient :

$$\begin{aligned} E &= e_1 e_2 e_3 e_4 \dots \\ &= f(m_1) f(m_2) f(m_3) f(m_4) \dots \end{aligned}$$

où la fonction $f(m)$ est une fonction ayant une fonction inverse. La clé est une permutation de l'alphabet (lorsque les substituts sont des lettres), par exemple

$$XGUACDTBFHRSLMQVYZWIEJ0KNP.$$

La première lettre X est le substitut de A, G est le substitut de B, etc.

2. Transposition (de période fixe d).

Le message est divisé en groupes de longueur d et une permutation est appliquée au premier groupe, la même permutation au second groupe, etc. La permutation est la clé et peut être représentée

par une permutation des d premiers entiers. Ainsi, pour $d = 5$, nous pourrions avoir 23154 comme permutation. Cela signifie que :

$$\begin{array}{rll} & m_1\, m_2\, m_3\, m_4\, m_5 & m_6\, m_7\, m_8\, m_9\, m_{10} \quad \ldots \\ \text{devient} & m_2\, m_3\, m_1\, m_5\, m_4 & m_7\, m_8\, m_6\, m_{10}\, m_9 \quad \ldots \end{array}$$

L'application séquentielle de deux ou plusieurs transpositions est appelée transposition composée. Si les périodes $d_1, d_2, \ldots, d_n$ sont égales, il est clair que le résultat est une transposition de période d, où d est le plus petit commun multiple de $d_1, d_2, \ldots, d_n$.

3. Vigenère et variations.

Dans le chiffrement de Vigenère, la clé est constituée d'une série de d lettres. Celles-ci sont écrites de manière répétée sous le message et les deux sont ajoutées modulo 26 (en considérant l'alphabet numéroté de $A = 0$ à $Z = 25$). Ainsi

$$e_i = m_e + k_i \pmod{26}$$

où k_i est de période d dans l'indice i. Par exemple, avec la clé GAH, on obtient

message	N	O	W	I	S	T	H	E	...
clé répétée	G	A	H	G	A	H	G	A	...
cryptogramme	T	O	D	O	S	A	N	E	...

Le chiffrement de Vigenère de la période est appelé "chiffrement de César". Il s'agit d'une simple substitution dans laquelle chaque lettre de M est avancée d'une valeur fixe dans l'alphabet. Cette valeur constitue la clé, qui peut être un nombre compris entre 0 et 25. Les chiffrements appelés "chiffrement de Beaufort", et "variantes de Beaufort", sont similaires au chiffrement de Vigenère et utilisent les équations suivantes :

$$e_i = k_i - m_i \pmod{26}$$

et

$$e_i = m_i - k_i \pmod{26}$$

respectivement. Le chiffrement de Beaufort de période de longueur 1 est appelé chiffrement de César inversé. L'application de deux ou plusieurs Vigenères successivement est appelée le chiffrement de Vigenère composé. Son équation est :

$$e_i = m_i + k_i + l_i + \ldots + s_i \pmod{26}$$

où $k_i, l_i, \ldots, s_i$ ont généralement des périodes différentes. La période de leur somme

$$k_i + l_i + \ldots + s_i$$

comme dans la transposition composée, est le plus petit commun multiple des périodes individuelles.

Lorsque le Vigenère est utilisé comme clé illimitée, sans répétition, on obtient le système de Vernam [7], avec

$$e_i = m_i + k_i \pmod{26}$$

k_i étant choisi aléatoirement et indépendamment parmi 0, 1, ..., 25. Si la clé est un texte significatif, nous obtenons le chiffrement par "clé courante".

4. Substitution de digrammes, de trigrammes et de N-grammes.

Plutôt que de substituer des lettres, on peut substituer des digrammes, des trigrammes, etc. La substitution de digrammes générale nécessite une clé constituée d'une permutation des 26^2 digrammes. Elle peut être représentée par un tableau dans lequel la ligne correspond à la première lettre du digramme et la colonne à la seconde, les entrées du tableau étant les substitutions (généralement aussi des digrammes).

5. Vigenère à alphabet mixte simple.

Il s'agit d'une substitution simple suivie d'un Vigenère.

$$\begin{aligned} e_i &= f(m_i) + k_i \\ m_i &= f^{-1}(e_i - k_i) \end{aligned}$$

L'"inverse" de ce système est un Vigenère suivi d'une simple substitution

$$\begin{aligned} e_i &= g(m_i + k_i) \\ m_i &= g^{-1}(e_i) - k_i \end{aligned}$$

6. Système matriciel. [8]

Une méthode de substitution de n-grammes consiste à opérer sur des n-grammes successifs avec une matrice possédant un inverse. Les lettres sont supposées numérotées de 0 à 25, ce qui en fait des éléments d'un anneau algébrique. À partir du n-gramme $m_1, m_2, \ldots, m_n$ du message, la matrice a_{ij} donne un n-gramme du cryptogramme. La matrice a_{ij} est la clé, et le déchiffrement s'effectue avec la matrice inverse. La matrice inverse existera si et seulement si le déterminant $|a_{ij}|$ possède un élément inverse dans l'anneau.

7. Le chiffre de Playfair

Il s'agit d'un type particulier de substitution de digrammes, régi par un alphabet mixte de 25 lettres, écrit dans un carré 5×5. (La lettre J est souvent omise en cryptographie ; elle est très rare et, lorsqu'elle apparaît, elle peut être remplacée par I.) Supposons que le carré clé soit comme suit

[7] G. S. Vernam, "*Cipher Printing Telegraph Systems for Secret Wire and Radio Telegraphic Communications*", Journal American Institute of Electrical Engineers, v. XLV, p. 109–115, 1926.

[8] Voir L. S. Hill, "*Cryptography in an Algebraic Alphabet*", American Math. Monthly, v. 36, n° 6, 1, 1929, p. 306–312 ; Voir aussi "*Concerning Certain Linear Transformation Apparatus of Cryptography*", v. 38, n° 3, 1931, p. 135–154.

:

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

Le substitut du digramme AC, par exemple, est la paire de lettres aux autres coins du rectangle défini par A et C, soit LO, le L étant pris en premier car il est au-dessus de A. Si les lettres du digramme sont sur une ligne horizontale comme RI, on utilise les lettres à leur droite DF ; RF devient DR. Si les lettres sont sur une ligne verticale, on utilise les lettres situées en dessous. Ainsi, PS devient UW. Si les lettres sont identiques, des zéros peuvent être utilisés pour les séparer ou l'une d'elles peut être omise, etc.

8. Substitution multiple d'alphabets mixtes.

Dans ce chiffrement, il existe un ensemble de substitutions simples qui sont utilisées séquentiellement. Si la période est de quatre, elle devient

$$m_1 m_2 m_3 m_4 m_5 m_6 \ldots$$

devient

$$f_1(m_1) f_2(m_2) f_3(m_3) f_4(m_4) f_1(m_5) f_2(m_6) \ldots$$

9. Chiffrement à clé automatique.

Un système de type Vigenère dans lequel le message lui-même ou le cryptogramme résultant sert de "clé" est appelé chiffrement à clé automatique. Le chiffrement commence par une "clé d'amorçage" (qui est la clé entière au sens où nous l'entendons) et se poursuit avec le message ou le cryptogramme décalé de la longueur de la clé d'amorçage, comme indiqué ci-dessous, où la clé d'amorçage est COMET. Le message utilisé comme "clé" :

Message	S	E	N	D	S	U	P	P	L	I	E	S	$\ldots$
Clé	C	O	M	E	T	S	E	N	D	S	U	P	$\ldots$
Cryptogramme	U	S	Z	H	L	M	T	C	O	A	Y	H	$\ldots$

Le cryptogramme utilisé comme "clé" [9] :

Message	S	E	N	D	S	U	P	P	L	I	E	S	$\ldots$
Clé	C	O	M	E	T	U	S	Z	H	L	O	H	$\ldots$
Cryptogramme	U	S	Z	H	L	O	H	O	S	T	S	$\ldots$	$\ldots$

10. Chiffrements fractionnaires.

[9] Ce système est trivial du point de vue du secret puisque, à l'exception des premières lettres, l'ennemi est en possession de la totalité de la "clé".

Dans ces cas, chaque lettre est d'abord chiffrée en deux ou plusieurs lettres ou chiffres, puis ces symboles sont mélangés (par exemple, par transposition). Le résultat peut ensuite être retraduit dans l'alphabet d'origine. Ainsi, en utilisant un alphabet mixte de 25 lettres comme clé, nous pouvons traduire les lettres en nombres quinaires à deux chiffres selon le tableau suivant :

	0	1	2	3	4
0	*L*	*Z*	*Q*	*C*	*P*
1	*A*	*G*	*N*	*O*	*U*
2	*R*	*D*	*M*	*I*	*F*
3	*K*	*Y*	*H*	*V*	*S*
4	*X*	*B*	*T*	*E*	*W*

Ainsi, *B* devient 41. Après transposition de la série de nombres résultante, ils sont pris par paires et retranscrits en lettres.

11. Codes.

Dans les codes, les mots (ou parfois les syllabes) sont remplacés par des groupes de lettres. Un chiffrement, quel qu'il soit, est parfois appliqué au résultat.

5. Évaluation d'un système de chiffrement

Plusieurs critères doivent être appliqués pour estimer la valeur d'un système de chiffrement proposé. Les plus importants sont :

1. Le niveau de secret.

Certains systèmes sont parfaits - l'ennemi n'a pas davantage d'information après avoir intercepté une quantité de matériau qu'auparavant. D'autres systèmes, bien que lui fournissant certaines informations, ne fournissent pas de "solution" unique aux cryptogrammes interceptés. Parmi les systèmes à résolution unique, il existe de grandes variations dans la quantité de travail nécessaire pour parvenir à cette résolution et dans la quantité de données à intercepter pour la rendre unique.

2. La taille de la clé.

La clé doit être transmise par des moyens non interceptables du point d'émission au point de réception. Elle doit parfois être mémorisée. Il est donc souhaitable que la clé soit la plus petite possible.

3. La complexité des opérations de chiffrement et de déchiffrement.

Le chiffrement et le déchiffrement doivent, bien entendu, être aussi simples que possible. S'ils sont réalisées manuellement, la complexité entraîne des pertes de temps, des erreurs, etc. S'ils sont réalisées mécaniquement, la prise en charge de la complexité nécessite des machines de grande taille et coûteuses.

4. La propagation des erreurs.

Dans certains types de chiffrements, une erreur d'une seule lettre lors du chiffrement ou de la transmission entraîne un grand nombre d'erreurs dans le texte déchiffré. Ces erreurs sont propagées par l'opération de déchiffrement, entraînant la perte d'une grande quantité d'informations et la nécessité de répéter fréquemment le cryptogramme. Il est naturellement souhaitable de minimiser cette multiplication des erreurs.

5. L'expansion du message.

Dans certains types de systèmes de chiffrement, la taille du message est augmentée par le processus de chiffrement. Cet effet indésirable peut être observé dans les systèmes où l'on tente de surcharger les statistiques du message par l'ajout de nombreuses valeurs nulles, ou lorsque plusieurs substituts sont utilisés. Il se produit également dans de nombreux systèmes de "dissimulation" (qui ne sont généralement pas des systèmes de chiffrement au sens de notre définition).

6. Algèbre des systèmes secrets

Si nous avons deux systèmes de chiffrement T et R, nous pouvons souvent les combiner de diverses manières pour former un nouveau système de chiffrement. Si T et R ont le même domaine (espace de messages), nous pouvons former une sorte de "somme pondérée",

$$S = pT + qR$$

où $p+q = 1$. Cette opération consiste d'abord à effectuer un choix préliminaire avec les probabilités p et q déterminant lequel de T ou R est utilisé. Ce choix fait partie de la clé de S. Une fois ce choix effectué, T ou R est utilisé tel que défini initialement. La clé totale de S doit préciser quelle opération (T ou R) est utilisée et quelle clé (de T ou R) doit être utilisée.

Si T est constitué des transformations $T_1, \ldots, T_m$ de probabilités $p_1, \ldots, p_m$ et que R est constitué de $R_1, \ldots, R_k$ de probabilités $q_1, \ldots, q_k$, alors $S = pT + qR$ est constitué des transformations $T_1, T_2, \ldots, T_m, R_1, \ldots, R_k$ de probabilités $pp_1, pp_2, \ldots, pp_m, qq_1, qq_2, \ldots, qq_k$ respectivement.

Plus généralement, on peut former la somme de plusieurs systèmes.

$$S = p_1T + p_2R + \ldots + p_mU \qquad\qquad \sum p_i = 1$$

On remarque que tout système T peut s'écrire comme une somme d'opérations fixes

$$T = p_1T_1 + p_2T_2 + \ldots + p_mT_m$$

T_i étant une opération de chiffrement définie de T correspondant au choix de clé i, de probabilité p_i.

Une deuxième façon de combiner deux systèmes de chiffrement consiste à prendre le "produit", représenté schématiquement sur la figure 3. Supposons que T et R soient deux systèmes et que le domaine (espace du langage) de R puisse être identifié à l'étendue (espace des cryptogrammes) de T. Nous pouvons alors appliquer d'abord T à notre langage, puis R

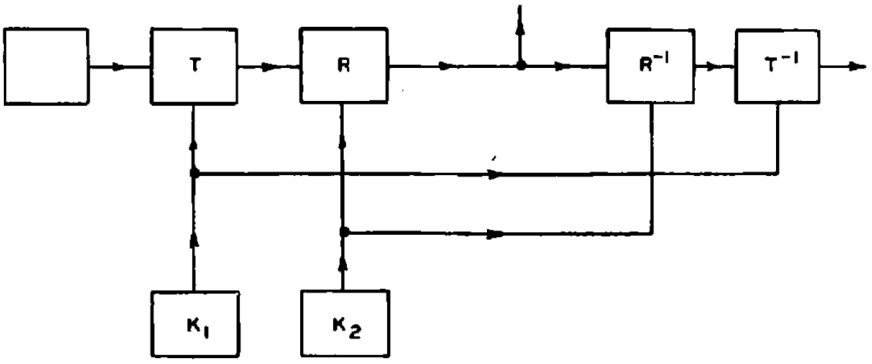

FIG. 3. Produit de deux systèmes

au résultat de ce chiffrement. Cela donne une opération résultante S que nous écrivons comme un produit

$$S = RT$$

La clé de S est constituée des clés de T et de R, supposées choisies selon leurs probabilités initiales et indépendamment. Ainsi, si les m clés de T sont choisies avec des probabilités

$$p_1 p_2 \dots p_m$$

et que les n clés de R ont des probabilités

$$p'_1 p'_2 \dots p'_n,$$

alors S a au plus des clés de probabilités $p_i p'_j$. Dans de nombreux cas, certaines transformations de produits $R_i T_j$ seront identiques et pourront être regroupées en additionnant leurs probabilités.

Le chiffrement par produit est souvent utilisé ; par exemple, on fait suivre une substitution par une transposition ou une transposition par un Vigenère, ou on applique un code au texte et on chiffre le résultat par substitution, transposition, fractionnement, etc.

Il convient de noter que la multiplication n'est pas commutative en général (on n'a pas toujours $RS = SR$), bien qu'elle le soit dans des cas particuliers, comme la substitution et la transposition. La multiplication représentant une opération, elle est par définition associative. Autrement dit, $R(ST) = (RS)T = RST$. De plus, on a les lois

$$p(p'T + q'R) + qS = pp'T + pq'R + qS$$

(loi associative pondérée pour l'addition)

$$T(pR + qS) = pTR + qTS$$
$$(pR + qS)T = pRT + qST$$

(lois de distribution à droite et à gauche)

et

$$p_1 T + p_2 T + p_3 R = (p_1 + p_2)T + p_3 R$$

Il convient de souligner que ces opérations combinant addition et multiplication s'appliquent aux systèmes secrets dans leur ensemble. Le produit de deux systèmes TR ne doit pas être confondu avec le produit des transformations des systèmes T_iR_j, qui apparaît également souvent dans cet ouvrage. Le premier TR est un système secret, c'est-à-dire un ensemble de transformations avec leurs probabilités associées ; le second est une transformation particulière. De plus, la somme de deux systèmes $pR + qT$ est un système ; la somme de deux transformations n'est pas définie. Les systèmes T et R peuvent commuter sans que les T_i et R_j individuels ne commutent. Par exemple, si R est un système de Beaufort d'une période donnée, toutes les clés sont également probables.

$$R_iR_j \neq R_jR_i$$

En général, mais bien sûr, RR ne dépend pas de son ordre ; en réalité,

$$RR = V$$

le Vigenère de la même période avec une clé aléatoire. En revanche, si les T_i et R_j individuels de deux systèmes T et R commutent, alors les systèmes commutent.

Un système dont les espaces M et E peuvent être identifiés, un cas très courant comme la transformation de suites de lettres en suites de lettres, peut être qualifié d'endomorphe. Un système endomorphe T peut être élevé à une puissance T^n.

Un système secret T dont le produit avec lui-même est égal à T, c'est-à-dire pour lequel

$$TT = T$$

sera dit idempotent. Par exemple, la substitution simple, la transposition de période p, le Vigenère de période p (tous avec chaque clé équiprobable) sont idempotents.

L'ensemble de tous les systèmes de chiffrement endomorphes définis dans un espace de messages fixe constitue une "variété algébrique", c'est-à-dire une sorte d'algèbre, utilisant les opérations d'addition et de multiplication. En fait, les propriétés de l'addition et de la multiplication que nous avons étudiées peuvent être résumées comme suit :

L'ensemble des chiffrements endomorphes sur un même espace de messages muni des deux opérations de combinaison d'addition pondérée et de multiplication forment une algèbre associative linéaire d'élément un, à la différence près que les coefficients d'une addition pondérée doivent être positifs et égaux à un.

Les opérations de combinaison nous permettent de construire de nombreux nouveaux types de systèmes de chiffrement à partir de certains, comme les exemples donnés. On peut également les utiliser pour décrire la situation à laquelle se trouve confronté un cryptanalyste lorsqu'il tente de résoudre un cryptogramme de type inconnu. Il s'agit en fait de résoudre un système secret de type

$$T = p_1A + p_2B + \ldots + p_rS + p'X \qquad\qquad \sum p = 1$$

où $A, B, \ldots, S$ sont des types de chiffrements connus, avec p_i leurs probabilités a priori dans cette situation, et $p'X$ correspond à la possibilité d'un type de chiffrement complètement nouveau et inconnu.

7. Chiffres purs et mixtes

Certains types de chiffrements, tels que la substitution simple, la transposition d'une période donnée, le Vigenère d'une période donnée, le Vigenère à alphabet mixte, etc. (tous avec chaque clé équiprobable), présentent une certaine homogénéité par rapport à la clé. Quelle que soit la clé, les processus de chiffrement, de déchiffrement et de décryptage sont essentiellement les mêmes. Ceci peut être comparé au chiffrement

$$pS + qT$$

où S est une substitution simple et T une transposition d'une période donnée. Dans ce cas, le système entier change pour le chiffrement, le déchiffrement et le décryptage, selon que la substitution ou la transposition est utilisée.

L'homogénéité de ces systèmes provient de la propriété de groupe : nous remarquons que, dans les exemples de chiffrements homogènes ci-dessus, le produit T_iT_j de deux transformations quelconques de l'ensemble est égal à une troisième transformation T_k de l'ensemble. En revanche, T_iS_j n'est égal à aucune transformation du chiffrement

$$pS + qT$$

qui ne contient que des substitutions et des transpositions, sans produit.

On pourrait donc définir un chiffrement "pur" comme un chiffrement dont les T_i forment un groupe. Ceci serait cependant trop restrictif, car cela nécessiterait que l'espace E soit identique à l'espace M, c'est-à-dire que le système soit endomorphe. La transposition fractionnaire est aussi homogène que la transposition ordinaire sans être endomorphe. La définition correcte est la suivante : un chiffrement T est pur si pour tout T_i, T_j, T_k il existe un T_s tel que

$$T_iT_j^{-1}T_k = T_s$$

et chaque clé est équiprobable. Sinon, le chiffrement est mixte. Les systèmes de la figure 2 sont mixtes. Le chiffrement de la figure 4 est pur si toutes les clés sont équiprobables.

Théorème 1. *Dans un chiffrement pur, les opérations $T_i^{-1}T_j$ qui transforment l'espace des messages en lui-même forment un groupe dont l'ordre est m, le nombre de clés différentes.*

Ceci est dû au fait que

$$T_j^{-1}T_kT_k^{-1}T_j = I,$$

de sorte que chaque élément possède un inverse. La loi associative est vraie puisqu'il s'agit d'opérations, et la propriété de groupe découle de

$$T_i^{-1}T_jT_k^{-1}T_l = T_s^{-1}T_kT_k^{-1}T_k = T_s^{-1}T_l$$

en utilisant notre hypothèse que $T_i^{-1}T_j = T_s^{-1}T_k$ pour un certain s.

L'opération $T_i^{-1}T_j$ consiste, bien sûr, à chiffrer le message avec la clé j, puis à le déchiffrer avec la clé i, ce qui nous ramène à l'espace des messages. Si T est endomorphe, c'est-à-dire que les T_i

transforment eux-mêmes l'espace Ω_M en lui-même (comme c'est le cas de la plupart des chiffrements, où l'espace des messages et l'espace des cryptogrammes sont tous deux constitués de suites de lettres), et que les T_i sont un groupe et sont équiprobables, alors T est pur, car

$$T_i T_j^{-1} T_k = T_i T_r = T_s.$$

Théorème 2. *Le produit de deux chiffrements purs qui commutent est pur.*

Cela est dû au fait que si T et R commutent $T_i R_j = R_l T_m$ pour tout i, j avec l, m convenables, et

$$\begin{aligned} T_i R_j (T_k R_l)^{-1} T_m R_n &= T_i R_j R_l^{-1} T_k^{-1} T_m R_n \\ &= R_u R_v^{-1} R_w T_r T_s^{-1} T_t \\ &= R_h T_u. \end{aligned}$$

La condition de commutation n'est cependant pas nécessaire pour que le produit soit un chiffrement pur.

Un système avec une seule clé, c'est-à-dire une seule opération définie T_1, est pur puisque le seul choix d'indices est

$$T_1 T_1^{-1} T_1 = T_1.$$

Ainsi, le développement d'un chiffrement général en une somme de transformations simples le présente également comme une somme de chiffrements purs.

L'examen de l'exemple de chiffrement pur présenté à la figure 4 révèle certaines propriétés. Les messages appartiennent à certains sous-ensembles que nous appellerons classes résiduelles, et les cryptogrammes possibles sont séparables selon les classes résiduelles correspondantes. Il y a au moins une ligne reliant chaque message d'une classe à chaque cryptogramme de la classe correspondante, et aucune ligne entre les classes qui ne correspondent pas. Le nombre de messages d'une classe est un diviseur du nombre total de clés. Le nombre de lignes "en parallèle" d'un message M vers un cryptogramme de la classe correspondante est égal au nombre de clés divisé par le nombre de messages de la classe contenant le message (ou le cryptogramme). L'annexe montre que ces affirmations sont généralement valables pour les chiffrements purs. En résumé, nous avons le :

Théorème 3. *Dans un système pur, les messages peuvent être séparés en un ensemble de "classes résiduelles" $C_1, C_2, \ldots, C_s$ et les cryptogrammes en un ensemble correspondant de classes résiduelles $C'_1, C'_2, \ldots, C'_s$ avec les propriétés suivantes* :

(1) *Les classes résiduelles de messages s'excluent mutuellement et contiennent collectivement tous les messages possibles. Il en va de même pour les classes résiduelles de cryptogrammes.*

(2) *Le chiffrement de tout message de C_i avec n'importe quelle clé produit un cryptogramme de C'_i. Le déchiffrement de tout cryptogramme de C'_i avec n'importe quelle clé produit un message dans C_i.*

(3) *Le nombre de messages dans C_i, disons φ_i, est égal au nombre de cryptogrammes dans C'_i et est un diviseur de k le nombre de clés.*

(4) *Chaque message dans C_i peut être chiffré en chaque cryptogramme de C'_i par exactement k/φ_i clés différentes. De même pour le déchiffrement.*

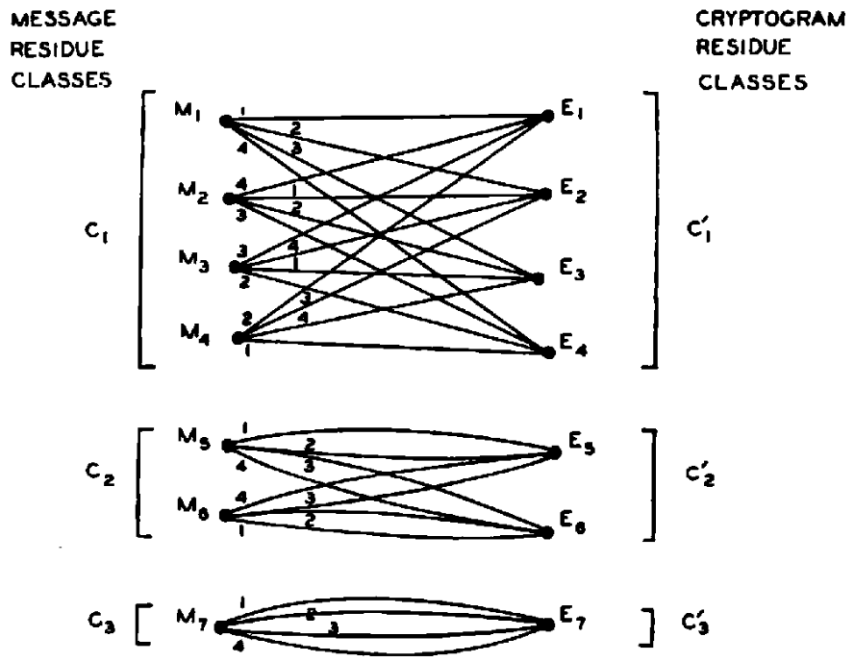

FIG. 4. Système pur

L'importance du concept de chiffrement pur (et la raison de son nom) réside dans le fait que, dans un tel chiffrement, toutes les clés sont essentiellement les mêmes. Quelle que soit la clé utilisée pour un message particulier, les probabilités a posteriori de tous les messages sont identiques. Pour illustrer cela, notons que deux clés différentes appliquées au même message conduisent à deux cryptogrammes de la même classe résiduelle, par exemple C'_i. Les deux cryptogrammes pourraient donc être déchiffrés chacun par $\dfrac{k}{\varphi_i}$ clés en chaque message de C_i et en aucun autre message possible. Toutes les clés étant équiprobables, les probabilités a posteriori des différents messages sont donc

$$P_E(M) = \frac{P(M)P_M(E)}{P(E)} = \frac{P(M)P_M(E)}{\sum_M P(M)P_M(E)} = \frac{P(M)}{P(C_i)}$$

où M est dans C_i, E est dans C'_i et la somme porte sur tous les messages de C_i. Si E et M ne sont pas dans des classes résiduelles correspondantes, $P_E(M) = 0$. De même, on peut montrer que les probabilités a posteriori des différentes clés ont la même valeur, mais que ces valeurs sont associées à des clés différentes lorsqu'une clé différente est utilisée. Le même ensemble de valeurs de $P_E(K)$ a subi une permutation entre les clés. On obtient donc comme résultat le

Théorème 4. *Dans un système pur, les probabilités a posteriori des différents messages $P_E(M)$ sont indépendantes de la clé choisie. Les probabilités a posteriori des clés $P_E(K)$ ont la même*

valeur, mais subissent une permutation avec un choix de clé différent.

On peut dire, en gros, que tout choix de clé conduit au même problème cryptanalytique dans un chiffrement pur. Puisque les différentes clés produisent toutes des cryptogrammes de la même classe résiduelle, cela signifie que tous les cryptogrammes de la même classe résiduelle sont cryptanalytiquement équivalents : ils conduisent aux mêmes probabilités a posteriori de messages et, à une permutation près, aux mêmes probabilités de clés.

À titre d'exemple, une simple substitution avec toutes les clés équiprobables constitue un chiffrement pur. La classe résiduelle correspondant à un cryptogramme E donné est l'ensemble de tous les cryptogrammes pouvant être obtenus à partir de E par les opérations $T_j T_k^{-1} E$. Dans ce cas, $T_j T_k^{-1}$ est elle-même une substitution ; par conséquent, toute substitution sur E donne un autre membre de la même classe résiduelle. Ainsi, si le cryptogramme est

$$E = X\ C\ P\ P\ G\ C\ F\ Q$$

alors

$$\begin{array}{lcccccccc} E_1 & = & R & D & H & H & G & D & S & N \\ E_2 & = & A & B & C & C & D & B & E & F \end{array}$$

etc. appartiennent à la même classe résiduelle. Il est évident, dans ce cas, que ces cryptogrammes sont essentiellement équivalents. Seul le motif de répétition des lettres importe dans une substitution simple à clé aléatoire, les lettres réelles étant des variables muettes. On pourrait même s'en passer complètement, en indiquant le motif de répétition comme suit :

Cette notation décrit la classe résiduelle, mais élimine toute information relative au membre spécifique de la classe. Elle conserve donc précisément l'information pertinente du point de vue cryptanalytique. Ceci est lié à une méthode d'attaque des chiffrements par substitution simple : la méthode des mots-mots.

Dans le chiffrement de type César, seules les premières différences mod 26 du cryptogramme sont significatives. Deux cryptogrammes ayant le même Δe_i appartiennent à la même classe résiduelle. On décrypte ce chiffrement en notant simplement les 26 membres de la classe résiduelle du message et en sélectionnant celui qui a du sens.

Le Vigenère de période e à clé aléatoire est un autre exemple de chiffrement pur. Ici, la classe résiduelle du message est constituée de toutes les séquences ayant les mêmes premières différences que le cryptogramme, pour des lettres séparées par une distance d. Pour $d = 3$, la classe résiduelle

est définie par

$$\begin{aligned} m_1 - m_4 &= e_1 - e_4 \\ m_2 - m_5 &= e_2 - e_5 \\ m_3 - m_6 &= e_3 - e_6 \\ m_4 - m_7 &= e_4 - e_7 \\ &\vdots \end{aligned}$$

où $E = e_1, e_2, \ldots$ est le cryptogramme et $m_1, m_2, \ldots$ est un M quelconque de la classe résiduelle correspondante.

Dans le chiffrement par transposition de période d à clé aléatoire, la classe résiduelle est constituée de tous les arrangements de e_i dans lesquels aucun e_i n'est déplacé hors de son bloc de longueur d, et deux e_i quelconques situés à une distance d restent à cette distance. Ceci est utilisé pour déchiffrer ces chiffrements comme suit : le cryptogramme est écrit en blocs successifs de longueur d, les uns sous les autres comme indiqué ci-dessous ($d = 5$) :

$$\begin{array}{ccccc} e_1 & e_2 & e_3 & e_4 & e_5 \\ e_6 & e_7 & e_8 & e_9 & e_{10} \\ e_{11} & e_{12} & \ldots & & \end{array}$$

Les colonnes sont ensuite découpées et réorganisées pour former un texte significatif. Une fois les colonnes découpées, la seule information restante est la classe résiduelle du cryptogramme.

Théorème 5. *Si T est pur alors $T_i T_j^{-1} T = T$ où $T_i T_j$ sont deux transformations quelconques de T. Inversement, si cela est vrai pour tout $T_i T_j$ dans un système T, alors T est pur.*

La première partie de ce théorème est évidente à partir de la définition d'un système pur. Pour démontrer la seconde partie, remarquons d'abord que, si $T_i T_j^{-1} T = T$, alors $T_i T_j^{-1} T_s$ est une transformation de T. Il reste à démontrer que toutes les clés sont équiprobables. On a $T = \sum_s p_s T_s$ et

$$\sum_s p_s T_i T_j^{-1} T_s = \sum_s p_s T_s.$$

Le terme de la somme de gauche avec $s = j$ donne $p_j T_i$. Le seul terme de T_i à droite est $p_i T_i$. Puisque tous les coefficients sont positifs, il s'ensuit que

$$p_j \leq p_i.$$

Le même argument est valable avec i et j intervertis, et par conséquent

$$p_j = p_i$$

et T est pur. Ainsi, la condition $T_i T_j^{-1} T = T$ pourrait être utilisée comme définition alternative d'un système pur.

8. Systèmes similaires

Deux systèmes secrets R et S seront dits similaires s'il existe une transformation A ayant un inverse A^{-1} tel que

$$R = AS$$

Cela signifie que chiffrer avec R revient à chiffrer avec S puis à transformer le résultat par l'opération A. Si nous écrivons RS pour signifier que R est similaire à S, alors il est clair que $R \approx S$ implique $S \approx R$. De plus, $R \approx S$ et $S \approx T$ impliquent $R \approx T$ et enfin $R \approx R$. On peut résumer ces hypothèses en disant que la similarité est une relation d'équivalence.

La signification cryptographique de la similarité est que si $R \approx S$, alors R et S sont équivalents du point de vue cryptanalytique. En effet, si un cryptanalyste intercepte un cryptogramme du système S, il peut le transformer en un cryptogramme du système R en lui appliquant simplement la transformation A. Un cryptogramme du système R est transformé en un cryptogramme de S en lui appliquant A^{-1}. Si R et S sont appliqués au même langage ou espace de messages, il existe une correspondance biunivoque entre les cryptogrammes obtenus. Les cryptogrammes correspondants donnent la même distribution de probabilités a posteriori pour tous les messages.

Si l'on dispose d'une méthode pour casser le système R, alors tout système S similaire à R peut être cassé par réduction à R par l'application de l'opération A. C'est un procédé fréquemment utilisé en cryptanalyse pratique.

À titre d'exemple trivial, une substitution simple où les substituts ne sont pas des lettres, mais des symboles arbitraires, est similaire à une substitution simple utilisant des substituts de lettres. Un deuxième exemple est celui des chiffrements de type César et César inversé. Ce dernier est parfois déchiffré par une première transformation en chiffrement de type César. Cela peut être réalisé en inversant l'alphabet du cryptogramme. Les chiffrements de Vigenère, de Beaufort et variante de Beaufort sont tous similaires, lorsque la clé est aléatoire. Le chiffrement "autoclé" (avec le message utilisé comme "clé") amorcé avec la clé $K_1K_2 \ldots K_d$ est similaire à un chiffrement de type Vigenère avec la clé alternativement ajoutée et soustraite modulo 26. La transformation dans ce cas consiste à "déchiffrer" l'autoclé avec une série de d "A" comme clé d'amorçage.

Partie II
Secret théorique
9. Introduction

Nous examinons maintenant les problèmes liés au "secret théorique" d'un système. Dans quelle mesure un système est-il immunisé contre la cryptanalyse lorsque le cryptanalyste dispose d'un temps et d'une main-d' œuvre illimités pour analyser les cryptogrammes ? Un cryptogramme possède-t-il une solution unique (même si sa recherche peut nécessiter un travail peu pratique) et, dans le cas contraire, combien de solutions raisonnables possède-t-il ? Quelle quantité de texte dans un système donné doit être interceptée avant que la solution ne devienne unique ? Existe-t-il des systèmes dont la solution ne devient jamais unique, quelle que soit la quantité de texte chiffré interceptée ? Existe-t-il des systèmes pour lesquels aucune information n'est transmise à l'ennemi, quelle que soit la quantité de texte interceptée ? Dans l'analyse de ces problèmes, les concepts

d'entropie, de redondance et autres développés dans "*Une théorie mathématique de la communication*" (ci-après MTC) trouveront une large application.

10. Secret parfait

Supposons que les messages possibles soient en nombre fini $M_1, \ldots, M_n$ et aient des probabilités a priori $P(M_1), \ldots, P(M_n)$, et que ceux-ci soient chiffrés dans les cryptogrammes possibles $E_1, \ldots, E_m$ par

$$E = T_i M.$$

Le cryptanalyste intercepte un E particulier et peut alors calculer, en principe du moins, les probabilités a posteriori pour les différents messages, $P_E(M)$. Il est naturel de définir le secret parfait par la condition que, pour tout E, les probabilités a posteriori soient égales aux probabilités a priori indépendamment de leurs valeurs. Dans ce cas, l'interception du message n'a fourni aucune information au cryptanalyste [10]. Toute action de sa part qui dépend de l'information contenue dans le cryptogramme ne peut être modifiée, car toutes ses probabilités quant à ce que contient le cryptogramme restent inchangées. En revanche, si la condition n'est pas satisfaite, il existera des situations où l'ennemi dispose de certaines probabilités a priori, et certains choix de clés et de messages pourront se produire, pour lesquels les probabilités de l'ennemi changeront. Ceci peut à son tour affecter ses actions, et donc le secret absolu n'est pas atteint. Par conséquent, la définition donnée est nécessairement requise par notre intuition de ce que devrait signifier le secret absolu.

Une condition nécessaire et suffisante pour un secret parfait peut être trouvée comme suit : on a, par le théorème de Bayes,

$$P_E(M) = \frac{P(M)P_M(E)}{P(E)}$$

où les symboles signifient :

- $P(M)$: probabilité a priori du message M.
- $P_M(E)$: probabilité conditionnelle du cryptogramme si le message est choisi, c'est-à-dire somme des probabilités de toutes les clés produisant le cryptogramme à partir du message.
- $P(E)$: probabilité d'obtenir un cryptogramme quelle qu'en soit la cause.
- $P_E(M)$: probabilité a posteriori du message si le cryptogramme est intercepté.

Pour un secret parfait, $P_E(M)$ doit être égal à $P(M)$ pour tout E et tout M. Par conséquent, soit $P(M) = 0$, mais cette solution est à exclure puisque l'égalité est exigée indépendamment des valeurs de $P(M)$, soit

$$P_M(E) = P(E)$$

pour tout M et E. Inversement, si $P_M(E) = P(E)$ alors

$$P_E(M) = P(M)$$

[10] Un puriste pourrait objecter que l'ennemi a obtenu une information dans la mesure où il sait qu'un message a été envoyé. On peut répondre à cela en ayant parmi les messages un "blanc" correspondant à "aucun message". Si aucun message n'est émis, le blanc est chiffré et envoyé sous forme de cryptogramme. Alors même ce minimum d'information restant est éliminé.

et on obtient un secret parfait. On a donc le résultat suivant :

Théorème 6. *Une condition nécessaire et suffisante pour un secret parfait est que*

$$P_M(E) = P(E)$$

pour tout M et tout E. C'est-à-dire que $P_M(E)$ doit être indépendant de M.

Autrement dit, la probabilité totale de toutes les clés transformant M en un cryptogramme E donné est égale à celle de toutes les clés transformant M_j en le même E, pour tous M_i, M_j et E.

Il doit donc y avoir autant de E que de M puisque, pour un i fixé, T_i donne une correspondance bijective entre tous les M et certains des E. Pour un secret parfait, $P_M(E) = P(E) \neq 0$ pour chacun de ces E et chacun de ces M. Il existe donc au moins une clé transformant tout M en chacun de ces E. Mais toutes les clés d'un M fixé vers différents E doivent être différentes, et donc le nombre de clés différentes est au moins aussi grand que le nombre de M. Il est possible d'obtenir un secret parfait avec seulement ce nombre de clés, comme on le montre par l'exemple suivant : soit les M_i numérotés de 1 à n et les E_i de même numéro, et en utilisant n clés, soit

$$T_i M_j = E_s$$

où $s = i + j \pmod{n}$. Dans ce cas, on constate que $P_E(M) = \dfrac{1}{n} = P(E)$ et que le secret est parfait. Un exemple est présenté à la figure 5 avec $s = i + j - 1 \pmod{5}$.

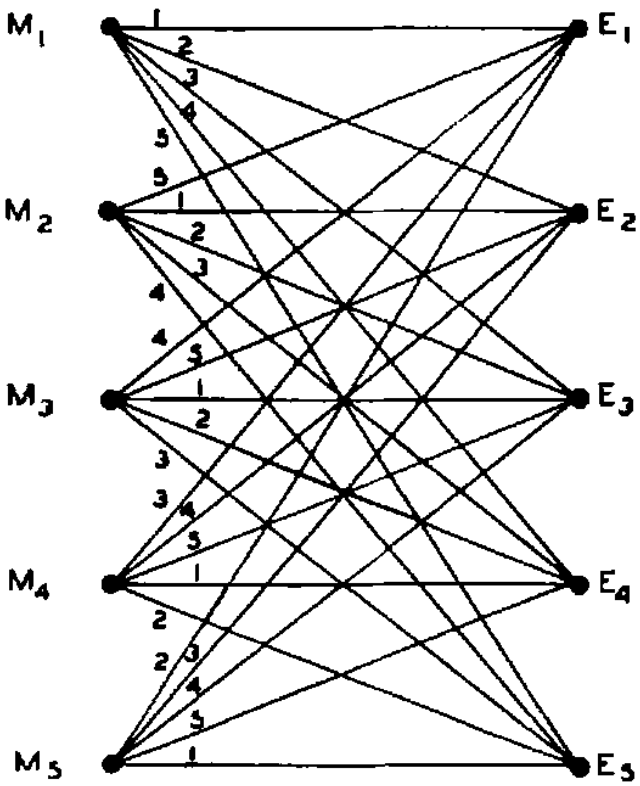

FIG. 5. Système parfait

Les systèmes parfaits dans lesquels le nombre de cryptogrammes, le nombre de messages et le nombre de clés sont tous égaux sont caractérisés par les propriétés suivantes : (1) chaque M est relié à chaque E par exactement une ligne ; (2) toutes les clés sont équiprobables. Ainsi, la représentation matricielle du système est un "carré latin".

Dans MTC, il a été démontré que l'information peut être facilement mesurée au moyen de l'entropie. Si nous disposons d'un ensemble de possibilités de probabilités $p_1, p_2, \ldots, p_n$, l'entropie H est donnée par :

$$H = -\sum p_i \log\ p_i.$$

Dans un système secret, deux choix statistiques sont impliqués : celui du message et celui de la clé. On peut mesurer la quantité d'information produite lorsqu'un message est choisi par $H(M)$:

$$H(M) = -\sum P(M) \log\ P(M),$$

la somme étant calculée sur tous les messages possibles. De même, il existe une incertitude associée au choix de la clé, donnée par :

$$H(K) = -\sum P(K) \log\ P(K).$$

Dans les systèmes parfaits du type décrit ci-dessus, la quantité d'information contenue dans le message est au plus de $\log n$ (ce qui se produit lorsque tous les messages sont équiprobables). Cette information ne peut être totalement masquée que si l'incertitude sur la clé est au moins de $\log n$. C'est le premier exemple d'un principe général qui apparaîtra fréquemment : il y a une limite à ce que nous pouvons obtenir avec une incertitude donnée sur la clé : la quantité d'incertitude que nous pouvons introduire dans la solution ne peut pas être supérieure à l'incertitude sur la clé. La situation est un peu plus compliquée si le nombre de messages est infini. Supposons, par exemple, que les messages soient générés sous forme de séquences infinies de lettres par un processus de Markov approprié. Il est clair qu'aucune clé finie ne garantit une confidentialité parfaite. Supposons alors que la source de la clé génère la clé de la même manière, c'est-à-dire sous forme d'une séquence infinie de symboles. Supposons également qu'une certaine longueur de clé L_K soit nécessaire pour chiffrer et déchiffrer un message de longueur L_M. Soit le logarithme du nombre de lettres de l'alphabet du message R_M et celui de l'alphabet de la clé R_K. Alors, dans le cas fini, il est évident qu'une confidentialité parfaite requiert

$$R_M L_M \leq R_K L_K.$$

Ce type de secret parfait est réalisé par le système Vernam.

Ces résultats ont été déduits sur la base de probabilités a priori inconnues ou arbitraires des messages. La clé requise pour une confidentialité parfaite dépend alors du nombre total de messages possibles.

On pourrait s'attendre à ce que, si l'espace des messages possède des statistiques connues et un taux moyen de génération d'information défini, au sens du MTC, la longueur de la clé nécessaire puisse être réduite en moyenne dans le rapport $\dfrac{R}{R_M}$, et c'est effectivement le cas. En fait, le message peut

être transmis à travers un transducteur qui élimine la redondance et réduit la longueur attendue dans ce rapport, puis un système de Vernam peut être appliqué au résultat. De toute évidence, la quantité de la clé utilisée par lettre du message est statistiquement réduite d'un facteur $\frac{R}{R_M}$ et, dans ce cas, la source de la clé et la source de l'information correspondent : un fragment de clé masque complètement un fragment d'information du message. Il est également facile de démontrer, par les méthodes utilisées dans l'article MTC, que c'est le meilleur résultat possible.

Les systèmes de chiffrement parfait ont leur place en pratique : ils peuvent être utilisés soit lorsqu'une importance primordiale est accordée au secret absolu, par exemple pour la correspondance entre les plus hauts niveaux de commandement, soit lorsque le nombre de messages possibles est faible. Ainsi, pour prendre un exemple extrême, si seulement deux messages "oui" ou "non" étaient anticipés, un système parfait serait approprié, avec peut-être la table de transformation suivante :

M	K	A	B
	oui	0	1
	non	1	0

L'inconvénient des systèmes parfaits pour les systèmes de correspondance volumineux réside, bien sûr, dans la quantité équivalente de clés à transmettre. Dans les sections suivantes, nous examinerons ce qui peut être réalisé avec des clés de taille réduite, en particulier avec des clés finies.

11. Équivoque

Supposons qu'un chiffrement par substitution simple ait été utilisé sur un texte anglais et que nous interceptions une certaine quantité de lettres du texte chiffré. Pour N assez grand, disons plus de 50 lettres, il existe presque toujours une solution unique au chiffrement ; c'est-à-dire qu'une seule séquence anglaise correcte se transforme en le texte intercepté par simple substitution. Avec un N plus petit, cependant, la probabilité d'obtenir plusieurs solutions est plus grande : avec $N = 15$, il y aura généralement un nombre important de fragments de texte possibles, tandis qu'avec $N = 8$, une bonne fraction (de l'ordre de 1/8) de toutes les séquences anglaises raisonnables de cette longueur sont possibles, car il y a rarement plus d'une lettre répétée parmi les 8. Avec $N = 1$, toute lettre est clairement possible et a la même probabilité a posteriori que sa probabilité a priori. Pour une lettre, le système est parfait.

Cela se produit généralement avec les chiffrements résolubles. Avant toute interception, nous pouvons imaginer les probabilités a priori attachées aux différents messages possibles, ainsi qu'aux différentes clés. À mesure que le contenu est intercepté, le cryptanalyste calcule les probabilités a posteriori ; et à mesure que N augmente, les probabilités de certains messages augmentent, et pour la plupart des autres, ces probabilités diminuent, jusqu'à ce qu'il n'en reste finalement qu'un, dont la probabilité est proche de 1, tandis que la probabilité totale de tous les autres est proche de zéro.

Ce calcul peut être réalisé pour des systèmes très simples. Le tableau 1 présente les probabilités a posteriori d'un chiffrement de type César appliqué à un texte anglais, la clé étant choisie aléatoirement parmi les 26 possibilités. Afin de permettre l'utilisation des tables de fréquence standard des lettres, des digrammes et des trigrammes, le texte a été démarré à un endroit aléatoire (en ouvrant un livre et en posant un crayon au hasard sur la page). Le message ainsi sélectionné

commence par "gmente de..." à l'intérieur du mot "augmente". Si le message devait commencer une phrase, un ensemble de probabilités différent devrait être utilisé, correspondant aux fréquences des lettres, des digrammes, etc., en début de phrase.

Table 1. probabilités a posteriori pour un cryptogramme de type César

déchiffrements	$N=1$	$N=2$	$N=3$	$N=4$	$N=5$
CREAS	.028	.0377	.1111	.3673	1
DSFBT	.038	.0314			
ETGCU	.131	.0881			
FUHDV	.029	.0189			
GVIEW	.020				
HWJFX	.053	.0063			
IXKGY	.063	.0126			
JYLHZ	.001				
KZMIA	.004				
LANJB	.034	.1321	.2500		
MBOKC	.025	.0222			
NCPLD	.071	.1195			
ODQME	.080	.0377			
PERNF	.020	.0818	.4389	.6327	
QFSOG	.001				
RGTPH	.068	.0126			
SHUQI	.061	.0881	.0056		
TIVRJ	.105	.2830	.1667		
UJWSK	.025				
VKXTL	.009				
WLYUM	.015	.0056			
XMZVN	.002				
YNAWO	.020				
ZOBXP	.001				
APCYQ	.082	.0503			
BQDZR	.014				
H (chiffres décimaux)	1.2425	.9686	.6034	.285	0

Le chiffrement de type César à clé aléatoire est un chiffrement pur et la clé particulière choisie n'affecte pas les probabilités a posteriori. Pour les déterminer, il suffit de lister les déchiffrements possibles par toutes les clés et de calculer leurs probabilités a priori. Les probabilités a posteriori sont celles-ci divisées par leur somme. Ces déchiffrements possibles sont trouvés par le processus standard de "dérouler l'alphabet" à partir du message et sont listés à gauche. Ils forment la classe résiduelle du message. Pour une lettre interceptée, les probabilités a posteriori sont égales aux probabilités a priori pour les lettres [11] et sont présentées dans la colonne intitulée $N = 1$. Pour deux lettres interceptées, les probabilités sont celles des digrammes ajustés à l'unité et sont indiquées dans la colonne $N = 2$.

Les fréquences des trigrammes ont également été tabulées et sont indiquées dans la colonne $N = 3$. Pour les séquences de quatre et cinq lettres, les probabilités ont été obtenues par multiplication à

[11] Les probabilités de ce tableau sont tirées des tables de fréquences données par Fletcher Pratt dans un livre "*Secret and Urgent*" publié par Blue Ribbon Books, New York, 1939. Bien qu'incomplètes, elles sont suffisantes pour les besoins actuels.

partir des fréquences des trigrammes, car, approximativement,

$$p(ijkl) = p(ijk)p_{jk}(l)$$

Notez qu'à trois lettres, le champ se réduit à quatre messages de probabilité assez élevée, les autres étant petits en comparaison. À quatre, il y a deux possibilités et à cinq, une seule : le déchiffrement correct.

En principe, cela pourrait être réalisé avec n'importe quel système, mais, à moins que la clé ne soit très petite, le nombre de possibilités est si important que le travail nécessaire empêche le calcul réel.

Cet ensemble de probabilités a posteriori décrit comment la connaissance du message et de la clé par le cryptanalyste devient progressivement plus précise à mesure que le matériau chiffré est obtenu. Cette description, cependant, est beaucoup trop complexe et difficile à obtenir pour nos besoins. Nous souhaitons une description simplifiée de cette approche de l'unicité des solutions possibles.

Une situation similaire se produit en théorie de la communication lorsqu'un signal transmis est perturbé par du bruit. Il est nécessaire d'établir une mesure appropriée de l'incertitude sur ce qui a été réellement transmis, en ne connaissant que la version perturbée du signal reçu. Dans l'article MTC, il a été démontré qu'une mesure mathématique naturelle de cette incertitude est l'entropie conditionnelle du signal transmis lorsque le signal reçu est connu. Cette entropie conditionnelle a été appelée, par commodité, l'équivoque.

Du point de vue du cryptanalyste, un système de chiffrement est quasiment identique à un système de communication bruité. Le message (signal transmis) est traité par un élément statistique, le système de chiffrement, avec sa clé statistiquement choisie. Le résultat de cette opération est le cryptogramme (analogue au signal perturbé) disponible pour l'analyse. Les principales différences entre les deux cas sont : premièrement, l'opération de transformation de chiffrement est généralement plus complexe que le bruit perturbateur d'un canal ; deuxièmement, la clé d'un système de chiffrement est généralement choisie parmi un ensemble fini de possibilités, tandis que le bruit d'un canal est le plus souvent introduit de manière continue, en fait choisi parmi un ensemble infini.

Compte tenu de ces considérations, il est naturel d'utiliser l'équivoque comme indice de secret théorique. Il convient de noter qu'il existe deux équivoques significatives : celle de la clé et celle du message. Ces valeurs seront notées respectivement par $H_E(K)$ et $H_E(M)$. Elles sont données par :

$$H_E(K) = \sum_{E,K} P(E,K) \ \log \ P_E(K)$$
$$H_E(M) = \sum_{E,M} P(E,M) \ \log \ P_E(K)$$

où E, M et K sont le cryptogramme, le message et la clé et

- $P(E,K)$: est la probabilité de la clé K et du cryptogramme$E$,
- $P_E(K)$: est la probabilité a posteriori de la clé K si le cryptogramme E est intercepté.

- $P(E, M)$ et $P_E(M)$: sont les probabilités similaires pour le message plutôt que pour la clé.

La somme dans $H_E(K)$ s'applique à tous les cryptogrammes possibles d'une certaine longueur (disons N lettres) et à toutes les clés. Pour $H_E(M)$, la somme s'applique à tous les messages et cryptogrammes de longueur N. Ainsi, $H_E(K)$ et $H_E(M)$ sont toutes deux des fonctions de N, le nombre de lettres interceptées. Ceci sera parfois indiqué explicitement en écrivant $H_E(K, N)$ et $H_E(M, N)$. Notez qu'il s'agit d'équivoques "totales" ; autrement dit, nous ne divisons pas par N pour obtenir le taux d'équivoque utilisé dans l'article MTC.

Les mêmes arguments généraux utilisés pour justifier l'équivoque comme mesure de l'incertitude en théorie de la communication s'appliquent ici. Notons qu'une équivoque nulle nécessite qu'un message (ou une clé) ait une probabilité unitaire, tous les autres étant nuls, ce qui correspond à une connaissance complète. Considérée comme une fonction de N, la diminution progressive de l'équivoque correspond à une connaissance croissante de la clé ou du message d'origine. Les deux courbes d'équivoque tracées en fonction de N seront appelées les caractéristiques d'équivoque du système de chiffrement en question.

Les valeurs de $H_E(K, N)$ et $H_E(M, N)$ pour le cryptogramme de type César considéré ci-dessus ont été calculées et sont données dans la dernière ligne du tableau 1. $H_E(K, N)$ et $H_E(M, N)$ sont égales dans ce cas et sont exprimées en chiffres décimaux (c'est-à-dire que le calcul utilise la base logarithmique 10). Il convient de noter que l'équivoque concerne ici un cryptogramme particulier, la somme ne portant que sur M (ou K), et non sur E. En général, la somme porterait sur tous les cryptogrammes interceptés possibles de longueur N et donnerait l'incertitude moyenne. Les difficultés de calcul sont prohibitives pour ce calcul général.

12. Propriétés de l'équivoque

On peut montrer que l'équivoque possède un certain nombre de propriétés intéressantes, dont la plupart correspondent à notre représentation intuitive du comportement d'une telle quantité. Nous allons d'abord montrer que l'équivoque de la clé ou d'une partie fixe d'un message diminue lorsque davantage de données chiffrées sont interceptées.

Théorème 7. *L'équivoque de la clé $H_E(K, N)$ est une fonction non croissante de N. L'équivoque des A premières lettres du message est une fonction non croissante du nombre N de lettres qui ont été interceptées. Si N lettres ont été interceptées, l'équivoque des N premières lettres du message est inférieure ou égale à celle de la clé. Ces équations peuvent s'écrire :*

$$H_E(K, S) \leq H_E(K, N) \qquad S \geq N$$
$$H_E(M, S) \leq H_E(M, N) \qquad S \geq N \ \ (H \text{ pour les premières lettres du texte})$$
$$H_E(M, N) \leq H_E(K, N)$$

La qualification concernant les lettres dans le deuxième résultat du théorème est telle que l'équivoque sera calculée par rapport à la quantité de messages interceptés. Si tel est le cas, l'équivoque du message peut (et c'est généralement le cas) augmenter pendant un certain temps, simplement parce que plus de lettres correspondent à une plus grande plage de messages possible. Les résultats du théorème correspondent à ce que l'on pourrait espérer d'un bon indice de confidentialité, car on ne

s'attendrait guère à être moins bien loti en moyenne après avoir intercepté davantage de données. Le fait que ces résultats puissent être prouvés justifie plus encore notre utilisation de la mesure d'équivoque.

Les résultats de ce théorème découlent de certaines propriétés de l'entropie conditionnelle démontrées dans l'article MTC. Ainsi, pour démontrer le premier ou le deuxième énoncé du théorème 7, nous avons pour tout événement aléatoire A et B

$$H(B) \geq H_A(B)$$

Si nous identifions B à la clé (connaissant les S premières lettres du cryptogramme) et A aux lettres restantes $N - S$, on obtient le premier résultat. De même, identifier B au message donne le deuxième résultat. Le dernier résultat découle de l'inégalité suivante

$$H_E(M) \leq H_E(K, M) = H_E(K) + H_{E,K}(M)$$

et du fait que $H_{E,K}(M) = 0$ puisque K et E déterminent de manière unique M.

Puisque le message et la clé sont choisis indépendamment, on a :

$$H(M, K) = H(M) + H(K)$$

De plus,

$$H(M, K) = H(E, K) = H(E) + H_E(K)$$

la première égalité résultant du fait que la connaissance de M et K ou de E et K équivaut à la connaissance des trois. En combinant ces deux formules, on obtient une formule pour l'équivoque de la clé :

$$H_E(K) = H(M) + H(K) - H(E),$$

En particulier, si $H(M) = H(E)$, alors l'équivoque de la clé, $H_E(K)$, est égale à l'incertitude a priori de la clé, $H(K)$. Ce phénomène se produit dans les systèmes parfaits décrits ci-dessus.

Une formule pour l'équivoque du message peut être trouvée de manière similaire. On a

$$H(M, E) = H(E) + H_E(M) = H(M) + H_M(E)$$
$$H_E(M) = H(M) + H_M(E) - H(E)$$

Soit un système produit $S = TR$, on peut s'attendre à ce que le second chiffrement diminue l'équivoque du message. On peut démontrer que cela est vrai comme suit : soit M, E_1, E_2 le message et les premier et second chiffrements, respectivement. Alors

$$P_{E_1 E_2}(M) = P_{E_1}(M).$$

Par conséquent

$$H_{E_1 E_2}(M) = H_{E_1}(M)$$

Puisque, pour toute variable aléatoire $x, y, z, H_{xy}(z) \leq H_y(z)$, on obtient le résultat souhaité, $H_{E_2}(M) \geq H_{E_1}(M)$.

Théorème 8. *L'équivoque dans le message d'un système produit $S = TR$ n'est pas inférieure à celle obtenue lorsque seul R est utilisé.*

Supposons maintenant que nous ayons un système T qui s'écrit comme la somme pondérée de plusieurs systèmes $R, S, \ldots, U$

$$T = p_1 R + p_2 S + \ldots + p_m U \qquad\qquad \sum p_i = 1$$

et que les systèmes $R, S, \ldots, U$ présentent des équivoques $H_1, H_2, \ldots, H_m$.

Théorème 9. *L'équivoque H d'une somme pondérée de systèmes est bornée par les inégalités*

$$\sum p_i H_i \leq H \leq \sum p_i H_i - \sum p_i \log p_i.$$

Ce sont les meilleures limites possibles. Les H peuvent être des équivoques de clé ou de message.

La limite supérieure est atteinte, par exemple, dans les systèmes fortement idéaux (décrits plus loin) où la décomposition se fait en transformations simples du système. La limite inférieure est atteinte si tous les systèmes $R, S, \ldots, U$ s'orientent vers des espaces cryptogrammes complètement différents. Ce théorème est également démontré par les inégalités générales régissant l'équivoque,

$$H_A(B) \leq H(B) \leq H(A) + H_A(B).$$

Identifions A au système particulier utilisé et B à la clé ou au message.

Il existe un théorème similaire pour les sommes pondérées de langages. Pour cela, on identifie A à la langue concernée.

Théorème 10. *Supposons qu'un système puisse s'appliquer aux langages $L_1, L_2, \ldots, L_m$ et possède des caractéristiques d'équivoque $H_1, H_2, \ldots, H_m$. Appliquée à la somme pondérée $\sum p_i L_i$, l'équivoque H est bornée par*

$$\sum p_i H_i \leq H \leq \sum p_i H_i - \sum p_i \log p_i.$$

Ces limites sont optimales et les équivoques en question peuvent concerner la clé ou le message.

La redondance totale D_N pour N lettres du message est définie par

$$D_N = \log G - H(M)$$

où G est le nombre total de messages de longueur N et $H(M)$ l'incertitude sur le choix de l'un d'eux. Dans un système de chiffrement où le nombre total de cryptogrammes possibles est égal au nombre de messages possibles de longueur N, $H(E) \leq \log G$. Par conséquent,

$$\begin{aligned} H_E(K) &= H(K) + H(M) - H(E) \\ &\geq H(K) - [\log G - H(M)]. \end{aligned}$$

D'où

$$H(K) - H_E(K) \leq D_N$$

Cela montre que, dans un système fermé, par exemple, la diminution de l'équivoque de la clé après l'interception de N lettres n'est pas supérieure à la redondance de N lettres du langage. Dans de tels systèmes, qui constituent la majorité des chiffrements, seule l'existence d'une redondance dans les messages originaux permet une solution.

Supposons maintenant que nous ayons un système pur. Soient les différentes classes résiduelles de messages $C_1, C_2, C_3, \ldots, C_r$, et l'ensemble correspondant de classes résiduelles de cryptogrammes $C'_1, C'_2, \ldots, C'_r$. La probabilité d'occurrence de chaque E dans C'_1 est la même :

$$P(E) = \frac{P(C_i)}{\varphi_i} \qquad\qquad E \text{ un élément de } C_i$$

où φ_i est le nombre de messages différents dans C_i. On a donc

$$\begin{aligned} H(E) &= -\sum_i \varphi_i \frac{P(C_i)}{\varphi_i} \log \frac{P(C_i)}{\varphi_i} \\ &= -\sum P(C_i) \log \frac{P(C_i)}{\varphi_i}. \end{aligned}$$

En effectuant la substitution pour $H_E(K)$ dans notre équation, on obtient le :

Théorème 11. *Pour un chiffrement pur*

$$H_E(K) = H(K) + H(M) + \sum_i P(C_i) \log \frac{P(C_i)}{\varphi_i}.$$

Ce résultat peut être utilisé pour calculer $H_E(K)$ dans certains cas intéressants.

13. Équivoque pour la substitution simple sur un langage à deux lettres

Nous allons maintenant calculer l'équivoque dans la clé ou le message lorsque la substitution simple est appliquée à une langue à deux lettres, avec des probabilités p et q pour 0 et 1, et des lettres successives choisies indépendamment. On a $\mathrm{H}_E(M) = H_E(K) = -\sum P(E) P_E(K) \log P_E(K)$
La probabilité que E contienne exactement s 0 dans une permutation particulière est :

$$\frac{1}{2}\left(p^s q^{N-s} + q^s p^{N-s}\right)$$

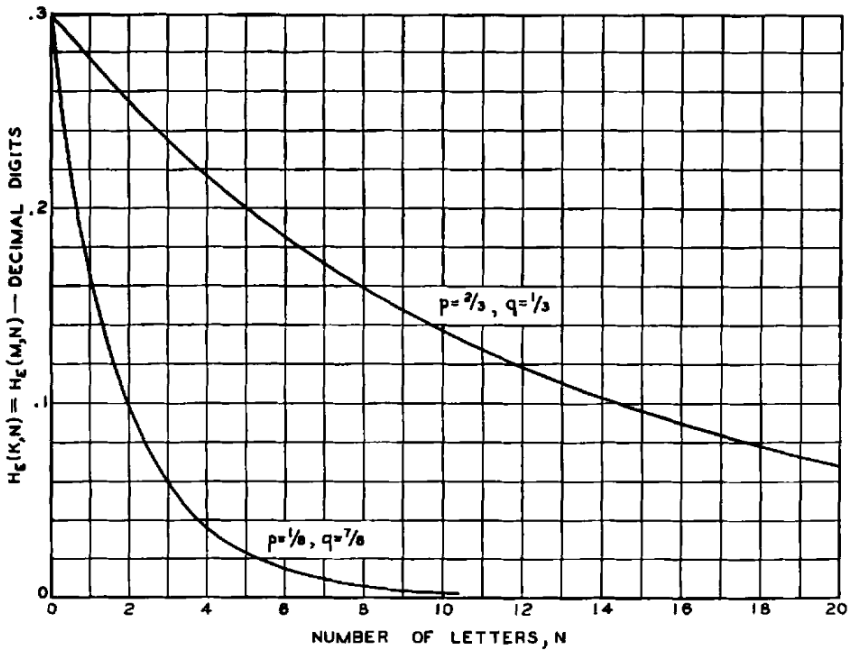

FIG. 6. Équivoque pour la substitution simple en langage à deux lettres

et les probabilités a posteriori des substitutions identité et inversion (les deux seules substitutions du système) sont respectivement :

$$P_E(0) = \frac{p^s q^{N-s}}{(p^s q^{N-s} + q^s p^{N-s})} \qquad P_E(1) = \frac{p^{N-s} q^s}{(p^s q^{N-s} + q^s p^{N-s})}.$$

Il y a $\binom{N}{s}$ termes pour chaque s et donc

$$H_E(K,N) = -\sum_s \binom{N}{s} p^s q^{N-s} \log \frac{p^s q^{N-s}}{(p^s q^{N-s} + q^s p^{N-s})}.$$

Pour $p = \frac{1}{3}, q = \frac{2}{3}$, et pour $p = \frac{1}{8}, q = \frac{7}{8}$, $H_E(K,N)$ a été calculé et est représenté sur la Fig. 6.

14. La caractéristique de l'équivoque pour un chiffrement "aléatoire"

Dans la section précédente, nous avons calculé l'équivoque d'une substitution simple appliquée à une langue à deux lettres. Il s'agit du type de chiffrement et de la structure de langue les plus simples possibles, mais les formules sont déjà si complexes qu'elles sont quasiment inutiles. Que faire des cas d'intérêt pratique, par exemple les transformations complexes d'un système de transposition fractionnaire appliqué à l'anglais, dont la structure statistique est extrêmement complexe ? Cette complexité suggère une méthode d'approche. Des problèmes suffisamment complexes peuvent souvent être résolus statistiquement. Pour faciliter cette tâche, on définit la notion de chiffrement "aléatoire".

On fait les hypothèses suivantes :

1. Le nombre de messages possibles de longueur N est $T = 2^{R_0 N}$, donc $R_0 = \log_2 G$, où G est le nombre de lettres de l'alphabet. Le nombre de cryptogrammes possibles de longueur N est également supposé égal à T.

2. Les messages possibles de longueur N peuvent être divisés en deux groupes : un groupe de probabilité a priori élevée et relativement uniforme, et un second groupe de probabilité totale négligeable. Le groupe de probabilité élevée contiendra $S = 2^{RN}$ messages, où $R = H(M)/N$, c'est-à-dire que R est l'entropie de la source du message par lettre.

3. L'opération de déchiffrement peut être considérée comme une série de lignes, comme dans les figures 2 et 4, reliant chaque E à différents M. Nous supposons différentes clés équiprobables, de sorte qu'il y aura k lignes sortant de chaque E. Pour le chiffrement aléatoire, nous supposons que les lignes sortant de chaque E renvoient à une sélection aléatoire des messages possibles. En réalité, un chiffrement aléatoire est un ensemble de chiffres et l'équivoque est l'équivoque moyenne de cet ensemble.

L'équivoque de la clé est définie par

$$H_E(K) = \sum P(E) P_E(K) \log\ P_E(K).$$

La probabilité qu'exactement m lignes reviennent d'un E particulier vers le groupe de messages à forte probabilité est

$$\binom{k}{m} \left(\frac{S}{T}\right)^m \left(1 - \frac{S}{T}\right)^{k-m}$$

Si un cryptogramme comportant m telles lignes est intercepté, l'équivoque est $\log\ m$. La probabilité d'un tel cryptogramme est $\dfrac{mT}{SK}$, car il peut être produit par m clés issues de messages à forte probabilité, chacune ayant une probabilité $\dfrac{T}{S}$. L'équivoque est donc :

$$H_E(K) = \frac{T}{Sk} \sum_{m=1}^{k} \binom{k}{m} \left(\frac{S}{T}\right)^m \left(1 - \frac{S}{T}\right)^{k-m} m \log\ m$$

Nous souhaitons trouver une approximation simple de ce résultat lorsque k est grand. Si l'espérance mathématique de m, à savoir $\overline{m} = Sk/T$, est $\gg T$, la variation de $\log m$ sur l'intervalle où la distribution binomiale suppose de grandes valeurs sera faible, et nous pouvons remplacer $\log m$ par $\log \overline{m}$. Ceci peut maintenant être exclu de la sommation, qui se réduit alors à $\overline{m}$. Ainsi, dans cette condition,

$$H_E(K) \doteq \log\ \frac{Sk}{T} = \log\ S - \log\ T + \log\ k$$

$$H_E(K) \doteq H(K) - DN,$$

où D est la redondance par lettre de la langue d'origine ($D = D_N/N$).

Si $\overline{m}$ est petit par rapport à k, la distribution binomiale peut être approximée par une distribution de Poisson :

$$\binom{k}{m} p^m q^{k-m} \doteq \frac{e^{-\lambda} \lambda^m}{m\,!}$$

où $\lambda = \dfrac{Sk}{T}$. Par conséquent

$$H_E(K) \doteq \frac{1}{\lambda} e^{-\lambda} \sum_{2}^{\infty} \frac{\lambda^m}{m\,!} m \log\, m.$$

Si l'on remplace m par $m+1$, on obtient :

$$H_E(K) \doteq e^{-\lambda} \sum_{1}^{\infty} \frac{\lambda^m}{m\,!}\ \log(m+1).$$

Ceci peut être utilisé dans la région où λ est proche de l'unité. Pour $\lambda \ll 1$, le seul terme important de la série est celui pour $m = 1$; en omettant les autres, on a :

$$\begin{aligned} H_E(K) &\doteq e^{-\lambda} \lambda \log\, 2 \\ &\doteq \lambda \log\, 2 \\ &\doteq 2^{-ND} k \log\, 2. \end{aligned}$$

En résumé : $H_E(K)$, considérée comme une fonction de N, le nombre de lettres interceptées, commence à $H(K)$ lorsque $N = 0$. Elle décroît linéairement avec une pente $-D$ jusqu'au voisinage de $N = \dfrac{H(K)}{D}$. Après une courte région de transition, $H_E(K)$ suit une exponentielle avec une distance de "demi-vie" $\dfrac{1}{D}$ si D est mesuré en bits par lettre. Ce comportement est illustré sur la figure 7, avec les courbes d'approximation.

Par un argument similaire, l'équivoque du message peut être calculée. On a

$$\begin{aligned} H_E(M) &= R_0 N \quad \text{for} \quad R_0 N \ll H_E(K) \\ H_E(M) &= H_E(K) \quad \text{for} \quad R_0 N \gg H_E(K) \\ H_E(M) &= H_E(K) - \varphi(N) \quad \text{for} \quad R_0 N \sim H_E(K) \end{aligned}$$

où $\varphi(N)$ est la fonction représentée sur la figure 7, avec une échelle N réduite d'un facteur $\dfrac{D}{R_0}$. Ainsi, $H_E(M)$ croît linéairement selon une pente R_0, jusqu'à presque croiser la droite $H_E(K)$. Après une transition arrondie, elle suit la courbe $H_E(K)$ vers le bas.

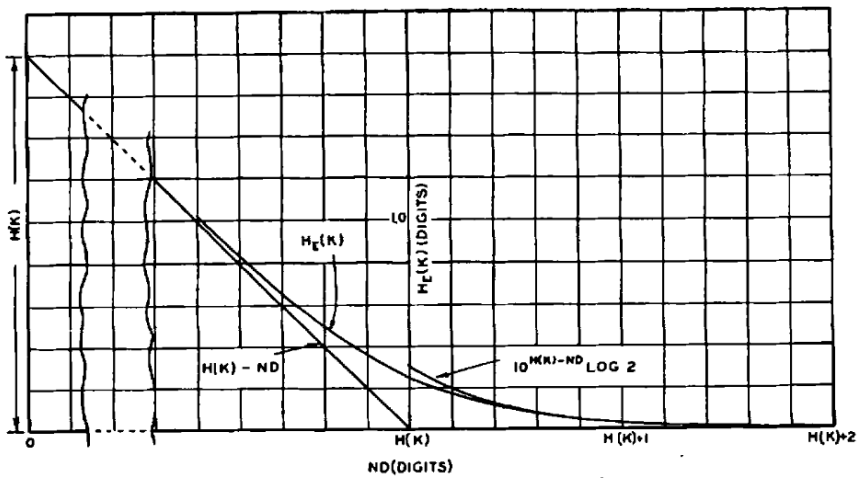

Fig. 7. Équivoque pour le chiffrement aléatoire

La figure 7 montre que les courbes d'équivoque se rapprochent assez brusquement de zéro. On peut donc, sans grande ambiguïté, parler d'un point où la solution devient unique. Ce nombre de lettres sera appelé distance d'unicité. Pour le chiffrement aléatoire, elle est approximativement de $H(K)/D$.

15. Application à des chiffrements standards

La plupart des chiffrements standards impliquent des opérations de chiffrement et de déchiffrement assez complexes. De plus, la structure statistique des langues naturelles est extrêmement complexe. Il est donc raisonnable de supposer que les formules dérivées du chiffrement aléatoire peuvent être appliquées dans de tels cas. Il est toutefois nécessaire d'appliquer certaines corrections dans certains cas. Les principaux points à observer sont les suivants :

1. Nous avons supposé, pour le chiffrement aléatoire, que les déchiffrements possibles d'un cryptogramme sont une sélection aléatoire parmi les messages possibles. Bien que cela ne soit pas strictement vrai dans les systèmes ordinaires, cela se rapproche de la réalité à mesure que la complexité des opérations de chiffrement et de la structure du langage augmente. Avec un chiffrement par transposition, il est clair que les fréquences des lettres sont préservées lors des opérations de déchiffrement. Cela signifie que les déchiffrements possibles sont choisis dans un groupe plus restreint, et non dans l'espace des messages entier, et la formule doit être modifiée. Au lieu de R_0, on utilise R_1, le taux d'entropie d'une langue avec des lettres indépendantes, mais avec des fréquences de lettres régulières. Dans d'autres cas, on observe une nette tendance à ramener les déchiffrements à des messages à forte probabilité. En l'absence de tendance claire de ce type et si le système est relativement complexe, il est alors raisonnable d'utiliser l'analyse du chiffrement aléatoire.

2. Dans de nombreux cas, la clé complète n'est pas utilisée pour chiffrer des messages courts. Par exemple, lors d'une substitution simple, seuls les messages assez longs contiendront toutes les lettres de l'alphabet et impliqueront donc la clé complète. Évidemment, l'hypothèse aléatoire ne tient pas pour un petit N dans ce cas, car toutes les clés qui ne diffèrent que par les lettres qui n'apparaissent pas encore dans le cryptogramme mènent au même message et ne sont pas distribuées aléatoirement. Cette erreur est facilement corrigée par l'utilisation d'une

"caractéristique d'apparence de clé". On utilise, pour un N donné, la quantité effective de clé attendue avec cette longueur de cryptogramme. Pour la plupart des chiffrements, cette estimation est facile.

3. Certains "effets finaux" dus au début défini du message produisent un écart par rapport aux caractéristiques aléatoires. Si l'on prend un point de départ aléatoire dans un texte anglais, la première lettre (lorsque l'on n'observe pas les lettres précédentes) peut être n'importe quelle lettre avec les probabilités habituelles des lettres. La lettre suivante est plus complètement spécifiée, car nous disposons alors de fréquences de digrammes. Cette diminution du choix de la valeur se poursuit pendant un certain temps. L'effet de cela sur la courbe est que la droite est déplacée et approximée par une courbe qui dépend de la façon dont la structure statistique de la langue se dissémine sur les lettres adjacentes. En première approximation, la courbe peut être corrigée en décalant la droite jusqu'au point moitié de la redondance, c'est-à-dire le nombre de lettres où la redondance du langage atteint la moitié de sa valeur finale.

Si l'on tient compte de ces trois effets, on peut faire des estimations raisonnables de la caractéristique d'équivoque et du point d'unicité. Le calcul peut être fait graphiquement comme indiqué dans la Fig. 8. On trace la caractéristique d'apparence de la clé et la courbe de redondance totale D_N (qui est généralement suffisamment bien représentée par la droite ND_∞). La différence entre celles-ci au voisinage de leur intersection est $H_E(M)$. Avec un chiffrement par substitution simple appliqué à l'anglais, ce calcul a donné les courbes présentées dans la Fig. 9. La caractéristique d'apparence de la clé dans ce cas a été estimée en comptant le nombre de lettres différentes apparaissant dans des passages anglais typiques de N lettres. Dans la mesure où des données expérimentales sur la substitution simple ont pu être trouvées, elles concordent très bien avec les courbes de la Fig. 9, compte tenu des diverses idéalisations et approximations qui ont été faites. Par exemple, le point d'unicité, à environ 27 lettres, peut être montré expérimentalement comme se situant entre les limites 20 et 30. Avec 30 lettres, un cryptogramme de ce type a presque toujours une solution unique, et avec 20, il est généralement facile de trouver plusieurs solutions.

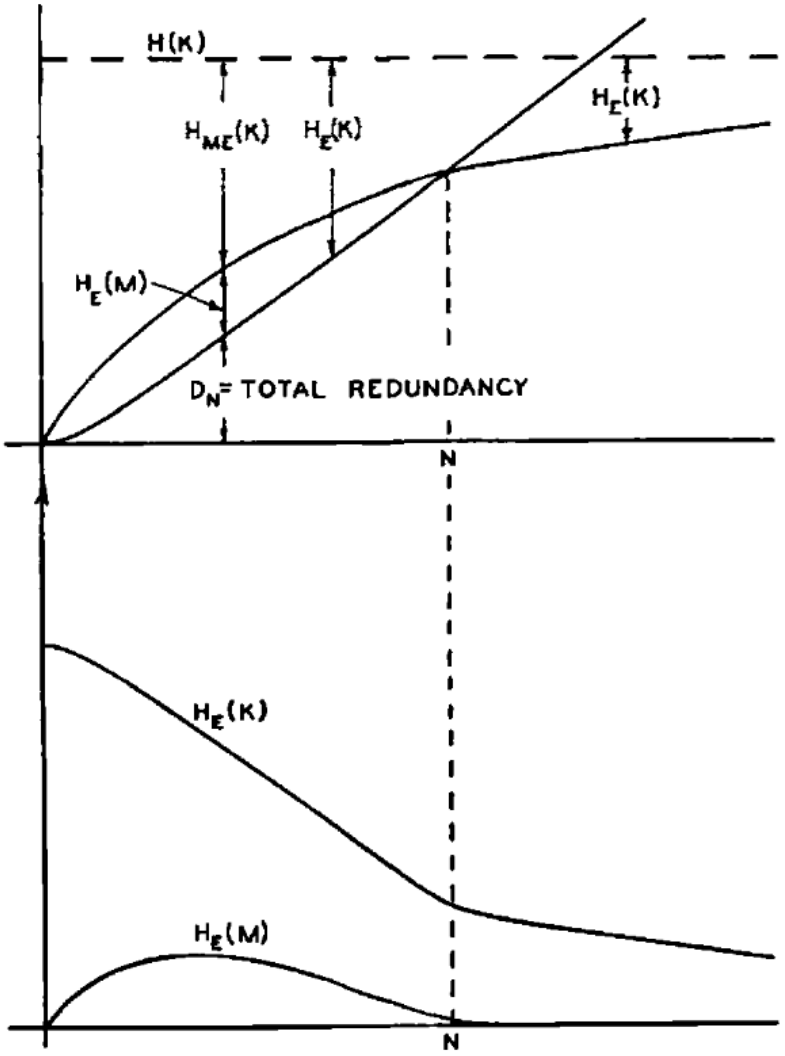

FIG. 8. Calcul graphique de l'équivoque

Avec une transposition de période d (clé aléatoire), $H(K) = \log d!$, soit environ $d \log d/e$ (en utilisant une approximation de Stirling pour $d!$). Si l'on prend $0,6$ chiffres décimaux par lettre comme redondance appropriée, en gardant à l'esprit la préservation des fréquences de lettres, on obtient environ $1,7\, d \log d/e$ comme distance d'unicité. Ceci est également assez bien vérifié expérimentalement. Notons que dans ce cas, $H_E(M)$ n'est défini que pour des multiples entiers de d.

Avec le chiffrement de Vigenère, le point d'unicité se situe à environ $2d$ lettres, ce qui est également correct. La caractéristique de Vigenère avec la même taille de clé comme substitution simple sera approximativement celle illustrée sur la figure 10.

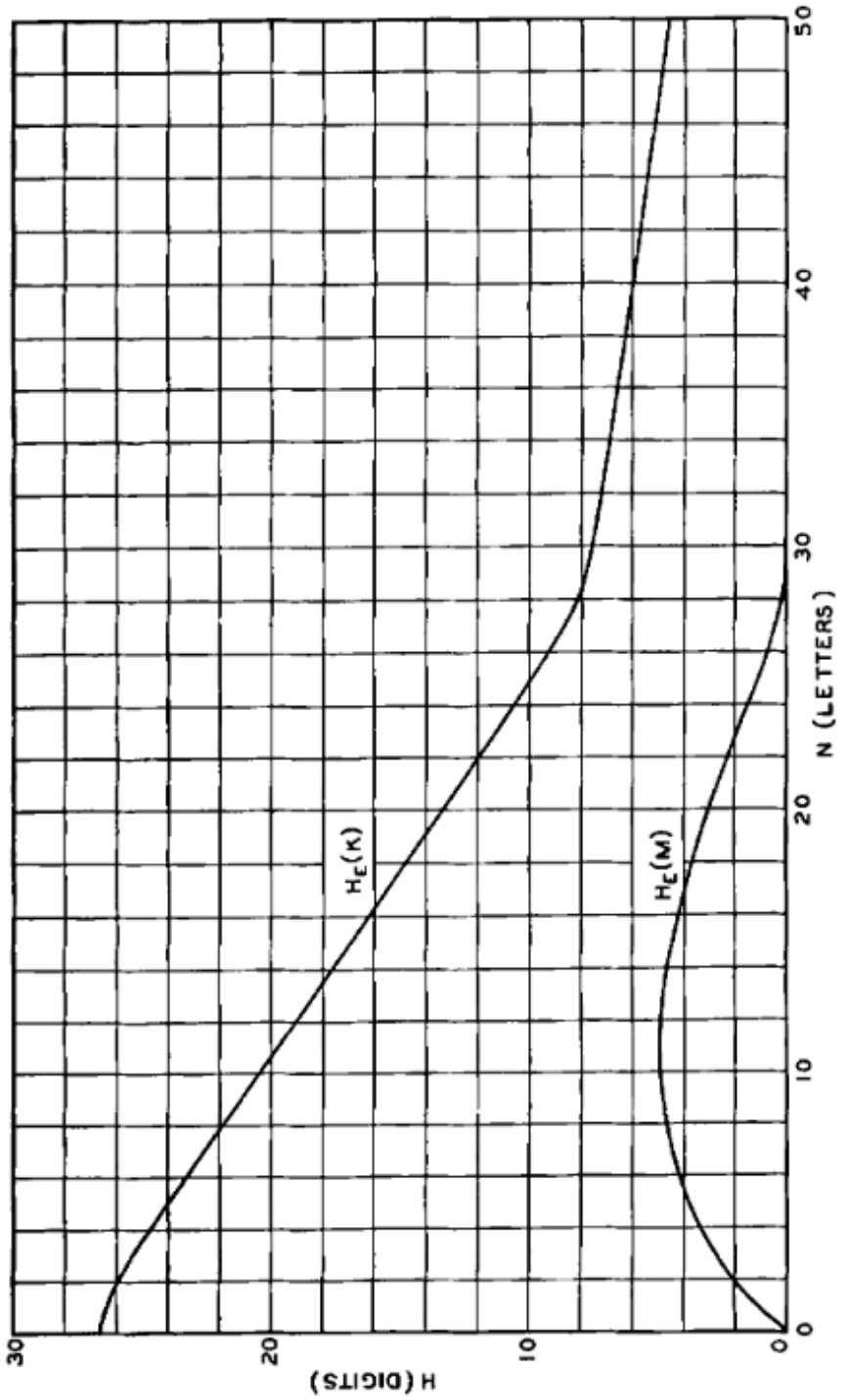

FIG. 9. Équivoque pour substitution simple pour l'anglais

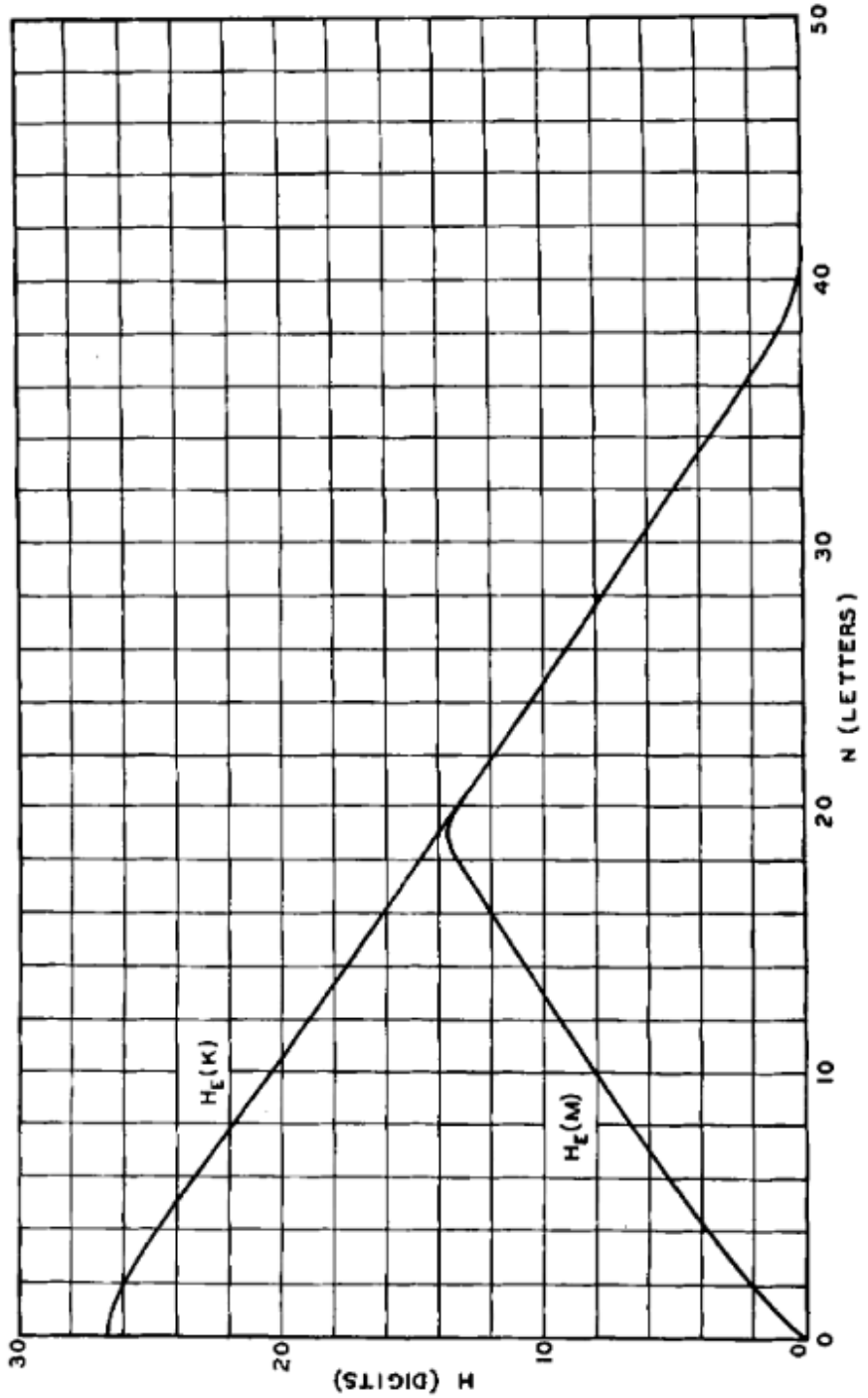

FIG. 10. Équivoque pour Vigenère pour l'anglais

Les cas des chiffrements de Vigenère, de Playfair et fractionnaire sont plus susceptibles de suivre les formules théoriques des chiffrements aléatoires que le chiffrement par substitution et transposition simples. Cela s'explique par leur plus grande complexité et par l'amélioration des caractéristiques de mélange des messages sur lesquels ils opèrent.

L'alphabet mixte Vigenère (chacun des d alphabets mélangés indépendamment et utilisés séquentiellement) a une taille de clé de

$$H(K) = d \log 26\,! = 26.3d$$

et son point d'unicité devrait se situer à environ $53d$ lettres.

Ces conclusions peuvent également être soumises à un test expérimental approximatif avec le chiffrement de type César. Dans le cryptogramme analysé dans le tableau 1, section 11, la fonction $H_E(K, N)$ a été calculée et est donnée ci-dessous, avec les valeurs pour un chiffrement aléatoire.

N	0	1	2	3	4	5
H (observé)	1.41	1.24	0.97	0.60	0.28	0
H (calculé)	1.41	1.25	0.98	0.54	0.15	0.03

L'accord semble assez bon, surtout si l'on se souvient que le H observé devrait en réalité être la moyenne de nombreux cryptogrammes différents, et que D pour les valeurs les plus élevées de N n'est qu'une estimation approximative.

Il apparaît donc que l'analyse du chiffrement aléatoire peut être utilisée pour estimer les caractéristiques d'équivoque et la distance d'unicité pour les types de chiffrements ordinaires.

16. Validité d'une solution de cryptogramme

Les formules d'équivoque sont pertinentes pour les questions qui se posent parfois en cryptographie concernant la validité d'une prétendue solution à un cryptogramme. Dans l'histoire de la cryptographie, de nombreux cryptogrammes, ou cryptogrammes possibles, ont été résolus par des analystes astucieux. Cependant, le processus était si complexe, ou le matériau utilisable était si peu conséquent, que la question s'est posée de savoir si le cryptanalyste avait "lu une solution" dans le cryptogramme. Voir, par exemple, les chiffrements de Bacon-Shakespeare et le manuscrit "Roger Bacon". [12].

En général, on peut dire que si un système et une clé proposés résolvent un cryptogramme pour une longueur de matériau utilisable considérablement supérieure à la distance d'unicité, la solution est fiable. Si le matériau est d'une taille du même ordre ou inférieur à la distance d'unicité, la solution est hautement suspecte.

Cet effet de redondance, produisant progressivement une solution unique à un chiffrement, peut être envisagé d'une autre manière, ce qui est utile. La redondance est essentiellement une série de conditions sur les lettres du message, qui garantissent son caractère statistiquement raisonnable. Ces conditions de cohérence produisent des conditions de cohérence correspondantes dans le cryptogramme. La clé confère une certaine liberté au cryptogramme, mais, à mesure que de plus en plus de lettres sont interceptées, les conditions de cohérence épuisent la liberté accordée par la clé. Finalement, il n'y a qu'un seul message et une seule clé qui satisfont toutes les conditions, et on obtient une solution unique. Dans le chiffrement aléatoire, les conditions de cohérence sont, en un sens, "orthogonales" au "grain de la clé" et produisent leur plein effet en éliminant les messages

[12] Voir Fletcher Pratt, loc. cit.

et les clés aussi rapidement que possible. C'est le cas habituel. Cependant, par une conception appropriée, il est possible d'"aligner" la redondance du langage sur le "grain de la clé" de telle sorte que les conditions de cohérence soient automatiquement satisfaites et que $H_E(K)$ ne s'approche pas de zéro. Ces systèmes "idéaux", qui seront considérés dans la section suivante, sont de nature telle que les transformations induisent les mêmes probabilités sur tout l'espace.

17. Systèmes secrets idéaux

Nous avons vu qu'un secret parfait nécessite une quantité infinie de clés si l'on autorise des messages de longueur illimitée. Avec une taille de clé finie, l'équivoque entre clé et message tend généralement vers zéro, mais pas nécessairement. En fait, il est possible que $H_E(K)$ reste constant à sa valeur initiale $H(K)$. Alors, quelle que soit la quantité de données interceptées, il n'existe pas de solution unique, mais plusieurs solutions de probabilité comparable. Nous définirons un système "idéal" comme un système dans lequel $H_E(K)$ et $H_E(M)$ ne tendent pas vers zéro lorsque $N \to \infty$. Un système "fortement idéal" est un système dans lequel $H_E(K)$ reste constant égal à $H(K)$.

Un exemple est une substitution simple sur une langue artificielle dans laquelle toutes les lettres sont équiprobables et les lettres successives choisies indépendamment. On constate facilement que $H_E(K) = H(K)$ et que $H_E(M)$ croît linéairement le long d'une droite de pente $\log G$ (où G est le nombre de lettres de l'alphabet) jusqu'à atteindre la droite $H(K)$, après quoi il reste constant à cette valeur.

Avec les langues naturelles, il est généralement possible d'approcher la caractéristique idéale : le point d'unicité peut être obtenu pour une grandeur aussi grande que souhaitée. Cependant, la complexité du système requis augmente généralement rapidement lorsque l'on tente d'y parvenir. Il n'est pas toujours possible d'atteindre réellement la caractéristique idéale avec un système de complexité finie.

Pour approcher l'équivoque idéale, on peut d'abord opérer sur le message avec un transducteur supprimant toutes les redondances. Ensuite, presque tout système de chiffrement simple (substitution, transposition, chiffrement de Vigenère, etc.) est satisfaisant. Plus le transducteur est élaboré et plus le résultat est proche de la forme souhaitée, plus le système de chiffrement approchera la caractéristique idéale.

Théorème 12. *Une condition nécessaire et suffisante pour que T soit fortement idéal est que, pour deux clés quelconques, $T_i^{-1}T_j$ soit une transformation préservant la mesure, de l'espace des messages vers lui-même.*

Ceci est vrai puisque la probabilité a posteriori de chaque clé est égale à sa probabilité a priori si et seulement si cette condition est satisfaite.

18. Exemples de systèmes secrets idéaux

Supposons que notre langage soit constitué d'une séquence de lettres, toutes choisies indépendamment et avec des probabilités égales. La redondance est alors nulle, et d'après le résultat de la section

12, $H_E(K) = H(K)$. La conséquence de cela est le

Théorème 13. *Si toutes les lettres sont équiprobables et indépendantes, tout chiffrement fermé est fortement idéal.*

L'équivoque du message augmentera en fonction de la caractéristique d'apparition de la clé, qui sera généralement proche de $H(K)$, bien que ce ne soit pas toujours le cas. Dans les cas de la substitution de n-grammes, de la transposition, du chiffrement de Vigenère et de ses variations, du chiffrement fractionnaire, etc., nous avons des systèmes fortement idéaux pour ce langage simple avec $H_E(M) \to H(K)$ lorsque $N \to \infty$.

Les systèmes secrets idéaux présentent plusieurs inconvénients.

1. Le système doit être étroitement adapté au langage. Cela nécessite une étude approfondie de la structure du langage par le concepteur. De plus, une modification de la structure statistique ou une sélection parmi les messages possibles, comme dans le cas des mots probables (mots attendus en présence de ce cryptogramme particulier), rend le système vulnérable à l'analyse.

2. La structure des langues naturelles est extrêmement complexe, ce qui implique une complexité des transformations nécessaires pour éliminer les redondances. Ainsi, toute machine capable d'effectuer cette opération doit nécessairement être très complexe, au moins en ce qui concerne le stockage de l'information, car il faut s'attendre à un "dictionnaire" d'une ampleur supérieure à celle d'un dictionnaire ordinaire.

3. En général, les transformations requises introduisent une propagation malencontreuse des erreurs. Une erreur dans la transmission d'une seule lettre produit une région de changements à proximité, d'une taille comparable à la longueur des effets statistiques dans la langue d'origine.

19. Remarques supplémentaires au sujet de l'équivoque et de la redondance

Nous avons estimé que la redondance de l'anglais normal était d'environ 10 chiffres décimaux par lettre, soit une redondance de 50 %. Ceci suppose l'omission des césures de mots. Il s'agit d'une valeur approximative basée sur une structure statistique s'étendant sur environ 10 lettres, et supposant que le texte soit de type ordinaire, comme un article de journal, une œuvre littéraire, etc. Nous pouvons noter ici une méthode d'estimation approximative de ce nombre, qui présente un intérêt cryptographique.

Un chiffrement à clé courante est un système de type Vernam où, au lieu d'une séquence aléatoire de lettres, la clé est un texte significatif. Il est maintenant connu que les chiffrements à clé courante peuvent généralement être résolus de manière unique. Cela montre que l'anglais peut être réduit d'un facteur deux à un et implique une redondance d'au moins 50 %. Ce chiffre ne peut toutefois pas être augmenté de manière significative, pour plusieurs raisons, à moins de prendre en compte la structure de "signification" à long terme de l'anglais.

Le chiffrement à clé courante peut être facilement amélioré pour aboutir à des systèmes de chiffrement impossibles à résoudre sans la clé. En utilisant, à la place d'un texte anglais, plusieurs textes différents comme clé, et en les ajoutant tous au message, on obtient une quantité de clé suffisante pour produire une équivoque positive élevée. Une autre méthode consiste à utiliser, par exemple, une lettre sur dix du texte comme clé. Les lettres intermédiaires sont omises et ne peuvent être utilisées à aucun autre endroit du message. L'effet est sensiblement le même, car ces lettres espacées sont quasiment indépendantes.

Le fait que les voyelles d'un passage puissent être omises sans perte essentielle suggère une méthode simple pour améliorer considérablement la plupart des systèmes de chiffrement. Il faut d'abord supprimer toutes les voyelles, ou autant de portions des messages que possible sans détruire le risque de reconstructions multiples, puis chiffrer le résidu. Comme cela réduit la redondance d'un facteur de 3 ou 4 pour 1, le point d'unicité sera décalé d'un facteur. C'est une façon d'approcher les systèmes idéaux : utiliser les connaissances en anglais du déchiffreur comme partie intégrante du système de déchiffrement.

20. Distribution de l'équivoque

Une description plus complète d'un système de chiffrement appliqué à une langue autre que celle fournie par les caractéristiques d'équivoque peut être établie en donnant la distribution d'équivoque. Pour les lettres interceptées, on considère la fraction de cryptogrammes pour lesquels l'équivoque de ces lettres, et non la moyenne, se situe entre certaines limites. Cela donne une fonction de distribution de densité

$$P(H_E(M), N)\, dH_E(M)$$

pour la probabilité que, pour N lettres, H soit compris entre les limites H et $H + dH$. L'équivoque moyenne que nous avons étudiée précédemment est la moyenne de cette distribution. La fonction $P(H_E(M), N)$ peut être représentée comme tracée selon une troisième dimension, orthogonale au papier, sur le plan $H_E(M), N$. Si le langage est pur, avec une faible portée d'influence, et que le chiffrement est pur, la fonction sera généralement une crête dans ce plan dont le point le plus élevé suit approximativement la moyenne $H_E(M)$, au moins jusqu'à proximité du point d'unicité. Dans ce cas, ou lorsque les conditions sont presque vérifiées, la courbe moyenne donne une image relativement complète du système.

En revanche, si le langage n'est pas pur, mais constitué d'un ensemble de composantes pures

$$L = \sum p_i L_i$$

si le système présente des courbes d'équivoque différentes, la distribution totale sera généralement constituée d'une série de crêtes. Il y en aura une pour chaque L_i pondérée selon son p_i. La caractéristique d'équivoque moyenne sera une ligne située quelque part au milieu de ces crêtes et on pourrait ne pas avoir une image très complète de la situation. Ceci est illustré sur la figure 11. Un effet similaire se produit si le système n'est pas pur, mais composé de plusieurs systèmes aux courbes différentes.

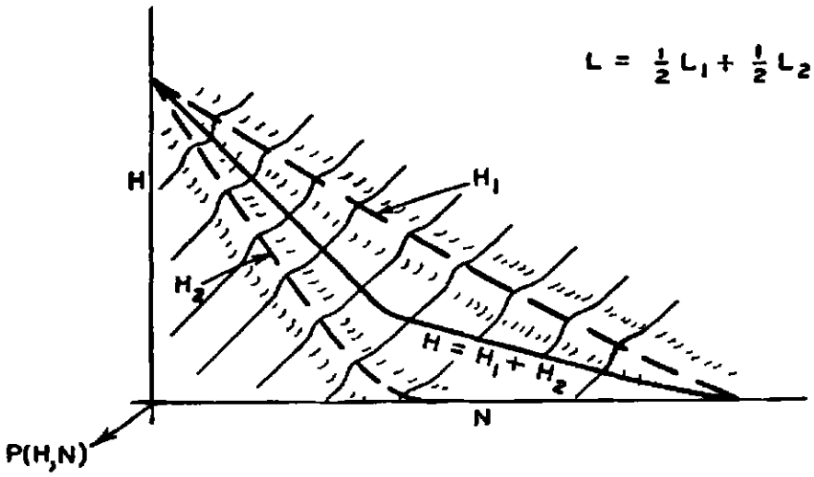

FIG. 11. Distribution de l'équivoque avec un langage mélangé $L = \frac{1}{2}L_1 + \frac{1}{2}L_2$

Le mélange de langues pures proches les unes des autres en termes de structure statistique a pour effet d'augmenter la largeur de la crête. Près du point d'unicité (voir la Fig. 11), ceci tend à augmenter l'équivoque moyenne, car l'équivoque ne peut pas devenir négative et la propagation est principalement positive. Nous nous attendons donc à ce que, dans cette région, les calculs basés sur le chiffrement aléatoire soient plutôt faibles.

Partie III
Secret pratique
21. La caractéristique du travail

Une fois le point d'unicité franchi à partir du matériau intercepté, il existe généralement une solution unique au cryptogramme. Le problème de l'isolement de cette solution unique de forte probabilité est celui de la cryptanalyse. Dans la région précédant le point d'unicité, on peut dire que le problème de la cryptanalyse consiste à isoler toutes les solutions possibles de forte probabilité (par rapport au reste) et à déterminer leurs différentes probabilités.

Bien qu'il soit toujours possible en principe de déterminer ces solutions (par exemple, en testant chaque clé possible), la quantité de travail requise varie considérablement selon les systèmes de chiffrement. La quantité moyenne de travail nécessaire pour déterminer la clé d'un cryptogramme de N lettres, $W(N)$, mesurée par exemple en heures-homme, peut être appelée la caractéristique de travail du système. Cette moyenne est calculée sur tous les messages et toutes les clés avec leurs probabilités appropriées. La fonction $W(N)$ est une mesure du degré de "secret pratique" offert par le système.

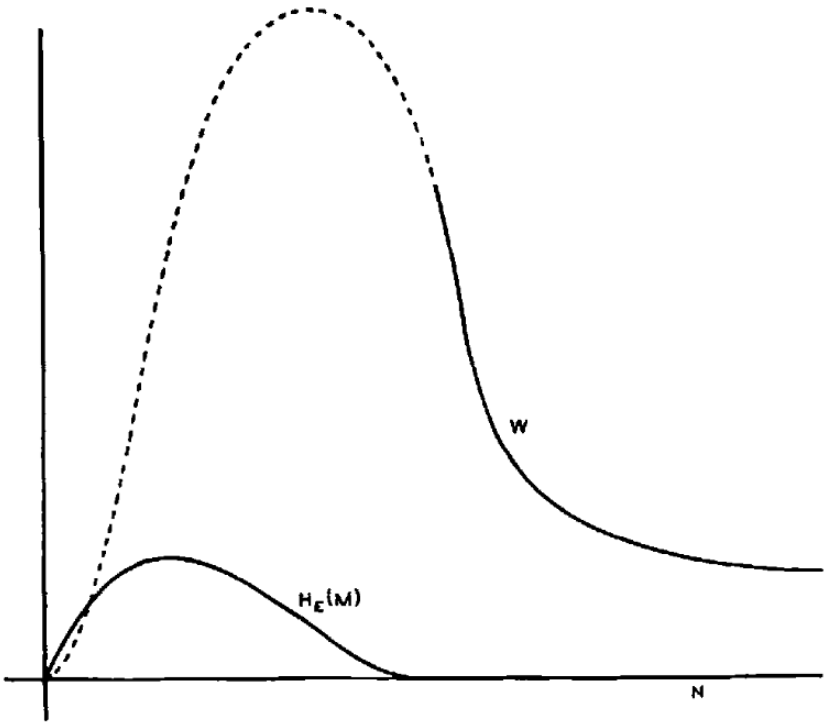

FIG. 12. Travail typique et caractéristiques de l'équivoque

Pour une substitution simple en anglais, les caractéristiques de travail et d'équivoque seraient similaires à celles illustrées sur la figure 12. La partie en pointillés de la figure 12 se situe dans la plage où il existe de nombreuses solutions possibles et celles-ci doivent toutes être déterminées. Dans la partie continue après le point d'unicité, une seule solution existe en général, mais si l'on ne dispose que du minimum de données nécessaires, un travail important doit être effectué pour l'isoler. À mesure que davantage de données sont disponibles, le travail diminue rapidement vers une valeur asymptotique, où les données supplémentaires ne réduisent plus ce travail.

On peut s'attendre au comportement illustré sur la figure 12 avec tout type de système de chiffrement où l'équivoque est proche de zéro. Cependant, le nombre d'heures de travail nécessaires varie considérablement selon les types de chiffrements, même lorsque les courbes $H_E(M)$ sont à peu près identiques. Un chiffrement de type Vigenère ou un chiffrement de type Vigenère composé, par exemple, avec la même taille de clé, auraient une caractéristique de travail bien meilleure (c'est-à-dire bien plus élevée). Un bon système de chiffrement pratique est celui dans lequel la courbe $W(N)$ reste suffisamment élevée, jusqu'au nombre de lettres que l'on s'attend à transmettre avec la clé, pour empêcher l'ennemi de réellement mettre en œuvre la solution, ou pour la retarder à un point tel que l'information est alors obsolète.

Nous examinerons dans les sections suivantes des moyens de conserver une fonction $W(N)$ de valeur élevée, même si $H_E(K)$ peut être pratiquement nul. Il s'agit essentiellement d'un problème de type "max min", comme c'est toujours le cas lorsque nous avons une bataille d'esprit [13]. En concevant un bon chiffrement, nous devons maximiser la quantité minimale de travail que l'ennemi doit faire

[13]Voir von Neumann et Morgenstern, loc. cit. La situation entre le concepteur du chiffrement et le cryptanalyste peut être considérée comme un "jeu" d'une structure très simple ; un jeu à somme nulle à deux personnes avec

pour le casser. Il ne suffit pas d'être sûr qu'aucune des méthodes standard de cryptanalyse ne fonctionne ; il faut aussi être certain qu'aucune méthode, quelle qu'elle soit, ne pourra facilement décrypter le système. C'est d'ailleurs la faiblesse de nombreux systèmes ; conçus pour résister à toutes les méthodes de résolution connues, ils ont ensuite donné naissance à de nouvelles techniques de cryptanalyse, les rendant vulnérables à l'analyse.

Le problème d'une bonne conception de chiffrement consiste essentiellement à identifier des problèmes complexes, sous réserve de certaines conditions. Il s'agit d'une situation plutôt inhabituelle, car on recherche généralement des problèmes simples et facilement solubles dans un domaine donné.

Comment être sûr qu'un système qui n'est pas idéal et qui possède donc une solution unique pour N suffisamment grand nécessitera un travail important pour être déchiffré par toutes les méthodes d'analyse ? Il existe deux approches pour résoudre ce problème : (1) étudier les méthodes de résolution possibles du cryptanalyste et tenter de les décrire en termes suffisamment généraux pour couvrir toutes les méthodes qu'il pourrait utiliser. Nous construisons ensuite notre système de manière à résister à cette méthode de résolution "générale". (2) construire notre chiffrement de telle sorte que son déchiffrement soit équivalent à (ou nécessite à un moment donné) la résolution d'un problème connu pour être laborieux. Ainsi, si nous pouvions montrer que la résolution d'un certain système nécessite au moins autant de travail que la résolution d'un système d'équations simultanées à grand nombre d'inconnues, de type complexe, alors nous aurions une sorte de borne inférieure pour la caractéristique de travail.

Les trois sections suivantes traitent de ces problèmes généraux. Il est difficile de définir les idées pertinentes impliquées avec suffisamment de précision pour obtenir des résultats sous forme de théorèmes mathématiques, mais on pense que les conclusions, sous forme de principes généraux, sont correctes.

22. Généralités sur la solution des cryptogrammes

Une fois la distance d'unicité dépassée dans le matériau intercepté, tout système peut en principe être résolu en essayant simplement chaque clé possible jusqu'à obtenir la solution unique, c'est-à-dire un message déchiffré "qui a du sens" dans la langue d'origine. Un calcul simple montre que cette méthode de résolution (que l'on pourrait qualifier d'essai-erreur complet) est totalement impraticable, sauf lorsque la clé est extrêmement petite.

Supposons, par exemple, que nous disposions d'une clé de 26 ! possibilités, soit environ $26,3$ chiffres décimaux, soit la même taille que pour une substitution simple en anglais. Il s'agit, à tous égards, d'une petite clé. Elle peut être écrite sur un petit bout de papier ou mémorisée en quelques minutes. Elle pourrait être enregistrée sur 27 commutateurs, chacun à dix positions, ou sur des commutateurs à deux positions.

---

des informations complètes et seulement deux "coups". Le concepteur du chiffrement choisit un système pour son "coup". Ensuite, le cryptanalyste est informé de ce choix et choisit une méthode d'analyse. La "valeur" du jeu est le travail moyen requis pour casser un cryptogramme dans le système par la méthode choisie.

Supposons également, pour donner au cryptanalyste tous les avantages possibles, qu'il construise un dispositif électronique permettant de tester des clés à raison d'une par microseconde (éventuellement en sélectionnant automatiquement les résultats par un test χ^2 de signification statistique). Il peut espérer atteindre la bonne clé à peu près à mi-chemin, et après un temps écoulé d'environ $2 \times 10^{26}/2 \times 60^2 \times 24 \times 365 \times 10^6$ ou 3×10^{12} ans.

En d'autres termes, même avec une petite clé, la méthode des essais et erreurs ne sera jamais utilisée pour résoudre les cryptogrammes, sauf dans le cas trivial où la clé est extrêmement petite, par exemple le César avec seulement 26 possibilités, soit $1,4$ chiffres. La méthode des essais et erreurs, si couramment utilisée en cryptographie, est d'un autre type, ou est complétée par d'autres moyens. Un système de confidentialité exigeant des essais et erreurs complets serait extrêmement sûr. Un tel système serait obtenu, semble-t-il, si les messages originaux significatifs, disons tous de 1000 lettres, étaient une sélection aléatoire parmi toutes les séquences possibles de 1000 lettres. Si l'un des chiffrements simples était appliqué à ce type de langage, il semble que peu d'améliorations par rapport à la méthode des essais et erreurs complets ne seraient possibles.

Les méthodes de cryptanalyse réellement utilisées impliquent souvent beaucoup d'essais et d'erreurs, mais d'une manière différente. Premièrement, les pistes progressent des hypothèses les plus probables vers les moins probables, et deuxièmement, chaque essai dispose d'un grand groupe de clés, et non d'une seule. Ainsi, l'espace des clés peut être divisé en, disons, 10 sous-ensembles, chacun contenant à peu près le même nombre de clés. En 10 essais au maximum, on détermine quel sous-ensemble est le bon. Ce sous-ensemble est ensuite divisé en plusieurs sous-ensembles secondaires et le processus est répété. Avec la même taille de clé ($26\,! \doteq 2 \times 10^{26}$), on s'attendrait à environ 26×5 ou 130 essais par rapport à une méthode d'essais et d'erreurs complète. La possibilité de choisir d'abord le sous-ensemble le plus probable pour le test améliorerait encore ce résultat. Si la division se faisait en deux sous-ensembles (le meilleur moyen de minimiser le nombre d'essais), seuls 88 essais seraient nécessaires. Alors que la méthode d'essais-erreurs complète nécessite des essais de l'ordre du nombre de clés, cette méthode d'essais-erreurs subdivisée ne nécessite que des essais de l'ordre de la taille de la clé en bits.

Cela reste vrai même lorsque les différentes clés ont des probabilités différentes. La procédure appropriée pour minimiser le nombre d'essais attendu consiste donc à diviser l'espace des clés en sous-ensembles d'équiprobabilité. Une fois le sous-ensemble approprié déterminé, celui-ci est à nouveau subdivisé en sous-ensembles d'équiprobabilité. Si ce processus peut être poursuivi, le nombre d'essais attendu pour chaque division en deux sous-ensembles sera de

$$h = \frac{H(K)}{\log\ 2}$$

Si chaque test S a des résultats possibles et que chacun d'eux correspond à la clé appartenant à l'un des S sous-ensembles d'équiprobabilité, alors on s'attend à

$$h = \frac{H(K)}{\log\ S}\ \text{ essais.}$$

La signification intuitive de ces résultats est à noter. Dans le test à deux sous-ensembles avec équiprobabilité, chaque test fournit une information sur la clé. Si les sous-ensembles ont des probabilités très différentes, comme lors du test d'une clé unique par essais-erreurs complets, seule

une petite quantité d'informations est obtenue. Ainsi, avec 26 ! clés équiprobables, un test de 1 ne produit que

$$-\left[\frac{26\,!-1}{26\,!}\log\frac{26\,!-1}{26\,!}+\frac{1}{26\,!}\log\frac{1}{26\,!}\right],$$

soit environ 10^{-26} bits d'information. La division en S sous-ensembles d'apparitions équiprobables borne supérieurement l'information obtenue à chaque essai par $\log S$, et le nombre espéré d'essais correspond à l'information totale à obtenir, soit $H(K)$ divisé par cette quantité.

La question ici est similaire à divers problèmes de pesée de pièces qui ont circulé récemment. Un exemple typique est le suivant : on sait qu'une pièce sur 27 est contrefaite et légèrement plus légère que les autres. Une balance de chimiste est disponible et la pièce contrefaite doit être isolée par une série de pesées. Quel est le nombre minimum de pesées nécessaires pour ce faire ? La bonne réponse est 3, obtenue en divisant d'abord les pièces en trois groupes de 9 chacun. Deux d'entre eux sont comparés sur la balance. Les trois résultats possibles déterminent l'ensemble de 9 contenant la contrefaçon. Cet ensemble est ensuite divisé en 3 sous-ensembles de 3 chacun et le processus continue. L'ensemble de pièces correspond au jeu de clés, la pièce contrefaite à la bonne clé et la procédure de découverte par pesée à un seul essai ou un test. L'incertitude initiale est de $\log_2 27$ bits, et chaque essai donne $\log_2 3$ bits d'information ; ainsi, lorsqu'il n'y a pas de "problème diophantien", $\log_2 27/\log_2 3$ ou 3 essais sont suffisants.

Cette méthode de résolution n'est réalisable que si l'espace des clés peut être divisé en un petit nombre de sous-ensembles, avec une méthode simple pour déterminer le sous-ensemble auquel appartient la clé correcte. Il n'est pas nécessaire de supposer une clé complète pour appliquer un test de cohérence et déterminer si l'hypothèse est justifiée ; une hypothèse sur une partie de la clé (ou sur sa présence dans une grande partie de l'espace des clés) peut être testée. Autrement dit, il est possible de résoudre la clé bit par bit.

La possibilité de cette méthode d'analyse constitue la faiblesse cruciale de la plupart des systèmes de chiffrement. Par exemple, en substitution simple, une hypothèse sur une seule lettre peut être vérifiée par rapport à sa fréquence, à la variété de ses contacts, à ses doubles ou inversions, etc. En déterminant une seule lettre, l'espace clé est réduit de 1,4 décimale par rapport aux 26 initiales. Le même effet est observé dans tous les types élémentaires de chiffrement. Dans le Vigenère, l'hypothèse de deux ou trois lettres de la clé est facilement vérifiée en déchiffrant d'autres parties du message avec ce fragment et en observant si la clarté émerge. Le Vigenère composé est bien meilleur de ce point de vue, si l'on suppose un nombre assez important de périodes composantes, produisant un taux de répétition supérieur à celui qui sera intercepté. Dans ce cas, on utilise autant de lettres clés pour chiffrer chaque lettre qu'il y a de périodes. Bien qu'il ne s'agisse que d'une fraction de la clé entière, un nombre important de lettres doit être utilisé avant qu'une vérification de cohérence puisse être appliquée.

Notre première conclusion, concernant la conception pratique d'un chiffrement à petite clé, est qu'une quantité considérable de clés doit être utilisée pour chiffrer chaque petit élément du message.

23. Méthodes statistiques

Il est possible de résoudre de nombreux types de chiffrements par analyse statistique. Prenons à nouveau l'exemple de la substitution simple. La première chose qu'un cryptanalyste fait avec un cryptogramme intercepté est de réaliser un comptage de fréquences. Si le cryptogramme contient, disons, 200 lettres, on peut supposer sans risque que peu, voire aucune, de ces lettres ne se trouvent hors de leurs groupes de fréquences, ce qui correspond à une division en quatre ensembles de limites de fréquences bien définies. Le logarithme du nombre de clés dans cette limite peut être calculé comme suit :

$$\log 2!9!9!6! = 14.28$$

Le simple comptage de fréquences réduit ainsi l'incertitude de la clé de 12 décimales, ce qui représente un gain considérable.

En général, une attaque statistique se déroule comme suit : une certaine statistique est mesurée sur le cryptogramme intercepté E. Cette statistique est telle que, pour tous les messages raisonnables M, elle prend approximativement la même valeur, S_K, cette valeur ne dépendant que de la clé particulière K utilisée. La valeur ainsi obtenue permet de limiter les clés possibles à celles qui donneraient des valeurs de S proches de celle observée. Une statistique indépendante de K ou variant autant avec M qu'avec K n'est pas utile pour limiter K. Ainsi, dans les chiffrements par transposition, le comptage de fréquence des lettres ne fournit aucune information sur K ; chaque K laisse cette statistique inchangée. Par conséquent, un comptage de fréquence ne peut être utilisé pour déchiffrer les chiffrements par transposition.

Plus précisément, on peut attribuer un "pouvoir de résolution" à une statistique S donnée. Pour chaque valeur de S, il existe une équivoque conditionnelle de la clé $H_S(K)$, équivoque lorsque S a sa valeur particulière, et c'est tout ce que l'on sait de la clé. La moyenne pondérée de ces valeurs

$$\sum P(S)H_S(K)$$

donne l'équivoque moyenne de la clé lorsque S est connue, $P(S)$ étant la probabilité a priori de la valeur particulière S. La taille de la clé $H(K)$, moins cette équivoque moyenne, mesure le "pouvoir de résolution" de la statistique S.

Dans un chiffrement fortement idéal, toutes les statistiques du cryptogramme sont indépendantes de la clé particulière utilisée. Il s'agit de la propriété de préservation de la mesure de $T_jT_k^{-1}$ sur l'espace E ou de $T_j^{-1}T_k$ sur l'espace mentionné ci-dessus.

Il existe de bonnes et de mauvaises statistiques, tout comme il existe de bonnes et de mauvaises méthodes d'essais-erreurs. En effet, le test d'une hypothèse par essais-erreurs est un type de statistique, et ce qui a été dit plus haut concernant les meilleurs types d'essais est généralement valable. Une bonne statistique pour résoudre un système doit avoir les propriétés suivantes :

1. Elle doit être simple à mesurer.
2. Elle doit dépendre davantage de la clé que du message si elle est censée résoudre la clé. La variation avec M ne doit pas masquer sa variation avec K.

3. Les valeurs de la statistique qui peuvent être "résolues" malgré le "flou" produit par la variation de M doivent diviser l'espace des clés en plusieurs sous-ensembles de probabilité comparable, la statistique spécifiant celui dans lequel se trouve la clé correcte. La statistique doit nous fournir des informations importantes sur la clé, et non une infime fraction de bit.

4. L'information fournie doit être simple et exploitable. Ainsi, les sous-ensembles dans lesquels la statistique localise la clé doivent être simples dans l'espace des clés.

Le comptage de fréquences pour une substitution simple est un exemple de très bonne statistique.

Deux méthodes (autres que le recours aux systèmes idéaux) semblent pouvoir contrecarrer une analyse statistique. On peut les appeler méthodes de diffusion et de confusion. Dans la méthode de diffusion, la structure statistique de M, qui conduit à sa redondance, est "dissipée" en statistiques à longue portée, c'est-à-dire en une structure statistique impliquant de longues combinaisons de lettres dans le cryptogramme. L'ennemi doit alors intercepter une quantité considérable de données pour cerner cette structure, car celle-ci n'est visible que dans des blocs de très faible probabilité individuelle. De plus, même avec suffisamment de données, le travail d'analyse requis est bien plus important, car la redondance a été diffusée sur un grand nombre de statistiques individuelles. Un exemple de diffusion de statistiques consiste à opérer sur un message $M = m_1, m_2, m_3, \ldots$ avec une opération de "moyenne", par exemple :

$$y_n = \sum_{i=1}^{s} m_{n+i} \pmod{26},$$

en ajoutant s lettres successives du message pour obtenir une lettre y_n. On peut montrer que la redondance de la séquence est la même que celle de la séquence m, mais la structure a été dissipée. Ainsi, les fréquences des lettres dans y seront plus proches de l'égalité que dans m, les fréquences des digrammes également, etc. En effet, toute opération réversible produisant une lettre en sortie pour chaque lettre en entrée et ne disposant pas d'une "mémoire" infinie a une sortie avec la même redondance que l'entrée. Les statistiques ne peuvent jamais être éliminées sans compression, mais elles peuvent être étalées.

La méthode de confusion consiste à rendre la relation entre les statistiques simples de E et la description simple de K très complexe et complexe. Dans le cas d'une substitution simple, il est facile de décrire la limitation de K imposée par les fréquences des lettres de E. Si la connexion est très complexe et confuse, l'adversaire peut toujours évaluer une statistique S_1, par exemple, qui limite la clé à une région de l'espace des clés. Cette limitation, cependant, concerne une région complexe R de l'espace, qui peut-être "plié à l'infini" de nombreuses fois, et il a du mal à l'utiliser. Une seconde statistique S_2 limite encore K à R_2, la situant ainsi dans la région d'intersection ; mais cela n'est pas très utile, car il est très difficile de déterminer précisément quelle est l'intersection.

Pour être plus précis, supposons que l'espace des clés possède certaines "coordonnées naturelles" $k_1, k_2, \ldots, k_p$ que le cryptanalyste adverse souhaite déterminer. Il mesure, disons, un ensemble de statistiques $s_1, s_2, \ldots, s_n$ et celles-ci sont suffisantes pour déterminer les k_i. Cependant, dans la méthode de confusion, les équations reliant ces ensembles de variables sont intriquées et complexes.

On a, disons,

$$\begin{aligned} f_1(k_1, k_2, \ldots, k_p) &= s_1 \\ f_2(k_1, k_2, \ldots, k_p) &= s_2 \\ &\vdots \\ f_n(k_1, k_2, \ldots, k_p) &= s_n \end{aligned}$$

et tous les k_i interviennent dans tous les f_i. Le cryptanalyste doit résoudre ce système simultanément, ce qui est une tâche difficile. Dans les cas simples (non confus), les fonctions n'impliquent qu'un petit nombre des k_i, ou du moins certains d'entre eux. On résout d'abord les équations les plus simples, en évaluant certaines d'entre elles et en les remplaçant dans les équations plus complexes.

La confusion ici vient du fait que, pour un bon système de chiffrement, des mesures doivent être prises pour diffuser ou brouiller la redondance (ou les deux).

24. La méthode du mot probable

L'un des outils les plus puissants pour déchiffrer les chiffrements est l'utilisation de mots probables. Ces mots peuvent être des mots ou des expressions attendus dans un message donné en raison de sa source, ou simplement des mots ou des syllabes courants présents dans n'importe quel texte de la langue, tels que “the”, “and”, “tion”, “that”, etc. en anglais.

En général, la méthode des mots probables est utilisée comme suit : en supposant qu'un mot probable soit en clair à un moment donné, la clé ou une partie de celle-ci est déterminée. Ceci permet de déchiffrer d'autres parties du cryptogramme et de réaliser un test de cohérence. Si les autres parties sont en clair, l'hypothèse est justifiée.

Parmi les chiffrements classiques, rares sont ceux qui utilisent une petite clé et résistent longtemps à une analyse de mots probables. L'étude de cette méthode permet de concevoir un test de chiffrement que l'on pourrait appeler le test décisif. Cela ne s'applique qu'aux chiffrements à clé courte (moins de 50 chiffres décimaux, par exemple), appliqués aux langues naturelles, et n'utilisant pas la méthode idéale pour obtenir le secret. Le test décisif est le suivant : est-il difficile de déterminer la clé ou une partie de celle-ci en connaissant un petit échantillon de message et le cryptogramme correspondant ? Tout système dans lequel cela est facile ne peut pas être très résistant, car le cryptanalyste peut toujours utiliser des mots probables, combinés à des essais et des erreurs, jusqu'à obtenir une solution cohérente.

Les conditions sur la taille de la clé réduisent le nombre d'essais et d'erreurs, et la condition relative aux systèmes idéaux est nécessaire, car ceux-ci fournissent automatiquement des vérifications de cohérence. L'existence de mots et d'expressions probables est implicite dans l'hypothèse des langues naturelles.

Notez que l'exigence d'une solution difficile dans ces conditions n'est pas, en soi, contradictoire avec l'exigence que le chiffrement et le déchiffrement soient des processus simples. En utilisant la notation fonctionnelle, nous avons pour le chiffrement

$$E = f(K, M)$$

et pour le déchiffrement

$$M = g(K, E).$$

Ces deux opérations peuvent être simples sur leurs arguments sans que la troisième équation

$$K = h(M, E)$$

soit simple.

Nous pouvons également souligner que, lors de l'étude d'un nouveau type de système de chiffrement, l'une des meilleures méthodes d'attaque consiste à déterminer la clé si une quantité suffisante de M et E était donnée.

Le principe de confusion peut (et doit) être utilisé pour compliquer la tâche du cryptanalyste utilisant les techniques de mots probables. Étant donné (ou en supposant) $M = m_1, m_2, \ldots, m_s$ et $E = e_1, e_2, \ldots, e_s$, le cryptanalyste peut établir des équations pour les différents éléments clés $k_1, k_2, \ldots, k_r$ (à savoir les équations de chiffrement).

$$\begin{aligned} e_1 &= f_1(m_1, m_2, \ldots, m_s \;\; k_1, \ldots, k_r) \\ e_2 &= f_2(m_1, m_2, \ldots, m_s \;\; k_1, \ldots, k_r) \\ &\vdots \\ e_s &= f_s(m_1, m_2, \ldots, m_s \;\; k_1, \ldots, k_r) \end{aligned}$$

On suppose que tout est connu, exceptés les k_i. Chacune de ces équations devrait donc être complexe en fonction des k_i et en comporter plusieurs. Sinon, l'ennemi peut résoudre les équations simples, puis les plus complexes par substitution.

Pour éviter toute confusion, il est souhaitable que plusieurs m_i interviennent dans les différents f_i, surtout si ces m_i ne sont pas adjacents et de ce fait sont moins corrélés. Cela introduit cependant un risque indésirable de propagation d'erreur : chaque e_i affectera alors généralement plusieurs m_i lors du déchiffrement, et l'erreur se propagera à tous ces éléments.

Nous concluons qu'une grande partie de la clé doit être utilisée de manière complexe pour obtenir une lettre du message à partir du cryptogramme afin de maintenir une caractéristique de travail nécessaire au décryptage élevée. De plus, une dépendance à plusieurs m_i non corrélés est souhaitable, si une certaine propagation d'erreur peut être tolérée. Les trois arguments de ces sections nous incitent à envisager des "transformations mixtes".

25. Transformations de mélange

Une notion qui s'est avérée précieuse dans certaines branches de la théorie des probabilités est le concept de transformation de mélange. Supposons que nous ayons un espace de probabilités ou de mesures Ω et une transformation préservant la mesure F de l'espace vers lui-même, c'est-à-dire une transformation telle que la mesure d'une région transformée FR soit égale à la mesure de la région initiale R. La transformation est appelée mélange si pour toute fonction définie sur l'espace

et toute région R l'intégrale de la fonction sur la région $F^n R$ tend, lorsque $n \to \infty$, vers l'intégrale de la fonction sur l'espace entier Ω multipliée par le volume de R. Cela signifie que toute région initiale R est mélangée avec une densité uniforme dans tout l'espace si F est appliqué un grand nombre de fois. En général, $F^n R$ devient une région constituée d'un grand nombre de filaments minces répartis dans Ω. À mesure que la densité augmente, les filaments deviennent plus fins et leur densité plus constante.

Une transformation de mélange, dans ce sens précis, ne peut se produire que dans un espace comportant un nombre infini de points, car dans un espace fini de points, la transformation doit être périodique. Cependant, en termes généraux, on peut considérer une transformation de mélange comme une transformation qui répartit toute région raisonnablement cohérente de l'espace de manière assez uniforme sur tout l'espace. Si la première région pouvait être décrite en termes simples, la seconde nécessiterait des termes très complexes.

En cryptographie, nous pouvons considérer tous les messages possibles de longueur N comme l'espace Ω et les messages à forte probabilité comme la région R. Ce dernier groupe possède une structure statistique assez simple. Si une transformation de mélange était appliquée, les messages à forte probabilité seraient répartis uniformément dans l'espace.

Les bonnes transformations de mélange sont souvent formées par la répétition de deux opérations simples non commutatives. Hopf [14] a montré, par exemple, que la pâte à pâtisserie peut être mélangée par une telle séquence d'opérations. La pâte est d'abord étalée en une fine plaque, puis repliée, puis roulée, puis repliée à nouveau, etc.

Dans une bonne transformation de mélange d'un espace de coordonnées naturelles $X_1, X_2, \ldots, X_S$, le point X_i est porté par la transformation en un point X_i' avec,

$$X_i' = f_1(X_1, X_2, \ldots, X_S) \qquad\qquad i = 1, 2, \ldots, S$$

et les fonctions f_i sont complexes, impliquant toutes les variables de manière "sensible". Une petite variation de l'une d'elles, X_3 par exemple, modifie considérablement tous les X_i'. Si X_3 traverse son domaine de variation possible, le point X_i' trace un long chemin sinueux autour de l'espace.

Différentes méthodes de mélange applicables aux suites statistiques du type de celles que l'on trouve dans les langues naturelles peuvent être imaginées. Une méthode assez efficace consiste à faire suivre une transposition préliminaire par une séquence de substitutions alternées et d'opérations linéaires simples, en ajoutant des lettres adjacentes mod 26 par exemple. On pourrait ainsi prendre

$$F = LSLSLT$$

où T est une transposition, L une opération linéaire et S une substitution.

[14] E. Hopf, "*On Causality, Statistics and Probability*", Journal of Math. and Physics, v. 13, p. 51-102, 1934.

26. Chiffres du type T_kFS_j

Supposons que F soit une bonne transformation de mélange applicable aux séquences de lettres, et que T_k et S_j soient deux familles simples de transformations, c'est-à-dire deux chiffrements simples, potentiellement identiques. Par souci de concision, nous pouvons les considérer comme de simples substitutions.

Il apparaît que le chiffrement TFS constitue un très bon système de chiffrement du point de vue de sa caractéristique de travail. En premier lieu, il apparaît clairement, en examinant nos arguments sur les méthodes statistiques, qu'aucune statistique simple ne fournira d'informations sur la clé — toute statistique significative dérivée de E doit être d'un type très complexe et très sensible — la redondance ayant été à la fois diffusée et brouillée par la transformation de mélange F. De plus, les mots probables conduisent à un système complexe d'équations impliquant toutes les parties de la clé (lorsque le mélange est bon), qui doit être résolu (i.e. toutes les équations du système doivent être résolues simultanément).

Il est intéressant de noter que si le chiffrement T est omis, le système restant est similaire à S et n'est donc pas plus fort. L'ennemi "démixe" simplement le cryptogramme par application de F^{-1}, puis le résout. Si S est omis, le système restant est beaucoup plus fort que T seul lorsque le mélange est bon, mais n'est toujours pas comparable à TFS.

Le principe de base des chiffrements simples séparés par une transformation de mélange peut bien sûr être étendu. Par exemple, on pourrait utiliser

$$T_kF_1S_jF_2R_i$$

avec deux combinaisons et trois chiffrements simples. On peut également simplifier en utilisant les mêmes chiffrements, et même les mêmes clés, ainsi que les mêmes transformations de combinaison. Cela pourrait simplifier la mécanisation de tels systèmes.

La transformation de combinaison qui sépare les deux (ou plusieurs) apparitions de la clé agit comme une sorte de barrière pour l'ennemi : il est facile de faire passer un élément connu par-dessus cette barrière, mais une inconnue (la clé) le fait plus difficilement.

En fournissant deux ensembles d'inconnues, la clé de S et celle de T, et en les séparant par la transformation de combinaison F, nous avons "intriqué" les inconnues ensemble d'une manière qui rend la résolution très difficile.

Bien que les systèmes construits selon ce principe soient extrêmement sûrs, ils présentent un grave inconvénient. Si la combinaison est bonne, la propagation des erreurs est mauvaise. Une erreur de transmission d'une lettre affectera plusieurs lettres lors du déchiffrement.

27. Incompatibilité du critère pour les bons systèmes

Les cinq critères pour de bons systèmes de chiffrement, présentés dans la section 5, semblent présenter une certaine incompatibilité lorsqu'ils sont appliqués à un langage naturel à la structure

statistique complexe. Avec des langages artificiels à structure statistique simple, il est possible de satisfaire simultanément toutes les exigences grâce aux chiffrements de type idéal. Dans les langages naturels, un compromis doit être trouvé et les évaluations doivent être équilibrées en fonction de l'application spécifique.

Si l'un des cinq critères est abandonné, les quatre autres peuvent être satisfaits relativement bien, comme le montrent les exemples suivants :

1. Si l'on omet la première exigence (niveau de secret), tout chiffrement simple, comme une simple substitution, fera l'affaire. Dans le cas extrême où cette condition est complètement omise, aucun chiffrement n'est requis et l'envoi est clair !

2. Si la taille de la clé n'est pas limitée, le système Vernam peut être utilisé.

3. Si la complexité de l'opération n'est pas limitée, divers types de processus de chiffrement extrêmement complexes peuvent être utilisés.

4. Si l'on omet la propagation de la condition d'erreur, des systèmes de type TFS seraient très efficaces, bien qu'un peu compliqués.

5. Si l'on autorise une large expansion du message, divers systèmes sont facilement conçus où le message "correct" est mélangé à de nombreux messages "incorrects" (désinformation). La clé détermine lequel de ces messages est correct.

Un argument très approximatif pour l'incompatibilité des cinq conditions peut être avancé comme suit : à partir de la condition 5, il faut utiliser un système de chiffrement essentiellement tel qu'étudié dans cet article, c'est-à-dire, un système faisant peu d'utilisation de valeurs nulles, etc. Les systèmes parfaits et idéaux sont exclus par la condition 2 et par les conditions 3 et 4, respectivement. Le secret élevé requis par la condition 1 doit alors provenir d'une caractéristique de travail de déchiffrement élevée, et non d'une caractéristique d'équivoque élevée. Si la clé est petite, le système simple et les erreurs non propagées, les méthodes des mots probables résoudront généralement le système assez facilement, car nous disposons alors d'un système d'équations relativement simple pour la clé.

Ce raisonnement est trop vague pour être concluant, mais l'idée générale semble tout à fait raisonnable. Si l'on pouvait attribuer une signification quantitative aux différents critères, on pourrait peut-être trouver une sorte d'équation d'échange les impliquant et donnant les ensembles de valeurs physiquement compatibles les plus performants. Les deux caractéristiques les plus difficiles à mesurer numériquement sont la complexité des opérations et la complexité de la structure statistique du langage.

Appendice

Preuve du théorème 3

Sélectionnons un message M_1 et regroupez tous les cryptogrammes pouvant être obtenus à partir de M_1 par une opération de chiffrement T_i. Soit cette classe de cryptogrammes C'_1. Regroupons avec

M_1 tous les messages pouvant être obtenus à partir de M_1 par $T_i^{-1}T_jM_1$, et appelons cette classe C_1. Le même C_1' serait obtenu si nous commencions par n'importe quel autre M de C_1 puisque

$$T_sT_j^{-1}T_iM_1 = T_lM_1$$

De même, le même résultat serait obtenu.

En choisissant un M absent de C_1 (s'il en existe un), on construit C_2 et C_2' de la même manière. En continuant ainsi, on obtient les classes résiduelles de propriétés (1) et (2). Soient M_1 et M_2 dans C_1 et supposons

$$M_2 = T_1T_2^{-1}M_1.$$

Si E_1 est dans C_1' et peut être obtenu à partir de M_1 par

$$E_1 = T_\alpha M_1 = T_\beta M_1 = \ldots = T_\eta M_1,$$

alors

$$\begin{aligned} E_1 &= T_\alpha T_2^{-1}T_1M_2 = T_\beta T_2^{-1}T_1M_2 = \ldots \\ &= T_\lambda M_2 = T_\mu M_2 \ldots \end{aligned}$$

Ainsi, chaque M_i de C_1 se transforme en E_1 par le même nombre de clés. De même, chaque E_i de C_1' est obtenu à partir de tout M de C_1 par le même nombre de clés. Il s'ensuit que ce nombre de clés est un diviseur du nombre total de clés et les propriétés (3) et (4) sont donc vérifiées.