CRYPTOLOGIE

PAGE Aurel
Chercheur
INRIA
aurel.page@inria.fr

Résumé
Cet atelier au colloque de l'IREM était basé sur un atelier que j'ai proposé de nombreuses fois à la Fête de la Science, dans le cadre du Parcours scientifique bordelais au centre Inria de Bordeaux, à des classes allant de la 3ème à la terminale en variant légèrement les sujets abordés. Ce texte retranscrit le contenu des différentes variantes de cet atelier, avec plus de détails.

I - BASES DE LA CRYPTOLOGIE

« À quoi vous fait penser le mot cryptologie ? » La plupart du temps, cette question suscite l'évocation des codes secrets et des téléphones portables. Lorsqu'on suggère le sigle HTTPS (HyperText Transfer Protocol Secure) et le symbole de cadenas des navigateurs web, tout le monde se souvient des sites internet à connexion sécurisée. Les cartes bleues et paiement sécurisés ou le vote électronique sont mentionnés plus rarement.

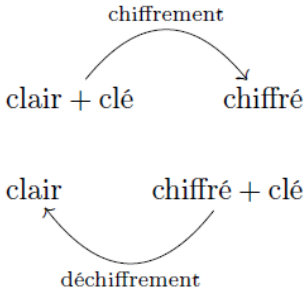
La cryptologie est **la science du secret** : le préfixe crypto- signifie « caché » et le suffixe -logie « science » ou « discours ». Elle regroupe l'ensemble des techniques permettant de transmettre ou conserver une information à destination de certaines personnes, et de la dissimuler aux autres. Elle possède deux composantes :

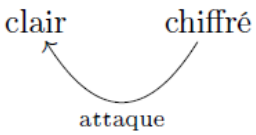
— La *cryptographie* : c'est la partie **défensive**, qui consiste à concevoir les procédés de dissimulation : le suffixe -graphie signifie « écriture » dans ce contexte.

— La *cryptanalyse* : c'est la partie **offensive**, qui consiste à essayer d'attaquer et de casser les procédés cryptographiques : le suffixe -lyse ou -analyse signifie « action de délier » ou « dissolution ».

Les deux composantes de cette discipline scientifique sont complémentaires et indissociables : il est souvent impossible de démontrer la sécurité d'un procédé cryptographique, mais on s'en convainc lorsque le procédé a résisté aux attaques de toute la communauté des experts pendant des années.

Vocabulaire de base de la cryptologie Dans le cadre de l'envoi sécurisé de messages, on peut résumer le procédé général ainsi :





Cela nous permet d'introduire un peu de vocabulaire :

- Le message sous une forme ordinaire, lisible par n'importe qui, est appelé le *clair*.
- Le message sous une forme illisible, indéchiffrable, est appelé le *chiffré*.
- Le procédé transformant le clair en chiffré s'appelle le *chiffrement*.
- Le procédé transformant le chiffré en clair s'appelle le *déchiffrement*.
- La *clé* est une information supplémentaire, secrète, qui est nécessaire pour procéder au chiffrement ou au déchiffrement.
- Lorsqu'on peut retrouver le clair à partir du chiffré sans la clé, on dit qu'on a une *attaque* contre ce procédé de chiffrement.

Remarquons que les terme de « code, codage, codé » sont réservés à un autre usage dans le jargon informatique et ne désignent pas une information cachée, contrairement à leur sens dans le langage courant.

Envoi sécurisé de messages On peut utiliser un tel procédé de chiffrement pour s'envoyer des messages de la manière suivante :

- Assia et Boaz se mettent d'accord sur une clé secrète commune.
- Assia chiffre son message à l'aide de leur clé.
- Assia envoie le chiffré à Boaz via un canal non sécurisé.
- Boaz reçoit le chiffré d'Assia et le déchiffre à l'aide de leur clé.
- Si l'indiscret Cyrille intercepte le chiffré, il ne peut pas retrouver le clair car il ne possède pas la clé.

Remarque : la même clé marche dans les deux sens, d'Assia vers Boaz ou de Boaz vers Assia. On parle de *cryptographie symétrique*, mais nous en reparlerons plus tard.

Les chiffres de substitution L'un des plus anciens procédés cryptographiques, utilisé au moins depuis l'Antiquité, consiste à remplacer chaque lettre par une autre dans l'alphabet : c'est ce qu'on appelle un *chiffre de substitution*.

Une version simple consiste à choisir un décalage dans l'alphabet (chiffre dit *de César*), par exemple :

$$a \to C, \quad b \to D, \quad c \to E, \quad \text{etc.}$$

Ici la clé est le décalage choisi (2 dans notre exemple), et permet de procéder au chiffrement comme suit :

$$\text{cesar (clair)} + \text{décalage } 2 \text{ (clé)} \dashrightarrow \text{EGUCT (chiffré)}$$

Le chiffre de César est-il sûr ? On s'aperçoit rapidement que non, car il y a seulement 26 clés différentes possibles. Si on possède un chiffré sans la clé, on peut essayer toutes les clés possibles, et un seul déchiffrement donnera un clair ayant du sens : on aura retrouvé le clair ainsi que la clé. Cette attaque est toujours considérée en premier dans l'analyse d'un procédé cryptographique, on l'appelle *attaque par énumération exhaustive*.

Un chiffre de substitution arbitraire est-il vulnérable face à une attaque par énumération exhaustive ? Pour répondre à cette question, il nous faut compter le nombre total de clés possibles. Le choix d'une clé consiste à choisir, pour chaque lettre, par quelle lettre de l'alphabet on la remplace. Pour la lettre 'a', on a 26 choix possibles. Pour chacun de ces 26 choix, il reste 25 choix possibles pour la lettre 'b' : on ne veut pas chiffrer 'a' et 'b' par la même lettre, car cela rendrait le déchiffrement impossible, même avec la clé. On arrive donc à 26×25 choix pour 'a' et 'b' combinés. En continuant ainsi pour toutes les lettres de l'alphabet, on compte un total de

$$26 \times 25 \times 24 \times \cdots \times 3 \times 2 \times 1 = 26! \approx 10^{27}$$

possibilités pour la clé (le symbole ! désigne la factorielle). Est-ce suffisant pour être protégé ? Pour cela il faut estimer la puissance de calcul disponible. Un ordinateur effectue de l'ordre de un milliard d'opérations par secondes. Le nombre d'ordinateurs en circulation est proche du nombre d'humains sur Terre (avec une répartition inégale), environ dix milliards. Le nombre de secondes dans une année est $3600 \times 24 \times 366 \approx 10^8$. Un an de calcul par l'humanité toute entière correspond donc environ à

$$10^9 \times 10^{10} \times 10^8 = 10^{9+10+8} = 10^{27} \text{opérations.}$$

On obtient une estimation similaire en remplaçant tous les ordinateurs personnels du monde par une dizaine des plus gros supercalculateurs du monde, d'une puissance d'environ un exaflop/s (10^{18} opérations par seconde). On peut donc estimer qu'il faudrait à l'humanité au moins 1 an de calcul pour tester toutes les clés possibles, ce qui paraît déjà être une bonne sécurité. Les crypto- systèmes réellement utilisés sont considérés comme sûrs si les meilleurs attaques connues nécessitent 10^{40} voire 10^{80} opérations suivant le niveau de sécurité visé.

II - ATTAQUE PAR ANALYSE DE FRÉQUENCES

Il faut attendre le IXe siècle pour qu'Al-Kindi, un savant arabe, propose une véritable technique de cryptanalyse contre les chiffres de substitution. Son attaque repose sur la remarque fondamentale suivante : dans une langue donnée, toutes les lettres n'apparaissent pas avec la même fréquence. Par exemple, en français, les fréquences approximatives sont les suivantes (en %) :

e	s	a	n	t	i	r	u	l	o
17.8	8.2	7.7	7.6	7.3	7.2	6.8	6.1	5.9	5.3

Si le chiffré est suffisamment long, on peut mesurer la fréquence d'apparition de chaque symbole, et s'en servir pour réduire le nombre de possibilités à un petit nombre. Les participants à l'activité ont utilisé cette technique pour retrouver le clair (et la clé) correspondant au chiffré suivant :

« U'PISYTFVQ ZY UD TREMFPHRDMKVY YCF ZY XDCAJYR JW XYCCDHY D U'DVZY Z'JWY TUY. PW TPWWDVF JWY XYFKPZY ZY TKVQQRYXYWF VWTDCCDIUY XDVC MDC FRYC MRDFVAJY. TYFFY FYTKWVAJY C'DMMYUUY UY XDCAJY SYFDIUY YF D YFY VWOYW- FYY MDR OYRWDX. YUUY TPWCVCFY D JFVUVCYR JWY TUY ZY UD XYXY UPWHJYJR AJY UY XYCCDHY D YWOPEYR. CV U'DFFDA- JDWF YCCDVY FPJFYC UYC TUYC MPCCVIUYC VU

OPVF FPJC UYC XYCCDHYC MPCCVIUYC. VU W'E D DJTJWY DFFDAJY MPJR RYFRP- JOYR UY XYCCDHY VWVFVDU XYXY DOYT JW FYXMC ZY TDUTJU VWQVWV. UY MRYXVYR MRPIUYXY ZY TYFFY FYTKWVAJY YCF UD HYWYRDFVPW Z'JWY TUY MDRQDVFYXYWF DUYDFPVRY. UY ZYJNVYXY YCF AJ'VU YCF ZVQQVTVUY Z'YWOPEYR JWY TUY FRYC UPWHJY ZY QDTPW CYTJRVCYY. UY FRPVCVYXY YCF AJ'VU WY QDJF MDC JFVUVCYR UD XYXY TUY MPJR ZYJN XYCCDHYC. »

Le texte chiffré était divisé en 9 phrases, réparties entre des groupes de 2 ou 3 participants. Chaque groupe comptait les occurrences des lettres, et les mesures étaient mises en commun de sorte à obtenir des fréquences globales sur le texte entier. Pour rendre l'activité réalisable en un temps court (20-30 minutes), seules les lettres fréquentes (C, D, F, J, U, V, W, Y) étaient comptées, le texte était préparé pour avoir des fréquences proches de la théorie, et la ponctuation et les espaces étaient conservés. De plus, après un temps de recherche indépendante, les groupes mettaient en commun les lettres qu'ils avaient trouvées, ce qui permettait à tout le monde de bénéficier des mots faciles à trouver de certaines phrases et de l'efficacité de certains groupes. La technique peut aussi être utilisée à la main sur des textes courts et sans ces aides, mais nécessite alors généralement quelques heures de tâtonnement. Les participants finissaient par reconstruire le texte clair suivant (ici reproduit sans majuscules ni accent puisqu'elles n'étaient pas prises en compte dans le chiffrement) :

« L'objectif de la cryptographie est de masquer un message à l'aide d'une clé. On connait une méthode de chiffrement incassable mais pas très pratique. Cette technique s'appelle le masque jetable et a été inventée par Vernam. Elle consiste à utiliser une clé de la même longueur que le message à envoyer. Si l'attaquant essaie toutes les clés possibles il voit tous les messages possibles. Il n'y a aucune attaque pour retrouver le message initial même avec un temps de calcul infini. Le premier problème de cette technique est la génération d'une clé parfaitement aléatoire. Le deuxième est qu'il est difficile d'envoyer une clé très longue de façon sécurisée. Le troisième est qu'il ne faut pas utiliser la même clé pour deux messages. »

Le texte évoque une autre technique de chiffrement, le *masque jetable* ou *chiffre de Vernam*. Elle consiste à chiffrer les lettres par un décalage (comme avec un chiffre de César), mais en utilisant un décalage différent pour chaque lettre. Plus précisément, la clé consiste en une suite de décalages, un pour chaque lettre du clair. On peut démontrer que cette technique est incassable : en effet, étant donné un chiffré, tout clair possible provient de ce chiffré via exactement une clé ; le chiffré ne contient donc aucune information sur le clair en l'absence de la clé. Un problème majeur de cette technique est que pour l'utiliser, il faut échanger au préalable une clé de la même longueur que le message à chiffrer. Nous reviendrons sur ce problème de l'échange de clé.

III - INFLUENCE DE LA DEUXIÈME GUERRE MONDIALE

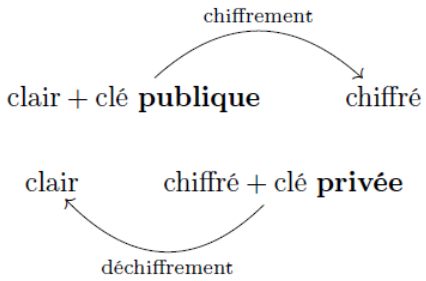
La cryptologie intéresse depuis longtemps les militaires. La deuxième guerre mondiale n'était pas une exception à cet égard : la cryptologie a joué un rôle clé dans son dénouement, et a connu des transformations majeures pendant cette période. Tout d'abord, une **automatisation** massive a été mise en place : du côté du chiffrement avec notamment les célèbres machines Enigma et Lorenz des allemands, mais aussi du côté de la cryptanalyse avec les « bombes » en Pologne puis à Bletchley Park, ancêtres de nos ordinateurs, qui permettaient d'analyser rapidement des milliers de possibilités. Un autre aspect important est celui de la

source de la confiance en la sécurité d'un procédé cryptographique. L'approche traditionnelle, qui semble de bon sens à première vue, a toujours été de **garder secret le procédé cryptographique** utilisé. Par exemple, durant la guerre, le fonctionnement des machines de chiffrement allemandes était gardé secret. Cependant, à plusieurs reprises des machines ou des manuels d'instructions ont été capturés, ce qui a permis aux services secrets de les étudier et de découvrir des faiblesses. A posteriori, il paraît illusoire d'avoir cru que ce type d'événement n'arriverait pas. Au contraire, dans la cryptographie moderne, les **procédés cryptographiques sont publics** et la sécurité doit dépendre seulement de la connaissance de la clé. C'est l'analyse scientifique et ouverte par la communauté des experts qui est la source de la confiance en la sécurité d'un cryptosystème.

Mais si la sécurité ne dépend que de la connaissance de la clé, une question reste en suspens, tout particulièrement à l'ère d'Internet : comment échange-t- on une clé avec une personne à l'autre bout de la planète, qu'on ne peut pas contacter physiquement ?

IV - CRYPTOGRAPHIE ASYMÉTRIQUE ET RSA

Principe La solution au problème du transfert de la clé a été résolue dans les années 70, avec l'invention par Diffie et Hellman de la cryptographie asymétrique. Leur idée est d'avoir deux clés différentes, une pour le chiffrement, et une pour le déchiffrement.



Le fonctionnement est le suivant. Supposons que Assia veut envoyer un message chiffré à Boaz.

- Boaz génère une paire de clés, l'une publique et l'autre privée.
- Boaz envoie sa clé publique à Assia via un canal non sécurisé.
- Assia **chiffre** son message à l'aide de la **clé publique**.
- Assia envoie le chiffré à Boaz via un canal non sécurisé.
- Boaz reçoit le chiffré d'Assia et le **déchiffre** à l'aide de sa **clé privée**.

La clé publique de Boaz peut être interceptée, mais cela n'a pas d'importance car elle ne permet pas de déchiffrer.

Dans leur article, Diffie et Hellman décrivent le principe d'un cryptosystème asymétrique, mais n'en proposent pas un concret. Le premier exemple est dû à Rivest, Shamir et Adleman, qui conçurent un système qui porte leur nom et est fréquemment utilisé : RSA.

Fonctionnement de RSA RSA est décrit comme un système de chiffrement qui **transforme un nombre en un autre nombre**. Il faut imaginer qu'on choisit un grand nombre N, et qu'on convient d'un encodage

public pour passer d'une suite de lettres à un nombre entre 0 et $N-1$ et inversement. Si le message à chiffrer est trop long, on le découpe en blocs qu'on chiffre indépendamment.

Par exemple pour $N = 24^2 = 576$ on peut encoder deux lettres dans chaque nombre entre 0 et $N-1$ par la règle : aa $\to 0$, ab $\to 1$, . . ., az $\to 25$, ba $\to 26$, bb $\to 27$, bc $\to 28$, . . ., zz $\to 575$. Insistons sur le fait que cet encodage est public et ne masque aucune information : il sert uniquement à transformer un problème de chiffrement de suites de lettres en un problème de chiffrement de nombres. On va donc définir un chiffrement sur les nombres entre 0 et $N-1$.

Le cryptosystème RSA est basé sur l'**arithmétique modulaire**. Cela consiste à choisir un entier N et à faire des opérations sur des nombres entiers, en éliminant systématiquement les multiples de N. Lorsqu'on calcule modulo N, on peut ainsi ramener tous les nombres dans l'intervalle $\{0, \ldots, N-1\}$.

Exemple : « arithmétique des horloges » modulo $N = 24$: 5h après 22h, il est 3h. On dit que $22 + 5 \equiv 3$ modulo 24.

Exemple : le calcul modulo $N = 10$ revient à ne garder que le dernier chiffre. $8 + 7 = 15 \equiv 5$ modulo 10.

L'arithmétique modulaire permet aussi de faire des multiplications.

Exemples : $5 \times 6 = 30 \equiv 6 \bmod 24$, $3 \times 8 = 24 \equiv 4 \bmod 10$.

Munis de cet outil, nous pouvons décrire le procédé de chiffrement.

- On choisit deux grands nombres premiers p et q, et on pose $N = pq$.
  Exemple : $p = 2$, $q = 5$ et $N = 10$.
- On choisit un entier e entre 1 et $(p-1)(q-1)$, premier avec $(p-1)(q-1)$.
  Exemple : $e = 3$ est premier avec $(p-1)(q-1) = 1 \times 4 = 4$.
- Avec l'algorithme d'Euclide, on calcule des entiers u, v tels que $eu + (p-1)(q-1)v = 1$.
  Exemple : $u = 3$, $v = -2$ vérifient $eu + (p-1)(q-1)v = 3 \times 3 - 4 \times 2 = 1$.
- La **clé publique** est (N, e) et la **clé privée** est u.
- Soit $m \in \{0, \ldots, N-1\}$ le clair à chiffrer. Le chiffré correspondant est $c \equiv m^e \bmod N$. On peut effectuer l'opération de chiffrement en connaissant uniquement la clé publique.
  Exemple : $m = 3$, d'où $c = m^e = 3^3 = 27 \equiv 7 \bmod 10$.
- Le déchiffrement du chiffré c consiste à calculer $c^u \bmod N$. Une conséquence du petit théorème de Fermat et du théorème des restes chinois est que pour tout message m qui n'est divisible ni par p ni par q, on a $c^u = (m^e)^u = m^{eu} \equiv m \bmod N$.
  Exemple : $c^u = 7^3 = 7 \times 49 \equiv 7 \times 9 = 63 \equiv 3 \bmod N$. On retrouve bien le message $m = 3$.

Pourquoi RSA est-il difficile à attaquer ? Étant donnés e et N, on ne sait trouver la clé privée u que si on sait trouver les facteurs premiers p et q de N.

- Il est **facile** de calculer $N = pq$ à partir de p et q (**multiplication**).
  Exemple : $859 \times 9001 = ?$ Vous pouvez effectuer cette multiplication à la main.
- Il est **difficile** de retrouver p et q à partir de N (**factorisation**).

Exemple : 3252911 = ? × ? Pour factoriser ce nombre naïvement, il faudrait essayer de le diviser par tous les nombres premiers jusqu'à 1800, cela donnerait 278 divisions à effectuer, ce qui serait très long!

Bien entendu, nous connaissons de meilleurs algorithmes que d'essayer toutes les divisions possibles, mais ils restent beaucoup plus lents qu'une multiplication. Actuellement les records de factorisation atteignent des nombres de 250 chiffres, alors qu'on peut sans problème multiplier des entiers de milliards de chiffres.

V- CRYPTOGRAPHIE POST-QUANTIQUE

Une partie de la recherche actuelle en cryptographie est consacrée au post-quantique. Malheureusement, l'ordinateur quantique est souvent une source de grande confusion. Essayons de clarifier les choses.

Qu'est-ce que l'ordinateur quantique ? C'est un modèle de calcul (une abstraction) inspiré des lois de la mécanique quantique, et qu'on espère pouvoir réaliser par un dispositif physique. On peut imaginer un ordinateur quantique comme un ordinateur ordinaire (classique) auquel est branché un appareil qui possède un état interne, inaccessible directement, mais sur lequel on peut faire certaines opérations. Les ordinateurs quantiques déjà construits ne sont que des approximations de ce modèle.

Quel est le lien avec la cryptographie ? L'algorithme quantique de Shor résout rapidement deux problèmes difficiles très utilisés en cryptographie : la factorisation et le problème du logarithme discret. Les ordinateurs quantiques existants sont très loin de pouvoir appliquer cet algorithme à des problèmes de taille cryptographique.

Mythe : un ordinateur quantique peut résoudre des problèmes que les ordinateurs classiques ne pourraient jamais résoudre, même avec un temps de calcul infini. C'est faux. Un ordinateur quantique peut être simulé par un ordinateur classique, mais avec un temps et une mémoire exponentiellement plus grands. Ils résolvent donc les mêmes problèmes, mais pas nécessairement avec la même efficacité.

Mythe : un ordinateur quantique résout des problèmes difficiles en essayant toutes les solutions en parallèle. C'est faux. L'état d'un ordinateur quantique peut encoder toutes les solutions possibles et leur appliquer certaines opérations, mais on ne peut pas accéder à cette information, on peut seulement obtenir une solution aléatoire. Toute la difficulté dans la conception d'algorithmes quantiques consiste à amplifier la probabilité d'obtenir une bonne solution en utilisant les opérations permises par la mécanique quantique.

Qu'est-ce que la cryptographie post-quantique ? Il s'agit de concevoir des procédés cryptographiques utilisables sur un ordinateur classique, mais qu'on pense résistants même à un attaquant disposant d'un ordinateur quantique.

VI - AUTRES APPLICATIONS DE LA CRYPTOGRAPHIE

Pour terminer, rappelons que la cryptographie moderne ne se limite pas au chiffrement des messages. Elle a de nombreuses autres applications, parmi lesquelles :

- le vote électronique ;
- les monnaies décentralisées ;

- le calcul sur des données chiffrées ;
- la signature, qui permet d'assurer l'authenticité (l'auteur est bien celui qui le prétend) et l'intégrité (non-modification) d'une donnée ;
- la mise en gage, dans laquelle une personne garde une information secrète temporairement puis la révèle, sans pouvoir la modifier entre-temps ;
- le partage de secret, dans lequel plusieurs personnes protègent une donnée, mais ne peuvent y accéder qu'ensemble ;
- les preuves zero-knowledge (l'expression française « preuve à divulgation nulle de connaissance » est très lourde), qui consistent à démontrer qu'on détient une information sans la révéler ;
- etc.

La cryptologie, qui est d'une importance cruciale pour notre mode de vie actuel, est une discipline scientifique florissante, qui a aussi le charme de faire émerger de très intéressants problèmes mathématiques, notamment en arithmétique.