Table of Contents

Cours 8 19/03/2020, "Cours à la maison" pour cause de Covid-19

1 VII. Cryptage RSA

Voila le "clou du spectacle" de ce cours, et je suis vraiment désolé de ne pas pouvoir vous le faire en "direct": **les codes secrets !** Ou plutôt, comment l'arithmétique permet de crypter nos transactions bancaires, nos messages intimes, les informations sensibles, etc.

Très intéressante lecture sur les codes secrets:

SIMON SINGH: Histoire des codes secrets , De l'Egypte des Pharaons à l'ordinateur quantique

1.1 vocabulaire

On utilisera la notion de 'texte en clair', qui contient les informations sensibles, donc en rouge, et 'texte chiffré' , auquel tout le monde a accès, donc en vert.

1.1.1 texte en clair (rouge) ===> texte chiffré (vert)

La **cryptographie** est l'ensemble des méthodes pour cacher le sens d'un message.

On a des (presque) synonymes

- coder/codage (système général pour dissimuler le sens d'un message)
- crypter/cryptage (idem)
- chiffrer/chiffrement (en principe réservé à des codes employant des chiffres)

le mot "chiffre" indique aussi le système de chiffrement (cf le "chiffre de César" ci-dessous).

La plupart des algorithmes de chiffrement dépendent d'un paramètre qu'on appelle **clef** (ou clé)

1.1.2 texte chiffré (vert) ===> texte en clair (rouge)

Si on a connaissance de l'algorithme et de la clef, la théorie qui permet de retrouver le texte en clair fait partie de la crytographie.

- déchiffrer/décoder/décrypter

Cryptanalyse: technique permettant de déchiffrer sans connaître à l'avance l'algorithme et/ou la clef. (on dit aussi **briser** (casser) un code).

À chaque fois qu'on invente un code, il est **essentiel** de réfléchir à toutes les attaques possibles, afin de s'en prémunir !

La guerre est incessante entre les nouveaux codes et les nouvelles attaques…

1.1.3 progatonistes

Pour décrire les méthodes employés, au lieu d'utiliser des personnes A et B, on aime leur donner un nom. Dans la littérature anglosaxone c'est traditionnellement:

Alice et Bob

En français on peut préférer Alice et Bruno ? ou Bernard ? À vous de trouver vos prénoms A/B préférés !

1.1.4 convention couleurs

- VERT = message auquel 'tout le monde' a accès
- ROUGE = message privé

1.2 un peu d'histoire

Depuis la nuit des temps on a cherché à transmettre des messages secrets, que ce soit pour des affaires d'états, des contrats commerciaux, …ou des intrigues amoureuses !

Parfois on se contente de **cacher** le message (par exemple, attesté chez les Grecs: écrire sur le crâne et laisser pousser les cheveux !)

Mais très vite on s'est rendu compte qu'il vaut mieux

- ne pas chercher à cacher (attire les soupçons)
- mais plutôt **coder** le message pour le rendre inintelligible aux "ennemis"

1.2.1 César

Un des premiers codes secrets célèbres!

Jules César (empereur romain -100 – -44 av JC) utilisait beaucoup des **chiffres de substitution**: remplacer une lettre par une autre (ou un autre symbole). Par exemple il raconte dans la "Guerre des Gaules" comment il a envoyé un cavalier gaulois avec un message secret ou les lettres latines étaient remplacées par des lettres grecques.

Mais le "chiffre de César" le plus connu consiste à décaler les lettres. Par exemple on décale de 3, ce qui donne (avec notre alphabet actuel):

A–> D

B–> E

C–> F

…

W –> Z

X –> A

Y –> B

Z –> C

Pour varier un peu, on peut changer la **clef** = nombre de décalage.

Mathématiquement: il s'agit d'une addition dans les **congruences modulo 26**: on associe d'abord les lettres aux nombres de 0 à 25, et alors:

- chiffrement = N –> N + 3 modulo 26
- déchiffrement = N –> N - 3 modulo 26 (ou N + 23)

Exercice écrire le code en python.

- Pour convertir une lettre (majuscule) en nombre de 0 à 25: `nb = ord(lettre)-65`
- Pour l'opération inverse `lettre = chr(nb+65)`

On peut bien entendu améliorer cette technique, pour obtenir ce qu'on appelle de façon générale un **alphabet de chiffrement**: au lieu d'un simple décalage, on s'autorise toutes les permutations possibles. (Il y a $26! = 403291461126605635584000000$ permutations possible des 26 lettres de l'alphabet ! Chacune de ces permutations devient la **clef** du code qui consiste à transformer les lettres du message par cette permutation.)

1.2.2 La crytpanalyse

1. Le chiffre de César est facile à "briser". Il suffit d'essayer tous les décalages possibles (ie toutes les clefs de 1 à 25).

2. On peut grandement améliorer le système en utilisant un **alphabet de chiffrement**: au lieu d'un simple décalage, on s'autorise toutes les permutations possibles:

   exemple

   ABCDEFGH.....

   RYUZDEXA.....

   Il y a 26! = 403291461126605635584000000 permutations possibles des 26 lettres de l'alphabet ! Chacune de ces permutations devient la **clef** du code qui consiste à transformer les lettres du message par cette permutation. Pour l'attaquant qui cherche la clef, s'il met une seconde par clef, ça prend tout de même un milliard de fois la durée de vie de l'univers pour tout tester…

   Pendant plusieurs siècles dans l'antiquité on a considéré donc ce chiffrement **inviolable**…

3. L'analyse fréquentielle.

   Vers 750 c'est l'âge d'or de la culture islamique. Les califes abbassides ne se préoccupent plus de guerre et de conquêtes, et développent **arts et sciences**, la vitalité économique (avec des messages secrets!) et l'éducation pour former de nombreux érudits: astronomes, linguistes, mathématiciens…

«Ils s'approprièrent les avancées des civilisations précédentes en se procurant des textes égyptiens, babyloniens, indiens, chinois, farses, syriens, armé-niens, hébreux et romains, qu'ils traduisirent en arabe.» Ils ont utilisé le **papier** chinois diffuser cette information.

Al Kindi au 9ème siècle: Son plus important traité, retrouvé seulement en 1987 dans les archives ottomanes d'Istanbul, est intitulé Manuscrit sur le déchiffrement des messages cryptographiques. Il décrit comment utiliser les **statistiques de fréquences des lettres** pour briser un code. Par exemple en français le **e** est la lettre la plus utilisée (14%), c'est donc facile de voir avec quelle symbole elle est codée…

En utilisant cette technique, pour peu qu'on dispose de messages assez longs, on voit que les alphabets de chiffrement sont en fait très faciles à casser !

1.2.3 Marie Stuart (1542 – 1587)

Voici un "joli" exemple de cryptage essentiellement monoalphabétique, au plus haut niveau de l'état !

Avec le développement de la cryptanalyse statistique, les cryptographes désespèrent de trouver des méthodes inattaquables… Marie Stuart en fera l'amère expérience…

Reine d'Écosse à l'âge de 9 mois, mariée au français François II.

« Wikipedia: François II (Fontainebleau, le 19 janvier 1544 - Orléans, le 5 décembre 1560) est roi de France du 10 juillet 1559 jusqu'à sa mort.

Fils aîné d'Henri II et de Catherine de Médicis, il monte sur le trône de France à l'âge de quinze ans après la mort accidentelle de son père le 10 juillet 1559. Son règne éphémère ne dure qu'un an et cinq mois mais constitue un prélude majeur au déclenchement des guerres de Religion.

Son règne est en effet marqué par une importante crise politique et religieuse. À son avènement, il confie les rênes du gouvernement aux Guise, les oncles de son épouse Marie Stuart, reine d'Écosse, partisans d'une politique de répression à l'égard des protestants. Après la conjuration d'Amboise, il entame la mise en place d'une conciliation à l'égard des réformés mais se montre implacable face aux émeutiers qui mettent à mal son autorité dans les provinces.

Son règne est également marqué par l'abandon de l'Écosse, du Brésil et, sous l'effet du traité du Cateau-Cambrésis signé par son père Henri II, de la Corse, de la Toscane, de la Savoie et de la quasi-totalité du Piémont. Il marque, au profit de l'Espagne, le point de départ de l'affaiblissement de l'influence française en Europe.»

1560 Mort de François II. « (Singh) À compter de cette date, la vie de Marie ne fut plus que tragédie. Elle regagna l'Écosse en 1561, et découvrit son pays transformé. Pendant sa longue absence, Marie avait grandi dans la foi catholique, alors que ses sujets écossais se tournaient de plus en plus vers l'Eglise protestante. Marie se plia aux vœux de la majorité, et les premières années de son règne furent assez heureuses, mais en 1565 elle épousa lord Henry Stewart, comte de Darnley : ce fut le premier pas vers sa chute. Darnley était un homme bas et brutal, et devant sa cruauté et son avidité les nobles écossais retirèrent à Marie leur soutien. Marie fut elle-même le témoin horrifié de la barbarie de son époux lorsqu'il tua sous ses yeux son secrétaire, David Riccio. Il devenait évident pour tout le monde qu'il fallait se débarrasser de Darnley. Il n'est pas établi si l'attentat fut fomenté par Marie ou par la noblesse écossaise mais, dans la nuit du 9 février 1567, une explosion détruisit la maison de Darnley. »

https://www.youtube.com/watch?v=4PGA1AsFxHw

L'histoire est passionante; on a même fait des films sur sa vie. Marie dut fuire l'Ecosse et demanda refuge à sa cousine, une certaine Elizabeth, reine d'Angleterre. Choix bizarre, car l'Angleterre était protestante ! Elle espérait que les liens familiaux lui seraient utiles. Au lieu de ça, dès son arrivée, Elisabeth la fit

mettre en prison ! Elle y resta de nombreuses années.

Marie avait des soutiens car, de par sa lignée, elle pouvait prétendre au trône d'Angleterre, ce qui embêtait beaucoup Elizabeth évidemment. Un complot s'est monté pour la libérer, (et peut-être également pour assassiner Elizabeth…). Elle fut mise au courant par message secret transmis par le gardien de sa prison.

Mais le redoutable Sir Francis Walsingham, Premier secrétaire de la reine Elizabeth, était très intelligent. Il souhaitait la mort de Marie; et pour cela il devait prouver qu'elle faisait partie du complot. Il créa donc une école de chiffrement et recruta un des meilleurs cryptanalistes d'Europe, Thomas Phelippes. Ce dernier était maître de l'analyse des fréquences, et il parvint rapidement à casser le code de Marie.

Il savait aussi qu'il ne fait jamais dire aux ennemis qu'on a brisé leur code ! À la place, il forgea un faux message secret pour Marie, en utilisant son propre code, lui demandant le nom des conspirateurs !

Ainsi, Marie fut jugée coupable et on lui coupa la tête à la hache, devant la reine…

RAPPEL: J'ai mis des illustrations sur la page du cours:

1.2.4 Le chiffre de Vigenère 1586 (ou Bellaso 1533)

Marie Stuart aurait (peut-être) échappé à la hache du bourreau si elle avait pu utiliser le chiffre de Vigenère. C'est un code **polyalphabétique**: une lettre peut être codée de plusieurs façons différentes! On se donne une **clef** (un mot ou même un texte entier) que les deux protagonistes se partagent.

Puis on code la première lettre du message avec la première lettre de la clef (en mode "César"), la deuxième avec la deuxième, etc. et lorsque la clef est épuisée on recommence au début.

Maths: c'est encore une addition de congruences, comme César, mais **à plusieurs dimensions** (on additionne un vecteur = la clef).

Mais ce n'est pas un chiffre parfait ! On peut le briser par analyse de fréquences si la clef est trop courte.

1.2.5 La machine Enigma

Le chiffre de Vigenère est très long à utiliser, ne convient pas dans les guerres "modernes" où il faut rapidement prévenir des attaques de l'ennemi. En 1918 un certain Arthur Scherbius, inventeur allemand, crée une machine sur le principe du disque à chiffrer, mais où le disque change de position après chaque lettre (donc polyalphabétique). Et pour compliquer le tout, plusieurs disques se combinent. D'abord un échec commercial, a fallu attendre 1923 avec la publication des faits de la première guerre mondiale, indiquant comment les transmissions Allemandes avait été décodées par l'Angleterre, pour que l'Allemagne se décide à investir dans Enigma. Avec raison: Enigma a été très utile aux Allemands durant la deuxième guerre mondiale.

On pensait Enigma inviolable… mais face à la menace, les sercices secrets Polonais, soutenus par la France, ont cherché sans relâche comment briser Enigma. Puis, avec la création de la "Government Code and Cypher School (GC&CS), l'École du code anglaise" installée à Bletchley Park, on a construit des **machines** extrêment sophistiquées appelées "bombes" (énormes et bruyantes) qui travaillaient sans relâche pour briser Enigma. Tout ceci en grande partie due à un certain **Alan Turing**… L'effort de cryptanalyse a grandement contribué à créer les ordinateurs ! (cf. machine suivante. Collossus, avec des composants électroniques)

Remarque: c'était le début de la mathématisation du cryptage. Auparavant les "experts" étaient plutôt linguistes.

1.2.6 Pourquoi crypter aujourd'hui ?

La cpryto a longtemps été motivée par les guerres ou affaires diplomatiques. Mais de nos jours elle est nécessaire partout!

- commerce en ligne
- cartes de crédit
- emails et appels téléphoniques passent par de nombreux ordinateurs et donc facilement interceptables.
- décrypter les messages des "terroristes"

Problème éthique: quel est le bon équilibre entre respect de la vie privée ou des transactions commerciales, et sécurité publique ?

1.3 Cryptographie à clef publique

(James Ellis 1969 de Bletchley Park, donc gardé secret jusqu'à récemment, redécouvert par Diffie Hellman Merkle (Stanford USA) en 1976)

L'adage "moins on cache, plus c'est sûr"… se poursuit. Cette fois la clef est publique ! Mais bien sûr elle ne permet pas de déchiffrer, seulement de chiffrer. On dit que c'est une clef **asymétrique**.

Principe de base: chacun a une clef publique et une clef privée. Alice veut envoyer un message à Bob. Elle utilise la clef publique de Bob pour chiffrer le message. Ainsi, Bob pourra le déchiffrer grâce à sa clef privée.

- le principe de chiffrement est connu tous
- les clefs publiques sont connues de tous

On notera e (comme encoding) la clef publique, et d (comme decoding) la clef privée.

1.4 5. chiffrement RSA (1977)

(1975 Ellis, Cocks et Williamson du GCHQ de Bletchley Park — donc secret, redécouvert par Rivest, Shamir et Adleman, MIT (USA))

R et S sont informaticiens, A est un matheux.

Une portion de message à transmettre est d'abord transformée en un nombre x. (plusieurs lettres doivent être groupées pour éviter l'analyse fréquentielle)

Puis on cherche une application C_e (chiffrement) qui dépend de la clef publique e qui envoie ce nombre sur un autre, de façon **injective**. (en effet il faut pouvoir l'inverser pour déchiffrer!) Mais bien sûr on veut que l'inverse soit difficile à trouver!

L'idée de RSA est d'utiliser la difficulté de décomposer un nombre en produit de (grand) facteurs premiers.

Ainsi, $n = pq$ où p et q sont des grands nombres premiers. Connaissant n, il est difficile (càd ça prend beaucoup de temps) de trouver p et q. L'application C_e sera "mettre à la puissance e modulo n". Comment trouver l'application de déchiffrement ?

1. a) Petit theorème de Fermat amélioré

Soient p et q des nombres premiers distincts et $n = pq$. Alors pour tout $a \in \mathbb{Z}$ premier avec n on a:

$$(*) \qquad a^{(p-1)(q-1)} \equiv 1 \pmod{n}$$

Preuve 1: On veut appliquer le Théorème d'Euler (cours précédent). Pour cela, il suffit de montrer que si p et q sont premiers distincts, $\varphi(pq) = (p-1)(q-1)$.

Lemme: si a,b premiers entre eux, alors $\varphi(ab) = \varphi(a)\varphi(b)$.

"Preuve rapide": on a une bijection entre $\Phi(ab)$ et $\Phi(a) \times \Phi(b)$, où on note $\Phi(n)$ = les entiers entre 1 et n qui sont premiers avec n. La bijection est $m \mapsto (m \mod a, m \mod b)$... [je ne fais pas les détails ici]

Preuve 2: On donne une preuve directe qui n'utilise pas Euler.

Lemme 1: si $a \equiv b \pmod{p}$ et $a \equiv b \pmod{q}$ (avec p et q premiers distincts) alors $a \equiv b \pmod{pq}$.

Preuve du lemme 1: $a - b = kp$ et $a - b = mq$ donc $kp = mq$, mais par Gauss ça implique que $p|m$ (et $q|k$). Donc $a - b = mq$ est bien multiple de pq.

On revient à la question(*). Grâce au lemme, il suffit de montrer que $a^{(p-1)(q-1)} \equiv 1 \pmod{p}$ et $a^{(p-1)(q-1)} \equiv 1 \pmod{q}$. Puisque a est premier avec n, il n'est pas multiple de q, et donc $a' := a^{p-1}$ non plus, et donc le petit théorème de Fermat donne $(a')^{q-1} \equiv 1 \pmod{q}$, soit

$$[a^{p-1}]^{q-1} \equiv 1 \pmod{q}$$

De même

$$[a^{q-1}]^{p-1} \equiv 1 \pmod{p}$$

ce qui donne bien ce qu'on veut (on peut appliquer le lemme).

1. b) Lemme de déchiffrement RSA

On fixe un entier de la forme $n = pq$, où p et q sont des **nombres premiers distincts** ($q \neq p$). On 'pose' $\varphi(n) = (p-1)(q-1)$ (pas besoin de savoir que c'est l'indicatrice d'Euler). On se donne e (comme **encoding**) premier avec $\varphi(n)$. Soit d (comme **decoding**) un inverse de e modulo $\varphi(n)$. Alors pour tout entier m (comme **message**),

$$\text{Si } x \equiv m^e \pmod{n}, \qquad \text{alors } m \equiv x^d \pmod{n}.$$

Preuve: Soit $x \equiv m^e \pmod{n}$.

- si m est premier avec n ça découle directement de Fermat amélioré. En effet $m^{\varphi(n)} \equiv 1 \pmod{n}$ donc

$$x^d = (m^e)^d = m^{ed} = m^{1+k\varphi(n)}$$

puisque d est un inverse de e modulo $\varphi(n)$.

Donc

$$x^d = m \times (m^{\varphi(n)})^d \equiv m \times 1^d \pmod{n}.$$

- sinon (m n'est pas premier avec n) on va utiliser le lemme 1. Quitte à remplacer m par un autre représentant de sa classe de congruence modulo n, on peut supposer $0 \leq m < n$. Puisque $n = pq$ et m n'est pas premier avec n, ils ont un facteur premier en commun, donc p ou q (mais pas les deux car $m < n$). p et q jouant le même rôle, supposons que p divise m, et donc $pgcd(m, q) = 1$. On va applique le lemme 1 à la quantité $(m^e)^d$:
  - modulo p : on a $m \equiv 0 \pmod{p}$ donc

$$(m^e)^d \equiv 0 \pmod{p} \equiv m$$

  - modulo q, comme précédemment:

$$(m^e)^d = m \times (m^{\varphi(n)})^k = m \times (m^{(q-1)})^{(p-1)k} \equiv m \times 1 \pmod{q}.$$

  Par le lemme 1 on a bien $(m^e)^d \equiv m \pmod{pq}$.

2. c) fonction à sens unique RSA.

Def fonction à sens unique: une fonction $m \mapsto f(m)$ telle que:

- connaissant m on peut suffisamment facilement calculer $f(m)$;
- mais étant donné un élément y dans l'image de f, il est très difficile de trouver le m tel que $y = f(m)$. ("difficile" veut dire par exemple: ça prendrait l'âge de l'univers avec les ordinateurs actuels)

Après la découverte du principe de clef publique, il fallait trouver une telle fonction "à sens unique"… ça a pris plusieurs années. Ellis, Cocks, Williamson et RSA ont l'idée de prendre la fonction "élever à une certaine puissance modulo n".

$$f_e(m) \equiv m^e \pmod{n}.$$

1. c'est "facile à calculer" car même si m^e est un nombre gigantesque, ici on ne cherche que sa classe modulo n: on peut calculer la puissance par des multiplications répétées et réduire modulo n à chaque étape… (cf tous les exos qu'on a faits! 1-6 feuille TD) En `python`:

   `pow(m,e,n)`

   message clair = m ====> message codé = $m^e \pmod{n}$

2. avec sa clef privée d on peut facilement décoder le message crypté

   message codé $y = m^e$ ====> message clair $m = y^d \pmod{n}$

   (appliquer le lemme RSA)

3. sans la clef privée, c'est très difficile de retrouver m. En gros, pour pouvoir retrouver m à partir de m^e, il faut un inverse de e modulo $\varphi(n)$. Facile (Bézout) si on connaît $\varphi(n)$. Mais comment calculer $\varphi(n)$?

   Calculer $\varphi(n)$ est aussi compliqué que de décomposer n en facteurs premiers $n = pq$. ('car'

$\varphi(n) = (p-1)(q-1)$, cf Exercice 13).

3. d) Pratique du cryptage RSA:

Supposons pour changer que Bob envoie un message à Alice.

1. Alice se prépare une clef publique (n, e) et une clef privée d:

   - elle choisit deux nombres premiers distincts (et grands: plusieurs centaines de chiffres), p et q et calcule $n = pq$ (la complexité de la multiplication de chiffres de taille n est $O(n^2)$). Elle calcule aussi $\varphi(n) = (p-1)(q-1)$.

     Exemple p=11, q=23, n = 253, $\varphi(n) = 220$.

   - elle choisit un exposant e premier avec $\varphi(n)$, exemple $e = 3$.
   - elle calcule un inverse de e modulo $\varphi(n)$ (algo Euclide), ici $d = 147$.
   - elle détruit soigneusement $\varphi(n)$.
   - Elle peut donc publier sa clef publique (n, e) et sa clef privée est $d = 147$.

2. Bob veut envoyer un message à Alice

   - En pratique il faut d'abord convertir le message en suite de chiffres

   (texte –> code ASCII ou UTF, image –> codage JPEG par exemple, etc)

   - Bob récupère la clef publique d'Alice {\color{green}{(n,e)}}
   - Puis il découpe son message en morceaux afin que chaque morceau soit un entier < n. Et il procède à l'envoi de chaque morceau m.

   - Chiffrement du message m.

     Exemple Bob envoie l'entier 123 à Alice. (on a bien 123<253)

     Il calcule le message chiffré:

     $$x \equiv m^e \pmod{n}$$

     Bien sûr pour cela il faut utiliser une méthode d' **exponentiation rapide** par exemple: écrire l'exposant en base 2:

     Exemple: comment calculer 5^{11} avec seulement 5 multiplications (et bien sûr réduire modulo n à chaque multiplication):

     11 = 8+2+1

     Il suffit de calculer 5^2 puis 5^4 puis 5^8 (3 multiplications) puis des les multiplier entre eux (2 multiplications).

   - Envoi du message.

     Bob peut envoyer son message x à Aice de façon publique.

3. Alice déchiffre le message reçu

   Le message x reçu par Alice est chiffré. Elle calcule simplement le message en clair m, grâce à sa clef privée d, par la formule

$$m \equiv x^d \pmod{n}$$

Le Lemme de Déchiffrement ci-dessus prouve qu'elle retrouve bien le message initial !

Pour un exemple plus réaliste avec des grands nombres premiers, cf la feuille `python`.

https://crypto.stackexchange.com/questions/13113/how-can-i-find-the-prime-numbers-used-in-rsa

http://doctrina.org/How-RSA-Works-With-Examples.html

FIN DU COURS!

Je vous incite à regarder (ou revoir, si vous l'avez déjà vu!) un beau film sur Alan Turing qui parle beaucoup de la machine Enigma:

Imitation Game

Voir en particulier l'épisode à **1.13.36**, où le code Enigma est finalement brisé!!

Author: San Vũ Ngọc
Created: 2021-03-27 sam. 19:49
Validate