Université Pierre et Marie Curie
Cours de cryptographie MM029 - 2009/10
Alain Kraus

Chapitre IV - Cryptosystèmes à clés publiques

Table des matières

1. Principe général

La cryptographie est l'étude de la science des communications par des messages codés qui ne pourront être lus que par leur destinataire. Un cryptosystème est un tel mode de communication. C'est un procédé de chiffrement et de déchiffrement permettant de transmettre ou de recevoir des informations secrètes. Une clé d'un cryptosystème est un code permettant de chiffrer un message lors de sa transmission, ou de le déchiffrer à la réception.

La cryptographie à clé publique est apparue en 1976 avec les travaux de W. Diffie et M. Hellman. Un cryptosystème à clé publique, appelé aussi asymétrique, repose sur l'existence d'une clé publique pour le chiffrement, et d'une clé secrète pour le déchiffrement. Ces deux clés sont distinctes. Un utilisateur A qui souhaite envoyer un message à un utilisateur B, chiffre son message au moyen de la clé publique de B, et ce dernier au moyen de sa clé secrète, qu'il est seul à connaître, est alors en mesure de déchiffrer le message envoyé. Deux utilisateurs d'un cryptosystème à clé publique peuvent donc s'échanger des messages chiffrés, via un canal non sécurisé, et sans posséder de secret en commun. Son efficacité est basée sur le fait qu'il est impossible en un temps raisonnable de déterminer la clé secrète à partir de la clé publique.

Le principe général peut se schématiser comme suit. Soit $\mathcal{M}$ un ensemble de chiffrements. Par exemple, on prend souvent pour $\mathcal{M}$ un ensemble $\mathbb{Z}/n\mathbb{Z}$ ou bien un corps fini. Soit A une personne souhaitant pouvoir se faire envoyer des messages chiffrés de $\mathcal{M}$ de façon confidentielle. Elle choisit une bijection $f_A : \mathcal{M} \to \mathcal{M}$ qui sera publique, telle que la bijection réciproque f_A^{-1} ne soit connue que d'elle même. L'idée essentielle étant qu'il

impossible pratiquement de déterminer f_A^{-1} connaissant f_A, le temps nécessaire à cette détermination étant beaucoup trop long. Supposons alors qu'une personne B envoie à A un message $x \in \mathcal{M}$. Pour cela, B envoie en clair l'élément $y = f_A(x)$. Afin de déchiffrer ce message, A calcule $f_A^{-1}(y)$, et retrouve ainsi le message x. La seule façon, a priori, qu'un intrus puisse identifier x est de connaître f_A^{-1}. On dit souvent que f_A est une fonction « à sens unique », vu la difficulté pratique d'expliciter sa fonction réciproque.

Nous allons voir des exemples de cryptosystèmes à clés publiques, dont l'efficacité est basée sur la difficulté de factoriser des « grands » entiers, ou bien sur la difficulté de résoudre le problème du logarithme discret dans des corps finis bien choisis. Ce sont les cryptosystèmes les plus sûrs. On verra aussi un exemple qui repose sur le problème appelé du sac à dos.

2. Cryptosystème RSA

Il a été découvert par Rivest, Shamir et Adleman en 1977. Son efficacité repose sur le fait que connaissant un entier n, qui est le produit de deux « grands » nombre premiers p et q distincts, il est généralement très difficile, voire impossible pratiquement, de déterminer p et q i.e. la factorisation de n. Ce système utilise le résultat suivant, qui est une conséquence du petit théorème de Fermat. Rappelons que φ désigne la fonction indicatrice d'Euler.

Proposition 4.1. *Soient p et q deux nombres premiers distincts. Posons $n = pq$. Soit t un entier naturel congru à 1 modulo $\varphi(n)$. Alors, on a*

$$a^t \equiv a \text{ mod. } n \quad \textit{quel que soit } a \in \mathbb{Z}.$$

Démonstration : Il existe un entier k tel que l'on ait $t = 1 + k\varphi(n)$. Soit a un entier relatif. Compte tenu de l'égalité $\varphi(n) = (p-1)(q-1)$, on obtient

$$a^t = a\Big(a^{(p-1)(q-1)}\Big)^k = a\Big(a^{p-1}\Big)^{(q-1)k}.$$

Si p ne divise pas a, on a $a^{p-1} \equiv 1$ mod. p, d'où l'on déduit que $a^t \equiv a$ mod. p. Si p divise a, cette congruence est aussi vérifiée. De même, on a $a^t \equiv a$ mod. q. Puisque p et q sont distincts, il en résulte que n divise $a^t - a$.

1. Principe

Chaque utilisateur procède de la façon suivante :

1) il choisit deux grands nombres premiers p et q, ayant chacun disons environ cent cinquante chiffres décimaux, et calcule $n = pq$.
2) Il choisit un entier e premier avec $\varphi(n)$ tel que $1 < e < \varphi(n)$. La classe de e modulo $\varphi(n)$ est donc inversible dans $\mathbb{Z}/\varphi(n)\mathbb{Z}$.

3) Il détermine l'entier d tel que $1 < d < \varphi(n)$ et $ed \equiv 1$ mod. $\varphi(n)$. La classe de d modulo $\varphi(n)$ est donc l'inverse de la classe de e dans $\big(\mathbb{Z}/\varphi(n)\mathbb{Z}\big)^*$. Ce calcul peut être effectué en utilisant l'algorithme d'Euclide.

4) Il publie ensuite le couple (e, n), qui est sa clé publique, et il conserve secret le couple $\big(d, \varphi(n)\big)$, qui est sa clé secrète.

Soit A un utilisateur dont la clé publique est (e, n) et la clé secrète est $\big(d, \varphi(n)\big)$. On dit que l'algorithme de chiffrement de A est l'application $f_A : \mathbb{Z}/n\mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$ définie pour tout $x \in \mathbb{Z}/n\mathbb{Z}$ par

$$f_A(x) = x^e.$$

C'est une bijection de $\mathbb{Z}/n\mathbb{Z}$ et d'après la proposition 4.1 on a

$$f_A^{-1}(x) = x^d.$$

On dit que f_A^{-1} est l'algorithme de déchiffrement de A. Si une personne B souhaite envoyer un message secret à A sous la forme d'un élément $x_0 \in \mathbb{Z}/n\mathbb{Z}$, il utilise la clé publique de A en lui envoyant l'élément $f_A(x_0)$. Afin d'obtenir x_0, il suffit alors pour A d'utiliser sa clé secrète en calculant $f_A^{-1}(x_0^e)$.

Exemple 4.1. Prenons $(e, n) = (239, 406121)$ comme clé publique. On a $n = pq$ avec $p = 101$ et $q = 4021$. Il est facile de vérifier que p et q sont premiers (si q n'était pas premier il devrait posséder un diviseur premier plus petit que 61, et ce n'est pas le cas). On obtient $\varphi(n) = 402000$. Afin de déterminer la clé secrète, il s'agit de calculer l'inverse de 239 modulo 402000. On utilise l'algorithme d'Euclide. Compte tenu du théorème 1.1, on peut effectuer ce calcul avec au plus douze divisions euclidiennes. En fait trois suffisent. Avec la présentation adoptée de cet algorithme, on obtient le tableau suivant :

	1682	119	2	
402000	239	2	1	0
1	0	1	-119	
0	1	-1682	200159	

La clé secrète est donc $\big(d, \varphi(n)\big) = (200159, 402000)$.

2. Cryptanalyse

Reprenons les notations précédentes. Une personne souhaitant retrouver x_0 à partir de $f_A(x_0)$ est confrontée, *a priori*, au problème de la détermination de $\varphi(n)$, ou ce qui revient au même, à celui de la factorisation de n. En effet :

Lemme 4.1. *Connaître de n et $\varphi(n)$ équivaut à connaître p et q.*

Démonstration : Supposons $\varphi(n)$ (et n) connus. Il s'agit d'expliciter p et q. On a

$$n = pq \quad \text{et} \quad p + q = n - \varphi(n) + 1.$$

Par suite, p et q sont les racines du polynôme $X^2 - \big(n - \varphi(n) + 1\big)X + n \in \mathbb{Z}[X]$. On obtient ainsi p et q. Inversement, si p et q sont connus, $\varphi(n)$ l'est aussi car $\varphi(n) = (p-1)(q-1)$.

On ne sait pas a priori déterminer x_0 sans identifier p et q. Cela étant, on ne dispose pas de preuve que la difficulté de déterminer x_0 soit équivalente à celle de la factorisation de n. Factoriser n est peut-être plus difficile que de trouver x_0.

Pour un utilisateur du système RSA, il s'agit donc de choisir p et q de sorte que, connaissant leur produit n, il n'y ait pas de circonstances numériques favorables à leur détermination. Voyons quelques remarques à ce sujet.

1) Tout d'abord p et q doivent être choisis assez grands, par exemple chacun avec environ cent cinquante chiffres décimaux. De plus, ils ne doivent pas être trop proches l'un de l'autre, il convient que l'un possède quelques chiffres décimaux de plus que l'autre. En effet, si $n = pq$, avec $p > q$, posons

$$s = \frac{p-q}{2} \quad \text{et} \quad t = \frac{p+q}{2}.$$

On a l'égalité

$$n = t^2 - s^2,$$

en particulier $t^2 - n$ est un carré. Si p et q sont proches, alors s et $t - \sqrt{n}$ sont petits. Dans ce cas, en effectuant des tests successifs, on peut trouver un entier $a > \sqrt{n}$, voisin de $\sqrt{n}$, tel que $a^2 - n$ soit un carré. Cela permet alors d'exprimer n comme une différence de deux carrés, et d'obtenir la factorisation de n.

2) Les entiers $p-1$ et $q-1$ ne doivent pas être facilement factorisables, par exemple il convient d'éviter que tous leurs diviseurs premiers soient petits. En effet, choisissons une constante $C > 0$ et notons S l'ensemble des nombres premiers plus petits que C. Soit T l'ensemble des entiers plus petits que n, dont tous les diviseurs premiers sont dans S. Supposons que $p-1$ soit dans T. Pour tout $a \in \mathbb{N}$ non divisible par p, on a

$$\operatorname{pgcd}\big(a^{p-1} - 1, n\big) \equiv 0 \bmod. \ p.$$

En calculant l'entier $\operatorname{pgcd}(a^t - 1, n)$, pour $t \in T$ et quelques entiers $a \in \mathbb{N}$ (par exemple $a = 2$), on peut ainsi espérer trouver la factorisation de n. Afin d'éviter cet inconvénient, on peut choisir deux grands nombres premiers ℓ_1 et ℓ_2, et prendre p et q de la forme $p = 1 + k\ell_1$ et $q = 1 + r\ell_2$ avec k et r petits.

3) On peut trouver la factorisation de n si le message envoyé $x_0 = \widetilde{x_0} + n\mathbb{Z}$ est tel que $\widetilde{x_0}$ ne soit pas premier à n. En effet, connaissant n et x_0^e, on peut déterminer le pgcd de n et $\widetilde{x_0^e}$. Si ces entiers ne sont pas premiers entre eux, on connaît alors un diviseur premier de n, et donc la factorisation de n. Ceci constitue une contrainte sur les messages à envoyer. Cela étant, la probabilité pour qu'un élément $a + n\mathbb{Z}$ de $\mathbb{Z}/n\mathbb{Z}$ choisi au hasard soit tel que $\text{pgcd}(a, n) \neq 1$ est

$$1 - \frac{\varphi(n)}{n} = \frac{1}{p} + \frac{1}{q} - \frac{1}{pq},$$

qui est donc très petite si p et q sont grands. Il est donc peu probable de se trouver dans cette situation.

4) Supposons que l'on connaisse un entier $m \geq 1$ tel que l'on ait

$$a^m \equiv 1 \text{ mod. } n \quad \text{pour tout } a \text{ tel que } 1 \leq a \leq n \text{ et } \text{pgcd}(a, n) = 1. \tag{1}$$

Il existe alors un algorithme probabiliste permettant de factoriser n. En explicitant cette condition avec $a = n - 1$, on constate d'abord que m est pair (car n est impair). On commence par tester si la condition (1) est encore satisfaite avec l'entier $\frac{m}{2}$. S'il existe un entier a compris entre 1 et n, premier avec n, tel que $a^{\frac{m}{2}} \not\equiv 1$ mod. n, alors il y a au moins $\frac{\varphi(n)}{2}$ tels entiers a, car l'ensemble

$$\left\{ x + n\mathbb{Z} \;\middle|\; x^{\frac{m}{2}} \equiv 1 \text{ mod. } n \right\}$$

est un sous-groupe de $\left(\mathbb{Z}/n\mathbb{Z}\right)^*$. Si après une dizaine de tests, on trouve que la congruence (1) est satisfaite avec $\frac{m}{2}$, alors on remplace m par $\frac{m}{2}$, la probabilité étant grande pour que la condition (1) soit effectivement réalisée avec $\frac{m}{2}$. On recommence ce processus jusqu'à trouver un entier m satisfaisant (1), mais pas $\frac{m}{2}$. Vérifions le lemme suivant.

Lemme 4.2. *Soit m un entier ≥ 1 tel que la condition (1) soit satisfaite par m, et ne le soit pas avec l'entier $\frac{m}{2}$. Alors, il y a exactement $\frac{\varphi(n)}{2}$ entiers a tels que l'on ait $1 \leq a \leq n$ et $\text{pgcd}(a, n) = 1$, et que $a^{\frac{m}{2}} - 1$ soit divisible par l'un des entiers p et q, mais pas par n.*

Ainsi, en choisissant au hasard un entier a compris entre 1 et n, et premier avec n, la détermination de l'entier

$$\text{pgcd}\left(a^{\frac{m}{2}} - 1, n\right)$$

permet, «avec une chance sur deux», d'obtenir un diviseur non trivial de n et donc sa factorisation. Comme on l'a déjà remarqué, la probabilité pour qu'un entier a choisi au hasard entre 1 et n ne soit pas premier avec n, est petite si p et q sont grands. Il suffit donc en pratique de choisir aléatoirement a, sans se préoccuper s'il est premier avec n ou pas, et de calculer le pgcd de $a^{\frac{m}{2}} - 1$ et n, pour obtenir rapidement la factorisation de n.

Démonstration du lemme 4.2 : D'après (1), pour tout a tel que $1 \leq a \leq n$ et $\text{pgcd}(a, n) = 1$, on a

$$a^{\frac{m}{2}} \equiv \pm 1 \text{ mod. } p \quad \text{et} \quad a^{\frac{m}{2}} \equiv \pm 1 \text{ mod. } q, \tag{2}$$

les signes dans ces deux congruences étant indépendants. Puisque $\frac{m}{2}$ ne satisfait pas la condition (1), $\frac{m}{2}$ n'est pas multiple des deux entiers $p-1$ et $q-1$ (cf. le petit théorème de Fermat). Deux cas peuvent alors se présenter.

(i) L'entier $\frac{m}{2}$ est multiple de l'un des entiers $p-1$ et $q-1$, par exemple on a

$$\frac{m}{2} \equiv 0 \text{ mod. } p-1 \quad \text{et} \quad \frac{m}{2} \not\equiv 0 \text{ mod. } q-1.$$

D'après (2), pour tout a entre 1 et n et premier avec n, on obtient

$$a^{\frac{m}{2}} \equiv 1 \text{ mod. } p \quad \text{et} \quad a^{\frac{m}{2}} \equiv \pm 1 \text{ mod. } q. \tag{3}$$

Il existe un entier b tel que $b^{\frac{m}{2}} \equiv -1$ mod. q. Sinon, en considérant un générateur de $\left(\mathbb{Z}/q\mathbb{Z}\right)^*$, on constate que $q-1$ devrait diviser $\frac{m}{2}$ (on peut aussi utiliser directement le fait que $\frac{m}{2}$ ne vérifie pas (1)). Soit $f : \left(\mathbb{Z}/n\mathbb{Z}\right)^* \to \left(\mathbb{Z}/q\mathbb{Z}\right)^*$ l'application définie par

$$f(x + n\mathbb{Z}) = x^{\frac{m}{2}} + q\mathbb{Z}.$$

C'est un morphisme de groupes. On a $f(b + n\mathbb{Z}) = -1$, donc l'image de f est $\left\{\pm 1\right\}$. Par suite, $\text{Ker}(f)$ est d'ordre $\frac{\varphi(n)}{2}$. Il en résulte que l'ensemble

$$\left\{x + n\mathbb{Z} \in \left(\mathbb{Z}/n\mathbb{Z}\right)^* \;\middle|\; x^{\frac{m}{2}} \equiv -1 \text{ mod. } q\right\}$$

est de cardinal $\frac{\varphi(n)}{2}$. La condition (3) entraîne alors le résultat dans ce cas.

(ii) L'entier $\frac{m}{2}$ n'est pas multiple de $p-1$ ni de $q-1$. Dans ce cas, pour la même raison que celle évoquée ci-dessus, il existe $b \in \mathbb{Z}$ tel que l'on ait

$$b^{\frac{m}{2}} \equiv -1 \text{ mod. } q.$$

Il existe $c \in \mathbb{Z}$ tel que $c \equiv 1$ mod. p et $c \equiv b$ mod. q (théorème chinois). On a alors

$$c^{\frac{m}{2}} \equiv 1 \text{ mod. } p \quad \text{et} \quad c^{\frac{m}{2}} \equiv -1 \text{ mod. } q. \tag{4}$$

Soit H le sous-groupe de $\left(\mathbb{Z}/n\mathbb{Z}\right)^*$ formé des éléments $x + n\mathbb{Z}$ tels que $x^{\frac{m}{2}} \equiv 1$ mod. p. Puisque $\frac{m}{2}$ n'est pas multiple de $p-1$, l'ordre de H est $\frac{\varphi(n)}{2}$ (même argument que celui utilisé dans le cas (i)). Soit $g : H \to \left(\mathbb{Z}/q\mathbb{Z}\right)^*$ le morphisme de groupes défini par

$$g(x + n\mathbb{Z}) = x^{\frac{m}{2}} + q\mathbb{Z}.$$

D'après (4), $c+n\mathbb{Z}$ appartient à H et $g(c+n\mathbb{Z})=-1$. Ainsi l'image de g est $\big\{\pm 1\big\}$, et son noyau est d'ordre

$$\frac{|H|}{2}=\frac{\varphi(n)}{4}.$$

Il y a donc exactement $\frac{\varphi(n)}{4}$ éléments $x+n\mathbb{Z}$ de $\big(\mathbb{Z}/n\mathbb{Z}\big)^*$ tels que l'on ait

$$x^{\frac{m}{2}}\equiv 1 \text{ mod. } p \quad \text{et} \quad x^{\frac{m}{2}}\equiv -1 \text{ mod. } q.$$

La même assertion vaut en échangeant p et q. Cela établit le résultat.

5) On verra dans le chapitre V que dans l'utilisation du cryptosystème RSA, il convient aussi de choisir p et q de sorte que le pgcd de $p-1$ et $q-1$ soit petit. On constatera que si ce pgcd est grand, on peut en pratique retrouver rapidement p et q.

3. Signature

L'algorithme RSA fournit un moyen de signer, ou d'authantifier, ses messages. Soit A un utilisateur ayant pour clé publique (e,n) et pour clé secrète $\big(d,\varphi(n)\big)$. Supposons que A souhaite envoyer à B un message $x\in\mathbb{Z}/n\mathbb{Z}$, sans se préoccuper de sa confidentialité, mais de sorte que B soit certain que c'est bien A qui lui a transmis x. Pour cela, A envoie à B le couple

$$(x,x^d)\in\mathbb{Z}/n\mathbb{Z}\times\mathbb{Z}/n\mathbb{Z}.$$

Avec la clé publique (e,n), B calcule alors

$$(x^d)^e=x^{de}=x.$$

Puisque A est seul à connaître d, B peut être a priori certain que c'est bien A l'expéditeur du message.

3. Algorithme de El Gamal

Cet algorithme, qui date de 1984, concerne le problème de la confidentialité des messages envoyés, et son efficacité est basée sur la difficulté de résoudre le problème du logarithme discret dans des corps finis bien choisis.

Le principe est le suivant. Une personne, Alice, souhaite permettre à quiconque de lui envoyer des messages confidentiels. Pour cela, elle choisit au départ un couple public (K,g), formé d'un corps fini K et d'un générateur g de K^*. Soit q le cardinal de K. Le procédé est alors le suivant :

1) Alice choisit aléatoirement un entier a tel que $1<a<q-1$, qui sera sa clé secrète. Elle calcule g^a qu'elle publie, et qui sera sa clé publique.

   La clé publique de l'algorithme est donc au départ le triplet (K,g,g^a).

2) Afin d'envoyer un message $m \in K$ à Alice, une personne Bob choisit aléatoirement un entier x tel que $1 < x < q-1$, et transmet à Alice le couple

$$(g^x, mg^{ax}). \tag{5}$$

C'est la phase d'encryptage du message m.

3) Afin de décrypter le message reçu, il s'agit donc de la phase de décryptage, Alice, connaissant a et g^x, détermine alors l'inverse dans K de l'élément $(g^x)^a$ i.e. g^{-ax} [(1)]. Elle effectue ensuite la multiplication de g^{-ax} par mg^{ax}, ce qui, vu l'égalité

$$g^{-ax}\big(mg^{ax}\big) = m, \tag{6}$$

lui permet de retrouver m.

Exemple 4.2. On prend $K = \mathbb{F}_{31}$. La classe de 3 est un générateur de $\mathbb{F}_{31}^*$. Supposons que la clé publique d'Alice soit le triplet

$$(\mathbb{F}_{31}, 3, 29).$$

Bob envoie à Alice le message $(17, 18)$. Afin de retrouver le message m que Bob veut faire parvenir à Alice, il s'agit de trouver le logarithme discret de base 3 de 29, autrement dit, le plus petit entier $a \geq 1$ tel que l'on ait

$$3^a \equiv 29 \text{ mod. } 31.$$

Dans $\mathbb{F}_{31}^*$, on a $3^3 = -4$, $3^6 = 16$, d'où $3^9 = -64 = 29$ et $a = 9$. Conformément à (5) et à la formule (6), vu que $17^{-1} = 11$, on a donc

$$m = 17^{-9}.18 = 11^9.18.$$

On a $11^2 = -3$, d'où $11^8 = 19$ et $11^9 = -8$, puis $m = 11$.

4. Protocole de Diffie-Hellman

On utilise aussi en cryptographie des cryptosystèmes, qui ne sont pas à clés publiques, qui sont plus rapides d'utilisation, mais moins efficaces, que ceux à clés publiques. Au cours du procédé d'utilisation choisi, il peut être alors opportun entre deux utilisateurs de se fabriquer une clé secrète commune, à partir d'une clé publique. La difficulté pour trouver

---

(1) Notons que pour tout $y \in K^*$, on a $y^{q-1} = 1$, donc l'inverse de y est $y^{-1} = y^{q-2}$. On a ainsi $g^{-ax} = (g^x)^{a(q-2)}$, de sorte qu'Alice peut trouver l'inverse de g^{ax} en élevant directement g^x à la puissance $a(q-2)$.

cette clé secrète, est alors analogue à celle pour décrypter un message dans l'utilisation d'un cryptosystème à clé publique.

Dans cette optique, le protocole de Diffie-Hellman, qui date de 1976, est le suivant. Deux personnes, Alice et Bob, souhaitent se construire une clé secrète commune, qu'ils seront les seuls à connaître, afin de communiquer sur un canal non sûr en utilisant cette clé pour chiffrer leur correspondance. Leur procédé de fabrication est basé sur le fait que le problème du logarithme discret soit difficile à résoudre dans certains corps finis. Soient K un corps fini de cardinal q, dans lequel le problème du logarithme discret soit a priori difficile à résoudre, et g un générateur de K^*. Le couple (K, g) est public.

1) Alice choisit secrètement et aléatoirement un entier a tel que $1 < a < q-1$, et elle transmet à Bob publiquement l'élément g^a.
2) Bob choisit aussi secrètement et aléatoirement un entier b tel que $1 < b < q-1$, et il transmet à Alice publiquement l'élément g^b.
3) Alice élève g^b à la puissance a, et elle obtient ainsi l'élément g^{ab}.
4) Bob élève g^a à la puissance b, obtenant de même g^{ab}.

Leur clé secrète commune est alors g^{ab}.

Ils sont les seuls à la connaître, car quiconque disposant du couple (K, g), ainsi que des éléments g^a et g^b, ne peut pas en déduire g^{ab}, sauf à déterminer, *a priori*, a ou b i.e. $\log_g(g^a)$ ou $\log_g(g^b)$. On ne connaît pas d'autres moyens pour déterminer g^{ab}.

5. Cryptosystème de Rabin

Ce cryptosystème à clé publique est basé, comme le système RSA, sur la difficulté de factoriser un entier qui est un produit de deux grands nombres premiers. Il a été inventé par Rabin en 1979. Contrairement au système RSA, on peut démontrer que la difficulté de « casser » ce cryptosystème est équivalente à celle du problème de la factorisation. Cela étant, il a le désavantage que le décryptage de sortie peut être issu de quatre messages envoyés distincts. Il faut donc déterminer quel est le bon par un procédé annexe.

1. Principe

Chaque utilisateur choisit deux nombres premiers impairs p et q, avec les mêmes précautions que pour le système RSA. Sa clé publique est l'entier $n = pq$, et sa clé secrète est (p, q). En notant $\left(\mathbb{Z}/n\mathbb{Z}\right)^{*2}$ le sous-groupe des carrés de $\left(\mathbb{Z}/n\mathbb{Z}\right)^*$, son algorithme de chiffrement est la fonction $f : \left(\mathbb{Z}/n\mathbb{Z}\right)^* \to \left(\mathbb{Z}/n\mathbb{Z}\right)^{*2}$ définie par

$$f(x) = x^2.$$

Le noyau de f étant d'ordre 4 (exemple 1.5 du chapitre I), $\left(\mathbb{Z}/n\mathbb{Z}\right)^{*2}$ est d'ordre $\frac{\varphi(n)}{4}$. Si l'utilisateur reçoit le message $m + n\mathbb{Z} \in \left(\mathbb{Z}/n\mathbb{Z}\right)^{*2}$, afin de le décrypter, il est amené à

résoudre l'équation

$$x^2 = m + n\mathbb{Z}. \tag{7}$$

Pour cela, il résoud les congruences

$$a^2 \equiv m \text{ mod. } p \quad \text{et} \quad b^2 \equiv m \text{ mod. } q. \tag{8}$$

Elles ont chacune deux solutions respectivement modulo p et q, que l'on peut obtenir en utilisant par exemple l'exercice 8 de la deuxième feuille d'exercices de travaux dirigés. L'équation (7) a ainsi quatre solutions, que l'on explicitent ensuite en résolvant quatre systèmes de congruences. Il suffit en fait d'en résoudre deux par un choix convenable des signes. Si $(a, b) \in \mathbb{Z}^2$ vérifie la condition (8), il existe $c \in \mathbb{Z}$, unique modulo n, tel que $c \equiv a$ mod. p et $c \equiv b$ mod. q (théorème chinois), d'où $c^2 \equiv m$ mod. n et $c + n\mathbb{Z}$ est une solution de (7).

Pour faciliter l'extraction des racines carrés dans $\mathbb{F}_p$ et $\mathbb{F}_q$, on peut choisir p et q congrus à 3 modulo 4, auquel cas

$$\pm m^{\frac{p+1}{4}} + p\mathbb{Z} \quad \text{et} \quad \pm\, m^{\frac{q+1}{4}} + q\mathbb{Z}$$

sont les deux racines carrées de $m+p\mathbb{Z}$ dans $\mathbb{F}_p$ et de $m+q\mathbb{Z}$ dans $\mathbb{F}_q$. Comme on le signalait plus haut, une fois que l'on a trouvé les quatre solutions de (7), il s'agit de déterminer laquelle est celle qui a été envoyée. On peut par exemple le faire si l'une d'elle représente un mot dans un langage choisi, et pas les autres.

Exemple 4.3. Prenons $n = 247$ et $m = 43 + 247\mathbb{Z}$. On a $n = pq$ avec $p = 13$ et $q = 19$. L'élément m est un carré modulo 247 vu c'est un carré modulo p et q. En effet,

$$\left(\frac{43}{13}\right) = \left(\frac{4}{13}\right) = 1 \quad \text{et} \quad \left(\frac{43}{19}\right) = \left(\frac{5}{19}\right) = \left(\frac{19}{5}\right) = \left(\frac{-1}{5}\right) = 1.$$

Déterminons l'ensemble S des solutions dans $\mathbb{Z}/247\mathbb{Z}$ de l'équation

$$x^2 = 43.$$

Modulo 13 les deux racines carrées de 43 sont ± 2. On a $43 \equiv 5$ mod. 19 et les deux racines carrés de 5 modulo 19 sont ± 9. On est alors amené à résoudre les deux systèmes de congruences

$$\begin{cases} x \equiv 2 \text{ mod. } 13 \\ x \equiv 9 \text{ mod. } 19 \end{cases} \quad \text{et} \quad \begin{cases} x \equiv 2 \text{ mod. } 13 \\ x \equiv -9 \text{ mod. } 19. \end{cases}$$

En utilisant l'égalité $3 \times 13 - 2 \times 19 = 1$, on obtient comme solutions particulières respectivement $x = 275$ et $x = 67$ (théorème chinois). En tenant compte des solutions opposées, on en déduit que l'on a

$$S = \left\{\overline{67}, \overline{180}, \overline{219}, \overline{275}\right\}.$$

2. Cryptanalyse

Vérifions que la difficulté de décryptage est équivalente à celle du problème de la factorisation de n, autrement dit, que savoir extraire les racines carrées dans $\mathbb{Z}/n\mathbb{Z}$ équivaut à savoir factoriser n. On a vu précédemment que la connaissance de p et q permet d'extraire les racines carrées modulo n.
Inversement, supposons que l'on dispose d'un algorithme permettant d'extraire les racines carrées modulo n. On choisit un élément $x + n\mathbb{Z} \in \mathbb{Z}/n\mathbb{Z}$ tel que x soit premier avec n et on détermine les racines carrées de $x^2 + n\mathbb{Z}$. Puisque x est premier avec n, il en existe quatre (il y en a deux modulo p et deux modulo q). Il existe donc $y \in \mathbb{Z}$ tel que l'on ait

$$y^2 \equiv x^2 \text{ mod. } n \quad \text{et} \quad y \not\equiv \pm x \text{ mod. } n.$$

Ainsi, n divise $(x + y)(x - y)$ sans diviser $x + y$ ni $x - y$. Posons

$$a = \text{pgcd}(x + y, n) \quad \text{et} \quad b = \text{pgcd}(x - y, n).$$

Les entiers a et b sont distincts de 1 et n. Ce sont donc des diviseurs non triviaux de n. On peut les calculer avec l'algorithme d'Euclide, et l'on obtient ainsi la factorisation de n. En fait, on a $\{a, b\} = \{p, q\}$, car si par exemple $p = a = b$, alors $p = q$ vu que q divise ab.

6. Problème du sac à dos

Un randonneur disposant d'un sac à dos de volume V, souhaite le remplir de façon optimale avec k objets de volumes différents $v_0, \cdots, v_{k-1}$. Il est donc confronté au problème de trouver, s'il en existe, un sous-ensemble I de $\{0, \cdots, k-1\}$ de sorte que l'on ait

$$V = \sum_{i \in I} v_i.$$

1. Formalisation du problème

Problème (Sac à dos). Soient V un entier ≥ 1 et $\{v_0, \cdots, v_{k-1}\}$ un ensemble de k entiers naturels non nuls distincts deux à deux. S'il en existe, trouver un k-uplet d'entiers $(a_0, \cdots, a_{k-1})$ où les a_i valent 0 ou 1, tel que l'on ait l'égalité

$$V = \sum_{i=0}^{k-1} a_i v_i. \tag{9}$$

Le problème est donc de trouver un entier n, dont l'écriture en base 2 est

$$n = (a_{k-1} a_{k-2} \cdots a_1 a_0)_2,$$

tel que l'égalité (9) soit satisfaite. Ce problème peut ne pas avoir de solution, en avoir une seule, ou bien plusieurs. Il est en général très difficile à résoudre, et on ne connaît pas

d'algorithme permettant d'y parvenir «en un temps raisonnable». L'efficacité de certains cryptosystèmes est basée sur ce fait.

2. Sac à dos super croissant

Un cas particulier du problème du sac à dos est celui du sac à dos super croissant. Dans ce cas, les v_i étant rangés par ordre croissant

$$0 < v_0 < v_1 \cdots < v_{k-1}, \tag{10}$$

la condition supplémentaire suivante est satisfaite : on a

$$v_i > \sum_{j=0}^{i-1} v_j \quad \text{pour tout } i \text{ tel que} \quad 0 \leq i \leq k-1. \tag{11}$$

Si les conditions (10) et (11) sont satisfaites, on dit que le système $(v_0, \cdots, v_{k-1})$ est super croissant. Il s'agit alors de résoudre le problème du sac à dos correspondant.

Contrairement au problème général, celui-ci est facile à résoudre. On procède comme suit. Soient V un entier naturel non nul et $(v_0, \cdots, v_{k-1})$ un système super croissant. Supposons qu'il existe une solution au problème, autrement dit qu'il existe un sous-ensemble I de $\{0, \cdots, k-1\}$ tel que

$$V = \sum_{i \in I} v_i.$$

On détermine, «en observant les v_i de façon décroissante», le premier qui soit inférieur ou égal à V, autrement dit, le plus grand des v_i plus petit que V. Notons le v_{i_1}. Compte tenu de la condition (11), i_1 est dans I. On remplace alors V par $V - v_{i_1}$, et on repère à nouveau le plus grand des v_i plus petit que $V - v_{i_1}$. Si v_{i_2} est cet entier, alors i_2 est aussi dans I. On recommence exhaustivement ce processus jusqu'à obtenir l'indice t tel que $V - (v_{i_1} + v_{i_2} + \cdots + v_{i_t})$ soit nul, auquel cas, on a $I = \{i_1, i_2, \cdots, i_t\}$. Il en résulte que si le problème a une solution, alors celle-ci est unique, et dans ce cas elle s'obtient de façon systématique par l'algorithme prédédent.

Remarque 4.1. Afin de se construire un système super croissant, on peut choisir k entiers naturels non nuls $z_0, \cdots, z_{k-1}$ et définir

$$v_0 = z_0 \quad \text{et} \quad v_i = z_i + v_{i-1} + v_{i-2} + \cdots + v_0 \quad \text{pour } i = 1, \cdots, k-1. \tag{12}$$

Le système $(v_0, \cdots, v_{k-1})$ est alors super croissant.

Exemple 4.4. Le système $(4, 8, 19, 49, 111)$ est super croissant. Avec $V = 61$, le problème du sac à dos correspondant a l'unique solution $61 = 49 + 8 + 4$. Avec $V = 120$, il n'y a pas de solution.

7. Cryptosystème de Merkle-Hellman

Il repose sur le problème du sac à dos. Les unités de message à transmettre sont des entiers binaires ayant disons k composantes. Par exemple, si l'on utilise l'alphabet usuel de vingt six lettres A,$\cdots$,Z, chaque lettre est codée par un entier binaire ayant cinq composantes, de A$= (00000)_2$ jusqu'à Z$= (11001)_2$.

Chaque utilisateur de ce cryptosystème, disons Alice, choisit une suite d'entiers super croissante $(v_0, v_1, \cdots, v_{k-1})$, et deux entiers m et a, de sorte que

$$m > \sum_{i=0}^{k-1} v_i \quad \text{avec} \quad 1 \leq a < m \text{ et } \operatorname{pgcd}(a, m) = 1. \tag{13}$$

Conformément à la remarque 4.1, elle peut choisir pour cela $k+1$ entiers naturels non nuls z_i plus petits qu'une borne convenable, définir les v_i par la condition (12), et prendre

$$m = z_k + \sum_{i=0}^{k-1} v_i.$$

En choisissant aléatoirement un entier naturel $a_0 < m$, elle peut prendre pour a le plus petit entier plus grand que a_0 et premier avec m. Elle détermine ensuite l'entier b tel que

$$1 \leq b < m \quad \text{et} \quad ab \equiv 1 \text{ mod. } m, \tag{14}$$

et pour tout $i = 0, \cdots, k-1$, l'entier w_i tel que

$$1 \leq w_i < m \quad \text{et} \quad w_i \equiv av_i \text{ mod. } m. \tag{15}$$

La clé secrète d'Alice est $\big((v_0, v_1, \cdots, v_{k-1}), m, a, b\big)$. Sa clé publique est la suite

$$(w_0, \cdots, w_{k-1}).$$

Supposons que Bob souhaite envoyer un message binaire $P = (\varepsilon_{k-1} \cdots \varepsilon_0)_2$ à Alice. Pour cela, il transmet à Alice l'entier

$$C = \sum_{i=0}^{k-1} \varepsilon_i w_i. \tag{16}$$

Afin de décrypter C, Alice procède comme suit. Elle calcule l'entier V tel que

$$0 \leq V < m \quad \text{et} \quad V \equiv bC \text{ mod. } m.$$

On a l'égalité

$$V = \sum_{i=0}^{k-1} \varepsilon_i v_i. \tag{17}$$

En effet, d'après les conditions (14) et (15), on a

$$bC = \sum_{i=0}^{k-1} \varepsilon_i b w_i \equiv \sum_{i=0}^{k-1} \varepsilon_i v_i \text{ mod. } m.$$

On obtient ainsi

$$V \equiv \sum_{i=0}^{k-1} \varepsilon_i v_i \text{ mod. } m.$$

Par ailleurs, on a $0 \leq V < m$, et d'après l'inégalité (13), on a

$$0 \leq \sum_{i=0}^{k-1} \varepsilon_i v_i \leq \sum_{i=0}^{k-1} v_i < m,$$

d'où l'égalité (17). À l'aide de l'algorithme du sac à dos super croissant, appliqué avec la suite $(v_0, v_1, \cdots, v_{k-1})$ et l'entier V, Alice peut alors retrouver le message P.

Remarque 4.2. Un intrus voulant décrypter ce message est confronté au problème du sac à dos, qui n'est pas super croissant, avec l'entier C et la suite $(w_0, \cdots, w_{k-1})$. Le fait d'avoir remplacé v_i par le plus petit résidu de av_i modulo m, a détruit la propriété de super croissance initiale. Cela étant, le problème du sac à dos avec C et la suite des w_i est d'un type très particulier, vu qu'il provient d'un problème de sac à dos super croissant via une transformation simple. Shamir en 1982, a en fait trouver un algorithme permettant de résoudre ce type de problèmes de façon efficace. Le cryptosystème envisagé ici ne peut donc pas être considéré comme sûr.

Exemple 4.5. Prenons $m = 83$, $a = 21$ et $(v_0, v_1, v_2, v_3, v_4) = (3, 5, 10, 19, 45)$ comme système super croissant. On a donc $b = 4$. Ce sont les données secrètes d'Alice. On vérifie que l'on a

$$(w_0, w_1, w_2, w_3, w_4) = (63, 22, 44, 67, 32).$$

C'est la clé publique d'Alice. Supposons que Bob souhaite lui transmettre le mot OUI, les lettres étant codées en binaire entre 0 à 25 (A est codé par 0 et Z par 25). Puisque O est la quinzième lettre de l'alphabet, il est codé par $14 = (01110)_2$. De même U est codé par $20 = (10100)_2$, et I est codé par $8 = (01000)_2$. Le message qu'il veut faire parvenir à Alice est donc

$$P = (01110)_2(10100)_2(01000)_2.$$

Pour cela, il envoie à Alice le message (formule (16))

$$C = (133)(76)(67).$$

Posons $C_1 = 133$, $C_2 = 76$ et $C_3 = 67$. Afin de décrypter C, Alice calcule les entiers V_1, V_2, V_3 tels que

$$0 \leq V_i < 83 \quad \text{et} \quad V_i \equiv bC_i \text{ mod. } 83.$$

Elle obtient

$$V_1 = 34, \quad V_2 = 55 \quad \text{et} \quad V_3 = 19,$$

puis les égalités (sac à dos super croissant)

$$34 = 5 + 10 + 19, \quad 55 = 10 + 45 \quad \text{et} \quad 19 = 19.$$

Elle en déduit successivement que $(\varepsilon_4\varepsilon_3\varepsilon_2\varepsilon_1\varepsilon_0)_2$ est $(01110)_2$, $(10100)_2$, $(01000)_2$, et elle retrouve ainsi le message P.