Attaque du chiffrement affine

R3.09 Cryptographie
T. Godin, L. Naert, T. Ferragut
IUT de Vannes, Département Informatique

Rappels

Préparation du protocole

Alice choisit [1] $\mathbf{a}, \mathbf{b}$ deux nombres de $\mathbb{Z}/\mathbb{Z}_n$, avec a un inversible de $\mathbb{Z}/\mathbb{Z}_n$ (c'est-à-dire $a \in (\mathbb{Z}/\mathbb{Z}_n)^*$). Elle partages ces clefs (le modulo n peut être partagé ou public).

Échange d'un message

Quand **Alice** veut envoyer un message $\mathbf{m}$ à **Bob**, elle utilise leurs *clef partagée* $(\mathbf{a}, \mathbf{b})$ pour envoyer le message chiffré $\mathbf{c} \equiv \mathbf{m}.\mathbf{a} + \mathbf{b} \ \mathbf{mod}(\mathbf{n})$.

Bob reçoit le message $\mathbf{c}$ et le *déchiffre* en utilisant $\mathbf{m} \equiv (\mathbf{c} - \mathbf{b}) .\mathbf{a^{-1}} \ \mathbf{mod}(\mathbf{n}) \equiv \left(\mathbf{c}.\mathbf{a^{-1}}\right) + \mathbf{a^{-1}}.(-\mathbf{b}) \ \mathbf{mod}(\mathbf{n})$.

En effet, l'inverse modulaire a^{-1} de a est calculable en temps raisonnable ($O(\ln n^2)$, par exemple en utilisant l'algorithme d'Euclide, et l'opposé $(-b)$ est calculable (en temps constant sous des hypothèses raisonnables).

Bob obtient donc bien le message $\mathbf{m}$ qu'**Alice** souhaitait lui envoyer.

---

1. Dans tout le texte, les informations publiques (ou non sécurisées, ce qui est presque la même chose en cryptographie) seront écrites en vert, les informations privées seront en bleu, les informations partagées en cyan et les informations provenant d'une attaque seront en rouge. Les calculs (*a priori* effectués par un ordinateur) seront `notés sur fond gris`.

Attaque par fréquence

On se place dans une situation où un adversaire, **Oscar**, espionne **Alice** et **Bob**.

Oscar dispose des informations visibles sur le réseau, en particulier de la clef publique, et il voit passer les messages chiffrés. Il peut donc calculer les fréquences d'apparition des lettres dans le message chiffré. supposons que x_c soit la lettre apparaissant le plus fréquemment tandis que y_c est celle ayant la deuxième plus grande fréquence d'apparition.

L'attaque par de l'hypothèse que ces lettres les plus fréquentes dans le chiffré correspondent aux lettres les plus fréquentes dans la langue d'origine du texte [2], en français ce serait donc le `e` et le `a`.

Il cherche alors α, β tels que

$$\begin{cases} x_c\alpha + \beta & = x_m \mod n \\ y_c\alpha + \beta & = y_m \mod n \end{cases}$$

Le système se réécrit :

$$\begin{cases} (x_c - y_c)\,\alpha & = x_m - y_m \mod n \\ y_c\alpha + \beta & = y_m \mod n \end{cases}$$

On a très envie de poser $\boxed{\alpha = (\mathbf{x_m} - \mathbf{y_m}) \cdot (\mathbf{x_c} - \mathbf{y_c})^{-1}}$ mais $x_c - y_c$ n'est pas forcément inversible dans $\mathbb{Z}/\mathbb{Z}_n$. **Oscar** doit donc travailler un peu plus.

Plus précisément, il va devoir utiliser un petit lemme

Lemme *Si a et c sont premiers entre eux, que b est un inversible de $\mathbb{Z}/\mathbb{Z}_n$ et $ab \equiv c \mod (n)$ alors a et n sont premiers entre eux.*

preuve : On sait que "a et n sont premiers entre eux" équivaut à $\exists u, v \in \mathbb{Z}, au + nv = 1$.

Par hypothèse, a et c sont premiers entre eux, donc $\exists x, y \in \mathbb{Z}, ax + cy = 1$.

L'égalité modulaire peut se réécrire dans $\mathbb{Z}$ comme $a = cd + qn$ Injectons cela dans l'égalité :

$$\begin{aligned} & ab = c + qn \\ \iff\ & aby = cy + qny \\ \iff\ & aby = (1 - ax) + qny \\ \iff\ & ax + aby + ny = 1 \\ \iff\ & a(x + by) + ny = 1 \end{aligned}$$

On a donc trouver un couple $u = x + by$ et $v = y$ tel que $au + nv = 1$. Donc a et n sont premiers entre eux.

Oscar va donc chercher a obtenir des nombres premiers entre eux. Pour cela il calcule

$\boxed{\mathbf{d} = \mathbf{pgcd}\,(\mathbf{x_m} - \mathbf{y_m}, \mathbf{x_c} - \mathbf{y_c})}$ puis

$\boxed{\mathbf{z_m} = \frac{\mathbf{x_m} - \mathbf{y_m}}{\mathbf{d}}}$ et $\boxed{\mathbf{z_c} = \frac{\mathbf{x_c} - \mathbf{y_c}}{\mathbf{d}}}$.

La division est bien définie dans $\mathbb{Z}$ car d est un diviseur de $x_m - y_m$ et $x_c - y_c$ (par construction).

2. On suppose donc en plus que la langue d'origine est connue. Il a d'ailleurs été utilisé des langues rares pour "chiffrer" des message, notamment pendant la seconde guerre mondiale en utilisant des langues amérindiennes mais aussi le Basque. Voir `https://en.wikipedia.org/wiki/Code_talker` pour plus de détails

On a alors l'équation modulaire $z_c\alpha = z_m$, avec z_c et z_m premiers entre-eux (par construction). Donc, par le lemme, $pgcd(\alpha, z_m) = 1$ donc z_m est un inversible de $\mathbb{Z}/\mathbb{Z}_n$

On peut alors résoudre l'équation avec

$$\begin{cases} \alpha & = z_m.z_c^{-1} \mod n \\ \beta & = y_m - y_c\alpha \mod n \end{cases}$$

Un exemple

Oscar reçoit le message $m =$ `xegtekclqnqjdumpqdmroqggewqepqkmrxwknamcrqgucjxegmr`

Les lettre apparaissant le plus souvent sont **q** et **e** (les lettres **16** et **4**).

Oscar consulte `https://fr.wikipedia.org/wiki/Fr%C3%A9quence_d%27apparition_des_lettres` et suppose donc que $e \rightsquigarrow r$ et $f \rightsquigarrow e$, il écrit donc les équations dans $\mathbb{Z}/\mathbb{Z}_{26}$:

$$\begin{cases} 16\alpha + \beta & = 4 \mod n \\ 4\alpha + \beta & = 0 \mod n \end{cases}$$

D'où $\mathbf{16-4})\alpha \equiv (\mathbf{4-0}) \mod (\mathbf{26})$, c'est-à-dire

$$\mathbf{12}\alpha \equiv \mathbf{4} \mod (\mathbf{26})$$

Malheureusement pour **Oscar** il ne peut pas "diviser" par 12 directement dans $\mathbb{Z}/\mathbb{Z}_{26}$, il calcule donc $\mathbf{pgcd(12,4) = 4}$, et récrit l'équation

$$\mathbf{3}\alpha \equiv \mathbf{1} \mod (\mathbf{26})$$

Il calcule alors

$$\mathbf{3}\alpha = \mathbf{3^{-1}} = \mathbf{9}$$

(en effet $3.9 = 27 = 26 + 1$)

Il peut alors calculer β par

$$\begin{array}{rrl} & 4.9 + \beta & = 0 \mod n \\ \iff & 10 + \beta & = 0 \mod n \\ \iff & \beta & = 0 - 10 \mod n \\ \iff & \beta & = 16 \mod n \end{array}$$

Oscar essaie donc de décrypter le message avec $(\alpha, \beta) = (\mathbf{9}, \mathbf{16})$.

Il obtient le texte `pasfaciledetrouverunmessageavecunpgcdquinesoitpasun`