Indistinguabilité

Soit $\mathbf{G}$ un groupe cyclique d'ordre q premier engendré par g. Soit $\mathcal{P}$ la distribution uniforme sur $(\mathbb{Z}_q, \mathbf{G})$. Soit $\mathcal{P}_{DL}$ la distribution uniforme sur l'ensemble $\{(r, g^r), r \in \mathbb{Z}_q\}$. Soit $\mathcal{P}_{NDL}$ la distribution uniforme sur l'ensemble $\{(r, g^{r'}), r \neq r' \in \mathbb{Z}_q\}$. On rappelle l'expression de la distance statistique entre deux variables aléatoires discrètes sur un ensemble A dénombrable :

$$\Delta(X, Y) = \frac{1}{2} \sum_{a \in A} |\Pr[X = a] - \Pr[Y = a]| \quad .$$

Question 1. *$\mathcal{P}_{NDL}$ et $\mathcal{P}$ sont-elles calculatoirement indistinguables ? Statistiquement indistinguables ? Les deux ?*

Solution. *Les deux, car statistiquement indistinguables (on peut aller vite là-dessus car il n'y a pas eu de problème dans le DM). C'est une bonne occasion de rappeler que indistinguabilité statistique implique calculatoire. Statistique est plus forte, car c'est vraiment de la théorie de l'information.*

Pour démontrer l'indistinguabilité statistique : notons qu'il y a q paires dans DL et $q^2 - q$ paires dans NDL. Notons $X = \mathcal{P}$ et $Y = \mathcal{P}_{NDL}$.

$$\begin{aligned}
\Delta(X, Y) &= \frac{1}{2} \sum_{(a,b) \in DL} |\Pr[X = (a, b)] - \Pr[Y = (a, b)]| \\
&\quad + \frac{1}{2} \sum_{(a,b) \in NDL} |\Pr[X = (a, b)] - \Pr[Y = (a, b)]| \\
&= \frac{1}{2} q \left(\frac{1}{q^2} - 0 \right) + \frac{1}{2} (q^2 - q) \left(\frac{1}{q^2 - q} - \frac{1}{q^2} \right) \\
&= \frac{1}{2q} - \frac{1}{2} \frac{q^2 - q}{q^2} + \frac{1}{2} = \frac{1}{q} \quad .
\end{aligned}$$

Question 2. *$\mathcal{P}_{DL}$ et $\mathcal{P}$ sont-elles calculatoirement indistinguables ? Statistiquement indistinguables ? Les deux ?*

Solution. *Distinguables calculatoirement, et c'est tout bête, car il suffit de calculer l'exponentiation et de vérifier si ça concorde. L'objet de cette question est juste de montrer que si D et D' sont indistinguables, alors ça n'implique pas que $D' \backslash D$ et D le sont (où les distributions sont notées ici comme des ensembles).*

Cryptosystème de Okamoto-Uchiyama

Cryptosystème de Okamoto-Uchiyama

$\mathsf{KeyGen}(1^n)$:

- Choisir deux entiers premiers p, q tels que $p \mid (q-1)$.
- Définir $N = p^2 q$.
- Choisir un générateur $g \in \mathbb{Z}_N^*$ tel que $g^{p-1} \neq 1 \mod p^2$
- Définir $h = g^N \mod N$.
- Clé publique : $\mathsf{pk} = (N, g, h)$; clé privée : $\mathsf{sk} = (p, q)$

$\mathsf{Enc}(\mathsf{pk}, m)$ $(m < p)$:

- $r \leftarrow U(\mathbb{Z}_N^*)$
- $c := g^m \cdot h^r \mod N$
- Renvoyer c

Dec ...

Soit $\Gamma = \{x \in \mathbb{Z}_{p^2}^*, x = 1 \mod p\}$.

Question 3. *Montrer que Γ est un sous-groupe de $(\mathbb{Z}_{p^2})^*$ d'ordre p.*

Solution. *Trivial (on fait ça vite).*

Soit la fonction $L : \mathbb{Z}_{p^2}^* \to \mathbb{Z}_p$ définie par :

$$L(x) := \frac{x-1}{p} \mod p.$$

Question 4. *Montrer que L est un isomorphisme entre Γ et le groupe additif $\mathbb{Z}_p$. En déduire que le problème du logarithme discret est facile dans Γ : sur une entrée $(x, y) \in \Gamma$ avec $L(x) \neq 0$ et $y = x^m \mod p^2$, on peut calculer efficacement m.*

Solution. *On fait ça vite aussi. Le calcul du DL : $m = L(y)/L(x) = (y-1)/(x-1) \mod p$.*

Question 5. *Montrer que $(h^r)^{p-1} = 1 \mod p^2$. En déduire l'algorithme de déchiffrement.*

Solution. *On a : $(h^r)^{p-1} = g^{Nr(p-1)} = (g^{p(p-1)})^{rpq} \mod p^2 = 1 \mod p^2$.*
L'algorithme de déchiffrement est donc de calculer $c^{p-1} = (g^{p-1})^m \mod p^2$ puis son logarithme discret en base g^{p-1}.

On va montrer la sécurité IND-CPA sous l'hypothèse :

> Pour $h = g^N \mod N$, la distribution $\{h^r \mod N, r \leftarrow U(\mathbb{Z}_N)\}$ (SUBGROUP) et la distribution $\{gh^{r'} \mod N, r' \leftarrow U(\mathbb{Z}_N)\}$ (RANDOM) sont calculatoirement indistinguables.

En d'autres termes, cette hypothèse suppose que le chiffrement de 0 et 1 sont indistinguables.

Question 6. *Montrer qu'étant donné une paire de messages (m_0, m_1) et $c = \mathsf{Enc}(\mathsf{pk}, b)$ où b est un bit inconnu, on peut facilement calculer un chiffré $c^* = \mathsf{Enc}(\mathsf{pk}, m_b)$ aléatoire valide.*

Solution. *On calcule :*

$$c^* = c^{m_1 - m_0} g^{m_0} \mod N$$

Si $c = \mathsf{Enc}(\mathsf{pk}, 0)$, *c est de la forme* $h^r \mod N$. *Donc* $c^* = h^{r(m_1-m_0)} g^{m_0} \mod N$ *est un chiffré valide de* m_0.

Si $c = \mathsf{Enc}(\mathsf{pk}, 1)$, *c est de la forme* $h^r g \mod N$. *Donc* $c^* = h^{r(m_1-m_0)} g^{m_1-m_0} g^{m_0} \mod N$ *est un chiffré valide de* m_1.

Pour finir on re-randomise le chiffré en multipliant par $h^{r'}$ *où* r' *est tiré uniformément au hasard (je ne suis pas sûr que ce soit complètement nécessaire, mais en tout cas ça nous assure de l'uniformité pour le chiffré final).*

Question 7. *Montrer la sécurité IND-CPA.*

Solution. *On part d'un adversaire IND-CPA* $\mathcal{A}$ *et on construit un distingueur* $\mathcal{D}$ *dans le jeu de distinguer entre les deux distributions données.*

Le distingueur $\mathcal{D}$ *interagit avec un challenger* $\mathcal{C}$, *selon les étapes suivantes :*

- $\mathcal{C}$ *choisit un entier* N *et un bit* b *u.a.r. : si* $b = 0$ *il prend le cas SUBGROUP, sinon le cas RANDOM*
- $\mathcal{D}$ *envoie des requêtes à* $\mathcal{C}$ *qui répond avec la distribution donnée*
- $\mathcal{D}$ *renvoie un bit* b'

On implémente $\mathcal{D}$ *en utilisant, de manière interne, l'algorithme* $\mathcal{A}$. *Rappelons que l'algorithme* $\mathcal{A}$ *doit « voir » un jeu IND-CPA, selon les étapes suivantes :*

- *Recevoir une clé publique du challenger*
- *Choisir deux messages* m_0, m_1, *les envoyer au challenger*
- *Recevoir un chiffré challenge* c^*
- *Renvoyer un bit* b''

Voici donc la manière dont nous implémentons $\mathcal{D}$ *:*

- $\mathcal{D}$ *reçoit la clé publique* N *de* $\mathcal{C}$ *et la transfère à* $\mathcal{A}$
- $\mathcal{D}$ *envoie une requête à* $\mathcal{C}$ *et récupère* $\mathsf{Enc}(\mathsf{pk}, 0)$ *(dans le cas SUBGROUP) et* $\mathsf{Enc}(\mathsf{pk}, 1)$ *(dans le cas RANDOM), en d'autres termes* $c = \mathsf{Enc}(\mathsf{pk}, b)$
- $\mathcal{A}$ *choisit deux messages* m_0, m_1 *et les envoie à* $\mathcal{D}$
- $\mathcal{D}$ *calcule* c^* *comme expliqué à la question précédente. Ici* $c^* = \mathsf{Enc}(\mathsf{pk}, m_b)$ *est bien un chiffré valide (et tiré au hasard parmi les chiffrés possibles), donc du point de vue de* $\mathcal{A}$ *rien n'a changé*
- $\mathcal{A}$ *renvoie un bit* b''
- $\mathcal{D}$ *renvoie* $b' = b''$

On a :

$$\Pr[\mathcal{D} \text{ wins Dist.}] = \Pr[b' = b] = \Pr[b'' = b] = \Pr[\mathcal{A} \text{ wins IND-CPA}] \ .$$

Sous l'hypothèse ci-dessus, pour tout distingueur $\mathcal{D}$ *:* $|\Pr[\mathcal{D} \text{ wins Dist.}] - 1/2| = \text{negl}$, *donc c'est vrai aussi pour tout adversaire* $\mathcal{A}$ *contre IND-CPA ; le schéma est donc IND-CPA-sûr.*

La preuve n'est pas tout à fait équivalente à ElGamal (où on a un facteur 1/2 qui apparaît) mais elle est intéressante.

Entrée : $x, N, e = \sum_{i=0}^{k-1} e_i 2^i$
Sortie : $x^e \mod N$

$R_0 \leftarrow 1, R_1 \leftarrow x$
for $i = k-1, k-2, \ldots, 0$ **do**
  if $e_i = 1$ **then**
    $R_0 \leftarrow R_0 \cdot R_1 \mod N$
    $R_1 \leftarrow R_1 \cdot R_1 \mod N$
  else
    $R_1 \leftarrow R_1 \cdot R_0 \mod N$
    $R_0 \leftarrow R_0 \cdot R_0 \mod N$
  end if
end for
Return R_0

Montgomery power ladder

On considère l'agorithme suivant (*Montgomery power ladder*).

Question 8. *Montrer que l'algorithme est correct. Quel est son intérêt ?*

Solution. *Soit* $f_i = \sum_{j=i}^{k-1} e_j 2^{j-i}$. *Alors avant la boucle* $R_0 = x^{f_{k-1}}$ *et* $R_1 = xR_0$.

L'hypothèse de récurrence est que à chaque itération de la boucle : $R_0 = x^{f_i}$ *et* $R_1 = x^{f_i+1}$.

En effet, si $e_i = 1$ *alors la mise à jour est :*

$$\begin{cases} R_0 \leftarrow R_0 R_1 = x^{2f_i+1} \\ R_1 \leftarrow R_1^2 = x^{2f_i+2} \end{cases} \tag{1}$$

si $e_i = 0$ *alors :*

$$\begin{cases} R_1 \leftarrow R_0 R_1 = x^{2f_i+1} \\ R_0 \leftarrow R_0^2 = x^{2f_i} \end{cases} \tag{2}$$

D'où le résultat par récurrence triviale.

L'intérêt est d'avoir un algorithme en temps constant, qui contient le même nombre d'opérations quel que soit l'exposant e.