On rappelle qu'on utilise la notation $x \leftarrow U(\mathcal{X})$ pour signifier que x est tiré uniformément au hasard dans $\mathcal{X}$. La difficulté des questions posées ci-dessous est, elle, non uniforme. Le barême final en tiendra compte.

Résistance aux collisions

Soit G un groupe cyclique d'ordre q premier et g un générateur. On considère une famille de fonctions de hachage définie par exponentiation dans le groupe G, dont on va prouver la résistance aux collisions en supposant la difficulté du problème du logarithme discret. On rappelle qu'une *collision* est une paire d'entrées *distinctes* ayant la même sortie.

On considère le jeu suivant $\mathcal{G}$ entre un challenger $\mathcal{C}$ et un adversaire $\mathcal{A}$:

- Le challenger $\mathcal{C}$ choisit $a \leftarrow U(\mathbb{Z}_q)$ et envoie $h := g^a$ à $\mathcal{A}$
- La valeur h définit une fonction de hachage $H : \mathbb{Z}_q^2 \to G$ par $H(x, y) := g^x h^y$
- $\mathcal{A}$ renvoie $(x, y), (x', y')$ et gagne le jeu s'il s'agit d'une collision de H

L'avantage de $\mathcal{A}$ est simplement défini par sa probabilité de gagner au jeu.

Question 1. *Soit $(x, y), (x', y')$ une collision de H. Montrer qu'on peut en déduire a efficacement.*

Solution. *Il suffit d'écrire l'égalité :*

$$g^x h^y = g^{x'} h^{y'} \implies g^{x-x'} = h^{y-y'} \implies h = g^{(x-x')(y-y')^{-1}}$$

ce qui donne $a = (x - x')(y - y')^{-1} \bmod q$.

Notons qu'on a supposé que $y \neq y'$, est-ce vrai ? Oui, car si on a $y = y'$ alors cela implique $x = x'$ dans $\mathbb{Z}_q$ (car $g^{x-x'} = 1$ et g est un générateur du groupe), on n'aurait donc pas une collision.

Question 2. *Écrire un jeu de sécurité $\mathcal{G}'$, joué entre un adversaire et un challenger, correspondant au problème du logarithme discret, et définir l'avantage de l'adversaire dans ce jeu.*

Solution. *Le jeu de sécurité $\mathcal{G}'$ est joué entre un adversaire $\mathcal{A}$ et un challenger $\mathcal{C}$. On décrit toutes les communications entre $\mathcal{A}$ et $\mathcal{C}$.*

1. *$\mathcal{C}$ choisit a, g, calcule g^a et l'envoie à $\mathcal{A}$*
2. *$\mathcal{A}$ renvoie une valeur a' et gagne si $g^{a'} = g^a$. Notez que c'est équivalent à $a' = a$ quand on travaille dans un groupe cyclique, donc les deux réponses sont possibles.*

L'avantage de $\mathcal{A}$ est sa probabilité de gagner au jeu. On peut aussi raffiner la définition en regardant $|\Pr[\mathcal{A}win] - 1/(q-1)|$, $1/(q-1)$ étant la probabilité de gagner quand on répond au hasard. Mais les deux définitions sont équivalentes, car le terme $1/(q-1)$ est négligeable (en effet, si on note n le paramètre de sécurité, c'est la taille de q, donc q est exponeniel en n).

Question 3. *Soit $\mathcal{A}$ un adversaire dans le jeu $\mathcal{G}$. Construire un adversaire $\mathcal{B}$ jouant au jeu $\mathcal{G}'$ et exprimer son avantage en fonction de celui de $\mathcal{A}$. Conclure.*

Solution. *On construit l'adversaire $\mathcal{B}$ pour le DLOG de la manière suivante.*

$\mathcal{B}$ fait jouer $\mathcal{A}$. Il lui envoie $h = g^a$. Ensuite $\mathcal{A}$ renvoie sa collision $(x, y), (x', y')$. $\mathcal{B}$ calcule un candidat pour le DLOG avec la formule de la question 1.

De son point de vue, $\mathcal{A}$ joue au jeu $\mathcal{G}$. Lorsque $\mathcal{A}$ renvoie une collision, $\mathcal{B}$ trouve a, et donc gagne au jeu $\mathcal{G}'$ du DLOG. Par conséquent, la probabilité que $\mathcal{B}$ gagne à $\mathcal{G}'$ est supérieure à la probabilité que $\mathcal{A}$ gagne à $\mathcal{G}$.

onc $\mathrm{Adv}(\mathcal{B}) \geq \mathrm{Adv}(\mathcal{A})$.

La conclusion est que si DL est difficile dans $\mathcal{G}$, la famille de fonctions de hachage est résistante aux collisions.

Auto-réducibilité aléatoire

Soit G un groupe cyclique d'ordre q premier et g un générateur. On suppose qu'il existe un algorithme $\mathcal{A}$ prenant en entrée un élément de G et renvoyant son logarithme discret, qui réussit avec probabilité ε :

$$\Pr_{u \leftarrow U(G)} [\mathcal{A}(u) = DL(u)] = \varepsilon \ .$$

Où cette probabilité est sur le choix uniforme de u et les choix aléatoires dans $\mathcal{A}$.

Question 4. *Soit $F \subseteq G$ l'ensemble des u tels que $\Pr[\mathcal{A}(u) = DL(u)] = 0$, c'est-à-dire que $\mathcal{A}$ échoue totalement sur l'entrée u. Quelle est la taille maximale de F ?*

Solution. *(Sketch) Proportion $1 - \varepsilon$ du groupe G.*

Question 5. *Soit $u \in G$ fixé. On tire $\sigma \leftarrow U(\mathbb{Z}_q)$. Justifier rapidement que $u \cdot g^\sigma$ est distribué uniformément au hasard dans G.*

Solution. *(Sketch) Soit r tq $u = g^r$. Alors $u \cdot g^\sigma = g^{\sigma + r}$ et $\sigma + r$ est uniforme, donc $u \cdot g^\sigma$ est uniforme.*

Question 6. *En déduire un algorithme $\mathcal{B}$ tel que, pour tout $u \in G$:*

$$\Pr[\mathcal{B}(u) = DL(u)] = \varepsilon$$

où la probabilité est prise sur les choix aléatoires dans $\mathcal{B}$ uniquement.

Solution. *L'algorithme consiste à tirer σ au hasard puis à lancer $\mathcal{A}$, et ensuite à corriger le DL qui est sorti. Alors* $\Pr[\mathcal{B}(u) = DL(u)] = \Pr_{u' \leftarrow U(G)}[\mathcal{A}(u') = DL(u')] = \varepsilon$.

J'ai vu l'algorithme suivant : on tire des dlogs au hasard et on en teste le nombre exact tel que la probabilité de réussite soit ε. Je confirme que c'est techniquement correct. Ce n'est bien entendu par du tout ce qui était demandé (l'algorithme obtenu étant très inefficace).

On a ainsi démontré une réduction du *pire cas* au *cas moyen* pour le problème du logarithme discret.

Problème de domaine dans ElGamal

Soient p, q des nombres premiers tels que $q := (p-1)/2$. On sait que $\mathbb{Z}_p^*$ est d'ordre $p-1 = 2q$ et contient donc un unique sous-groupe d'ordre q noté G. On admet que ce sous-groupe est l'ensemble des éléments de $\mathbb{Z}_p^*$ qui sont des carrés modulo p, i.e., de la forme g^a où g est un générateur de $\mathbb{Z}_p^*$ et a est pair.

On considère la version suivante du cryptosystème ElGamal.

Cryptosystème ElGamal dans $\mathbb{Z}_p^*$

KeyGen :

- Choisir $a \leftarrow U(\mathbb{Z}_p^*)$ (différent de 1) et calculer $g = a^2 \bmod p$ un générateur de G
- Choisir $x \leftarrow U(\mathbb{Z}_q)$ et calculer $h = g^x \bmod p$
- $\mathsf{pk} = h$, $\mathsf{sk} = x$

$\mathsf{Enc}(\mathsf{pk}, m)$ où $m \in \mathbb{Z}_p^*$:

- Choisir $r \in \mathbb{Z}_q$
- Renvoyer $(g^r \bmod p, m \cdot h^r \bmod p)$

Question 7. *Montrer qu'il existe un algorithme efficace qui, étant donné un élément de $\mathbb{Z}_p^*$, détermine si celui-ci est un carré modulo p.*

Indice : utiliser une exponentiation.

Solution. *On propose d'utiliser une exponentiation.*

Si a est un carré mod p ssi il est de la forme $g^{2x'}$ ssi $a^{(p-1)/2} = 1 \mod p$. Sinon on aurait $g^{(2x'+1)(p-1)/2} = 1 \mod p \implies g^{(p-1)/2} = 1 \mod p$ impossible car g est un générateur de $\mathbb{Z}_p^*$.

Autre preuve : $g^{aq} = 1$ implique que l'ordre de g divise aq ssi $2q$ divise aq ssi 2 divise a.

Question 8. *Montrer que ce schéma n'est pas IND-CPA, et expliquer pourquoi la preuve de sécurité de ElGamal (vue en cours) ne fonctionne pas ici.*

Solution. *On chiffre deux messages dont l'un est un carré et l'autre un non-carré. On peut donc distinguer les chiffrés.*

Le problème c'est que l'hypothèse DDH (nécessaire pour la preuve de ElGamal) n'est pas vérifiée dans le groupe $\mathbb{Z}_p^*$, car on peut différencier les éléments qui sont des carrés ou qui ne le sont pas (et donc distinguer grâce à la parité des exposants).

Pour sauver le cryptosystème il faudrait en fait travailler dans le sous-groupe des carrés modulo p.

Question 9. *On suppose maintenant que m est encodé comme une chaîne de bits de même taille que p : $m \in \{0,1\}^{|p|}$. On modifie la fonction de chiffrement en :*

$$\mathsf{Enc}(\mathsf{pk}, m) = (g^r \bmod p, m \oplus (h^r \bmod p))$$

où $\oplus$ est un XOR de représentations binaires. Le schéma est-il IND-CPA ?

Solution. *Les éléments g et h sont des carrés mod p. Donc si on chiffre $m = 0$ on a toujours un carré mod p dans la deuxième composante, tandis que si on chiffre $m = 1$ on a un nombre qui est un carré mod p avec proba 1/2 (proba de tomber dans le sous-groupe). Donc on arrive à distinguer avec bonne probabilité.*

Chiffrement à clé publique anonyme

Soit G un groupe cyclique d'ordre q premier et g un générateur. On suppose que le problème DDH est difficile dans G.

On considère les trois jeux de sécurité suivants $\mathcal{G}_0, \mathcal{G}_1, \mathcal{G}_2$ joués entre un challenger et un distingueur $\mathcal{D}$. Dans les trois jeux, le challenger tire $r, x, y, t \leftarrow U(\mathbb{Z}_q)$.

- Dans $\mathcal{G}_0$, il envoie (g^r, g^x, g^y, g^{xr})
- Dans $\mathcal{G}_1$, il envoie (g^r, g^x, g^y, g^t)
- Dans $\mathcal{G}_2$, il envoie (g^r, g^x, g^y, g^{yr})

Le distingueur renvoie ensuite un bit b.

Question 10. *Montrer que, sous l'hypothèse DDH dans G, les paires de jeux $(\mathcal{G}_0, \mathcal{G}_1)$ et $(\mathcal{G}_1, \mathcal{G}_2)$ sont indistinguables. Qu'en est-il de $(\mathcal{G}_0, \mathcal{G}_2)$?*

Solution. *Soit $\mathcal{D}$ un distingueur pour les paires de jeux $(\mathcal{G}_0, \mathcal{G}_1)$. On construit un distingueur $\mathcal{D}'$ pout DDH de la manière suivante :*

- *$\mathcal{D}'$ utilise $\mathcal{D}$*
- *Sur le triplet venant du challenger $g^r, g^r, g^{xr}/g^t$, $\mathcal{D}'$ insère un g^y tiré uniformément au hasard et l'envoie à $\mathcal{D}$*
- *Lorsque $\mathcal{D}$ renvoie un bit b, $\mathcal{D}'$ renvoie b*

On a donc :

$$\begin{cases} \Pr\left[\mathcal{D} \xleftarrow{\mathcal{G}_1} 1\right] = \Pr\left[\mathcal{D}' \xleftarrow{RAND} 1\right] \\ \Pr\left[\mathcal{D} \xleftarrow{\mathcal{G}_0} 1\right] = \Pr\left[\mathcal{D}' \xleftarrow{DDH} 1\right] \end{cases}$$

Et par conséquent :

$$\mathrm{Adv}(\mathcal{D}') = \mathrm{Adv}(\mathcal{D})$$

Où on prend l'avantage de $\mathcal{D}'$ pour distinguer entre RAND et DDH et l'avantage de $\mathcal{D}$ pour distinguer les jeux $(\mathcal{G}_0, \mathcal{G}_1)$.

La preuve pour les jeux $(\mathcal{G}_1, \mathcal{G}_2)$ est quasiment identique, on change juste l'endroit où on insère le nouvel élément en transformant le triplet en quadruplet.

On a donc montré que sous l'hypothèse DDH, pour tout adversaire PPT $\mathcal{D}$:

$$\begin{cases} |\Pr\left[\mathcal{D} \xleftarrow{\mathcal{G}_1} 1\right] - \Pr\left[\mathcal{D} \xleftarrow{\mathcal{G}_0} 1\right]| = \mathsf{negl} \\ |\Pr\left[\mathcal{D} \xleftarrow{\mathcal{G}_2} 1\right] - \Pr\left[\mathcal{D} \xleftarrow{\mathcal{G}_1} 1\right]| = \mathsf{negl} \end{cases}$$

On conclut par l'inégalité triangulaire : les jeux $\mathcal{G}_0, \mathcal{G}_2$ sont indistinguables sous l'hypothèse DDH.

On considère le jeu suivant $\mathcal{G}$ joué entre un challenger $\mathcal{C}$ et un adversaire $\mathcal{A}$.

1. Le challenger tire $r, x, y \leftarrow U(\mathbb{Z}_q)$ et un bit $b \in \{0, 1\}$
2. Si $b = 0$, le challenger envoie (g^r, g^x, g^y, g^{xr}), sinon il envoie (g^r, g^x, g^y, g^{yr})
3. L'adversaire renvoie un bit b' et gagne si $b = b'$

Question 11. *Déduire de la question précédente que, sous l'hypothèse DDH, tout adversaire PPT a un avantage négligeable dans le jeu $\mathcal{G}$.*

Solution. *Il suffit de remarquer que les cas $b = 0$ et $b = 1$ correspondent aux jeux $\mathcal{G}_0$ et $\mathcal{G}_2$. On a alors parfaitement le droit d'invoquer le résultat d'équivalence des notions d'indistinguabilité à un ou deux jeux vu au tout premier cours (!). Ici, les deux distributions sont (g^r, g^x, g^y, g^{xr}) et $g(^r, g^x, g^y, g^{yr})$. Comme tout distinguer PPT a un avantage négligeable pour distinguer $(\mathcal{G}_0, \mathcal{G}_2)$, tout adversaire a aussi un avantage négligeable pour gagner au jeu $\mathcal{G}$.*

On peut aussi refaire la preuve en direct. Il peut être utile de rappeler le calcul. On prend le même algorithme $\mathcal{D}$ pour les deux définitions et on relie son avantage dans les deux définitions :

$$\begin{aligned}
\mathrm{Adv}^{\mathcal{G}_0,\mathcal{G}_1}(\mathcal{D}) &= \left|\Pr\left[\mathcal{D} \xrightarrow{G_0} 1\right] - \Pr\left[\mathcal{D} \xrightarrow{G_1} 1\right]\right| \quad \textit{(dans la paire de jeux)} \\
&= \left|\Pr_{\mathcal{G}}\left[\mathcal{D} \to 1 | b = 0\right] - \Pr_{\mathcal{G}}\left[\mathcal{D} \to 1 | b = 1\right]\right| \quad \textit{(dans le jeu } \mathcal{G}\textit{)} \\
&= |\Pr\left[b' = 1 | b = 0\right] - \Pr\left[b' = 1 | b = 1\right]| \\
&= |1 - \Pr\left[b' = 0 | b = 0\right] - \Pr\left[b' = 1 | b = 1\right]| \\
&= |1 - 2\Pr\left[b' = 0 \wedge b = 0\right] - 2\Pr\left[b' = 1 \wedge b = 1\right]| \\
&= |1 - 2\Pr_{\mathcal{G}}\left[Win\right]| = \frac{1}{2}\mathrm{Adv}^{\mathcal{G}}(\mathcal{D}) \quad .
\end{aligned}$$

On considère le cryptosystème ElGamal (vu en cours). Plusieurs utilisateurs ont publié leur clé publique. Alice veut envoyer un message sans qu'il soit possible de déterminer à quelle personne il a été envoyé. On formalise cette propriété d'*anonymité* par le jeu de sécurité suivant entre un challenger $\mathcal{C}$ et un adversaire $\mathcal{A}$.

- $\mathcal{C}$ produit deux paires de clés $(\mathsf{pk}_0, \mathsf{sk}_0)$ et $(\mathsf{pk}_1, \mathsf{sk}_1)$, et un bit $b \in \{0, 1\}$
- $\mathcal{A}$ sélectionne un message m et l'envoie à $\mathcal{C}$
- $\mathcal{C}$ renvoie $\mathsf{Enc}(\mathsf{pk}_b, m)$
- $\mathcal{A}$ renvoie un bit b' et gagne si $b = b'$

Question 12. *En utilisant la question précédente, montrer que le chiffrement ElGamal est anonyme.*

Solution. *Soit $\mathcal{A}$ un adversaire pour la propriété d'anonymité. On fabrique un adversaire $\mathcal{A}'$ pour le jeu $\mathcal{G}$ de la manière suivante. $\mathcal{A}'$ agit comme challenger dans le jeu d'anonymité et exécute $\mathcal{A}$.*

- *$\mathcal{A}'$ reçoit un triplet qui est soit $(g^r, g^x, g^y, t = g^{xr})$ ou $(g^r, g^x, g^y, t = g^{yr})$.*
- *$\mathcal{A}$ sélectionne un message m et l'envoie à $\mathcal{A}'$*
- *$\mathcal{A}'$ envoie g^x, g^y à $\mathcal{A}$ (j'ai oublié de le mettre dans ma définition, mais oui les clés publiques doivent être connues de $\mathcal{A}$)*
- *$\mathcal{A}'$ calcule $(g^r, t \cdot m)$ et l'envoie à $\mathcal{A}$*
- *$\mathcal{A}$ renvoie un bit b' et $\mathcal{A}'$ renvoi ce bit également*

Alors l'avantage de $\mathcal{A}'$ dans le jeu d'anonymité est le même que celui de $\mathcal{A}$ dans le jeu $\mathcal{G}$. Par conséquent tout adversaire PPT a un avantage négligeable dans le jeu d'anonymité sous l'hypothèse DDH.

Question 13. *On considère le cryptosystème* padded RSA *vu en cours, avec une génération de clés utilisant deux nombres premiers aléatoires de ℓ bits exactement (c'est-à-dire $2^{\ell-1} \leq P, Q < 2^\ell$), et un chiffrement de la forme $\mathsf{Enc}(x) = (r\|x)^e \bmod N$ où $\|$ indique une concaténation de bit-strings entre le clair x et la valeur aléatoire r.*

Justifier que padded RSA n'est pas anonyme.

Solution. *C'est à cause de la taille des nombres.*

Une preuve exacte ferait intervenir la distribution des nombres premiers, on ne rentrera pas dans de tels détails. Supposons qu'il existe une probabilité au moins inverse-polynomiale en ℓ (c'est le cas) que P, Q soient plus grands que $2^{\ell-1}(2 - 1/4)$ et une probabilité au moins inverse-polynomiale qu'ils soient plus petits que $2^{\ell-1}(1 + 1/4)$.

Alors dans le premier cas, on a $N \geq 2^{(2\ell-2)}(49/16) := N_g$ et dans le deuxième cas, on a $N \leq 2^{(2\ell-2)}(25/16) := N_l$.

Dans les deux cas, le message est tiré uniformément au hasard dans $[0; N]$. Voici donc notre adversaire :

- *Connaissant les deux clés publiques N_1, N_2, si on n'a pas $N_1 \leq 2^{(2\ell-2)}(25/16)$ et $N_2 \geq 2^{(2\ell-2)}(49/16)$ alors on répond au hasard ;*
- *Si le chiffré est dans l'intervalle $[N_g; +\infty]$ on répond 2 ;*
- *Sinon on répond au hasard.*

Il suffit de constater que le cas dans lequel on ne répond pas au hasard arrive avec probabilité au moins inverse-poly. En effet, c'est le cas des conditions sur les clés, et de plus, comme $N \leq 2^{2\ell}$, il y a un overlap constant entre l'intervalle $[N_g; N]$ et l'intervalle $[0; N]$, donc en chiffrant avec N_2 on a probabilité constante de tomber dans cet intervalle. Dans ce cas, l'adversaire réussit avec probabilité 1.

C'est l'argument central de cette preuve heuristique : il suffit de montrer qu'il existe un cas qui n'est pas rare (arrive avec proba non négligeable) dans lequel on distingue avec proba 1.

Un peu de cryptographie symétrique

On considère un chiffrement symétrique (un « chiffrement à bloc », que nous verrons en cours) utilisant une clé k de n bits, une permutation *publique* $\Pi : \{0,1\}^n \to \{0,1\}^n$, et avec des clairs / chiffrés de n bits. La construction est :

$$E_k(x) = \Pi(k \oplus x) \oplus k \ .$$

On essaie de récupérer la clé k dans une attaque à *clair choisi* où l'attaquant peut obtenir le chiffré sur toute entrée de son choix.

Question 14. *Montrer comment attaquer efficacement le chiffrement si :*

- *On enlève le premier XOR de clé ;*
- *On enlève le deuxième XOR de clé ;*
- *On utilise comme permutation une fonction linéaire, de la forme : $\Pi(y) = My$ où M est une matrice Booléenne inversible aléatoire de dimensions $n \times n$.*

Solution. *La permutation est publique donc on est capable de l'inverser (oui, ce n'était pas clair dans l'énoncé, désolé).*

Les deux attaques en enlevant le XOR de clé sont triviales.

Avec une fonction linéaire, on veut résoudre un système de la forme :

$$y = M(k \oplus x) \oplus k \implies y = Mx \oplus (M \oplus I)k \ .$$

Est-ce que $M \oplus I$ est inversible ? Pas forcément (par exemple si $M = I$). Cela dit comme on a supposé M aléatoire, avec probabilité constante la matrice $M \oplus I$ (qui est

aussi aléatoire) sera inversible. Si elle ne l'est pas, cela veut en fait dire que le chiffrement ne dépend pas d'une partie de la clé.

On peut supposer que la matrice est inversible et on retrouve tout k d'un coup. Sinon, on résout le système linéaire et on retrouve une partie de k, et pour le reste, on fait une recherche exhaustive.