Des éléments de solution sont affichés en vert.

Théorème de Shannon

Le but de cet exercice est de prouver le résultat suivant de Shannon.

Theorem 1. *Soit* $\mathsf{KeyGen}, \mathsf{Enc}, \mathsf{Dec}$ *un chiffrement symétrique tel que* $|\mathcal{M}| = |\mathcal{C}| = |\mathcal{K}|$*. Le schéma est parfaitement sûr si et seulement si :*

1. *Toute clé* $k \in \mathcal{K}$ *est choisie avec probabilité* $1/|\mathcal{K}|$ *par* KeyGen
2. *Pour tout* $m \in \mathcal{M}$ *et tout* $c \in \mathcal{C}$*, il existe une unique clé* $k \in \mathcal{K}$ *telle que* $\mathsf{Enc}(m, k) = c$*.*

D'abord, nous justifions que l'hypothèse faite sur les espaces est raisonnable.

Question 1. *On rappelle la définition de la sécurité parfaite : pour toute variable aléatoire* M*,* m *et* c *:*

$$\Pr[M = m | \mathsf{Enc}(\mathsf{KeyGen}, M) = c] = \Pr[M = m] \quad .$$

Montrer que si le chiffrement est correct, on a $|\mathcal{C}| \geq |\mathcal{M}|$ *; et que si le chiffrement est parfaitement sûr, on a* $|\mathcal{K}| \geq |\mathcal{C}|$*.*

Solution. *Si le chiffrement est correct, la fonction* $\mathsf{Enc}(k, \cdot)$ *est injective pour tout* $k \in \mathcal{K}$*, donc* $|\mathcal{C}| \geq |\mathcal{M}|$*. Dans la suite, on note* $C = \mathsf{Enc}(\mathsf{KeyGen}, M)$ *(qui est aussi une variable aléatoire).*

Si nous fixons $m \in \mathcal{M}$*, alors par la propriété de sécurité parfaite* $\Pr[M = m | C = c] = \Pr[M = m]$*. De plus* $\Pr[M = m | C = c] = \Pr[C = c | M = m] \times \Pr[M = m] / \Pr[C = c]$*, donc inversement* $\Pr[C = c | M = m] = \Pr[C = c] > 0$ *pour tout* c *(tous les chiffrements sont accessibles).*

Donc pour tout m *il doit exister une clé qui envoie* m *sur* c*, i.e.,* $|\mathcal{K}| \geq |\mathcal{C}|$*.*

On va utiliser la définition suivante de la sécurité parfaite :

> Un chiffrement symétrique est parfaitement sûr si, pour tous $m_1, m_2, c \in \mathcal{M} \times \mathcal{M} \times \mathcal{C}$:
>
> $$\Pr_{k \leftarrow \mathsf{KeyGen}}[\mathsf{Enc}(k, m_1) = c] = \Pr_{k \leftarrow \mathsf{KeyGen}}[\mathsf{Enc}(k, m_2) = c] \quad . \tag{1}$$

Question 2. *Montrer l'équivalence avec l'autre définition.*

Solution. ***Première implication.*** *Supposons que pour tous* m_1, m_2, c *:*

$$\Pr_{k \leftarrow \mathsf{KeyGen}}[\mathsf{Enc}(k, m_1) = c] = \Pr_{k \leftarrow \mathsf{KeyGen}}[\mathsf{Enc}(k, m_2) = c]$$

On a, pour toute variable aléatoire M*,* m *et* c *:*

$$\begin{aligned}
\Pr[M = m | \mathsf{Enc}(\mathsf{KeyGen}, M) = c] &= \frac{\Pr[M = m \wedge \mathsf{Enc}(\mathsf{KeyGen}, M) = c]}{\Pr[\mathsf{Enc}(\mathsf{KeyGen}, M) = c]} \\
&= \frac{\Pr[M = m \wedge \mathsf{Enc}(\mathsf{KeyGen}, m) = c]}{\Pr[\mathsf{Enc}(\mathsf{KeyGen}, M) = c]} \\
&= \Pr[M = m] \frac{\Pr[\mathsf{Enc}(\mathsf{KeyGen}, m) = c]}{\Pr[\mathsf{Enc}(\mathsf{KeyGen}, M) = c]}
\end{aligned}$$

On montre que :

$$\begin{aligned}
\Pr[\mathsf{Enc}(\mathsf{KeyGen}, M) = c] &= \sum_{m' \in \mathcal{M}} \Pr[M = m'] \Pr[\mathsf{Enc}(\mathsf{KeyGen}, m) = c] \\
&= \sum_{m' \in \mathcal{M}} \Pr[M = m'] \left(\Pr[\mathsf{Enc}(\mathsf{KeyGen}, m) = c]\right) \text{ par hypothèse} \\
&= \Pr[\mathsf{Enc}(\mathsf{KeyGen}, m) = c] \underbrace{\sum_{m' \in \mathcal{M}} \Pr[M = m']}_{=1}
\end{aligned}$$

Donc :

$$\Pr[M = m | \mathsf{Enc}(\mathsf{KeyGen}, M) = c] = \Pr[M = m] \quad .$$

Deuxième implication. *Supposons que pour toute variable aléatoire M, m et c :*

$$\Pr[M = m | \mathsf{Enc}(\mathsf{KeyGen}, M) = c] = \Pr[M = m] \quad .$$

Soit m_1, m_2, c et M la distribution uniforme sur $\{m_1, m_2\}$. On a :

$$\begin{cases} \Pr[M = m_1 | \mathsf{Enc}(\mathsf{KeyGen}, M) = c] = \Pr[M = m_1] = \frac{1}{2} \\ \Pr[M = m_2 | \mathsf{Enc}(\mathsf{KeyGen}, M) = c] = \Pr[M = m_2] = \frac{1}{2} \end{cases}$$

Donc :

$$\begin{gathered} \Pr[M = m_1 | \mathsf{Enc}(\mathsf{KeyGen}, M) = c] = \Pr[M = m_2 | \mathsf{Enc}(\mathsf{KeyGen}, M) = c] \\ \frac{\Pr[M = m_1] \Pr[\mathsf{Enc}(\mathsf{KeyGen}, m_1) = c]}{\Pr[\mathsf{Enc}(\mathsf{KeyGen}, M) = c]} = \frac{\Pr[M = m_2] \Pr[\mathsf{Enc}(\mathsf{KeyGen}, m_2) = c]}{\Pr[\mathsf{Enc}(\mathsf{KeyGen}, M) = c]} \\ \Pr[\mathsf{Enc}(\mathsf{KeyGen}, m_1) = c] = \Pr[\mathsf{Enc}(\mathsf{KeyGen}, m_2) = c] \quad . \end{gathered}$$

Question 3. *Montrer que les conditions 1. et 2. sont suffisantes pour la sécurité parfaite.*

Solution. *En utilisant la deuxième définition c'est assez clair.*

En effet pour tout m, c : $\Pr_k[\mathsf{Enc}(k, m) = c] = \frac{1}{|\mathcal{K}|}$ car c'est la probabilité que k (qui est choisie uniformément au hasard par KeyGen) atteigne la valeur unique telle que $\mathsf{Enc}(k, m) = c$.

Question 4. *Montrer le sens inverse de la preuve.*

Solution. *Soit $m \in \mathcal{M}$. Nous savons que pour tout c, il existe au moins une clé k telle que $\mathsf{Enc}(m, k) = c$. Par conséquent :*

$$|\{\mathsf{Enc}(k, m), k \in \mathcal{K}\}| = |\mathcal{C}|$$

mais comme $|\mathcal{C}| = |\mathcal{K}|$ par hypothèse, on a :

$$|\{\mathsf{Enc}(k, m), k \in \mathcal{K}\}| = |\mathcal{K}|$$

ce qui donne l'unicité.

On revient ensuite à la propriété de sécurité parfaite. On a :

$$\Pr[M = m | \mathsf{Enc}(K, M) = c] = \Pr[M = m] = \Pr[\mathsf{Enc}(K, M) = c | M = m] \frac{\Pr[M = m]}{\Pr[\mathsf{Enc}(K, M) = c]}$$

Donc pour tout m, c : $\Pr[\mathsf{Enc}(K, M) = c | M = m] = \Pr[\mathsf{Enc}(K, M) = c]$ (sur les variables aléatoires K, du KeyGen, et M, du message). Or il existe exactement une clé k telle que $\mathsf{Enc}(k, m) = c$, donc $\Pr[\mathsf{Enc}(K, M) = c | M = m] = \Pr[K = k]$. Et donc pour toute clé k :

$$\Pr[K = k] = \Pr[\mathsf{Enc}(K, M) = c] \quad .$$

Ici c est une constante, et $\Pr[\mathsf{Enc}(K, M) = c]$ est donc aussi une constante. Les clés sont donc toutes utilisées avec la même probabilité, qui doit donc être $1/|\mathcal{K}|$.

Distance Statistique

On rappelle que la distance statistique entre deux variables aléatoires discrètes sur un espace dénombrable A est définie par :

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

Question 5. *Soit X et Y deux variables aléatoires discrètes sur un ensemble dénombrable A, et soit Z une troisième variable sur un ensemble B (possiblement différent). Montrer que si Z est indépendante de X et Y, alors $\Delta((X,Z),(Y,Z)) = \Delta(X,Y)$.*

Solution.

$$\begin{aligned}
\Delta((X,Z),(Y,Z)) &= \frac{1}{2}\sum_{a,b\in A,B} |\Pr[X,Z=a,b] - \Pr[Y,Z=b]| \\
&= \frac{1}{2}\sum_{a,b\in A,B} \Pr[Z=b]\,|\Pr[X=a] - \Pr[Y=a]| \\
&= \frac{1}{2}\underbrace{\sum_{b\in B}\Pr[Z=b]}_{=1}\underbrace{\sum_{a\in A}|\Pr[X=a]-\Pr[Y=a]|}_{=\Delta(X,Y)} \ .
\end{aligned}$$

Question 6. *Soit $(X_1,\dots,X_k)$ et $(Y_1,\dots,Y_k)$ deux listes de variables aléatoires totalement indépendantes. Montrer que :*

$$\Delta((X_i)_i,(Y_i)_i) \le \sum_{i=1}^{k}\Delta(X_i,Y_i) \ . \tag{2}$$

Question 7. *Soit X et Y deux variables aléatoires à valeurs dans un ensemble A. Montrer que pour toute fonction (potentiellement randomisée) f de domaine A, $\Delta(f(X),f(Y)) \le \Delta(X,Y)$.*

Solution. *Soit $B = f(A)$ l'image de A par f.*

$$\begin{aligned}
\Delta(f(X),f(Y)) &= \frac{1}{2}\sum_{b\in B}|\Pr[f(X)=b] - \Pr[f(Y)=b]| \\
&= \frac{1}{2}\sum_{b\in B}\left|\sum_{a\in A}\underbrace{\Pr[f(X)=b|X=a]}_{=\Pr[f(a)=b]}\Pr[X=a] - \underbrace{\Pr[f(Y)=b|Y=a]}_{=\Pr[f(a)=b]}\Pr[Y=a]\right| \\
&= \frac{1}{2}\sum_{b\in B}\left|\sum_{a\in A}\Pr[f(a)=b]\,(\Pr[X=a]-\Pr[Y=a])\right| \\
&\le \frac{1}{2}\sum_{a\in A}\sum_{b\in B}\Pr[f(a)=b]\,|\Pr[X=a]-\Pr[Y=a]| \\
&= \frac{1}{2}\sum_{a\in A}\left(\underbrace{\sum_{b\in B}\Pr[f(a)=b]}_{=1}\right)|\Pr[X=a]-\Pr[Y=a]| \ .
\end{aligned}$$

Question 8. *En déduire que l'avantage de tout adversaire $\mathcal{A}$ pour distinguer entre X et Y* en une requête *est inférieur à $\Delta(X,Y)$.*

Solution. *Dans les jeux consistant à distinguer, l'adversaire $\mathcal{A}$ est une fonction randomisée f à valeurs dans 0, 1. Dans le jeu G_0 on lui donne accès à X, dans le jeu G_1 on lui donne accès à Y.*

L'avantage est la différence entre la probabilité de renvoyer 1 dans les deux jeux, donc :

$$\mathrm{Adv}(\mathcal{A}) = |\Pr[f(X) = 1] - \Pr[f(Y) = 1]|$$

Notons également que $|\Pr[f(X) = 0] - \Pr[f(Y) = 0]| = |1 - \Pr[f(X)] = 1 - 1 + \Pr[f(Y) = 1]| = \mathrm{Adv}(\mathcal{A})$*. Donc :*

$$\Delta(f(X), f(Y)) = \frac{1}{2}(\mathrm{Adv}(\mathcal{A}) + \mathrm{Adv}(\mathcal{A})) = \mathrm{Adv}(\mathcal{A}) \quad .$$

On a donc que pour tout distingueur $\mathcal{A}$ (à une requête) : $\mathrm{Adv}(\mathcal{A}) \leq \Delta(X, Y)$.

Question 9. *Montrer que si $\Delta(X, Y) = \mathrm{negl}(n)$, les distributions X et Y sont calculatoirement indistinguables.*

Solution. *Dans les jeux de distingueur l'adversaire a droit à un nombre multiple de requêtes. Les résultats de ces requêtes sont des* copies indépendantes *de X ou de Y, donc on peut utiliser la question 9 en combinaison avec la question 10.*

Un adversaire efficace doit s'exécuter en temps $\mathrm{poly}(n)$*, donc ne peut faire que* $\mathrm{poly}(n)$ *requêtes. Soit $k = \mathrm{poly}(n)$ le nombre de requêtes effectuées (on va le supposer constant pour simplifier). On a donc :*

$$\mathrm{Adv}(\mathcal{A}) \leq \Delta((X_1, \ldots, X_k), (Y_1, \ldots, Y_k)) \leq k\Delta(X, Y) = \mathrm{poly}(n)\,\mathrm{negl}(n) = \mathrm{negl}(n) \quad .$$

C'est vrai pour tout adversaire PPT, ce qui nous permet de conclure.

Chiffrement de Vigenère

Le chiffrement de Vigenère sur un alphabet fini strictement ordonné Σ est une version relaxée et généralisée du OTP, où la clé peut être plus petite que les messages considérés, et on utilise l'addition modulaire au lieu du XOR (qui est l'addition modulo 2). Pour chiffrer / déchiffrer, on répète la clé ($abc \to abcabcabc\ldots$) pour obtenir une clé aussi longue que le message / chiffré, et on ajoute (modulairement) la clé au message / chiffré, caractère par caractère.

Question 10. *Définir formellement le schéma.*

Solution. *Définir* $\mathsf{KeyGen}, \mathsf{Enc}, \mathsf{Dec}$*. Pour KeyGen, on prend en entrée la longueur du mot, et on choisit le mot au hasard (par exemple).*

Nous considérons le jeu de sécurité suivant sur un schéma de chiffrement Π avec un adversaire $\mathcal{A}$, noté G^{EAV}.

1. $\mathcal{A}$ choisit deux messages m_0, m_1 de $\mathcal{P}$
2. La clé k est générée par KeyGen, un bit $b \hookleftarrow U(0, 1)$ est choisi
3. $c = \mathsf{Enc}(m_b, k)$ est donné à $\mathcal{A}$
4. $\mathcal{A}$ renvoie un bit b'

La sortie du jeu, notée $G^{EAV}(\mathcal{A}, n)$ est $\top$ si $b = b'$ et $\bot$ sinon. On peut donc remarquer que $G^{EAV}(\mathcal{A}, n)$ est une variable aléatoire à valeurs dans $\{\top, \bot\}$.

Formellement :

Definition 1 (Sécurité EAV)**.** Un schéma de chiffrement symétrique est EAV (*secure against eavesdropping*) si pour tout adversaire PPT $\mathcal{A}$ son avantage $\mathrm{Adv}^{EAV}(\mathcal{A}) = |\Pr[G^{EAV}(\mathcal{A}, n) \leftarrow \top] - 1/2|$ est négligeable en n. Le jeu de sécurité est donné par :

1. $(m_0, m_1) \in \mathcal{M}^2 \leftarrow \mathcal{A}(1^n)$
2. $k \leftarrow \mathsf{KeyGen}(1^n)$, $b \hookleftarrow U(0,1)$
3. $b' \leftarrow \mathcal{A}(\mathsf{Enc}(k, m_b))$
4. Si $b = b'$ alors $\top$ sinon $\bot$

Dans la suite, le chiffrement de Vigenère est noté Π, et l'alphabet latin Σ.

Question 11. *Montrer qu'un adversaire $\mathcal{A}$ participant au jeu EAV peut retrouver la clé k.*

Solution. *Il suffit d'appeler le chiffrement sur une entrée contenant un bloc de zéros, de longueur la taille de la clé ($k + 0 = 0$ sur chaque caractère correspondant). L'adversaire choisit donc deux messages quelconques avec un blos de zéros.*

Question 12. *Montrer que le chiffrement de Vigenère n'est pas EAV-sûr.*

Solution. *Puisque l'adversaire peut retrouver la clé, il peut déchiffrer, donc distinguer le chiffrement de m_0 de celui de m_1.*

L'attaque s'exécute donc ainsi : l'adversaire produit deux messages $0\|m'_0$ et $0\|m'_1$ avec un bloc de zéros et deux blocs arbitraires différents, le challenger chiffre, l'adversaire retrouve la clé grâce au bloc de zéros, ensuite il déchiffre c, et distingue entre les deux.