Sécurité des Chiffres contre l'espionnage(Cours #2)

Louis Salvail[1⋆]

Université de Montréal (DIRO), QC, Canada
salvail@iro.umontreal.ca

Abstract. Dans cette section, nous regardons comment définir la sécurité des chiffres à clé secrète contre l'espionnage. Nous donnons une définition basée sur la distinguabilité du chiffrement de deux messages. Nous montrons des définitions de sécurité équivalentes comme la sécurité sémantique.

⋆ Notes de cours pour *Introduction à la cryptographie–IFT6180*. Ces notes sont une version abrégée de la Section 3.2 du livre *Introduction to Modern Cryptography*, Jonathan Katz et Yehuda Lindell, Chapman & Hall/CRC, 2008.

1 TDP, TPP, probabilités négligeables

Soient $\mathbb{N}$ l'ensemble des entiers naturels, soit $\mathbb{R}$ l'ensemble des réels et soit $\mathbb{R}^+$ l'ensemble des nombres réels non-négatifs. Un algorithme A est dit *efficace* s'il roule en temps polynômial $p(\cdot)$ sur tous les inputs de taille n. Autrement dit, si pour chaque $x \in \{0,1\}^*$, $A(x)$ termine en au plus $p(|x|)$ étapes élémentaires, où $|x|$ désigne la taille de x en bits. Un algorithme est dit *probabiliste* s'il a accès à un ruban de bits aléatoires uniformément distribués en plus de l'input x pour accomplir sa tâche. Nous dénoterons par *TDP* l'ensemble des algos polynômiaux déterministes et par *TPP* l'ensemble des algo polynômiaux probabilistes. Nous dénoterons par la suite

$$\mathsf{poly}(n) := \{p : \mathbb{N} \to \mathbb{R}^+ \mid p(\cdot) \text{ est un polynôme non-négatif}\} \ .$$

Notez que $(\forall f(\cdot), g(\cdot) \in \mathsf{poly}(n))[f(n) + g(n), f(n) \cdot g(n), f(g(n)) \in \mathsf{poly}(n)]$. Nous dirons d'une fonction $f : \mathbb{N} \to \mathbb{R}$ qu'elle est *négligeable* si pour chaque polynôme $p(\cdot)$, il existe $n_0 \in N$ tel que pour chaque $n > n_0$, $f(n) < 1/p(n)$. L'ensemble des fontions négligeables est dénoté par $\mathsf{neg}(n)$. Notez que pour chaque $\delta(\cdot), \epsilon(\cdot) \in \mathsf{neg}(n)$, $\delta(n) + \epsilon(n) \in \mathsf{neg}(n)$ et pour chaque $p(\cdot) \in \mathsf{poly}(n)$, $p(n) \cdot \delta(n) \in \mathsf{neg}(n)$.

2 Chiffres à clé secrète

Un chiffre à clé secrète $\Pi = (\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$ pour des messages $m \in \mathcal{M}$ est tel que:

1. L'algorithme Gen est l'algo TPP de génération de clé. C'est-à-dire que

   $$k := \mathsf{Gen}(1^n),$$

   où n est le facteur de sécurité. Nous supposons sans perte de généralité que $|k| \geq n$. D'une certaine façon, n représente la *longueur* de k.
2. L'algorithme Enc est l'algo TPP de chiffrement, pour tout $m \in \mathcal{M}$:

   $$c := \mathsf{Enc}_k(m).$$

   Enc fonctionne en TPP borné supérieurement par $\mathsf{poly}(|k| + |m|)$ en autant que $k := \mathsf{Gen}(1^n)$ pour un certain entier naturel n.
3. L'algorithme Dec est l'algo TDP de décodage (ou de déchiffrement):

   $$m := \mathsf{Dec}_k(\mathsf{Enc}_k(m)).$$

   dont le temps d'exécution est borné supérieurement par $\mathsf{poly}(|k| + |c|)$.

Nous définissons maintenant une expérience de distinguabilité comme celle définie pour les chiffres parfaitement sûrs. Nous l'adaptons simplement pour prendre en compte:

- Des clés secrètes qui sont choisies en fonction du paramètre de sécurité n, et
- Les adversaires fonctionnant en TPP en fonction du paramètre de sécurité n.

$\mathsf{PrivK}^{\mathsf{esp}}_{\mathcal{A},\Pi}(n)$

1. L'adversaire $\mathcal{A}$ se voit donner 1^n et produit $m_0, m_1 \in \mathcal{M}$ avec $|m_0| = |m_1|$. $\mathcal{A}$ annonce m_0 et m_1. Si Π est de longueur fixe $\ell(n)$ alors $|m_0| = |m_1| = \ell(n)$.
2. $k := \mathsf{Gen}(1^n)$, $b \in_R \{0,1\}$, $c := \mathsf{Enc}_k(m_b)$. Le cryptogramme (i.e. le cryptogramme défi) c est retourné à $\mathcal{A}$.
3. $\mathcal{A}$ produit $b' \in \{0,1\}$.
4. $\mathsf{PrivK}^{\mathsf{esp}}_{\mathcal{A},\Pi}(n) := (b = b')$.

Definition 2.1. *Le chiffre à clé secrète Π admet un* chiffrement indistinguable en présence d'espion *si pour chaque TPP $\mathcal{A}$ il existe une fonction négligeable* $\mathsf{neg}(\cdot)$ *telle que:*

$$\Pr\left(\mathsf{PrivK}^{\mathsf{esp}}_{\mathcal{A},\Pi}(n) = 1\right) \leq \frac{1}{2} + \mathsf{neg}(n),$$

où la probabilité est calculée sur les choix de $\mathcal{A}$ et ceux de l'expérience.

- Le fait que la sécurité soit seulement garantie contre les espions ayant accès qu'à un seul cryptogramme est implicite car $\mathsf{PrivK}^{\mathsf{esp}}_{\mathcal{A},\Pi}(n)$ ne donne accès qu'à un seul cryptogramme à $\mathcal{A}$.
- Il y a une défininition de sécurité équivalente qui dit que le comportement de $\mathcal{A}$ ne change pas en fonction de la valeur de $b \in \{0,1\}$ choisie dans $\mathsf{PrivK}^{\mathsf{esp}}_{\mathcal{A},\Pi}(n)$. Soit $\mathsf{PrivK}^{\mathsf{esp}}_{\mathcal{A},\Pi}(n,b)$ le résultat de l'expérience lorsque le bit b est choisi.
- Soit $\mathsf{out}(\mathsf{PrivK}^{\mathsf{esp}}_{\mathcal{A},\Pi}(n,b))$ la variable aléatoire pour le bit b' produit par $\mathcal{A}$ dans $\mathsf{PrivK}^{\mathsf{esp}}_{\mathcal{A},\Pi}(n)$ lorsque le bit b a été choisi.

Considérez la définitions de sécurité suivante:

Definition 2.2. *Le chiffre à clé secrète Π admet un* chiffrement indistinguable en présence d'espion *si pour chaque TPP $\mathcal{A}$ il existe une fonction négligeable* $\mathsf{neg}(\cdot)$ *telle que:*

$$|\Pr\left(\mathsf{out}(\mathsf{PrivK}^{\mathsf{esp}}_{\mathcal{A},\Pi}(n,0)) = 1\right) - \Pr\left(\mathsf{out}(\mathsf{PrivK}^{\mathsf{esp}}_{\mathcal{A},\Pi}(n,1)) = 1\right)| \leq \mathsf{neg}(n),$$

où la probabilité est calculée sur les choix de $\mathcal{A}$ et ceux de l'expérience.

Autrement dit, la définition 2.2 demande que pour tout adversaire $\mathcal{A}$, son comportement est *statistiquement* indépendant du choix de b. Il n'est pas trop difficile de montrer que:

Lemma 2.3. *Π est une chiffre à clé secrète qui satisfait la définition 2.1 si et seulement s'il satisfait la définition 2.2.*

3 Sécurité Sémantique

Historiquement, une des premières notion satisfaisante de sécurité est appelée la *sécurité sémantique*. La sécurité sémantique considère une distribution arbitraire sur les messages clairs en plus de permettre à l'adversaire d'obtenir de l'information externe (un historique) sur le message clair avant l'espionnage. On dit d'un chiffre qu'il est sémantiquement sûr si l'adversaire ne peut apprendre aucune fonction TPP f appliquée au message $m \in \mathcal{M}$ à partir de $c = \mathsf{Enc}_k(m)$ qu'il ne peut apprendre sans la connaissance de c.

Definition 3.1. *Un chiffre $\Pi = (\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$ est sémantiquement sûr en présence d'espion si pour tout TPP $\mathcal{B}$, il existe un algo TPP $\mathcal{B}'$ t.q. pour chaque famille de distribution efficacement échantillonnable $X = (X_1, X_2, \ldots)$ et pour chaque fonctions TPP $f(\cdot)$ et $h(\cdot)$, il existe une fonction négligeable $\mathsf{neg}(\cdot)$ telle que:*

$$|\Pr(\mathcal{B}(1^n, \mathsf{Enc}_k(m), h(m)) = f(m)) - \Pr(\mathcal{B}'(1^n, h(m)) = f(m))| \leq \mathsf{neg}(n),$$

où $m \in_R X_n$ et les probs sont calculées en fonction des choix de m, k ainsi que les choix de $\mathcal{B}$ et $\mathcal{B}'$ ainsi que ceux de l'algo de chiffrement.

Un définition un peu différente peut être formulée comme suit:

Definition 3.2. *Un chiffre $\Pi = (\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$ est sémantiquement sûr en présence d'espion si pour tout TPP $\mathcal{B}$ et pour chaque famille de distributions efficacement échantillonnable $X = (X_1, X_2, \ldots)$ et pour chaque fonction TPP $f(\cdot)$ et $h(\cdot)$, il existe une fonction négligeable $\mathsf{neg}(\cdot)$ telle que:*

$$\Pr(\mathcal{B}(1^n, \mathsf{Enc}_k(m), h(m)) = f(m)) - \Pr(\mathcal{B}(1^n, \mathsf{Enc}_k(m'), h(m)) = f(m)) \leq \mathsf{neg}(n),$$

où $m, m' \in_R X_n$ et les probs sont calculées en fonction des choix de m, k ainsi que les choix de $\mathcal{B}$ ainsi que ceux de l'algo de chiffrement.

Il n'est pas trop difficile de montrer le lemme suivant:

Lemma 3.3. *Un chiffre Π satisfait la définition 3.1 si et seulement s'il satisfait la définition 3.2.*

Il est facile de voir que donner $\mathsf{Enc}_k(m')$ ne peut pas aider à évaluer $f(m)$. Même donner m' en clair n'aiderait pas. Dans une direction, la preuve est simple tandis que dans l'autre direction c'est un peu plus subtil.

La fonction $h(\cdot)$ est la fonction qui caractérise l'historique de l'adversaire au sujet du message m dont le chiffrement lui est fourni. La fonction $f(\cdot)$, quant à elle, détermine l'information que l'adversaire obtient au sujet de m. La sécurité sémantique établie qu'un chiffre est sûr s'il n'est pas possible à l'adversaire d'obtenir plus d'information $f(\cdot)$ (avec $f(\cdot)$ TPP) sur le message m lorsque $\mathsf{Enc}_k(m)$ est donné à l'adversaire que lorsque qu'il ne lui est pas donné. Ceci doit demeurer vrai pour toutes les historiques $h(\cdot)$, en autant que $h(\cdot)$ soit TPP. L'algorithme $\mathcal{B}'$ est souvent appelé *simulateur*.

Il est possible de montrer que la sécurité sémantique est la même chose que la sécurité exprimée à la définition 2.1:

Theorem 3.4 (Goldwasser-Micali1982). *Π est une chiffre à clé secrète qui satisfait la définition 2.1 si et seulement s'il satisfait la définition 3.1.*

Proof. Nous montrons chaque implication séparemment:

- (définition 3.1 $\Rightarrow$ définition 2.1) Soit $\mathcal{A}$ un adversaire TPP arbitraire dans l'expérience $\mathsf{PrivK}^{\mathsf{esp}}_{\mathcal{A},\Pi}(n)$. Considérons maintenant une simulation de $\mathcal{A}$ jusqu'au moment où $m_0, m_1 \in \mathcal{M}$ sont produits. Soit $X_n = \{m_0, m_1\}$ avec $\Pr(X_n = m_0) = \Pr(X_n = m_1) = \frac{1}{2}$. Soit $h(m) = \nu(m_0, m_1)$ où $\nu(m_0, m_1)$ est l'état de la mémoire (i.e. la *vue*) de $\mathcal{A}$ lorsque m_0 et m_1 sont produits. Par construction, $h(\cdot)$ est TPP. Soit $f(m_b) = b$ pour $b \in \{0, 1\}$

et $f(m) = 2$ si $m \notin \{m_0, m_1\}$. La fonction $f(\cdot)$ est clairement TPP. Maintenant, considérons $\mathcal{B}$ simulant $\mathcal{A}$ à partir de la vue $\nu(m_0, m_1)$ pour produire b'. Puisque nous supposons que Π satisfait la définition 3.1, nous avons donc qu'il existe un TPP $\mathcal{B}'$ tel que

$$|\Pr\left(\mathcal{B}(1^n, \mathsf{Enc}_k(m), \nu(m_0, m_1)) = b\right) - \Pr\left(\mathcal{B}'(1^n, \nu(m_0, m_1)) = b\right)| \leq \mathsf{neg}(n), \quad (1)$$

avec $m \in_R X_n = \{(\frac{1}{2}, m_0), (\frac{1}{2}, m_1)\}$. D'autre part, nous avons par construction:

$$\Pr\left(\mathcal{B}(1^n, \mathsf{Enc}_k(m), \nu(m_0, m_1)) = b\right) = \Pr\left(\mathsf{PrivK}^{\mathsf{esp}}_{\mathcal{A},\Pi}(n) = 1\right). \quad (2)$$

De plus, il est clair que

$$\Pr\left(\mathcal{B}'(1^n, \nu(m_0, m_1)) = b\right) = \frac{1}{2}, \quad (3)$$

puisque b est indépendant de la vue $\nu(m_0, m_1)$. Nous concluons donc à partir de (1), (2) et (3) que:

$$\left|\Pr\left(\mathsf{PrivK}^{\mathsf{esp}}_{\mathcal{A},\Pi}(n) = 1\right) - \frac{1}{2}\right| = \left|\Pr\left(\mathcal{B}(1^n, \mathsf{Enc}_k(m), \nu(m_0, m_1)) = b\right) - \frac{1}{2}\right| \leq \mathsf{neg}(n). \quad (4)$$

Évidemment, (4) implique que

$$\Pr\left(\mathsf{PrivK}^{\mathsf{esp}}_{\mathcal{A},\Pi}(n) = 1\right) \leq \frac{1}{2} + \mathsf{neg}(n).$$

Ceci conclue la preuve de la première implication.

– (définition 2.1 $\Rightarrow$ définition 3.1) Nous montrons la contraposé. Supposons donc que pour un chiffre Π, il existe une famille de distribution efficacement échantillonable $X = (X_i)_i$, des fonctions $h(\cdot)$ et $f(\cdot)$ efficacement évaluables, ainsi que l'algo TPP $\mathcal{B}$ tels que pour chaque TPP $\mathcal{B}'$:

$$|\Pr\left(\mathcal{B}(1^n, \mathsf{Enc}_k(m), h(m)) = f(m)\right) - \Pr\left(\mathcal{B}'(1^n, h(m)) = f(m)\right)| \geq \frac{1}{p(n)}, \quad (5)$$

où $p(n) \in \mathsf{poly}(n)$ et $m \in_R X_n$. Par le lemme 3.3, nous savons que (5) implique que pour tout TPP $\mathcal{B}$:

$$|\Pr\left(\mathcal{B}(1^n, \mathsf{Enc}_k(m), h(m)) = f(m)\right) - \Pr\left(\mathcal{B}(1^n, \mathsf{Enc}_k(m'), h(m)) = f(m)\right)| \geq \frac{1}{p(n)}. \quad (6)$$

Nous construisons un adversaire $\mathcal{A}$ dans l'expérience $\mathsf{PrivK}^{\mathsf{esp}}_{\mathcal{A},\Pi}(n)$ qui, pour un polynôme positif $p'(n)$, est tel que:

$$\Pr\left(\mathsf{PrivK}^{\mathsf{esp}}_{\mathcal{A},\Pi}(n) = 1\right) \geq \frac{1}{2} + \frac{1}{p'(n)}. \quad (7)$$

Voici la description de $\mathcal{A}$:

1. $\mathcal{A}$ reçoit 1^n.

2. Tirer $m_0, m_1 \in_R X_n$. Soumettre (m_0, m_1) à l'oracle de chiffrement.
3. $k = \mathsf{Gen}(1^n)$, $b \in_R \{0,1\}$, retourne $c := \mathsf{Enc}_k(m_b)$ à $\mathcal{A}$.
4. $\mathcal{A}$ simule $\mathcal{B}$ sur entrée $(1^n, c, h(m_0))$ pour produire r.
5. Si $r = f(m_0)$ alors $\mathcal{A}$ pose $b' = 0$,
6. Sinon $\mathcal{A}$ pose $b' \in_R \{0,1\}$.
7. $\mathcal{A} \uparrow b'$.

Il est clair que $\mathcal{A}$ est TPP si l'étape 2 peut être complétée en TPP. Supposons pour le moment que tel est bien le cas. Nous le montrerons formellement par la suite. Nous avons:

$$\begin{aligned}
\Pr\left(\mathsf{PrivK}^{\mathsf{esp}}_{\mathcal{A},\Pi}(n) = 1\right) &= \frac{1}{2}\Pr\left(\mathcal{B}(1^n, \mathsf{Enc}_k(m_0), h(m_0)) = f(m_0)\right) \\
&\quad + \frac{1}{4}\Big(\Pr\left(\mathcal{B}(1^n, \mathsf{Enc}_k(m_0), h(m_0)) \neq f(m_0)\right) \\
&\quad\quad + \Pr\left(\mathcal{B}(1^n, \mathsf{Enc}_k(m_1), h(m_0)) \neq f(m_0)\right)\Big) \\
&= \frac{1}{2}\Pr\left(\mathcal{B}(1^n, \mathsf{Enc}_k(m_0), h(m_0)) = f(m_0)\right) \\
&\quad + \frac{1}{4}\Big(2 - \Pr\left(\mathcal{B}(1^n, \mathsf{Enc}_k(m_0), h(m_0)) = f(m_0)\right) \\
&\quad\quad - \Pr\left(\mathcal{B}(1^n, \mathsf{Enc}_k(m_1), h(m_0)) = f(m_0)\right)\Big) \\
&= \frac{1}{2} + \frac{1}{4}\Big(\Pr\left(\mathcal{B}(1^n, \mathsf{Enc}_k(m_0), h(m_0)) = f(m_0)\right) \\
&\quad\quad - \Pr\left(\mathcal{B}(1^n, \mathsf{Enc}_k(m_1), h(m_0)) = f(m_0)\right)\Big) \qquad (8) \\
&\geq \frac{1}{2} + \frac{1}{4p(n)}. \qquad (9)
\end{aligned}$$

L'équation (9) est une conséquence de (6). Le résultat est obtenu facilement en posant $p'(n) = 4p(n)$ puisque $\mathcal{A}$ brise efficacement Π dans l'expérience $\mathsf{PrivK}^{\mathsf{esp}}_{\mathcal{A},\Pi}(n)$ avec probabilité $\frac{1}{2} + \frac{1}{q(n)}$. Notez que $\mathcal{A}$ est efficace puisque $h(\cdot)$, $f(\cdot)$ sont évaluables en TPP et X_n efficacement échantillonnable.

□

Notez que nous n'avons montré équivalente que les définitions de sécurité contre l'espionnage d'un seul cryptogramme. Les deux définitions modifiées pour tenir compte des attaques CPA sont aussi équivalentes. La sécurité sémantique demeure la même sauf que l'algo $\mathcal{B}$ peut choisir des messages clairs et les faire chiffrer par un oracle de chiffrement (toujours à partir de la même clé). L'expérience $\mathsf{PrivK}^{\mathsf{cpa}}_{\mathcal{A},\Pi}(n)$, quant à elle, va être définie plus tard dans le cours.