école normale supérieure **2024-2025**
Département d'informatique

Introduction à la cryptologie
TD n° 1 : Top modèle

Exercice 1 (Plus on est de fous, plus on rit)**.**

Definition 1. *Let $\epsilon \in [0;1]^{\mathbb{N}\times\mathbb{N}}$. A scheme $\Pi_{\mathcal{S}}$ is ϵ-secure if for all probabilistic algorithm $\mathcal{A}$, which takes a time of computation at most t:* $\mathbb{P}\left[\text{Inforg}_{\mathcal{A},\Pi}(\lambda) = 1\right] \leq \epsilon(t,\lambda)$.

Definition 2. *Let $\epsilon \in [0;1]^{\mathbb{N}\times\mathbb{N}}$. A scheme $\Pi_{\mathcal{S}}$ is ϵ-secure in a multi-user context if for all probabilistic algorithm $\mathcal{A}$, which takes a time of computation at most t and does at most q queries* OCreate*:*

$$\mathbb{P}\left[\text{InforgMU}_{\mathcal{A},\Pi}(\lambda) = 1\right] \leq \epsilon(t,q,\lambda).$$

Show that if $\Pi_{\mathcal{S}}$ is ϵ-secure (definition 1), then it is $q\epsilon$-secure in a multi-user context (definition 2).
Hint : We could try to guess which user among the q's will be the "target" and imagine that this one has the key of the original single-user game (and the other $(q-1)$ keys are artificially built).

$\underline{\text{InforgMU}_{\mathcal{A},\Pi}(\lambda)}$:	$\underline{OCreate\,()}$:	$\underline{OSign\,(vk,m)}$:
$C \leftarrow U \leftarrow Q \leftarrow \emptyset$	$(sk,vk) \leftarrow \Pi.\texttt{KeyGen}\left(1^{\lambda}\right)$	$Q := Q \cup \{(vk,m)\}$
$(vk,m,\sigma) \leftarrow \mathcal{A}^{OCreate(),OCorrupt(\cdot),OSign(\cdot,\cdot)}$	$U := U \cup \{(sk,vk)\}$	Si $\exists sk/(sk,vk) \in U$
Retourner $\Pi.\texttt{Verify}\,(vk,m,\sigma)$	Retourner vk	$\sigma \leftarrow \Pi.\texttt{Sign}(sk,m)$
$\wedge\,((vk,m) \notin Q) \wedge ((vk,\cdot) \in U)$	$\underline{OCorrupt\,(vk)}$:	Retourner σ
	$Q := Q \cup (\{vk\} \times \mathcal{M})$	Sinon Retourner $\perp$
	Si $\exists sk/(sk,vk) \in U$	
	Retourner sk	
	Sinon Retourner $\perp$	

Figure 1: Jeu de sécurité d'inforgeabilité dans un contexte multi-utilisateurs.

Solution: We start by recalling the game $\text{Inforg}_{A,\Pi}$ for the security of a scheme Π in the single user context.

$\underline{\text{Inforg}_{\mathcal{A},\Pi}(\lambda)}$:	$\underline{OSign_{sk}\,(m)}$:
$Q \leftarrow \emptyset$	$Q \leftarrow Q \cup \{m\}$
$(s_k, v_k) \leftarrow \Pi.\texttt{KeyGen}(1^{\lambda})$	Return $\Pi.\texttt{Sign}(sk,m)$
$(m,\sigma) \leftarrow \mathcal{A}^{OSign_{s_k}(\cdot)}(v_k)$	
Return $\Pi.\texttt{Verify}\,(vk,m,\sigma) \wedge (m \notin Q)$	

Figure 2: Security game for unforgeability in a single user context

Suppose that $\Pi_{\mathcal{S}}$ is ε-secure according to definition 1. For all probabilistic algorithm $\mathcal{A}$ with runnning time $\leq t$,

$$\mathbb{P}\left[Inforg_{\mathcal{A},\Pi_{\mathcal{S}}}(\lambda) = 1\right] \leq \epsilon(t,\lambda).$$

We show that $\Pi_{\mathcal{S}}$ is $q\varepsilon$-secure according to definition 2.
The sketch of the proof is the following. Let $\mathcal{A}$ be a probabilistic algorithm with computation time $\leq t$ that does at most q queries to *OCreate*. The goal is to bound $p_{\mathcal{A}} = \mathbb{P}\left[\text{InforgMU}_{\mathcal{A},\Pi_{\mathcal{S}}}(\lambda) = 1\right]$ by $q\epsilon$. To do so, we will build a probabilistic algorithm $\mathcal{B}$ from $\mathcal{A}$ for the game $\text{Inforg}_{\mathcal{B},\Pi_{\mathcal{S}}}(\lambda)$ and express the probability $p_{\mathcal{B}} = \mathbb{P}\left[\text{Inforg}_{\mathcal{B},\Pi_{\mathcal{S}}}(\lambda) = 1\right]$ as $f(p_{\mathcal{A}})$ for some function f. Yet by assumption, we have

$$p_{\mathcal{B}} = f(p_{\mathcal{A}}) \leq \varepsilon.$$

And hopefully, this inequality will give us $p_{\mathcal{A}} \leq q\varepsilon$. Note that in this example, we would like to have $f(x) = 1/q \cdot x$ in order to conclude.
We build the algorithm $\mathcal{B}^{OSign_{sk}(\cdot)}(vk)$ as follows.

$\mathcal{B}^{OSign_{sk}(\cdot)}(vk)$:	$\overline{OCreate\,()}$:	$\overline{OSign\left(\bar{vk}, m\right)}$:
$C_{\mathcal{A}} \leftarrow U_{\mathcal{A}} \leftarrow Q_{\mathcal{A}} \leftarrow \emptyset$ Sample $1 \leq j \leq q$ uniformly at random Run $\mathcal{A}^{OCreate(),OCorrupt(\cdot),OSign(\cdot,\cdot)}$ by implementing the three oracles $OCreate\,()$, $OCorrupt\,(\cdot)$, $OSign\,(\cdot,\cdot)$ as specified on the right-hand side of this figure and get $(vk_{\mathcal{A}}, m_{\mathcal{A}}, \sigma_{\mathcal{A}})$ as result If $vk_{\mathcal{A}} = vk$ Return $(m_{\mathcal{A}}, \sigma_{\mathcal{A}})$ Else Return $\perp$	If it is the j-th call of $\mathcal{A}$ to $OCreate$: Return vk Else: $(\bar{sk}, \bar{vk}) \leftarrow \Pi.\texttt{KeyGen}(1^{\lambda})$ $U_{\mathcal{A}} \leftarrow U_{\mathcal{A}} \cup \{(\bar{sk}, \bar{vk})\}$ Return $\bar{vk}$	$Q_{\mathcal{A}} := Q_A \cup \{(\bar{vk}, m)\}$ If $\bar{vk} = vk$ Return $OSign_{sk}(vk, m)$ If $\exists \bar{sk}/(\bar{sk}, \bar{vk}) \in \bar{U}$ $\sigma \leftarrow \Pi.\texttt{Sign}(\bar{sk}, m)$ Return σ Else Return $\perp$
	$\overline{OCorrupt\left(\bar{vk}\right)}$: $Q_{\mathcal{A}} := Q_A \cup \left(\{\bar{vk}\} \times \mathcal{M}\right)$ If $\exists \bar{sk}/(\bar{sk}, \bar{vk}) \in U_{\mathcal{A}}$ Return $\bar{sk}$ Else Return $\perp$	

Figure 3: Security game for unforgeability in a single user context

In the view of $\mathcal{A}$, the simulation is perfect, thus $\mathcal{A}$ the output of this simulation of $\mathcal{A}$ satifies

$$\mathbb{P}\left[\text{InforgMU}_{\text{Simu}(\mathcal{A}),\Pi_{\mathcal{S}}}(\lambda) = 1\right] = p_{\mathcal{A}}.$$

Moreover, let $1 \leq i \leq q$ be such that $vk_{\mathcal{A}}$ has been created at the i-th call to the oracle $OCreate$, then $vk_A = vk$ iff $i = j$ and thus this happens with probability $1/q$. Finally, $\text{Inforg}_{\mathcal{B},\Pi_{\mathcal{S}}} = 1$ if and only if $vk_{\mathcal{A}} = vk$ and $\text{InforgMU}_{\mathcal{A},\Pi_{\mathcal{S}}} = 1$ and these two events are independent. Hence we get

$$p_{\mathcal{B}} = \frac{1}{q} \cdot p_{\mathcal{A}}.$$

However, by assumption, $p_{\mathcal{B}} \leq \varepsilon$. Therefore, $p_A \leq q\varepsilon$ and we conclude that the scheme $\Pi_{\mathcal{S}}$ is $q\varepsilon$-secure in a multi-user context. □

Exercice 2. Let $f \in \mathbb{R}^{\mathbb{N}}$, show the following properties are equivalent:

1. f is negligeable in all function of $\left(x \mapsto \frac{1}{P(x)}\right)_{P \in \mathbb{Z}[X]} \subset \mathbb{R}^{\mathbb{N}}$.
2. $|f|$ is asymptotically upper bounded by all function of $\left(x \mapsto \frac{1}{x^k}\right)_{k \in \mathbb{N}} \subset \mathbb{R}^{\mathbb{N}}$.

Then show this property define a sub-$\mathbb{Z}[X]$-module of $\mathbb{R}^{\mathbb{N}}$.

Exercice 3. Is f a negligeable function

1. Let $c \in \mathbb{N} : f(\lambda) = \frac{1}{\binom{\lambda}{c}}$.
2. $f(\lambda) = \frac{1}{e^{\log^2(\lambda)}}$.
3. $f(\lambda) = \frac{1}{e^{\log^2(\lambda)}}$.
4. $f(\lambda) = \frac{\lambda!}{\lambda^{\lambda}}$.

Exercice 4.

Definition 3 (Statistical distance)**.** *Let X and Y be two discrete random variables over a countable set A. The statistical distance between X and Y is the quantity*

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

The statistical distance verifies usual properties of distance function, i.e., it is a positive definite binary symmetric function that satisfies the triangle inequality:

- $\Delta(X, Y) \geq 0$, with equality if and only if X and Y are identically distributed,
- $\Delta(X, Y) = \Delta(Y, X)$,
- $\Delta(X, Z) \leq \Delta(X, Y) + \Delta(Y, Z)$.

1. Show that if $\Delta(X, Y) = 0$, then for any deterministic adversary $\mathcal{A}$, we have $\mathrm{Adv}_{\mathcal{A}}(X, Y) := \Delta(\mathcal{A}(X), \mathcal{A}(Y)) = 0$.

   Solution: By definition, $\mathrm{Adv}_{\mathcal{A}}(X, Y) = |\mathbb{P}_{a \leftarrow X}[\mathcal{A}(a) = 1] - \mathbb{P}_{a \leftarrow Y}[\mathcal{A}(a) = 1]|$. Since $\Delta(X, Y) = 0$, we directly obtain that $\mathbb{P}[X = a] = \mathbb{P}[Y = a]$ for all $a \in S$, or in other words, X and Y are identically distributed. As a result, $\mathbb{P}_{a \leftarrow X}[\mathcal{A}(a) = 1] = \mathbb{P}_{a \leftarrow Y}[\mathcal{A}(a) = 1]$ and thus $\mathrm{Adv}_{\mathcal{A}}(X, Y) = 0$. □

   In the next question, we will prove the *data processing inequality* for the statistical distance.

   Let X, Y be two random variables over a common set A.

2. Let $f : A \to S$ be a deterministic function with domain S. Show that

$$\Delta(f(X), f(Y)) \leq \Delta(X, Y).$$

   Solution: We write the definition of Δ.

$$\Delta(f(X), f(Y)) = \frac{1}{2} \sum_{s \in S} |\mathbb{P}(f(X) = s) - \mathbb{P}(f(Y) = s)|$$

   Then decompose the event $f(X) = s$ into something more explicit.

$$\Delta(f(X), f(Y)) = \frac{1}{2} \sum_{s \in S} \left| \sum_{a \in f^{-1}(s)} \mathbb{P}(X = a) - \sum_{a \in f^{-1}(s)} \mathbb{P}(Y = a) \right|$$

   Now use the triangle inequality.

$$\Delta(f(X), f(Y)) \leq \frac{1}{2} \sum_{s \in S} \sum_{a \in f^{-1}(s)} |\mathbb{P}(X = a) - \mathbb{P}(Y = a)|$$

   Finally, recall that $\sqcup_{s \in S} f^{-1}(s) = A$, and this ends the proof. □

3. Let Z be another random variable with domain $\mathcal{Z}$, statistically independent from X and Y. Show that

$$\Delta((X, Z), (Y, Z)) = \Delta(X, Y).$$

   Solution: Once again, we write the definition of the statistical distance.

$$\begin{aligned} \Delta((X, Z), (Y, Z)) &= \sum_{(a,z) \in A \times \mathcal{Z}} |\mathbb{P}(X = a \wedge Z = z) - \mathbb{P}(Y = a \wedge Z = z)| \\ &= \sum_{(a,z) \in A \times \mathcal{Z}} |\mathbb{P}(Z = z) \cdot (\mathbb{P}(X = a) - \mathbb{P}(Y = a))| \\ &= \sum_{z \in \mathcal{Z}} \mathbb{P}(Z = z) \cdot \sum_{a \in A} |\mathbb{P}(X = a) - \mathbb{P}(Y = a)|. \end{aligned}$$

   This is exactly $\Delta(X, Y)$. □

4. Let f be a (possibly probabilistic) function with domain S. Define f' a deterministic function and R a random variable independent from X and Y such that for any input x, we have $f'(x, R) = f(x)$. The random variable R is the internal randomness of f. Using f' and R, show that $\Delta(f(X), f(Y)) = \Delta(f'(X, R), f'(Y, R)) \leq \Delta(X, Y)$.

Solution: We apply the two previous results: $\Delta(f(X), f(Y)) \leq \Delta((X, R), (Y, R)) = \Delta(X, Y)$. □

5. Show that for any (possibly probabilistic) adversary $\mathcal{A}$, we have $\mathrm{Adv}_{\mathcal{A}}(X, Y) \leq \Delta(X, Y)$.

Solution: This follows from the definition of the advantage, and from the above property ($\mathcal{A}$ is a function):

$$\mathrm{Adv}_{\mathcal{A}}(X, Y) = |\mathbb{P}[\mathcal{A}(X) = 1] - \mathbb{P}[\mathcal{A}(Y) = 1]| = \frac{1}{2} \sum_{b \in \{0,1\}} |\mathbb{P}[\mathcal{A}(X) = b] - \mathbb{P}[\mathcal{A}(Y) = b]| = \Delta(\mathcal{A}(X), \mathcal{A}(Y)) \leq \Delta(X, Y).$$

□

Exercice 5. We consider two distributions D_0 and D_1 over $\{0, 1\}^n$. You found a distinguisher $\mathcal{A}$ on internet. However, you cannot find anywhere in the documentation its performances!

1. Assuming that you have access to as many samples as you like from D_0 and D_1 (you can for instance assume that you can sample yourself from these distributions), how would you estimate the advantage of $\mathcal{A}$? *Hint: use the Chernoff Bound:* $\mathbb{P}(|X - np| \geq nt) \leq 2\exp(-2nt^2)$*, where* X *follows a binomial distribution with parameters* (n, p).

Solution: Let $\mathsf{Exp}\ b$ for $b \in \{0, 1\}$ denote the experience "sample from the distribution D_b". Run N times $\mathsf{Exp}\ 0$ and $\mathsf{Exp}\ 1$ for a number N to be determined later. This gives us $b_1^{(1)}, \dots b_1^{(N)}$ and $b_2^{(1)}, \dots, b_2^{(N)}$, $2N$ results. Define

$$\bar{b}_1 := \frac{\sum_{i=1}^{N} b_1^{(i)}}{N} \text{ and } \bar{b}_2 := \frac{\sum_{i=1}^{N} b_2^{(i)}}{N}.$$

Then let p_b be the probability that $\mathcal{A}$ outputs 1 at the end of $\mathsf{Exp}\ b$. The Chernoff bound gives

$$\mathbb{P}(|\bar{b_i} - p_i| \geq \varepsilon) \leq 2\exp(-2N\varepsilon^2),$$

for any accuracy $\varepsilon > 0$. Then notice the following sequence of inequalities:

$$\mathsf{Adv}(\mathcal{A}) = |p_1 - p_0| \leq |p_1 - \bar{b}_1| + |\bar{b}_1 - \bar{b}_0| + |\bar{b}_0 - p_0| \leq 2\varepsilon + |\bar{b}_1 - \bar{b}_0|,$$

where the last inequality holds with probability at least $1 - 4\exp(-2N\varepsilon^2)$. The same sequence can be written by reversing the roles of p_b and $\bar{b}_p$. This gives us $|\mathsf{Adv}(\mathcal{A}) - |\bar{b}_1 - \bar{b}_0|| \leq 2\epsilon$ with probability at least $1 - 4\exp(-2N\varepsilon^2)$.

Assuming that you want to compute the advantage with accuracy $\frac{1}{\lambda^c}$ and probability 0.95, set $\varepsilon := \frac{1}{2\lambda^c}$ and N such that $1 - 4\exp(N/(2\lambda^{2c})) \geq 0.95$ i.e. $N/\lambda^{2c} \geq 2\ln(80) \approx 8.76$. □

By convention, you want to design a distinguisher such that it outputs 1 when it thinks the sample comes from D_1 and 0 otherwise. However, because of the definition of advantage, it is also possible to design distinguishers that do the reverse, and still have the same advantage. For instance, the above distinguisher $\mathcal{A}$ may often be "wrong". This could be troublesome if your aim is to use its output to do further computations. Luckily, there exists a way to transform $\mathcal{A}$ into a distinguisher that is more often right than wrong, whatever it previously did.

2. The definition of advantage given in class may be called Absolute Advantage, for the purpose of this exercise. In this question, we define the Positive Advantage of $\mathcal{A}$ as

$$\mathsf{Adv}_P(\mathcal{A}) := \mathbb{P}(\mathcal{A} \xrightarrow{Exp_1} 1) - \mathbb{P}(\mathcal{A} \xrightarrow{Exp_0} 1).$$

Given a distinguisher $\mathcal{A}$ with Absolute Advantage ε, we build a distinguisher $\mathcal{A}'$ that does the following:

(a) Upon receiving a sample $y \hookleftarrow D_b$, it runs $b' \leftarrow \mathcal{A}(y)$.

(b) It samples $x_0 \hookleftarrow D_0$ and $x_1 \hookleftarrow D_1$ and runs $b_0 \leftarrow \mathcal{A}(x_0)$ and $b_1 \leftarrow \mathcal{A}(x_1)$.

(c) It returns b' if $b_0 = 0$ and $b_1 = 1$. It returns $1 - b'$ if $b_0 = 1$ and $b_1 = 0$.

(d) In any other cases, it returns a uniform bit.

Prove that the Positive Advantage of $\mathcal{A}'$ is ε^2.

Solution: To clarify notation $\mathbb{P}(\mathcal{A} \xrightarrow{Exp_b} 1)$ for any $b \in \{0,1\}$ stands for "the probability that $\mathcal{A}$ outputs 1 knowing that we are in $\mathsf{Exp}\ b$", namely y has been sampled using D_b. The probability that $\mathcal{A}$ outputs 1 in experience $\mathsf{Exp}\ b$ is $p_1(1-p_0)p_b + p_0(1-p_1)(1-p_b) + \frac{1}{2}(p_0p_1 + (1-p_0)(1-p_1))$. The positive advantage of $\mathcal{A}'$ is then:

$$\begin{aligned}
\mathsf{Adv}_P(\mathcal{A}') &= p_1(1-p_0)(p_1-p_0) + p_0(1-p_1)(p_0-p_1) \\
&= (p_1-p_0)\cdot(p_1(1-p_0) - p_0(1-p_1)) \\
&= (p_1-p_0)\cdot(p_1 - p_0p_1 - p_0 + p_0p_1) \\
&= \varepsilon^2.
\end{aligned}$$

□