Propriétés de sécurité des fonctions de hachage

Soit $h : \{0,1\}^* \to \{0,1\}^n$ une fonction de hachage que l'on suppose résistante aux collisions. Soit h' la fonction suivante :

$$h' : \begin{cases} \{0,1\}^* & \to & \{0,1\}^{n+1} \\ x & \mapsto & \begin{cases} 0\|x \text{ si } |x| = n \\ 1\|h(x) \text{ sinon} \end{cases} \end{cases}$$

Question 1. *Montrer que h' est résistante aux collisions.*

Question 2. *Montrer que h' n'est pas résistante aux préimages.*

Variations de Merkle-Damgård

On cherche à concevoir une fonction de hachage sûre basée sur la construction Merkle-Damgård. Dans la suite de cet exercice, on supposera que les blocs de message sont tous complets. De plus, on utilisera une construction de Merkle-Damgård à *deux* fonctions de compression (h, h'), pour laquelle *aucun padding n'est nécessaire*. Un exemple est représenté sur la figure 1. Les deux fonctions $h, h' : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^n$ prennent en entrée un bloc de n bits et une valeur de chaînage de n bits, et sont considérées comme des fonctions aléatoires indépendantes.

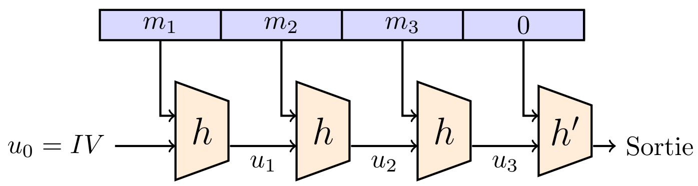

FIGURE 1 – Fonction MD typiquement considérée pour cet exercice.

La complexité en temps des algorithmes sera comptée en évaluations des fonctions h et h'. Lorsque des algorithmes sont demandés, vous pouvez les écrire sous forme de pseudocode peu détaillé.

Étant donné une fonction de compression $h : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^n$, l'itération de h sur les blocs $m_1, \ldots, m_\ell$ en partant de IV est notée :

$$h^*(IV, m_1, \ldots, m_\ell) \ ,$$

par exemple :

$$h^*(IV, m_1, m_2, m_3) = h(h(h(IV, m_1), m_2), m_3) \ .$$

Ainsi la fonction Merkle-Damgård complète, notée H, a pour expression :

$$H(m_1, \ldots, m_\ell) = h'(h^*(IV, m_1, \ldots, m_\ell), 0) \ .$$

Question 3. *On commence avec une taille de bloc (et de chaînage) de 128 bits. Quel est le niveau de sécurité de la fonction H contre les collisions ? (Il n'est pas nécessaire de détailler l'algorithme d'attaque).*

Question 4. *On modifie maintenant la fonction h'. Elle se comporte toujours comme une fonction aléatoire, mais sa sortie est étendue à 256 bits :*

$$h' : \{0,1\}^{128} \times \{0,1\}^{128} \to \{0,1\}^{256}$$

Quel est le niveau de sécurité de la nouvelle fonction H contre les collisions ? Justifiez cette réponse par un algorithme d'attaque simple dont vous estimerez la complexité.

Question 5. *On utilise maintenant deux fonctions de compression $h, h' : \{0,1\}^{256} \times \{0,1\}^{256} \to \{0,1\}^{256}$. On définit une nouvelle construction basée sur Merkle-Damgård, donnée par l'algorithme suivant :*

Entrée : *ℓ blocs de message $(m_1, \dots, m_\ell)$ de 512 bits chacun*

1. *Séparer les blocs en moitiés comme suit : $m'_1 \| m''_1 = m_1, \dots, m'_\ell \| m''_\ell = m_\ell$*
2. *Calculer $t_0 = h'(h^*(0, m'_1, \dots, m'_\ell), 0)$*
3. *Calculer $t_1 = h'(h^*(1, m''_1, \dots, m''_\ell), 0)$*
4. *Renvoyer $t_0 \oplus t_1$*

Donnez un algorithme d'attaque en préimage contre cette fonction et estimez sa complexité. Est-elle plus ou moins sûre qu'une fonction Merkle-Damgård classique à 256 bits de sortie ?

Dans la suite de cet exercice, on définit une fonction de Merkle-Damgård *avec checksum*, de la manière suivante :

$$HC(m_1, \dots, m_\ell) = h'\left(h^*(0, m_1, \dots, m_\ell), \bigoplus_{i=1}^{\ell} m_i\right)$$

Le dernier bloc, précedemment 0, est maintenant le XOR de tous les blocs de message (la checksum). Blocs et valeurs de chaînages sont de taille n, et nous nous intéressons exclusivement à des complexités asymptotiques en n.
Nous allons montrer une attaque en seconde préimage sur cette fonction.

Question 6. *La première étape est de construire une multicollision à $2n$ blocs, c'est-à-dire une série de $2n$ paires de blocs (m_i^0, m_i^1) tels que :*

$$\exists U, \forall b_1, \dots, b_{2n}, h^*(0, m_1^{b_1}, \dots, m_{2n}^{b_{2n}})) = U$$

Donner un algorithme pour cette étape, et sa complexité asymptotique.

Nous admettons le résultat suivant.
Étant donnée une cible $t \in \{0,1\}^n$ quelconque, il existe (avec très grande probabilité) un choix de blocs $(b_1, \dots, b_{2n})$ dans les paires de la multicollision, tel que :

$$\bigoplus_{i=1}^{2n} m_i^{b_i} = t \ .$$

De plus ce choix peut être calculé en temps $\mathcal{O}(n^3)$ à l'aide d'une résolution de système linéaire.

Question 7. *Soit $P = (p_1, \dots, p_{2^k})$ un message de longueur 2^k. Soient :*

$$u_0 := IV, u_1 := h(u_0, p_1), \dots, u_{2^k} := h(u_{2^k-1}, p_{2^k}),$$

les valeurs de chaînage dans la fonction. Donner un algorithme simple pour trouver un bloc m^* tel que $h(U, m^*) \in \{u_1, \dots, u_{2^k}\}$, et donner sa complexité asymptotique.

Question 8. *Soit $m_1, \dots, m_{2n}, m^*$ un message de $2n+1$ blocs tel que pour un certain i :*

- $h^*(IV, m_1, \dots, m_{2n}, m^*) = h^*(IV, p_1, \dots, p_i)$
- $m_1 \oplus m_2 \oplus \dots \oplus m_{2n} \oplus m^* = p_1 \oplus \dots \oplus p_i$

Montrer que $m_1, \dots, m_{2n}, m^*, p_{i+1}, \dots, p_{2^k}$ est une seconde préimage du message P originel.

Question 9. *En déduire une attaque en seconde préimage sur Merkle-Damgård avec checksum, et donner sa complexité.*