Cryptographie Fonctions de hash

Léo Colisson Palais Master CSI 2024 – 2025

leo.colisson-palais@univ-grenoble-alpes.fr
https://leo.colisson.me/teaching.html

Fonctions de hash

Qu'est-ce qu'une fonction de hachage?

Fonction de hachage

Une fonction de hachage est une fonction $h: \{0, 1\}^* \to \{0, 1\}^n$.

Telle qu'elle, définition peu utile, mais en fonction des applications on peut ajouter **plusieurs propriétés** :

- **Résistance aux collisions**: il est difficile de trouver une collision, c'est-à-dire $x \neq x'$ tel que h(x) = h(x')
- Résistance à la première préimage (à sens unique = one-way): étant donné y, il est difficile de trouver x tel que h(x) = y
- Résistance à la seconde préimage: étant donné un x aléatoire, il est difficile de trouver x' ≠ x tel que h(x) = h(x')
- Masquage (hiding): étant donné h(r||x) pour un r suffisamment long et aléatoire, il est difficile de retrouver x
- Universalité: Hypothèse sur l'uniformité de la distribution de sortie











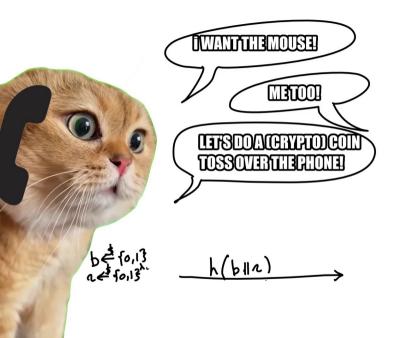


DWANDTHE MOUSE

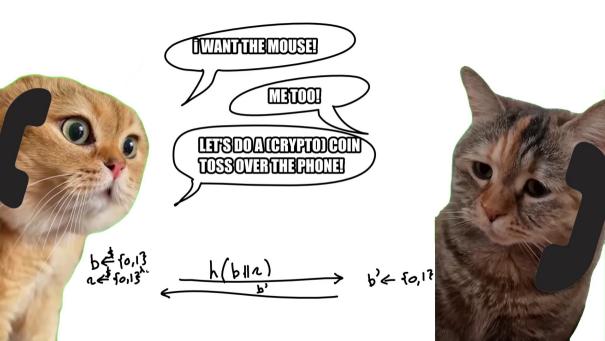
METOOL

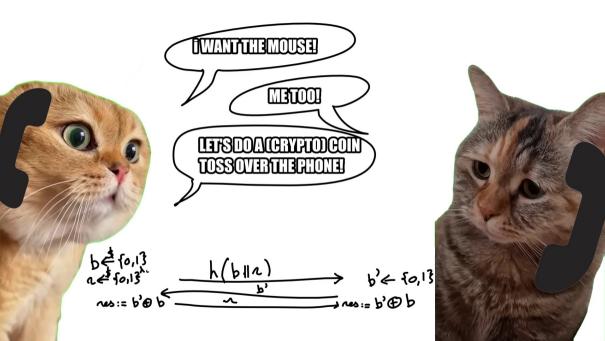
LETSDOACERYPTODEOIN TOSSOVERTHECHONEL















- A Non
- B Oui, pour se protéger contre le chat gauche malveillant
- Oui, pour se protéger contre le chat droit malveillant

Avons-nous besoin de la résistance aux collisions ici?





- B Oui, pour se protéger contre le chat gauche malveillant 🗸 Si le chat gauche connaît r, r' tels que h(0||r) = h(1||r') et souhaite que le résultat soit 0 : il révèle r si b'=0, sinon r' (res $=b'\oplus b=0$)
- Oui, pour se protéger contre le chat droit malveillant 💥



The loser

Me who knew a collision for h





Avons-nous besoin de la propriété de masquage ici?

- A Non
- B Oui, pour se protéger contre le chat gauche malveillant
- Oui, pour se protéger contre le chat droit malveillant

Avons-nous besoin de la propriété de masquage ici?





B Oui, pour se protéger contre le chat gauche malveillant



Oui, pour se protéger contre le chat droit malveillant 🗹 Si le chat droit peut retrouver b à partir de h(b||r) et souhaite que le résultat soit 0, il suffit qu'il envoie $b' \coloneqq b$ (res $= b' \oplus b = 0$).

Applications des fonctions de hachage

Fonctions de hachage = nombreuses applications :

- Vérifier efficacement l'intégrité d'un fichier (empreinte)
- Authentification (HMAC, NMAC, Envelope MAC...)
- Constructions IND-CCA
- Stockage sécurisé des mots de passe
- Organiser, retrouver et/ou mettre en cache des données de manière efficace et/ou sécurisée (git, nix, ...)
- Blockchain (preuve de travail)
- Engagement (=commitments) cryptographiques
- Pile ou face sécurisé
- Preuves à divulgation nulle de connaissance (=*Zero-Knowledge proofs*)
- Calcul multipartite

Comment définir formellement la résistance aux collisions?

Première tentative : h est résistante aux collisions ssi :

 $\frac{\mathcal{L}_0}{\underset{\text{else: return false}}{\underline{\mathsf{TEST}(x,x'):}}} \times \frac{\mathcal{L}_1}{\underset{\text{return false}}{\underline{\mathsf{TEST}(x,x'):}}} \times \frac{\mathcal{L}_1}{\underset{\text{return false}}{\underline{\mathsf{TEST}(x,x'):}}}$

?

Est-ce une définition raisonnable?



- 1 Oui
- 2 Non, car essentiellement toutes les fonctions seraient résistantes aux collisions
- 3 Non, car essentiellement aucune fonction ne serait résistante aux collisions

Est-ce une définition raisonnable?

- 1 Oui
- Non, car essentiellement toutes les fonctions seraient résistantes aux collisions
- Non, car essentiellement aucune fonction ne serait résistante aux collisions

Très **subtile** problème d'ordre des quantificateurs. Cette définition dit : h résistante aux collisions ssi $\forall A$,

 $|\Pr[A \diamond \mathcal{L}_0 = 1] - \Pr[A \diamond \mathcal{L}_1 = 1]| \leq \mathsf{negl}(\lambda)$. Puisque A apparaît après h, A peut dépendre arbitrairement de h. Ainsi A pourrait simplement coder

en dur une collision (x, x'), comme :

retourner TEST(X, X') Il est vraiment

difficile de trouver le code de \mathcal{A} , mais $\overline{\mathcal{A}}$ s'exécute toujours en temps polynomial !

 \Rightarrow Nous aimerions fixer h après A: "grain de sel" (salt) publique



Grain de sel = valeur publique aléatoire échantillonnée pour "personnaliser" la fonction h.

Résistance aux collisions (version 1)

Une fonction de hachage $h\colon \{0,1\}^*\times \{0,1\}^* \to \{0,1\}^*$ est résistante aux collisions si :

$$\mathcal{L}^n_{\mathsf{cr-real}}$$
 $s \not \triangleq \{0,1\}^\lambda$
$$\frac{\mathsf{GETSALT}():}{\mathsf{return}\ s}$$

$$\frac{\mathsf{TEST}(x,x'):}{\mathsf{if}\ x \neq x'} \mathsf{and}\ h(s,x) = h(s,x'): \mathsf{return}\ \mathsf{true}$$
 else: $\mathsf{return}\ \mathsf{false}$

 $\mathcal{L}_{\mathsf{cr-fake}}^{h}$ $s \overset{\$}{\leftarrow} \{0,1\}^{\lambda}$ $\frac{\mathsf{GETSALT}():}{\mathsf{return}\ s}$ $\frac{\mathsf{TEST}(x,x'):}{\mathsf{Return}\ \mathsf{false}}$

Problème : cette définition est **rarement utile** car nous ne vérifions jamais explicitement s'il y a une collision : nous supposons simplement qu'il n'y en a pas.

Plutôt **utilisé dans les réductions** : si \mathcal{A} peut distinguer \mathcal{L}_0 de \mathcal{L}_1 , alors nous pouvons construire \mathcal{A}' (appelant \mathcal{A} en interne) qui trouve une collision contre h (il est alors trivial de distinguer $\mathcal{L}^h_{\text{cr-real}}$ de $\mathcal{L}^h_{\text{cr-fake}}$). Par conséquent, cette définition équivalente pourrait être plus facile à utiliser :

Résistance aux collisions (version 2)

Une fonction de hachage h est résistante aux collisions si pour tout $\mathcal A$ borné polynomialement en temps de calcul :

$$\Pr_{\substack{s \leftarrow \{0,1\}^{\lambda} \ (x,x') \leftarrow \mathcal{A}(s)}} \left[h(s,x) = h(s,x') \right] \leq \mathsf{negl}(\lambda)$$

Spécificité du hachage de mot de passe

Hachage de mot de passe



Alice crée un site web et, pour fournir une sécurité supplémentaire, elle décide de stocker les mots de passe des utilisateurs en les chiffrant avec AES en mode CTR. Est-ce une bonne idée. pourquoi?

- Oui
- B Non

Hachage de mot de passe

Alice crée un site web et, pour fournir une sécurité supplémentaire, elle décide de stocker les mots de passe des utilisateurs en les chiffrant avec AES en mode CTR. Est-ce une bonne idée. pourquoi?



- Oui
- B Non ✓ Pour vérifier les mots de passe, elle a besoin de la clé de déchiffrement, et cette clé restera sur le serveur. Si le serveur est corrompu (base de données volée...), la clé sera probablement aussi volée, révélant les mots de passe.

Hachage des mots de passe

Vous devriez toujours <u>hacher</u> les mots de passe que vous stockez dans une base de données!



Hachage des mots de passe

Si nous ne pouvons pas "déchiffrer" le mot de passe (s_{Alice}, h_{Alice}) (fonction de hachage), comment pouvons-nous vérifier si le mot de passe p est correct?

Hachage des mots de passe

Si nous ne pouvons pas "déchiffrer" le mot de passe (s_{Alice}, h_{Alice}) (fonction de hachage), comment pouvons-nous vérifier si le mot de passe p est correct?

 \Rightarrow Vérifier si $h(s_{Alice}, p) = h_{Alice}!$

Grain de sel

Le grain de sel : utile en théorie... Mais le salt est aussi utile en pratique! (changer le hachage pour chaque mot de passe) Sinon

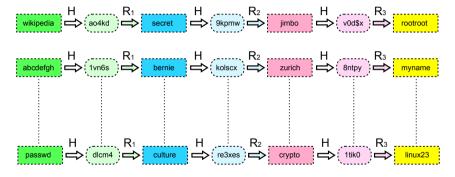
- Facile de voir si deux utilisateurs ont des mots de passe identiques
- Limiter les attaques par pré-calcul

Tables arc-en-ciel (rainbow tables)

Comment (essayer de) récupérer un mot de passe haché sans salt?

- Méthode 1: force brute, recommencer de zéro pour chaque nouveau mot de passe
 - \Rightarrow inefficace en temps $O(\#_{\mathsf{mots}\ \mathsf{de}\ \mathsf{passe}})$, efficace en espace O(1)
- **Méthode 2**: force brute et stockage pour réutilisation ultérieure \Rightarrow efficace en temps une fois la table générée $O(\log\#_{\mathsf{mots}\ \mathsf{de}\ \mathsf{passe}})$, mais nécessite un énorme stockage $O(\#_{\mathsf{mots}\ \mathsf{de}\ \mathsf{passe}})$
- Méthode 3: tables arc-en-ciel = compromis temps/espace \Rightarrow par exemple, temps modéré $O(\sqrt{\#_{\mathsf{mots de passe}}})$, stockage modéré $O(\sqrt{\#_{\mathsf{mots de passe}}})$

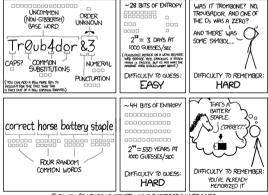
Tables arc-en-ciel



Fonction de réduction différente pour chaque colonne = évite les longues chaînes de collision

Le salage suffit-il?

Le salage est nécessaire (coût d'attaque de n mots de passe = $n \times$ coût d'attaque d'un mot de passe), mais ce n'est pas suffisant : faible entropie des mots de passe = peu de mots de passe utilisés :



THROUGH 20 YEARS OF EFFORT, WE'VE SUCCESSFULLY TRAINED EVERYONE TO USE PASSWORDS THAT ARE HARD FOR HUMANS TO REMEMBER, BUT EASY FOR COMPUTERS TO GUESS.

Comment mitiger encore plus la faible sécurité des mots de passe ?

En temps qu'administrateur : éviter également attaques de fishing ⇒ utiliser authentification à 2 facteurs (2FA)

= RFC 6238 = Time-based One-Time Password TOTP \approx mapToFewDigits(HMAC(k, rounded time))

En temps qu'utilisateur: ne pas ré-utiliser des mots de passes entre sites webs (un seul site corrompu/malicieux et ton seul mot de passe est connu !) ⇒ utilise un gestionnaire de mots de passe pour éviter d'oublier tous tes longs mots de passe !

Recommandations pour le hachage de mots de passe

Mitigation : limiter les attaques par force brute avec des fonctions de hachage lentes (idéalement sur n'importe quel matériel = les fonctions résistantes à la mémoire sont de bons

(idéalement sur n'importe quel matériel = les fonctions résistantes à la mémoire sont de bons candidats : supposons que la mémoire est également coûteuse/rapide partout)



Argon2 (idéalement Argon2id) Scrypt (si Argon2 n'est pas disponible) Bcrypt (systèmes hérités)

PBKDF2 si la conformité FIPS-140 est requise



NTLM (Windows: trop rapide)

MD5 (cassé)

SHA1 (cassé)

SHA256 (trop rapide)

+ il est bon d'ajouter **poivre** (HMAC du hachage, avec une clé stockée en dehors de la base de données en cas d'injection SQL/accès de sauvegarde)

Recommandations pour le hachage de mots de passe

Mitigation : limiter les attaques par force brute avec des fonctions de hachage lentes

(idéalement sur n'importe quel matériel = les fonctions résistantes à la mémoire sont de bons candidats : supposons que la mémoire e dans le Password Storage Cheat Sheet, testing quide...



Argon2 (idéalement Argon2id)

Scrypt (si Argon2 n'est pas disponible)

Bcrypt (systèmes hérités)

PBKDF2 si la conformité FIPS-140 est requise

XÀ ne pas utiliser!

NTLM (Windows: trop rapide)

MD5 (cassé)

SHA1 (cassé)

SHA256 (trop rapide)

+ il est bon d'ajouter **poivre** (HMAC du hachage, avec une clé stockée en dehors de la base de données en cas d'injection SQL/accès de sauvegarde)

Construction de fonctions de hachage

Merkle-Damgård

Fonction de compression de taille fixe

Transformation Merkle-Damgård

Fonction de hachage

Merkle-Damgård est utilisé dans :

- MD5 (cassé)
- SHA-1 (cassé)
- SHA-2 (toujours sûr)

Merkle-Damgård

Construction de Merkle-Damgård

Soit $h: \{0,1\}^{n+t} \to \{0,1\}^n$ une fonction de compression. Alors la transformation de Merkle-Damgård de h est $\mathsf{MD}_h: \{0,1\}^* \to \{0,1\}^n$ avec :

```
 \frac{\text{MDPAD}_t(x)}{\ell := |x|, \text{ as length-}t \text{ binary number} } \\ \ell := |x|, \text{ as length-}t \text{ binary number} \\ \text{while } |x| \text{ not a multiple of }t: \\ x := x \parallel 0 \\ \text{return } x \parallel \ell   return x \parallel \ell   x := x \parallel 0 \\ \text{output } y_{k+1}   x := x \parallel 1 \\ x := x \parallel 1
```

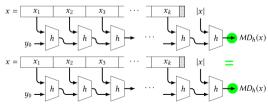
(en réalité, h(x) est défini seulement si $x < 2^t$ ici, mais nous pouvons améliorer la partie de padding)

Merkle-Damgård

Théorème (Merkle-Damgård est résistant aux collisions)

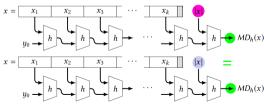
Soit $h: \{0,1\}^{n+t} \to \{0,1\}^n$ une fonction de compression résistante aux collisions. Alors la transformation de Merkle-Damgård MD_h est résistante aux collisions.

Preuve. Par contradiction, supposons que nous ayons trouvé une collision $x \neq x'$ contre MD_h , nous voulons trouver une collision contre h:



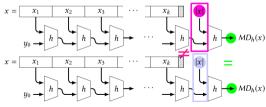
Théorème (Merkle-Damgård est résistant aux collisions)

Soit $h: \{0,1\}^{n+t} \to \{0,1\}^n$ une fonction de compression résistante aux collisions. Alors la transformation de Merkle-Damgård MD_h est résistante aux collisions.



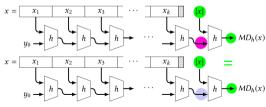
Théorème (Merkle-Damgård est résistant aux collisions)

Soit $h: \{0,1\}^{n+t} \to \{0,1\}^n$ une fonction de compression résistante aux collisions. Alors la transformation de Merkle-Damgård MD_h est résistante aux collisions.



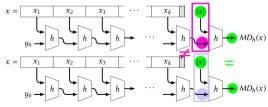
Théorème (Merkle-Damgård est résistant aux collisions)

Soit $h: \{0,1\}^{n+t} \to \{0,1\}^n$ une fonction de compression résistante aux collisions. Alors la transformation de Merkle-Damgård MD_h est résistante aux collisions.



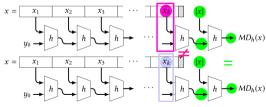
Théorème (Merkle-Damgård est résistant aux collisions)

Soit $h: \{0,1\}^{n+t} \to \{0,1\}^n$ une fonction de compression résistante aux collisions. Alors la transformation de Merkle-Damgård MD_h est résistante aux collisions.



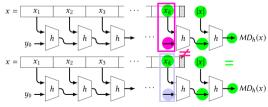
Théorème (Merkle-Damgård est résistant aux collisions)

Soit $h: \{0,1\}^{n+t} \to \{0,1\}^n$ une fonction de compression résistante aux collisions. Alors la transformation de Merkle-Damgård MD_h est résistante aux collisions.



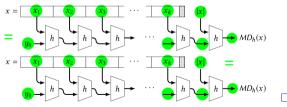
Théorème (Merkle-Damgård est résistant aux collisions)

Soit $h: \{0,1\}^{n+t} \to \{0,1\}^n$ une fonction de compression résistante aux collisions. Alors la transformation de Merkle-Damgård MD_h est résistante aux collisions.



Théorème (Merkle-Damgård est résistant aux collisions)

Soit $h: \{0,1\}^{n+t} \to \{0,1\}^n$ une fonction de compression résistante aux collisions. Alors la transformation de Merkle-Damgård MD_h est résistante aux collisions.



Attaque par extension de longueur

Objectif: Obtenir un code d'authentification de message (MAC, = "signature", voir le prochain cours) à partir de fonctions de hachage via h(k||m).

Problème: Attaque par extension de longueur: Avec la construction MD, il est possible d'obtenir h(k||m') à partir de h(k||m) (= signer un message différent sans connaître k)

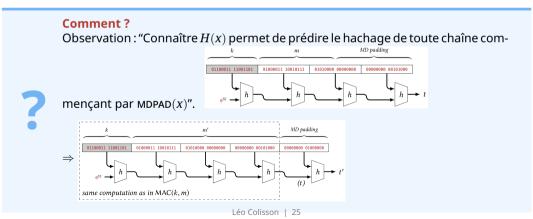
Comment?



Attaque par extension de longueur

Objectif: Obtenir un code d'authentification de message (MAC, = "signature", voir le prochain cours) à partir de fonctions de hachage via h(k||m).

Problème: Attaque par extension de longueur: Avec la construction MD, il est possible d'obtenir h(k||m') à partir de h(k||m) (= signer un message différent sans connaître k)

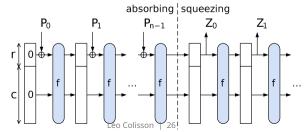


Mitigation de l'attaque par extension de longueur

L'attaque par extension de longueur est possible car le hachage **contient**l'état interne complet

Solutions :

- Construction Wide-pipe: Ne renvoyer, par exemple, la moitié du hachage final. Utilisée dans SHA-512/224 et SHA-512/256 (famille SHA-2), tandis que SHA-512 et SHA-256 sont vulnérables à cette attaque.
- Construction éponge (=Sponge): deux phases d'absorption et d'extraction, utilisée dans SHA-3



Construction d'une fonction de compression

Comment obtenir des fonctions de compression?

Construire une fonction de compression $h: \{0,1\}^{n+t} \to \{0,1\}^n$:

- À partir de zéro
- À partir d'un chiffrement par blocs (par exemple, AES) \mathcal{E} , choisir ce qui définit la clé/message, ajouter une rétroaction (sinon inversible) :
 - Davies–Meyer : $h(x||k) := \mathcal{E}_k(x) \oplus x$ (par exemple, utilisé dans SHA-2 avec un chiffrement personnalisé)
 - Matyas-Meyer-Oseas : $h(x||x') \coloneqq \mathcal{E}_{g(x)}(x') \oplus x'$
 - Miyaguchi-Preneel : $h(x||x') \coloneqq \mathcal{E}_{g(x)}(x') \oplus x' \oplus x$
 - Hirose

Les preuves sont généralement faites dans le **modèle Ideal Cipher Model** pour modéliser $\mathcal E$

Modèles de sécurité

Modèles idéaux

Dans le modèle de chiffrement idéal (resp. modèle d'oracle aléatoire), nous supposons qu'une fonction se comporte comme une permutation (resp. fonction) uniformément échantillonnée aléatoirement, à laquelle les parties (y compris l'attaquant) ne peuvent accéder que de manière boîte noire (pour les chiffrements idéaux, les parties peuvent également demander l'inverse de la fonction). Mais :

- En pratique, nous devons l'instancier avec une permutation réelle (resp. fonction), par exemple AES (resp. SHA-3) :
 - ⇒ nous avons alors une **sécurité heuristique** (pas de réduction)
- Il existe des schémas (pathologiques) sécurisés dans le ROM [Bellare, Boldyreva, Palacio 03] mais impossibles à instancier
- Pourtant, aucune construction non pathologique n'est connue pour être sécurisée dans le ROM mais non sécurisée en pratique

Modèle idéalisé ≠ modèle standard

Microsoft avait besoin d'une fonction de hachage pour la vérification d'intégrité ROM pour la XBOX :

- Ils ont utilisé l'algorithme de chiffrement Tiny Encryption Algorithm (TEA, chiffrement par blocs) comme chiffrement de base avec Davies-Meyer ¹
- Problème : pour tout k, il est facile de trouver k' tel que $\mathsf{TEA}_k(m) = \mathsf{TEA}_{k'}(m)$ (comme inverser un bit de k), et \Rightarrow Trivial d'obtenir une collision :

$$\mathsf{DM} - \mathsf{TEA}(x \| k') = \mathsf{TEA}_{k'}(x) = \mathsf{TEA}_k(x) = \mathsf{DM} - \mathsf{TEA}(x \| k)$$

 Pourtant, TEA est toujours un bon PRP (une fois que nous choisissons un k aléatoire)!

¹Détails de l'attaque dans [Steil, 2005] https://events.ccc.de/congress/2005/fahrplan/events/559.en.html

Modèle de l'oracle aléatoire

Modèle de l'oracle aléatoire (ROM)

Un protocole est dit défini dans le modèle de l'oracle aléatoire (ROM) si toutes les parties (y compris les parties honnêtes lors de la définition du protocole et les adversaires) ont un accès oracle à H défini comme suit :

Random Oracle

```
T \coloneqq 	ext{empty assoc. array} \ rac{H(x \in \{ m{0}, m{1} \}^*):}{	ext{if } T[x] 	ext{ undefined:}} \ T[x] \overset{\$}{\leftarrow} \{ m{0}, m{1} \}^{	ext{out}} \ 	ext{return } T[x]
```

Modèle de l'oracle aléatoire

Remarques:

 Échantillonnage paresseux plutôt que de déterminer complètement H au début = nécessaire dans les réductions pour avoir un adversaire polynomial :

Attaquant contre le problème YYY (poly(
$$\lambda$$
))

A (poly(λ))

ROM (poly(λ))

• ROM \neq PRF !!! Dans le modèle ROM, les parties ont seulement un accès oracle à H, alors que dans le cas d'une PRF, les parties peuvent également voir le "code" de H. Cela permet de nouvelles techniques de preuve !

Modèle du chiffrement idéal

Modèle du chiffrement idéal = choisir une permutation aléatoire pour chaque clé

Modèle du chiffrement idéal (ICM)

Un protocole est dit défini dans le modèle du chiffrement idéal (ICM) si toutes les parties (y compris les parties honnêtes lors de la définition du protocole et les adversaires) ont un accès oracle à F et F^{-1} définis ainsi :

```
Ideal-Cipher Model
T := \text{empty assoc. array of assoc. array}
F(k \in \{0, 1\}^{\lambda}, x \in \{0, 1\}^{blen}):
  if T[k][x] undefined:
      T[k][x] \stackrel{\$}{\leftarrow} \{ 0, 1 \}^{\text{blen}} \setminus T[k]. \text{values}
  return T[k][x]
F^{-1}(k \in \{0, 1\}^{\lambda}, y \in \{0, 1\}^{blen}):
  if \exists x \text{ s.t. } T[k][x] = y:
      return x
  else:
      x \stackrel{\$}{\leftarrow} \{ 0, 1 \}^{\text{blen}} \setminus T[k]. \text{kevs}
      T[k][x] := y
```

Fonctions de hash principales

Comparaison des principales fonctions de hachage

Pour résumer : utilisez SHA-3, ou SHA-2 (mais pas pour les MAC). **N'utilisez jamais MD5**, SHA-0, SHA-1

Algorithm and variant		Output size (bits)	Internal state size (bits)	Block size (bits)	Rounds	Operations	Security against collision attacks (bits)	Security against length extension attacks (bits)	Performance on Skylake (median cpb) [61]		
									Long messages	8 bytes	First published
MD5 (as reference)		128	128 (4 × 32)	512	4 (16 operations in each round)	And, Xor, Or, Rot, Add (mod 2 ³²)	≤ 18 (collisions found) ^[62]	0	4.99	55.00	1992
SHA-0		160	160 (5 × 32)	512	80	And, Xor, Or, Rot, Add (mod 2 ³²)	< 34 (collisions found)	0	≈ SHA-1	≈ SHA-1	1993
SHA-1							< 63 (collisions found) ^[63]		3.47	52.00	1995
SHA-2	SHA-224 SHA-256	224 256	256 (8 × 32)	512	64	And, Xor, Or, Rot, Shr, Add (mod 2 ³²)	112 128	32 0	7.62 7.63	84.50 85.25	2004 2001
	SHA-384	384	512 (8 × 64)	1024	80	And, Xor, Or, Rot, Shr, Add (mod 2 ⁶⁴)	192	128	5.12	135.75	2001
	SHA-512	512					256	0[64]	5.06	135.50	2001
	SHA-512/224 SHA-512/256	224 256					112 128	288 256	≈ SHA-384	≈ SHA-384	2012
SHA-3	SHA3-224 SHA3-256 SHA3-384 SHA3-512	224 256 384 512	1600 (5 × 5 × 64)	1152 1088 832 576 1344 1088	24[65]	And, Xor, Rot, Not	112 128 192 256	448 512 768 1024	8.12 8.59 11.06 15.88	154.25 155.50 164.00 164.00	2015
	SHAKE128 SHAKE256	d (arbitrary) d (arbitrary)					min(d/2, 128) min(d/2, 256)	256 512	7.08 8.59	155.25 155.50	