Cas d'etude d'une attaque cryptographique - cryptanalyse de MD4 -

Duco van Amstel

ENS Lyon - Université Lyon 1

24/10/2011

Structure générale du cours

Prochains points abordés

La raison d'être du hashage

Cadre :

- Transfert d'un fichier
- Partage de fichiers en Pair-à-Pair
- Stockage longue durée
- ... ou sur un support vulnérable

La raison d'être du hashage

Cadre :

- Transfert d'un fichier
- Partage de fichiers en Pair-à-Pair
- Stockage longue durée
- ... ou sur un support vulnérable

Vérification de l'intégrité d'un message

- Y a-t-il eu des erreurs de transmission ?
- Quelqu'un a-t-il modifié le message ?

Capturer l'essence d'une information...

Intuition : enregistrer une caractéristique essentielle du message

Capturer l'essence d'une information...

Intuition : enregistrer une caractéristique essentielle du message

Idée naïves :

- Enregistrer une copie
- Enregistrer une partie du message

Capturer l'essence d'une information...

Intuition : enregistrer une caractéristique essentielle du message

Idée naïves :

- Enregistrer une copie
- Enregistrer une partie du message

Methodes de vérification simple :

- Checksum = bit paritaire
- CRC = Cyclic Redundancy Checksum
- Codes correcteurs d'erreur

...cryptographiquement

Methode cryptographique : Fonctions « trapdoor » à sens unique

- En connaissant `M` on obtient facilement `trapdoor(M)`
- En connaissant `trapdoor(M)` il est difficile de retrouver M

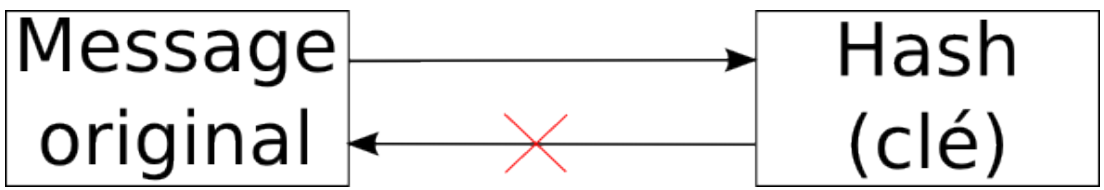
Ex : fonctions cryptographiques

...cryptographiquement

Methode cryptographique : Fonctions « trapdoor » à sens unique

- En connaissant `M` on obtient facilement `trapdoor(M)`
- En connaissant `trapdoor(M)` il est difficile de retrouver M

Ex : fonctions cryptographiques



Portrait-robot de l'attaquant λ

Son identité : *le même type d'individu que dans tous les problèmes de cryptographie. Charlie, Trudy, etc...*

Portrait-robot de l'attaquant λ

Son identité : *le même type d'individu que dans tous les problèmes de cryptographie. Charlie, Trudy, etc...*

Son but : *faire passer un message modifié pour authentique, modifier un message en préservant le hash*

Portrait-robot de l'attaquant λ

Son identité : *le même type d'individu que dans tous les problèmes de cryptographie. Charlie, Trudy, etc...*

Son but : *faire passer un message modifié pour authentique, modifier un message en préservant le hash*

Ses cibles : *trafic bancaire, documents administratifs, correspondance confidentielle, etc...*

Portrait-robot de l'attaquant λ

Son identité : *le même type d'individu que dans tous les problèmes de cryptographie. Charlie, Trudy, etc...*

Son but : *faire passer un message modifié pour authentique, modifier un message en préservant le hash*

Ses cibles : *trafic bancaire, documents administratifs, correspondance confidentielle, etc...*

Ses moyens : *les collisions*

Zoom sur les collisions

Trois attaques de fonctions de hachage : trois « niveaux » de sécurité

Collision

Trouver deux messages `M` et `M´` tels que `hash(M)` $=$ `hash(M´)`

Zoom sur les collisions

Trois attaques de fonctions de hachage : trois « niveaux » de sécurité

Collision

Trouver deux messages `M` et `M´` tels que `hash(M)` $=$ `hash(M´)`

Attaque avec première pré-image

En connaissant par avance un hash `h` d'un message à contenu inconnu trouver un message `M` tel que `hash(M)` $=$ `h`

Zoom sur les collisions

Trois attaques de fonctions de hachage : trois « niveaux » de sécurité

Collision

Trouver deux messages `M` et `M´` tels que `hash(M)` = `hash(M´)`

Attaque avec première pré-image

En connaissant par avance un hash `h` d'un message à contenu inconnu trouver un message `M` tel que `hash(M)` = `h`

Attaque avec seconde pré-image

En connaissant par avance un message `M` et le hash `h` correspondant trouver un message `M´` tel que `hash(M´)` = `h`

Prochains points abordés

Block après block

Fait concrèt : données de taille aléatoire

Traitement block par block $\longrightarrow$ permet d'avoir une fonction unique

Block après block

Fait concrèt : données de taille aléatoire

Traitement block par block $\longrightarrow$ permet d'avoir une fonction unique

Padding

La methode du « padding » permet d'obtenir un message avec une longueur qui est un multiple exact de la taille de block utilisée dans un algorithme

Block après block

Fait concrèt : données de taille aléatoire

Traitement block par block $\longrightarrow$ permet d'avoir une fonction unique

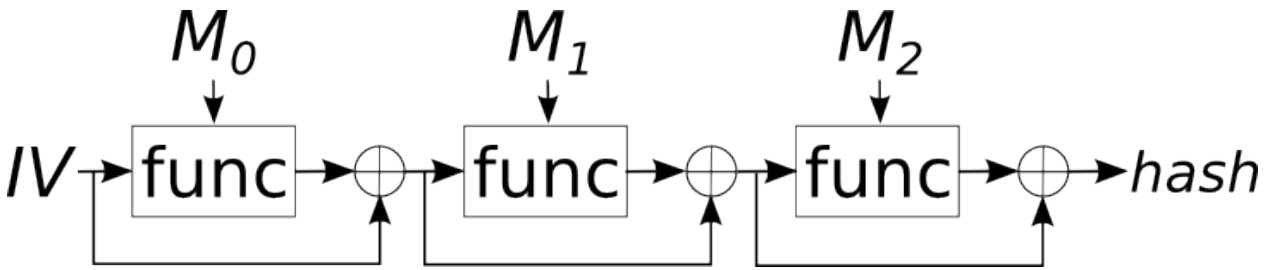
Padding

La methode du « padding » permet d'obtenir un message avec une longueur qui est un multiple exact de la taille de block utilisée dans un algorithme

Exemple : message de longueur 320 bits, blocks de 512 bits
> ajout d'un 1 suivi de 0's jusqu'à obtenir une longueur de 448 bits
> ajout de la longueur du message initial (320) codée sur 64 bits

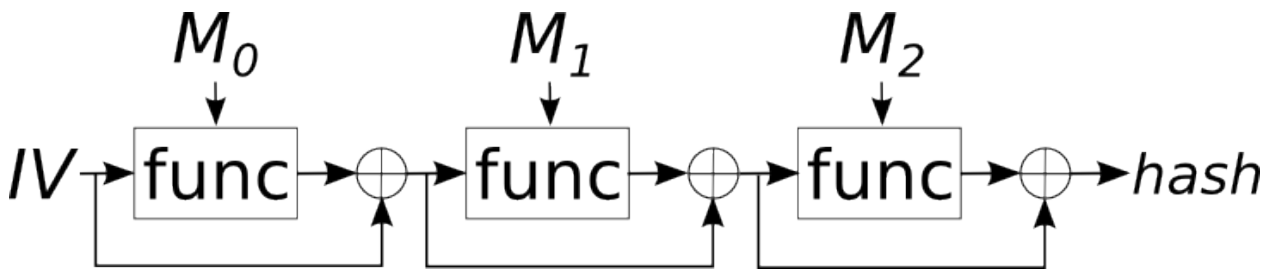
Merkle-Damgård

On enchaîne les blocks de façon itérative



Merkle-Damgård

On enchaîne les blocks de façon itérative



L'algorithme de hachage doit donc spécifier :

- un IV (vecteur d'initialisation)
- une taille de block
- un methode pour obtenir un message de longueur adéquate

Paramètres du MD4

Les faits divers :

- Message Digest 4
- Définie en 1990 par Ronald Rivest
- Version simplifiée de son successeur MD5
- Cassé pour la première fois en 1998

Paramètres du MD4

Les faits divers :

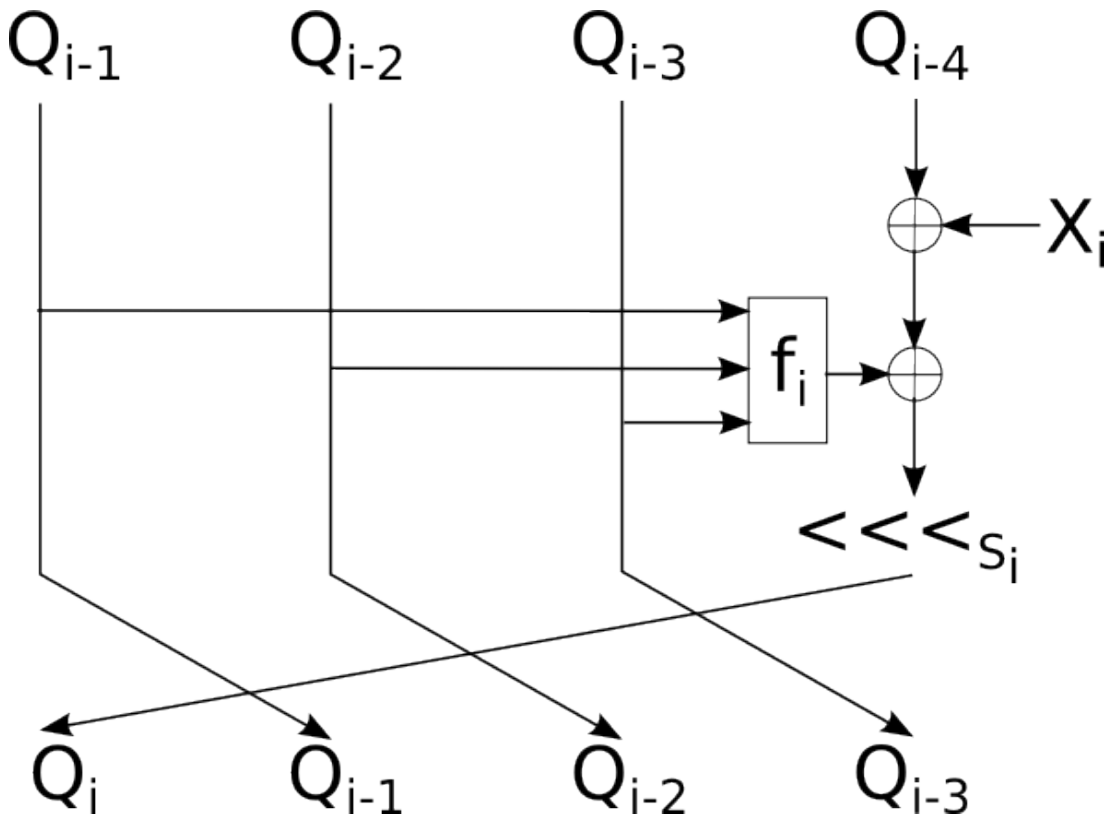
- Message Digest 4
- Définie en 1990 par Ronald Rivest
- Version simplifiée de son successeur MD5
- Cassé pour la première fois en 1998

Les paramètres de l'algorithme :

- Hash de 128 bits de long
- Blocks composés de 16 mots de 32-bits, 512 bit au total
- $IV = \{$`0x67452301, 0xefcdab89, 0x98badcfe, 0x10325476`$\}$
- Trois constantes `K1`, `K2` et `K3`

Confusion

Les Q_i sont des variables intermédiaires
Les X_i les sous-mots du block
Les f_i sont des fonctions logiques nommées F, G et H



Diffusion

MD4 : 48 itérations du schéma présenté

16 sous-mots $\Rightarrow$ chacun apparaît plusieurs fois dans les X_i
Variation du s_i : diffusion éfficace de la confusion de chaque tour

Tour i	Fonction	Rotation	Sous-mot
1	F	3	0
2	F	7	1
⋮	⋮	⋮	⋮
48	H	15	15

Choix, vote & parité

Les fonctions logiques utilisées sont :

$$\begin{aligned} F(A,B,C) &= (A \wedge B) \vee (\neg A \wedge C) \\ G(A,B,C) &= (A \wedge B) \vee (B \wedge C) \vee (C \wedge A) \\ H(A,B,C) &= A \oplus B \oplus C \end{aligned}$$

Choix, vote & parité

Les fonctions logiques utilisées sont :

$$\begin{aligned} F(A,B,C) &= (A \wedge B) \vee (\neg A \wedge C) \\ G(A,B,C) &= (A \wedge B) \vee (B \wedge C) \vee (C \wedge A) \\ H(A,B,C) &= A \oplus B \oplus C \end{aligned}$$

F prend un bit de A pour sélectionner le bit de B ou de C qui correspond
i.e : F(01001101, 10110011, 00111010) = 00110011

G est un vote majoritaire entre A, B et C
i.e : G(01001101, 10110011, 00111010) = 00111011

H est un comptage de parité
i.e : H(01001101, 10110011, 00111010) = 11000100

Prochains points abordés

Les boîte à outils du hacker

Plusieurs façons d'aborder la cryptanalyse d'une fonction de hashage :

1. Calculer des hash de blocks au hasard : attaque par force brute
2. Invariants de la fonction : exploiter une faiblesse mathématique
3. Etudier la propagation d'une minime différence donnée en entrée : cryptanalyse différentielle

Les boîte à outils du hacker

Plusieurs façons d'aborder la cryptanalyse d'une fonction de hashage :

1. Calculer des hash de blocks au hasard : attaque par force brute
2. Invariants de la fonction : exploiter une faiblesse mathématique
3. Etudier la propagation d'une minime différence donnée en entrée : cryptanalyse différentielle

Cryptanalyse différentielle

Observer et contrôler la divergence dans le calcul des hash de deux blocks différents. Les deux blocks diffèrent de façon minimaliste : typiquement 1 à 5 bits sur un block de 512 bits.

Les boîte à outils du hacker

Plusieurs façons d'aborder la cryptanalyse d'une fonction de hashage :

1. Calculer des hash de blocks au hasard : attaque par force brute
2. Invariants de la fonction : exploiter une faiblesse mathématique
3. Etudier la propagation d'une minime différence donnée en entrée : cryptanalyse différentielle

Cryptanalyse différentielle

Observer et contrôler la divergence dans le calcul des hash de deux blocks différents. Les deux blocks diffèrent de façon minimaliste : typiquement 1 à 5 bits sur un block de 512 bits.

Attaque étudiée ici :

- Présentée en 1998 par Hans Dobbertin
- Première attaque éfficace connue sur MD4
- Signifie le passage complèt à MD5 et SHA-1

Une attaque par collision

Calcul de deux messages avec le même hash

$\Rightarrow$ Attaque par collision

Une attaque par collision

Calcul de deux messages avec le même hash

$\Rightarrow$ Attaque par collision

Rappel : Attaque par collision

Trouver deux messages `M` et `M´` tels que `hash(M)` $=$ `hash(M´)`

Attaque de Dobbertin :

- Pas de conditions particulières au départ
- Différence sur un seul sous-mot de 32-bits du block

Attaque en trois phases

Rappel : 48 itérations
Notations : M, X_i et Q_i (resp. M', X'_i et Q'_i), $\Delta Q_i = Q'_i - Q_i$

Attaque en trois phases

Rappel : 48 itérations
Notations : M, X_i et Q_i (resp. M', X'_i et Q'_i), $\Delta Q_i = Q'_i - Q_i$

Première phase :
Trouver Q_{19}, Q_{18}, Q_{17} et Q_{16} tels que

$$(\Delta Q_{19}, \Delta Q_{18}, \Delta Q_{17}, \Delta Q_{16}) = (2^{25}, -2^5, 0, 0)$$

et

$$(\Delta Q_{35}, \Delta Q_{34}, \Delta Q_{33}, \Delta Q_{32}) = (0, 0, 0, 0)$$

Attaque en trois phases

Rappel : 48 itérations
Notations : M, X_i et Q_i (resp. M', X'_i et Q'_i), $\Delta Q_i = Q'_i - Q_i$

Première phase :
Trouver Q_{19}, Q_{18}, Q_{17} et Q_{16} tels que

$$(\Delta Q_{19}, \Delta Q_{18}, \Delta Q_{17}, \Delta Q_{16}) = (2^{25}, -2^5, 0, 0)$$

et

$$(\Delta Q_{35}, \Delta Q_{34}, \Delta Q_{33}, \Delta Q_{32}) = (0, 0, 0, 0)$$

Deuxième phase :
Résoudre un système d'équations de façon itérative
$\longrightarrow$ Déterminer les Q_{11} à Q_{15}

Attaque en trois phases, suite.

Troisième phase :
Déduire les X_i du message M et donc le message M'

Choix du moindre effort

La différentielle employée est : $X'_{12} = X_{12} + 1$

Choix du moindre effort

La différentielle employée est : $X'_{12} = X_{12} + 1$

Étude du tableau d'utilisation des X_i et leur réutilisation :

i	0	1	2	3	4	5	6	7
Tours d'utilisation	33	40	35	42	31	38	32	40
i	8	9	10	11	12	13	14	15
Tours d'utilisation	26	33	28	35	24	31	26	33

Choix du moindre effort

La différentielle employée est : $X'_{12} = X_{12} + 1$

Étude du tableau d'utilisation des X_i et leur réutilisation :

i	0	1	2	3	4	5	6	7
Tours d'utilisation	33	40	35	42	31	38	32	40
i	8	9	10	11	12	13	14	15
Tours d'utilisation	26	33	28	35	24	31	26	33

Trouver les $Q_i \equiv$ brute-force local
Respecter un « chemin » de ΔQ_i prédéfinie

Chemin différentiel

	Δ_i						
i	ΔQ_i	ΔQ_{i-1}	ΔQ_{i-2}	ΔQ_{i-3}	f_i	s_i	p_i
19	2^{25}	-2^5	0	0	*	*	*
20	0	2^{25}	-2^5	0	G	3	1
21	0	0	2^{25}	-2^5	G	5	1/9
22	-2^{14}	0	0	2^{25}	G	9	1/3
23	2^6	-2^{14}	0	0	G	13	1/3
24	0	2^6	-2^{14}	0	G	3	1/9
25	0	0	2^6	-2^{14}	G	5	1/9
26	-2^{23}	0	0	2^6	G	9	1/3
27	2^{19}	-2^{23}	0	0	G	13	1/3
28	0	2^{19}	-2^{23}	0	G	3	1/9
29	0	0	2^{19}	-2^{23}	G	5	1/9
30	-1	0	0	2^{19}	G	9	1/3
31	1	-1	0	0	G	13	1/3
32	0	1	-1	0	H	3	1/3
33	0	0	1	-1	H	9	1/3
34	0	0	0	1	H	11	1/3
35	0	0	0	0	H	15	1

Raffinement de solutions partielles

La définition de la fonction MD4 donne plusieurs équations

- Résolution difficile en principe
- On procède petit à petit
- Détermination des bits un par un

Raffinement de solutions partielles

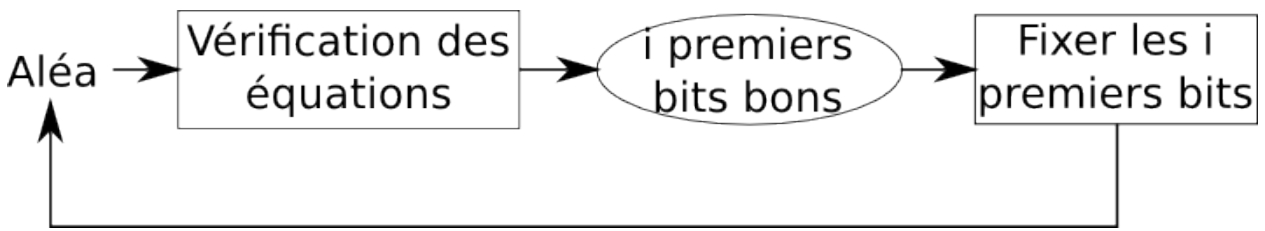
La définition de la fonction MD4 donne plusieurs équations

- Résolution difficile en principe
- On procède petit à petit
- Détermination des bits un par un

Il s'agit de la contribution la plus importante de Dobbertin

$\longrightarrow$ Il utilise la structure itérative de MD4

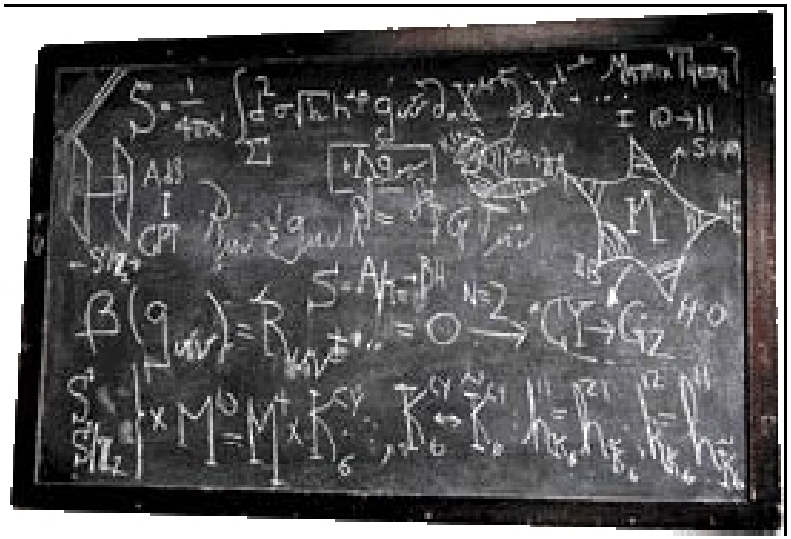
$\longrightarrow$ Une méthode de raffinement de résultats déja obtenus



Prochains points abordés

Warning

ATTENTION ! Ce qui va suivre est sensiblement plus mathématique !



Chance de réussite par étape

Dans le tableau présenté nous distinguons trois cas :

Chance de réussite par étape

Dans le tableau présenté nous distinguons trois cas :

1. Sur le modèle de l'étape $28 \longrightarrow 29$

$$\Delta_{28} = (0, 2^{19}, -2^{23}, 0) \quad | \quad \Delta Q_{29} = 0$$

Chance de réussite par étape

Dans le tableau présenté nous distinguons trois cas :

1. Sur le modèle de l'étape $28 \longrightarrow 29$

$$\Delta_{28} = (0, 2^{19}, -2^{23}, 0) \quad | \quad \Delta Q_{29} = 0$$

D'après la définition de MD4

$$Q_{29} = (Q_{25} + G(Q_{28}, Q_{27}, Q_{26}) + X_7) <<< 5$$

$$Q'_{29} = (Q'_{25} + G(Q'_{28}, Q'_{27}, Q'_{26}) + X_7) <<< 5$$

Chance de réussite par étape

Dans le tableau présenté nous distinguons trois cas :

1. Sur le modèle de l'étape $28 \longrightarrow 29$

$$\Delta_{28} = (0, 2^{19}, -2^{23}, 0) \quad | \quad \Delta Q_{29} = 0$$

D'après la définition de MD4

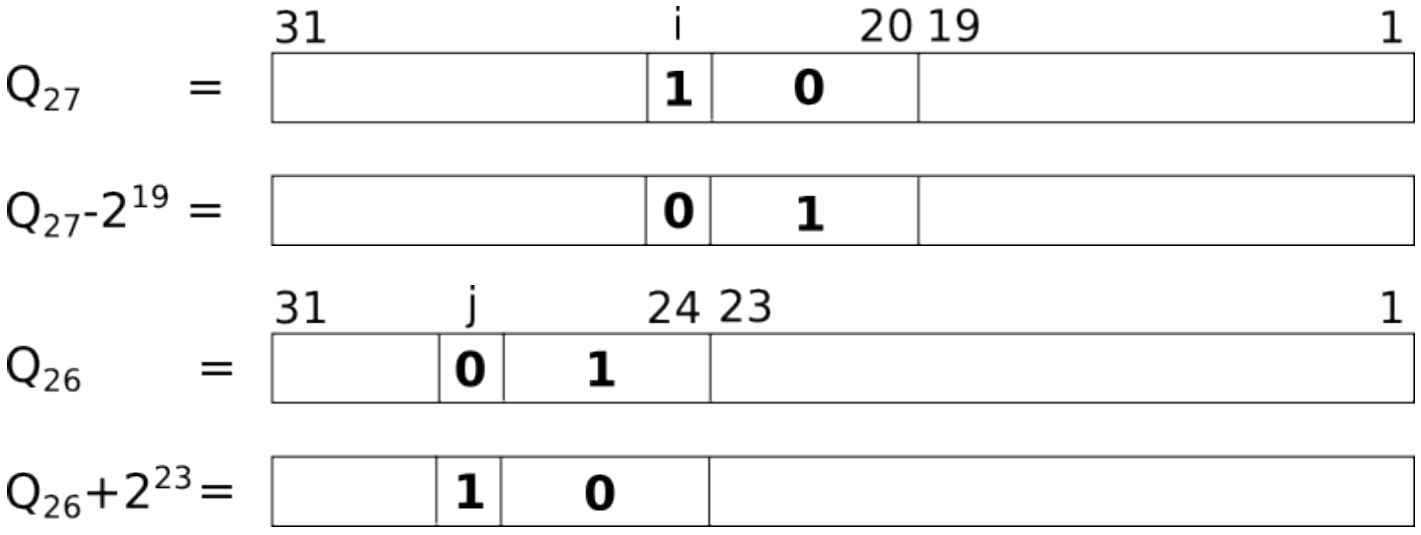
$$Q_{29} = (Q_{25} + G(Q_{28}, Q_{27}, Q_{26}) + X_7) <<< 5$$

$$Q'_{29} = (Q'_{25} + G(Q'_{28}, Q'_{27}, Q'_{26}) + X_7) <<< 5$$

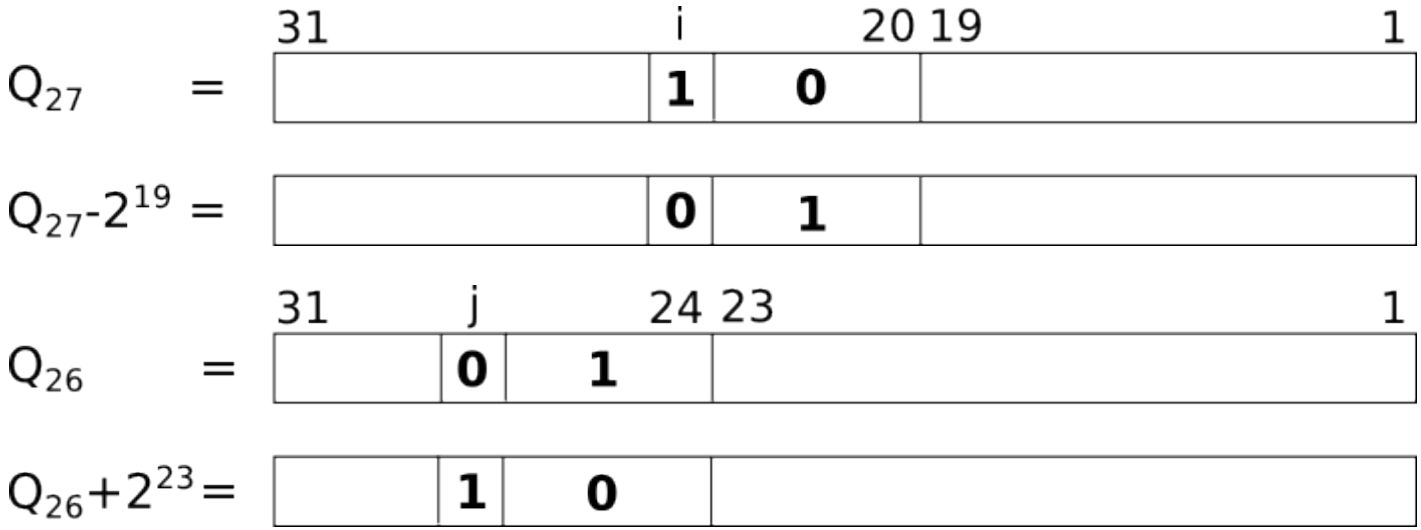
Or $\Delta Q_{25} = 0$ et $X_7 = X'_7$. De là on cherchera la probabilité que

$$G(Q_{28}, Q_{27}, Q_{26}) = G(Q_{28}, Q_{27} - 2^{19}, Q_{26} + 2^{23})$$

Première étape



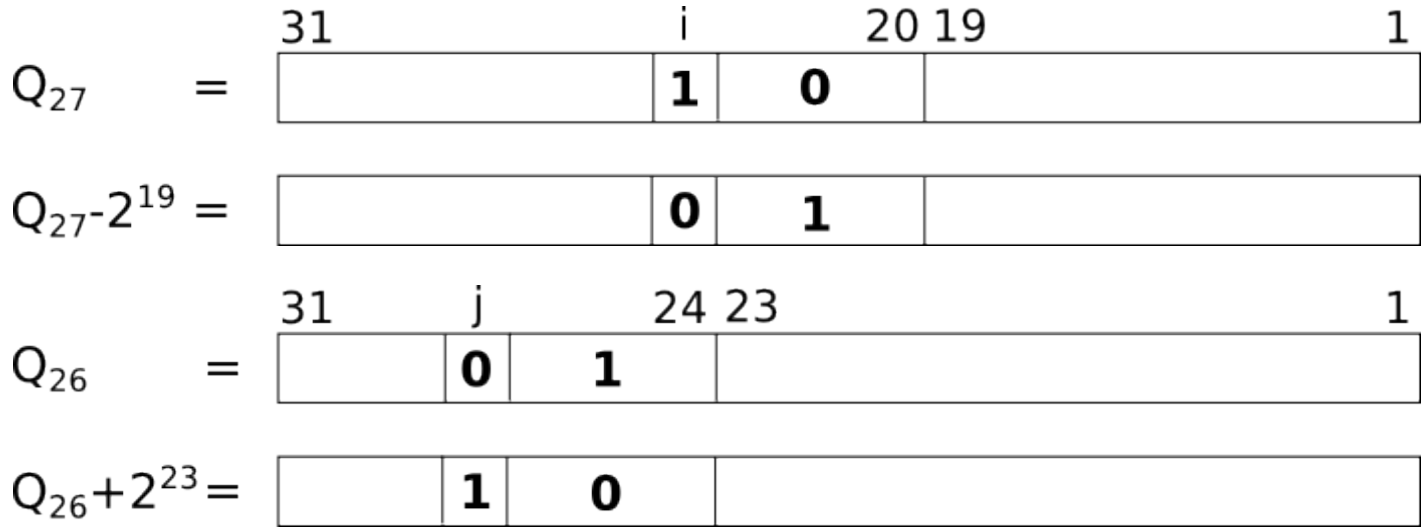
Première étape



Définissons les probabilités des valeurs de i et j :

$$p_i(X = i) = (1/2)^{i-19} \quad | \quad p_j(X = j) = (1/2)^{j-23}$$

Première étape



Définissons les probabilités des valeurs de i et j :

$$p_i(X = i) = (1/2)^{i-19} \quad | \quad p_j(X = j) = (1/2)^{j-23}$$

De même la probabilité que l'égalité soit respectée à i et j fixés :

$$p(X = vrai|i,j) = \begin{cases} (1/2)^{(i-19)+(j-23)} & si\ i < 24 \\ (1/2)^4 & si\ i = j \\ 0 & sinon \end{cases}$$

Première étape, suite.

De là on conclût par :

$$p(X = vrai) = \sum_{i,j} (p_i(X = i) \cdot p_j(X = j) \cdot p(X = vrai|i, j))$$

$$p(X = vrai) = \sum_{i=20}^{23} \sum_{j=24}^{33} \left(\frac{1}{2}\right)^{2(i-19)+2(j-23)} + \sum_{i=24}^{33} \left(\frac{1}{2}\right)^{(i-19)+(i-23)+4}$$

Première étape, suite.

De là on conclût par :

$$p(X = vrai) = \sum_{i,j} (p_i(X = i) \cdot p_j(X = j) \cdot p(X = vrai|i, j))$$

$$p(X = vrai) = \sum_{i=20}^{23} \sum_{j=24}^{33} \left(\frac{1}{2}\right)^{2(i-19)+2(j-23)} + \sum_{i=24}^{33} \left(\frac{1}{2}\right)^{(i-19)+(i-23)+4}$$

Soit :

$$p(X = vrai) = \frac{1}{9}$$

Deuxième étape

2. Sur le modèle de l'étape $30 \longrightarrow 31$

Ceci revient :

$$G(Q_{30}, Q_{29}, Q_{28}) = G(Q_{30} + 1, Q_{29}, Q_{28})$$

Deuxième étape

2. Sur le modèle de l'étape $30 \longrightarrow 31$

Ceci revient :

$$G(Q_{30}, Q_{29}, Q_{28}) = G(Q_{30} + 1, Q_{29}, Q_{28})$$

Ici le calcul est simplifié avec un seul changement (soit $Q_{30} \longrightarrow Q_{30} + 1$)

$$p_i(X = i) = \left(\frac{1}{2}\right)^i$$

Deuxième étape

2. Sur le modèle de l'étape $30 \longrightarrow 31$

Ceci revient :

$$G(Q_{30}, Q_{29}, Q_{28}) = G(Q_{30} + 1, Q_{29}, Q_{28})$$

Ici le calcul est simplifié avec un seul changement (soit $Q_{30} \longrightarrow Q_{30} + 1$)

$$p_i(X = i) = \left(\frac{1}{2}\right)^i$$

Ce qui donne :

$$p(X = vrai|i) = \left(\frac{1}{2}\right)^i$$

$$p(X = vrai) = \sum_{i=1}^{33} p_i(X = i) \cdot p(X = vrai|i) = \sum_{i=1} 33 \left(\frac{1}{2}\right)^{2i} = \frac{1}{3}$$

Troisiéme étape

3. Sur le modèle de l'étape $33 \longrightarrow 34$

Ici on recherche la probabilité que :

$$H(Q_{33}, Q_{32}, Q_{31}) + 1 = H(Q_{33}, Q_{32}, Q_{31} + 1)$$

Troisiéme étape

3. Sur le modèle de l'étape $33 \longrightarrow 34$

Ici on recherche la probabilité que :

$$H(Q_{33}, Q_{32}, Q_{31}) + 1 = H(Q_{33}, Q_{32}, Q_{31} + 1)$$

Pour le calcul de probabilité encore une fois :

$$p(X = vrai|i) = \left(\frac{1}{2}\right)^i$$

$$p(X = vrai) = \sum_{i=1}^{33} p_i(X = i) \cdot p(X = vrai|i) = \sum_{i=1} 33 \left(\frac{1}{2}\right)^{2i} = \frac{1}{3}$$

Contraintes fortes

Détermination des Q_{12} à Q_{15}

Contraintes fortes

Détermination des Q_{12} à Q_{15}

$\Rightarrow$ Résolution d'équations

Contraintes fortes

Détermination des Q_{12} à Q_{15}
$\Rightarrow$ Résolution d'équations

Conditions au départ :

- $\Delta_{19} = (0, 0, 0, 0)$
- $\Delta_{11} = (0, 0, 0, 0)$

Contraintes fortes

Détermination des Q_{12} à Q_{15}
$\Rightarrow$ Résolution d'équations

Conditions au départ :

- $\Delta_{19} = (0, 0, 0, 0)$
- $\Delta_{11} = (0, 0, 0, 0)$

Reste a exprimer ces contraintes sous une forme plaisante...

Equations non-linéaires

$$
\begin{aligned}
1 &= (Q'_{12} <<< 29) - (Q_{12} <<< 29) \\
F(Q'_{12}, Q_{11}, Q_{10}) - F(Q_{12}, Q_{11}, Q_{10}) &= (Q'_{13} <<< 25) - (Q_{13} <<< 25) \\
F(Q'_{13}, Q'_{12}, Q_{11}) - F(Q_{13}, Q_{12}, Q_{11}) &= (Q'_{14} <<< 21) - (Q_{14} <<< 21) \\
F(Q'_{14}, Q'_{13}, Q'_{12}) - F(Q_{14}, Q_{13}, Q_{12}) &= (Q'_{15} <<< 13) - (Q_{15} <<< 13) \\
G(Q'_{15}, Q'_{14}, Q'_{13}) - G(Q_{15}, Q_{14}, Q_{13}) &= Q_{12} - Q'_{12} \\
G(Q_{16}, Q'_{15}, Q'_{14}) - G(Q_{16}, Q_{15}, Q_{14}) &= Q_{13} - Q'_{13} \\
G(Q_{17}, Q_{16}, Q'_{15}) - G(Q_{17}, Q_{16}, Q_{15}) &= Q_{14} - Q'_{14} + (Q'_{18} <<< 23) - (Q_{18} <<< 23) \\
G(Q'_{18}, Q_{17}, Q_{16}) - G(Q_{18}, Q_{17}, Q_{16}) &= Q_{15} - Q'_{15} + (Q'_{19} <<< 19) - (Q_{19} <<< 19)
\end{aligned}
$$

Système final à résoudre

$$Q'_{15} = Q_{15} - G(Q'_{18}, Q_{17}, Q_{16}) + G(Q_{18}, Q_{17}, Q_{16}) + (Q'_{19} <<< 19) - (Q_{19} <<< 19) - 1$$
$$Q'_{14} = Q_{15} - G(Q'_{17}, Q_{16}, Q'_{15}) + G(Q_{17}, Q_{16}, Q_{15}) + (Q'_{18} <<< 23) - (Q_{18} <<< 23)$$
$$Q_{13} = (Q_{14} <<< 21) - (Q'_{14} <<< 21)$$
$$Q'_{13} = Q_{13} - G(Q_{16}, Q'_{15}, Q'_{14}) + G(Q_{16}, Q_{15}, Q_{14})$$
$$Q_{10} = (Q'_{13} <<< 25) - (Q_{13} <<< 25)$$

avec :

$$G(Q_{15}, Q_{14}, Q_{13}) - G(Q'_{15}, Q'_{14}, Q'_{13}) = 1$$
$$F(Q'_{14}, Q'_{13}, 0) - F(Q_{14}, Q_{13}, -1) - (Q'_{13} <<< 15) + (Q_{13} <<< 15) = 0$$
$$G(Q_{19}, Q_{18}, Q_{17}) = G(Q_{19}, Q_{18}, Q_{17})$$

Un produit de probabilités

La chance de réussite de la phase différentielle est égal au produit des probabilité réussite des étapes individuelles.
Coût total :

$$p = 1/2^{30}$$

Un produit de probabilités

La chance de réussite de la phase différentielle est égal au produit des probabilité réussite des étapes individuelles.
Coût total :

$$p = 1/2^{30}$$

Probabilité de résolution des équations :

$$p = 1/2^{6}$$

Un produit de probabilités

La chance de réussite de la phase différentielle est égal au produit des probabilité réussite des étapes individuelles.
Coût total :

$$p = 1/2^{30}$$

Probabilité de résolution des équations :

$$p = 1/2^{6}$$

Complexité fixe et minimaliste pour les calculs de finition

Une attaque réaliste

En théorie : équivalent au calcul de 2^{36} hash MD4

En pratique : 2^{22}