Université de Bordeaux

Master 2 : Cryptologie et Sécurité Informatique

Projet de fin d'études

Wave - Un procédé de signature à base de codes correcteurs

Suzanne Lansade
Eva Palandjian

Encadrant:
Gilles Zemor

Février, 2020

Contents

Introduction

- passage au post-quantique
- appel d'offre NIST
-> tableau : aucun code correcteur en signatures
- dur de trouver l'ensemble des syndromes facilement décodable
- dur de créer une fonction de hachage qui envoie m dans l'ensemble des syndromes possibles
- problème du décodage NP-complet
- mot y de syndrome s est associé à un unique mot de code c le plus proche de y
quand on chiffre -> ne pose pas de problème
quand on signe -> pose un problème car il est dur de trouver un syndrome de cette sorte
la solution Wave est d'enlever la restriction au mot le plus proche
Nous allons détailler le schéma de signature Wave et détailler sa sécurité.

1 Le schéma de signature Wave

Pour répondre aux problèmes :

- Des codes UV-généralisés
- Des fonctions GPV en moyenne
- Un schéma de signature de type hash et signe utilisant ces codes

1.1 La famille de codes (U,U+V)-généralisés

Définition des codes (U,U+V)-généralisés:
- Comment les créer FAIT
- Choix des paramètres a,b,c,d À DETAILLER
- Liens entre les matrices des codes U et V et du code UV FAIT
- Les dimensions et différents paramètres EN COURS
- Calcul du hull ==> q > 2 TODO
- ...?

Définition 1.1. Soient U et V deux codes de même longueur $n/2$ et de dimension respectives k_u et k_v. Un code $(U, U+V)$ est un code de longueur n et de dimension $k = k_u + k_v$ et tel que :

$$(U, U+V) = \{(u, u+v) \text{ tel que } u \in U \text{ et } v \in V\}$$

Définition 1.2. (codes $(U, U+V)$-généralisés) Soient n un entier pair et a,b,c,d quatres vecteurs de $\mathbb{F}_q^{n/2}$ tels que pour tout $i \in 1, n/2$:

$$a_i c_i \neq 0$$

$$a_i d_i - b_i c_i \neq 0$$

Soient U et V deux codes définis comme précédemment. Le code $(U, U+V)$-généralisé correspond à l'ensemble :

$$\{(a.u + b.v, c.u + d.v) \text{ tel que } u \in U \text{ et } v \in V\}$$

où $x.y$ est le produit coordonnée par coordonnée des x_i et y_i.

Remarque 1.3. Dans la suite, on prend a,b,c,d tels que

$$a_i d_i - b_i c_i = 1 \text{ pour tout } i \in 1, n/2.$$

Proposition 1.4. Soient U, V, a, b, c et d définis comme précédemment. Soit UV le code $(U, U+V)$-généralisé associé. Alors

$$k = \dim UV = k_u + k_v.$$

De plus soient $G_U \in \mathbb{F}_q^{k_u \times n/2}$ (respectivement $G_V \in \mathbb{F}_q^{k_v \times n/2}$) et $H_U \in \mathbb{F}_q^{(n/2-k_u)\times n/2}$ (respectivement $H_V \in \mathbb{F}_q^{(n/2-k_v)\times n/2}$) les matrices génératrices et de parité des codes U et V. Soient A, B, C, D de $\mathbb{F}_q^{n\times n}$ les matrices diagonales de diagonales respectives les vecteurs a, b, c et d.
Alors la matrice de $\mathbb{F}_q^{(k_u+k_v)\times n}$:

$$G := \left(\begin{array}{c|c} G_u A & G_u C \\ \hline G_v B & G_V D \end{array}\right)$$

et la matrice $\mathbb{F}_q^{(n-k_u-k_v)\times n}$:

$$H := \left(\begin{array}{c|c} H_u D & -H_u B \\ \hline -H_v C & H_V A \end{array}\right)$$

sont des matrices génératrices et de parité du code UV.

Preuve. Remarquons d'abord que G engendre bien le code UV. Remarquons aussi que

$$\left(\begin{array}{c|c} G_u A & G_u C \\ \hline G_v B & G_V D \end{array}\right) = \left(\begin{array}{c|c} G_u & 0 \\ \hline 0 & G_V \end{array}\right) \left(\begin{array}{c|c} A & C \\ \hline B & D \end{array}\right)$$

Par définition des matrices G_V et G_U, la matrice $\left(\begin{array}{c|c} G_u & 0 \\ \hline 0 & G_V \end{array}\right)$ est de rang $k_u + k_v$. De plus les matrices A, B, C, D étant diagonales, le déterminant de la matrice $\left(\begin{array}{c|c} A & C \\ \hline B & D \end{array}\right)$ est le produit des $(a_i d_i - b_i c_i)$ pour $i \in 1, n/2$, et donc non-nul par définition des vecteurs a, b, c, d. On a donc bien $k = k_u + k_v$.
On remarque aussi que $GH^T = 0$ et que H est de rang plein par le même raisonnement que précédemment, ce qui conclut la preuve. □

1.2 Le principe de signature

Un schéma hash et signe utilisant la fonction syndrome comme fonction à sens unique :
- Définition des fonctions GPVM, un couple (Trapdoor, InvertAlg) où trapdoor est un algo poly proba renvoyant une matrice de parité et la trappe associée, et où InvertAlg est un algo poly proba prenant en entrée la trappe et renvoyant l'inverse de la fonction syndrome.
De plus, ces fonctions sont (1) bien distribuées, (2) sans fuite d'info en moyenne, (3) sens unique sans la trappe
- Le schéma : un algo signe et un algo verify.

Notre schéma de signature utilisera donc les codes $(U, U+V)$-généralisés et la fonction syndrôme comme fonction à sens unique, sous l'hypothèse de la difficulté de résoudre le problème du décodage.

Nous allons définir la notion de fonctions GPV en moyenne (GPVM). Pour cela, introduisons d'abord la notion de distance statistique.

Définition 1.5. Soient X et Y deux variables aléatoires à valeurs dans le même espace ϵ. Soient $\mathcal{D}_X$ et $\mathcal{D}_Y$ leurs distributions respectives. On définit la distance statistique entre ces deux distributions comme :

$$\rho(\mathcal{D}_X, \mathcal{D}_Y) := \frac{1}{2} \sum_{x \in \epsilon} |\mathcal{D}_X(x)\mathcal{D}_Y(x)|.$$

Définition 1.6. (Fonctions GPVM). On appelle fonction GPV en moyenne une paire d'algorithmes (`Trapdoor`,`InvertAlg`) ainsi qu'un triplet de fonctions $(n(\lambda), k(\lambda), \omega(\lambda))$ en fonction d'un paramètre de sécurité λ, tels que :

- `Trapdoor` est un algorithme probabiliste et polynomial en 1^λ et renvoyant le couple (H, T) où $H \in \mathbb{F}_q^{(n-k)\times n}$ de rang $n-k$ et T est la trappe associée.
- `InvertAlg` est un algorithme probabiliste et polynomial prenant en entrée la trappe T et un syndrôme $s \in \mathbb{F}_q^{n-k}$, et renvoyant $e \in \mathbb{F}_q^n$ de poids ω tel que $eH^T = s$.

De plus, pour *presque toutes* matrice H renvoyée par `Trapdoor`, la fonction est :

1. bien distribuée :
   $\rho(eH^T, s) \in \text{negl}(\lambda)$ où e est pris uniformément dans l'ensemble des mots de poids ω et de longueur n et s est pris uniformément dans $\mathbb{F}_q^{n-k}$.
2. sans fuite d'information *en moyenne* :
   $\rho(\texttt{InvertAlg}(s, T), e) \in \text{negl}(\lambda)$ où e est pris uniformément dans l'ensemble des mots de poids ω et de longueur n et s est pris uniformément dans $\mathbb{F}_q^{n-k}$.
3. À sens unique sans la trappe :
   Pour tout algorithme probabiliste polynomial $\mathcal{A}$, on a
   $$\mathbb{P}(\mathcal{A}(H, s) = e | eH^T = s) \in \text{negl}(\lambda).$$

C'est une définition relaxée des fonctions GPV.

Nous pouvons maintenant définir notre système de signature.

$\texttt{Sign}^{sk}(s)$:
  e $\leftarrow$ `InvertAlg`(s,T)
  `renvoie` e

$\texttt{Verify}^{pk}(s, e')$:
  `Si` $e'H^T = s$ `et` $|e'| = \omega$
    `renvoie 1`
  `renvoie 0`

1.3 Le décodage avec trappe

Détail de l'algorithme invertAlg avec utilisation de la trappe:
- Conditions sur le poids de e:
–> facile
–> facile avec trappe
–> difficile
- Inverser le syndrome sur le code UV <==> inverser le syndrome sur U et sur V et prendre son image par Phi.
- Prendre un ev par un algo de décodage quelconque, utiliser les propriétés du code UV pour en déduire un eu, vérifier le poids de e, recommencer.
- Différences gros poids et petits poids

En partant de l'hypothèse que la matrice de parité $\mathbf{H}$ du code $(U, U+V)$-généralisé ressemble à une matrice aléatoire, la difficulté de créer une fausse signature sans connaître la trappe $\mathbf{T}$ est exactement celle de résoudre le problème du décodage d'un code aléatoire, que l'on sait difficile. Nous allons expliciter dans cette section l'algorithme d'inversion de la fonction syndrôme, et discuter sa difficulté en fonction du poids ω de e.

Notons $\mathcal{S}_{\omega,n}$ l'ensemble des mots de poids ω et de longueur n. On notera $\mathcal{S}_\omega$ s'il n'y a pas d'ambiguité sur la longueur. On rappelle que l'agorithme `InvertAlg` cherche à inverser la fonction syndrôme :

$$\begin{array}{rccc} f_{\omega,\mathbf{H}} & : & \mathcal{S}_{\omega,n} & \rightarrow & \mathbb{F}_q^{n-k} \\ & & \mathbf{e} & \mapsto & \mathbf{eH^T} \end{array}$$

On rappelle que la fonction $f_{\omega,\mathbf{H}}$ avec $\mathbf{H} \in \mathbb{F}_q^{(n-k)\times n}$ s'inverse génériquement si $\omega \in \{\omega_{easy}^-, \omega_{easy}^+\}$, où :

$$\omega_{easy}^- := \frac{q-1}{q}(n-k) \qquad \text{et} \qquad \omega_{easy}^+\} := k + \frac{q-1}{q}(n-k).$$

Preuve. TODO □

On rappelle aussi que la fonction $f_{\omega,\mathbf{H}}$ admet un inverse pour toute entrée $s \in \mathbb{F}_q^{n-k}$ si $\omega \in \{\omega^-, \omega^+\}$, où :

$$\omega^- :=??? \qquad \text{et} \qquad \omega^+\} :=???.$$

Preuve. TODO □

Nous voulons donc un moyen d'inverser la fonction syndrôme pour $\omega \in \{\omega^-_{UV}, \omega^+_{UV}\}$ avec ω^-_{UV} et ω^+_{UV} tels que :

$$\{\omega^-_{easy}, \omega^+_{easy}\} \subsetneq \{\omega^-_{UV}, \omega^+_{UV}\} \subset \{\omega^-, \omega^+\}$$

INSERER SCHEMA !!

Afin d'expliciter le décodage, introduisons la fonction :

$$\begin{array}{rccc} \varphi_{\mathbf{a},\mathbf{b},\mathbf{c},\mathbf{d}} & : & \mathbb{F}_q^{n/2} \times \mathbb{F}_q^{n/2} & \to & \mathbb{F}_q^{n/2} \times \mathbb{F}_q^{n/2} \\ & & (\mathbf{x}, \mathbf{y}) & \mapsto & (\mathbf{a}.\mathbf{x} + \mathbf{b}.\mathbf{y}, \mathbf{c}.\mathbf{x} + \mathbf{d}.\mathbf{y}) \end{array}$$

Si cette fonction respecte les conditions sur les vecteurs $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}$ définies dans la définition 1.2, on dit qu'elle est UV-normalisée. Dans ce cas on peut vérifier qu'elle est bijective d'inverse :

$$\begin{array}{rccc} \varphi^{-1}_{\mathbf{a},\mathbf{b},\mathbf{c},\mathbf{d}} & : & \mathbb{F}_q^{n/2} \times \mathbb{F}_q^{n/2} & \to & \mathbb{F}_q^{n/2} \times \mathbb{F}_q^{n/2} \\ & & (\mathbf{x}, \mathbf{y}) & \mapsto & (\mathbf{d}.\mathbf{x} - \mathbf{b}.\mathbf{y}, -\mathbf{c}.\mathbf{x} + \mathbf{a}.\mathbf{y}) \end{array}$$

Ainsi, pour chaque vecteur $\mathbf{e}$ de $\mathbb{F}_q^n$, on peut associer deux vecteurs $\mathbf{e_U}$ et $\mathbf{e_V}$ de $\mathbb{F}_q^{n/2}$ tels que

$$(\mathbf{e_U}, \mathbf{e_V}) = \varphi^{-1}_{\mathbf{a},\mathbf{b},\mathbf{c},\mathbf{d}}(\mathbf{e}).$$

Proposition 1.7. Inverser $f_{\omega,\mathbf{H}}$ pour un certain $\mathbf{s} \in F_q^{n-k}$ est équivalent à trouver $\mathbf{e} \in \mathbb{F}_q^n$ tel que:

$$\mathbf{e}_U \mathbf{H}_U^T = \mathbf{s}^U \qquad \text{et} \qquad \mathbf{e}_V \mathbf{H}_V^T = \mathbf{s}^V$$

où $\mathbf{s} = (\mathbf{s}^U, \mathbf{s}^V)$ avec $\mathbf{s}^U \in \mathbb{F}_q^{n/2-k_U}$ et $\mathbf{s}^V \in \mathbb{F}_q^{n/2-k_V}$.

Preuve. TODO □

Ainsi, on on aura :

`InvertAlg`$(\mathbf{s}, \mathbf{T})$:
$(\mathbf{s}_U, \mathbf{s}_V) = s$
$\mathbf{e}_U =$ `DECODE_U`$(\mathbf{s}_U)$
$\mathbf{e}_V =$ `DECODE_V`$(\mathbf{s}_V)$
`renvoie` $\varphi_{\mathbf{a,b,c,d}}(\mathbf{e_U}, \mathbf{e_V})$

Si l'on choisit un algorithme générique pour `DECODE_U` et `DECODE_V`, alors nous obtiendrons un vecteur $\mathbf{e}$ de poids $\omega\ in\{\omega^-_{easy}, \omega^+_{easy}\}$. Non allons montrer comment utiliser les propriétés des codes $(U, U+V)$-généréralisés pour permettre un décodage hors de cet intervalle.

Remarque 1.8. Pour tout $\mathbf{e} = \varphi_{\mathbf{a,b,c,d}}(\mathbf{e_U}, \mathbf{e_V})$, on a pour tout $i \in \{1, n/2\}$:

$$\begin{cases} a_i \mathbf{e}_U(i) + b_i \mathbf{e}_V(i) &= \mathbf{e}(i) \\ c_i \mathbf{e}_U(i) + d_i \mathbf{e}_V(i) &= \mathbf{e}(i + n/2) \end{cases}$$

Choisir la valeur de $\mathbf{e}_U$ en fonction de la valeur de $\mathbf{e}_V$ nous permettras donc d'influer sur le poids de $\mathbf{e}$. On aura alors :

`InvertAlg`$(\mathbf{s}, \mathbf{T})$:
$(\mathbf{s}_U, \mathbf{s}_V) = s$
$\mathbf{e}_V =$ `DECODE_V`$(\mathbf{s}_V)$
$\mathbf{e}_U =$ `DECODE_U`$(\mathbf{s}_U, \mathbf{e}_V)$
`renvoie` $\varphi_{\mathbf{a,b,c,d}}(\mathbf{e_U}, \mathbf{e_V})$

Remarque 1.9. Fonction de hash

1.4 Implémentation et choix de paramètres

TODO

2 Uniformisation des signatures et syndromes

2.1 Une fuite d'information

- Malheureusement, fuite d'information en raison des correspondance entre e[i] et e[i+n/2] !
- Calcul de proba
- Pourquoi c'est problématique

2.2 La méthode du rejet

Idée générale : On attend pour que les sorties aient l'air uniforme (soient suffisemment proches de l'uniforme):
- On choisit ev de façon a ce qu'il soit uniforme dans son ensemble
- On met des conditions de rejet sur eu en fonction de ev pour que eu ait l'air uniforme
- On obtient un e qui a l'air uniforme

2.3 Estimation du nombre de rejet

TODO

2.4 Une famille de fonctions uniformément distribuée

On a donc le point (2) de la definition des fonctions GPV qui est obtenu dans la section précédente. On va montrer le point (1), à savoir, notre famille de fonctions syndrômes est uniformément distribuée avec les codes (U,U+V)-généralisés

3 Sécurité du schéma

3.1 Sécurité EUF-CMA

3.1.1 Définitions

3.1.2 Réduction au problème DOOM

3.1.3 Et la fonction de hachage ?

3.2 Indistinguabilité des codes (U,U+V)-généralisés

Distinguer une matrice de parité d'un code (U,U+V)-généralisé d'une matrice de parité aléatoire.
Réduction à un problème NP-complet.
Utilisation de S et P pour masquer les propriétés de la matrice.

Conclusion