Amélioration de la complexité de résolution du problème de décodage par syndrome en métrique rang

Philippe Gaborit[1] Olivier Ruatta[1] **Julien Schrek**[1]

[1]Université de Limoges,XLIM-DMI

8 Octobre 2012

Plan

Codage et Cryptographie

La cryptographie à besoin de problèmes diffciles.

- factorisation
- logarithme discret
- problème SVP dans les réseaux
- problème de décodage par syndrome

En code et cryptographie, la sécurité des cryptosystèmes est reliée au problème de décodage par syndrome.

Problème de décodage par syndrome

Syndrome decoding

Trouver x de petit poids tel que $Hx^t = y$ avec H une matrice aléatoire connue et y un vecteur connu.

Les cryptosystèmes qui utilisent les codes correcteurs basent leur sécurité sur ce problème :

- chiffrement de McEliece/Niederreiter
- signature CFS
- authentification basé sur le protocol de Stern

Problème de décodage par syndrome

Problème étudié depuis des dixaines d'années en théorie des codes.

Caractéristiques :

- NP-Complet
- Est basé sur des opérations rapides
- A priori résistant à l'ordinateur quantique

Meilleures attaques connues : Information Set Decoding et variations : FS '09, MMT '11, BJMT '12

Les attaques semblent converger, ce qui semble aller vers une stabilisation de la difficulté du problème.

Métrique rang : notations

La métrique rang se définie dans le cadre des extentions de corps finis.

- $\mathbb{F}_q$ un corps fini avec q une puissance de nombre premier.
- $\mathbb{F}_{q^m}$ une extention de corps de degré m de $\mathbb{F}_q$.

$\mathbb{F}_{q^m}$ peut être vu comme un espace vectoriel sur $\mathbb{F}_q$.

- $\mathcal{C}$ un code linéaire sur $\mathbb{F}_{q^m}$ de dimension k et de longueur n.
- G la matrice génératrice $k \times n$ du code $\mathcal{C}$.
- H la matrice $n \times (n-k)$ de contrôle de $\mathcal{C}$, $GH = 0$.

Métrique rang

Les mots du code $\mathcal{C}$ sont des n-uplets dont les coordonnées appartiennent à $\mathbb{F}_{q^m}$.

$$v = (v_1, \ldots, v_n)$$

avec $v_i \in \mathbb{F}_{q^m}$.
Chacunes de ces coordonnées v_i est un vecteur à coordonnées dans $\mathbb{F}_q$.

Métrique rang

Dire que v est de rang r signifie que les v_i engendrent un sous e.v. de $\mathbb{F}_{q^m}$ de dimension r.

Problème de décodage par syndrome en métrique rang

L'énoncé du problème reste le même en métrique rang.

Syndrome decoding

Trouver e de petit poids tel que $He^t = y$ avec H une matrice aléatoire connue et y un vecteur connu.

- induit de petites tailles de clés
- pas prouvé NP-Complet

Métrique rang et cryptographie

On peut définir le même type d'algorithmes cryptographiques que pour la métrique de Hamming :

- cryptosystème GPT '91 : analogue de McEliece
- authentification (et signature) par zero-knowledge : Chen '95 (cassé), Gaborit-Schrek-Zemor '11

Complexité des attaques connues

Complexité des attaques de A.Ourivski et T.Johannson '02 :

- énumération des bases : $\leq (k+r)^3 q^{(r-1)(m-r)+2}$
  (amélioration sur la partie polynomiale de Chabaud-Stern '96)
- énumération des coordonées : $\leq (k+r)^3 r^3 q^{(r-1)(k+1)}$

$\Rightarrow$ **Ces attaques ne dépendent pas de n.**

Analogie sur le support entre métriques Hamming et Rang

- Support d'un mot d'un code sur $\mathbb{F}_{q^m}$

Le support d'un mot $x = (x_1, x_2, \ldots, x_n)$ est l'ensemble des positions $x_i \neq 0$

Comment retrouver un mot de petit poids associé à un syndrome ?

1) trouver le support du mot (ou plutot essayer de deviner le support)

2) resoudre un système pour retrouver le mot exact pour un support donné

• Support d'un mot en métrique rang

Le support d'un mot $x = (x_1, x_2, \ldots, x_n)$ de rang r est l'espace vectoriel E de dim r tel que $\forall x_i, x_i \in E$.

Comment retrouver un mot associé à un syndrome donné ?

1) trouver le support (ou plutot essayer de deviner le support)

2) résoudre un système à partir des équations du syndrome la valeur exacte des $x_i \in E$

Remarque : à la différence du support en métrique de Hamming, en général pour la métrique rang : $x_i \neq 0$.

Attaque sur le support : analogie sur les attaques Hamming/Rang

- Attaque en métrique de Hamming pour retrouver le support

- une approche consisterait à essayer TOUS les supports : tous les ensembles de positions de poids w

$\Rightarrow$ Bien sûr on ne fait jamais ca ! ! ! !

- Attaque en métrique rang pour retrouver le support

L'analogue de cette attaque en métrique rang : essayer tous les supports possibles, ie tous les espaces vectoriels de dimension r, puis résoudre un système.

$\Rightarrow$ c'est l'attaque de Chabaud-Stern ('96) - améliorée par OJ '02

Par analogie avec le cas Hamming : on voit bien que ce n'est pas optimal ! ! ! ! En particulier la complexité de l'exposant ne dépend pas de n !

Amélioration : ISD pour métrique rang

• Information Set Decoding en métrique de Hamming (version simple)
- taille du syndrome : $n-k \rightarrow n-k$ équations
- on prend au hasard $n-k$ colonnes, si elles contiennent le support de l'erreur, on peut résoudre
• Analogue en métrique rang :
- taille du syndrome : $n-k \rightarrow (n-k)m$ équations sur $\mathbb{F}_q$
- on prend au hasard un sur-espace E' de E de dimension r'
$\rightarrow$ on peut résoudre si $nr' \leq (n-k)m$
$\rightarrow$ comme pour ISD en Hamming : améliore la complexité car plus facile à trouver.

Attaque sur le support

Détail :

Recherche d'une erreur e' de rang : $r' \geq r$ avec $r'n \leq m(n-k)$.

$$e' = \beta U \rightarrow HU^t\beta^t = Hy^t$$

avec β une base de rang r' dans $\mathbb{F}_{q^m}$ et U une matrice $r' \times n$ dans $\mathbb{F}_q$.

Opérations :

- Plus de supports à tester : $q^{(r-1)(m-r)} \rightarrow q^{(r'-1)(m-r')}$
- Meilleure Probabilité de trouver : $\frac{1}{q^{(r-1)(m-r)}} \rightarrow \frac{q^{(r'-m)}}{q^{(r-1)(m-r)}}$

La complexité est donc :

$$min((n-k)^3 m^3 q^{r\frac{\lfloor km \rfloor}{n}}, (n-k)^3 m^3 q^{(r-1)\frac{\lfloor (k+1)m \rfloor}{n}})$$

Attaque sur le support

Conclusion sur la première attaque

- Amélioration des deux attaques de OJ '02.
- attaques exponentielles dans le cas général
- Complexité : $min((n-k)^3 m^3 q^{r\frac{\lfloor km \rfloor}{n}}, (n-k)^3 m^3 q^{(r-1)\frac{\lfloor (k+1)m \rfloor}{n}})$

Comparaison avec les anciennes complexités :

- énumération des bases : $\leq (k+r)^3 q^{(r-1)(m-r)+2}$
- énumération des coordonées : $\leq (k+r)^3 r^3 q^{(r-1)(k+1)}$

Remarque : dans le cas ou $n = m$ même complexité exponentielle que l'énumeration des coordonnées de OJ '02

Attaques algèbriques pour la métrique rang

difficulté : comment traduire en équations le fait qu'un vecteur appartienne à un sev de dimension donnée de $\mathbb{F}_{q^m}$?

• Plusieurs possibilités qui se ramenent à resoudre un système d'équations sur le petit corps $\mathbb{F}_q$

Pb : casser la structure du gros corps implique plus d'inconnues : au moins $n \times r$

En pratique devient vite ingérable pour la résolution dès que $r \geq 3$ (Levy-dit-Vehel-Perret '06) car la complexité est exponentielle (en general) dans le nombre d'inconnues.

• Nécessité de trouver une nouvelle mise en équation plus adaptée au problème

Attaque par q-polynômes

q-polynôme

Un q-polynôme est un polynôme de la forme $P(x) = \sum_{i=0}^{r} p_i x^{q^i}$ *avec* $p_i \in \mathbb{F}_{q^m}$.

- Linéarité : $P(\alpha x + \beta y) = \alpha P(x) + \beta P(y)$ avec $x, y \in \mathbb{F}_{q^m}$ et $\alpha, \beta \in \mathbb{F}_q$.
- $\forall B$ base de r vecteurs de $\mathbb{F}_{q^m}$, $\exists! P$ q-polynôme unitaire de degré q^r tel que $\forall b \in B$, $P(b) = 0$.

On peut donc définir un sous-espace vectoriel de dimension r à l'aide d'un polynôme de q-degré r.

Attaque par q-polynômes

Reformulation du problème :

$$c + e = y$$

avec c un mot du code $\mathcal{C}$, e un mot de poids r et y connu.
Il existe un polynôme P de q-degré r tel que

$$P(c - y) = 0$$

de plus il existe $x \in \mathbb{F}_{q^m}^k$ tel que $c = xG$, ce qui donne :

$$(\sum_{i=0}^{r} p_i(xG_1 - y_1)^{q^i}, \ldots, \sum_{i=0}^{r} p_i(xG_n - y_n)^{q^i})$$

avec $x \in \mathbb{F}_{q^m}^k$, G_j la j-ième colonne de G et $y \in \mathbb{F}_{q^m}^n$ connu.

Attaque par q-polynômes

Avantages : moins d'inconnues, des équations creuses, semi-régulier
Inconvénient : des équations de degré $q^r + 1$
Deux types de méthode pour le résoudre :

- Linéarisation et Linéarisation hybride
- Bases de gröbner et Bases de gröbner hybrides

Méthode hybride : énumération partielle des inconues

Linéarisation

- **Traiter les monômes comme des inconnues indépendantes.**

$$\sum_{i=0}^{r} p_i \left(\sum_{t=1}^{k} x_t g_{j,t} - y_j\right)^{q^i}, \; j\text{-ième équation}$$

- x_t : les k inconnues dans $\mathbb{F}_{q^m}$.
- $g_{j,t}$: les $n \times k$ coéfficients connus de la matrice génératrice G.
- y_j : les n éléments de $\mathbb{F}_{q^m}$ connus dans le problème.
- p_i : les $r+1$ coéfficients inconnus du polynôme, avec $p_r = 1$.

Attaque : linéarisation des monomes $x_t^{q^i} p_i$ et p_i.

attaque par linéarisation

Si $n \geq (r+1)(k+1) - 1$, *la complexité de cette attaque est en* $((r+1)(k+1) - 1)^3$ *opérations dans* $\mathbb{F}_{q^m}$.

Linéarisation hybride

Idée : deviner une coordonnée de l'erreur e afin de diminuer le nombre d'inconnues du problème.
Problème :

$$c + e = y$$

avec $c_j = \sum_{t=1}^{k} x_t g_{t,j}$ car $c \in \mathcal{C}$.
En combinant les équations ont obtient pour chaque coordonnée j :

$$\sum_{t=1}^{k} x_t g_{j,t} + e_j = y_j$$

Linéarisation hybride

$$\sum_{t=1}^{k} x_t g_{j,t} + e_j = y_j$$

- Connaître e_j donne une équation supplémentaire sur les x_t
- Chaque équation supprime r monômes $(x_t^{q^i} p_i)$.
- Le nombre de valeurs que peut prendre l'erreur est : q^r.

Deviner une erreur est moins coûteux que de deviner une inconnue.

attaque hybride

Si $\lceil \frac{(r+1)(k+1)-(n+1)}{r} \rceil \leq k$ *cette attaque permet de résoudre le problème en* $r^3 k^3 q^{r \lceil \frac{(r+1)(k+1)-(n+1)}{r} \rceil}$.

Bases de Gröbner

Les bases de Gröbner permettent de résoudre des systèmes d'équations non linéaires.

Nous avons $k + r$ inconnues et n équations creuses de degré $q^r + 1$ de la forme :

$$\sum_{i=0}^{r} p_i \left(\sum_{t=1}^{k} x_t g_{j,t} - y_j\right)^{q^i}$$

Utilisation de l'algorithme $F4$ dans MAGMA pour obtenir des bases de Gröbner à partir des équations du problème.

méthode hybride : efficace car le rapport inconnue/équation se ressent bien pour les bases de Gröbner

Application à la cryptanalyse

• **Cryptosystème GPT '91** (utilise les codes de Gabidulin - analogue de McEliece avec RS)
Intêret : petite taille de clés (quelques milliers de bits)
Différents types d'attaques :

- attaques sur le mot : on essaie de décoder avec les attaques génériques (type OJ)
- attaques structurelles : on essaie de retrouver la structure cachée

Application à la cryptanalyse

• nombreuses variations (type de masquage) sur le système
- en 2005 Overbeke donne une attaque structurelle très efficace qui casse tous les paramètres proposés
- 2008 et après : réparations et nouveaux types de masquage resistant à l'attaque d'Overbeke
2 familles de paramètres connus :
- Loidreau (PQC '10)
- Haitham Rashwan, Ernst M. Gabidulin, Bahram Honary ISIT '09,'10
$\rightarrow$ on montre que nos attaques génériques cassent toutes les nouveaux paramètres proposés

Résultats

Paramètres proposés par Pierre Loidreau (PQC '10)

Code $[n, k, r]_m$	OJ1	OJ2	Over	ES	L	LH	HGb
$[64, 12, 6]_{24}$	2^{104}	2^{85}	2^{80}	2^{50}	non	2^{48}	2 heures
$[76, 12, 6]_{24}$	2^{104}	2^{85}	2^{80}	2^{49}	non	2^{36}	1 seconde

Paramètres proposés par Haitham Rashwan, Ernst M. Gabidulin, Bahram Honary ISIT '09,'10

Code $[n, k, r]_m$	Over	ES	L	LH	HGb
$[28, 14, 3]_{24}$	2^{80}	2^{55}	non	2^{49}	2 jours
$[28, 14, 4]_{24}$	2^{80}	2^{70}	non	2^{65}	pas fini
$[20, 10, 4]_{24}$	2^{80}	2^{56}	non	2^{51}	5 jours
$[20, 12, 4]_{24}$	2^{80}	2^{60}	non	2^{60}	pas fini

Conclusion

- Les 2 attaques connues ont été généralisées en une seule
- Les attaques dépendent maintenant de tous les paramètres du problème
- Une meilleure compréhension du problème de décodage par syndrome en métrique rang
- Une nouvelle formulation de ce problème qui permet des attaques algèbriques plus efficaces
- Cryptanalyse de tous les paramètres proposés pour réparer le cryptosystème de Gabidulin