Étude du problème MP-LWE

Sacha Ben-Arous, sous la direction d'Alice Pellet-Mary

Résumé

Insérer abstract ici

Table des matières

1 Introduction

Les réseaux euclidiens sont une construction algébrique permettant entre autres de définir des problèmes mathématiques dont la résolution algorithmique est conjecturée difficile, même pour des ordinateurs quantiques. Cela les rend donc particulièrement intéressants pour construire des protocoles sûrs en cryptographie post-quantique.
Deux exemples fondamentaux de problèmes sur les réseaux sont le *Small Integer Solutions problem* (SIS) introduit par Ajtai en 1996, et le *Learning With Errors problem* (LWE), découvert par Regev en 2005.
Le problème LWE est privilégié dans la construction de schémas de chiffrement car il été prouvé [Reg09] que résoudre une instance aléatoire de LWE est aussi dur que résoudre la pire instance d'un problème dur (cf suite). On dit que LWE bénéficie d'une réduction *pire-cas moyen-cas*.
Cependant, si l'on utilise la version standard de LWE pour construire des protocoles, ces derniers auront une efficacité très réduite à cause d'opérations faisant intervenir de grandes matrices aléatoires.
La variante *Polynomial Learning With Errors*, proposée par Stehlé *et al.* [SSTX09], résout ce problème en utilisant des réseaux structurés : les calculs matriciels correspondent alors à des produits de polynômes, calculables très efficacement. Cependant, ce gain d'efficacité se fait au détriment de garanties de sécurité : les polynômes sont manipulés dans $\mathbb{Z}_p[X]/f$ et la complexité de la variante est directement liée au f choisi.
Ainsi, afin de se débarasser de cette dépendance, Roşca *et al.* introduisent le *Middle-Product Learning With Errors problem* (MP-LWE) [RSSS17] dont l'intérêt est d'être aussi dur que des classes exponentiellement grandes de problèmes PLWE(f) (dont on espère qu'elles contiennent au moins une instance difficile), tout en conservant l'efficacité des réseaux structurés.

Durant mon stage, j'ai donc travaillé sur la variante MP-LWE du problème de Regev, et plus particulièrement sur les effets de la réduction de PLWE vers MP-LWE.

2 Préliminaires

Définition 2.1 Un *réseau euclidien* de $\mathbb{R}^m$ est l'ensemble des combinaisons à coefficients entiers de vecteurs linéairements indépendants $b_1, \dots, b_n$, que l'on note :

$$\mathcal{L}(b_1, \dots, b_n) := \left\{ \sum_{i=1}^{n} x_i b_i, x_i \in \mathbb{Z} \right\}$$

Le réseau est alors de *dimension* n, et la famille des $(b_i)_{1 \leq i \leq n}$ est appelée *base* de ce réseau.

En notant $B := [b_1, \dots, b_n]$ la matrice dont les colonnes sont formées par les $(b_i)_{1 \leq i \leq n}$, on considérera de manière équivalente :

$$\mathcal{L}(B) := \{Bx, x \in \mathbb{Z}^n\}$$

Dans ce qui suit, on notera $\mathbb{Z}_q$ (resp. $\mathbb{R}_q$) le quotient $\mathbb{Z}/q\mathbb{Z}$ (resp. $\mathbb{R}/q\mathbb{Z}$).

2.1 Problème LWE et variantes

Définition 2.2 (Distribution LWE) Soient $q \geq 2$, $m \geq 1$, χ une distribution de probabilité sur $\mathbb{T} := \mathbb{R}/\mathbb{Z}$. À partir de $s \in \mathbb{Z}_q^m$, on définit $\mathcal{D}_{q,\chi,m}$ la distribution sur $\mathbb{Z}_q^m \times \mathbb{R}_q^m$ obtenue en choisissant $a \hookleftarrow \mathcal{U}(\mathbb{Z}_q^m)$, $e \hookleftarrow \chi^m$, et qui renvoie $(a, b) := (a, \langle a, s \rangle + e)$.

Définition 2.3 (Problèmes LWE) Soient $q \geq 2$, $m \geq 1$, χ. Munis de la distribution $\mathcal{D}_{q,\chi,m}$ précédente, on peut alors définir deux problèmes :

- **LWE-Décisionnel** consiste à un distinguer un nombre arbitraire d'échantillons de $\mathcal{D}_{q,\chi,m}$ et le même nombre d'échantillons de $\mathcal{U}(\mathbb{Z}_q^m) \times \mathcal{U}(\mathbb{R}_q^m)$.
- **LWE-Calculatoire**

Définition 2.4 Dans un réseau Λ, on note $\lambda_i := \inf\{r, \dim(\text{Vect}(\Lambda \cap \mathcal{B}(0,r)) = i\}$ le i-ème minimum du réseau. En particulier, on a $\lambda_1 = \inf\{\|v\|, v \in \mathcal{L}(B)/\{0\}\}$

Rq : λ_i s'interprète comme le rayon de la plus petite boule centrée à l'origine qui contient i vecteurs indépendants du réseau.
Le calcul effectif de ces minimums semble très dur : les meilleurs algorithmes polynomiaux (en la dimension du réseau) connus calculant λ_1 ont un facteur d'approximation exponentiel.

On peut alors considérer le problème algorithmique suivant, qui apparait lors de l'étude des réseaux reliés aux protocoles utilisant LWE :

Définition 2.5 Soit B une base d'un réseau de dimension n, et $\delta \in \mathbb{R}^+$. Une instance du problème *Bounded Distance Decoding* est un vecteur $t \in \mathbb{R}^m$ de la forme $t = x+e$, où $x \in \mathcal{L}(B)$ et $\|e\| \leq \lambda_1/\delta$. Le problème consiste à retrouver x (ou e) à partir de t.

Conjecture 2.1 Dans un réseau de dimension n, pour δ polynomial en n, le problème BDD est conjecturé exponentiellement (en n) dur à résoudre, même sur des ordinateurs quantiques.

3 Développements

3.1 Preuve de correction

On se propose tout d'abord de détailler la preuve de correction du schéma de chiffrement proposé dans [RSSS17] :

On rappelle le cadre de la preuve : $s \hookleftarrow \mathcal{U}(\mathbb{Z}^{n+d+k-1}[X])$, pour $i \leq t$ on a $a_i \hookleftarrow \mathcal{U}(\mathbb{Z}^n[X])$; $e_i \hookleftarrow \lfloor D_{\alpha q} \rceil [X]^{<k+d}$ et finalement $r_i \hookleftarrow \mathcal{U}(\{0,1\}^{<k+1}[X])$. On note $e_i(j)$ le j-ème coefficient de e_i. Le but est d'avoir avec bonne probabilité que :

$$\|\mu + 2\sum_{i \leq t} r_i \odot_d e_i\|_\infty < q/2$$

où μ est le message à chiffrer.

Théorème 3.1 Si $\alpha \leq \frac{1}{64(k+1)t\sqrt{\lambda}}$ et $t(k+d) \leq e^\lambda$, alors tout texte μ, avec probabilité $\geq 1 - 2^{-\Omega(\lambda)}$ sur `(sk,pk)` $\hookleftarrow$ `KeyGen`, on a `Decrypt(sk,Encrypt(pk,mu))= mu`

Preuve : On commence par donner une borne classique sur la distribution gaussienne :

$$\begin{aligned}
\mathbb{P}_{X\hookleftarrow\mathcal{N}(\sigma^2)}(|X|\geq M) &= \frac{2}{\sigma\sqrt{2\pi}}\int_M^\infty e^{-\frac{t^2}{2\sigma^2}}\,\mathrm{d}t \\
&\leq \frac{2\sigma^2}{\sigma\sqrt{2\pi}}\int_M^\infty \frac{t}{\sigma^2 M}e^{-\frac{t^2}{2\sigma^2}}\,\mathrm{d}t \\
&\leq \frac{2\sigma}{M\sqrt{2\pi}}[-e^{-\frac{t^2}{2\sigma^2}}]_M^\infty \\
&\leq \frac{2\sigma}{M\sqrt{2\pi}}e^{-\frac{M^2}{2\sigma^2}}
\end{aligned}$$

De plus, si on note $E := \max_{i,j}|e_i(j)|$, on a que :

$$\begin{aligned}
|(r_i \odot_d e_i)_{l\text{-eme}}| &= |(r_i \times e_i)_{l+k\text{-eme}}| \\
&= |\sum_{j=0}^{k+l} r_i(j)e_i(k+l-j)| \\
&\leq (k+1)E
\end{aligned}$$

Donc $|(\sum_{i\leq t} r_i \odot_d e_i)_{l\text{-eme}}| \leq t(k+1)E$.

Alors, on obtient que :

$$\begin{aligned}
\mathbb{P}(\|\mu + 2\sum_{i\leq t} r_i \odot_d e_i\|_\infty \geq q/2) &\leq \mathbb{P}(E \geq \frac{q}{4t(k+1)}) \\
&\leq \mathbb{P}(\bigcup_{i,j}|e_i(j)| \geq \frac{q}{4t(k+1)}) \\
&\leq (k+d)t\ \mathbb{P}_{X\hookleftarrow\mathcal{N}(\alpha^2q^2)}(X \geq \frac{q}{4t(k+1)}) \\
&\leq \frac{8t^2(k+1)(k+d)\alpha q}{q\sqrt{2\pi}}e^{-\frac{q^2}{32(\alpha q(k+1)t)^2}} \\
&\leq \frac{8t^2(k+1)(k+d)\alpha}{\sqrt{2\pi}}e^{-\frac{1}{32(\alpha(k+1)t)^2}}
\end{aligned}$$

En appliquant les hypothèses sur les paramètres de sécurité, on obtient finalement :

$$\begin{aligned}
\mathbb{P}(\|\mu + 2\sum_{i\leq t} r_i \odot_d e_i\|_\infty \geq q/2) &\leq \frac{t(k+d)}{8\sqrt{2\pi\lambda}}e^{-2\lambda} \\
&\leq e^{-\lambda-\log(8\sqrt{2\pi\lambda})} \\
&\leq 2^{-\Omega(\lambda)}
\end{aligned}$$

□

Références

[Reg09] Oded Regev. “On lattices, learning with errors, random linear codes, and cryptography”. In : *Journal of the ACM* (2009), p. 1-40.

[SSTX09] Damien Stehlé, Ron Steinfeld, Keisuke Tanaka et Keita Xagawa. “Efficient public key encryption based on ideal lattices”. In : *International Conference on the Theory and Application of Cryptology and Information Security* (2009), p. 617-635.

[RSSS17] Miruna Roşca, Amin Sakzad, Damien Stehlé et Ron Steinfeld. “Middle-Product Learning With Errors”. In : *Annual International Cryptology Conference* (2017), p. 283-297.