Université Paris 13 Année 2021–2022

M2 Mathématiques des données

Codes correcteurs – Feuille de TD 4

Éléments de solutions

10 décembre 2021

Exercice 1. Attaque de Sidelnikov-Shestakov.

Dans cet exercice, on propose d'étudier l'attaque de Sidelnikov et Shestakov [1] sur une variation du cryptosystème de McEliece employant des codes de Reed-Solomon généralisés.

Originellement, un code de Reed-Solomon généralisé peut être introduit comme suit. On considère $\boldsymbol{x} = (x_1, \dots, x_n)$ des éléments deux-à-deux distincts de $\mathbb{F}_q$, et $\boldsymbol{y} = (y_1, \dots, y_n)$ des éléments non-nuls de $\mathbb{F}_q$. Ensuite, on définit

$$\mathrm{GRS}_k(\boldsymbol{x}, \boldsymbol{y}) = \{(y_1 f(x_1), \dots, y_n f(x_n)) \mid f \in \mathbb{F}_q[X], \deg f < k\} \subseteq \mathbb{F}_q^n .$$

Par la suite, on notera $\mathrm{ev}_{\boldsymbol{x},\boldsymbol{y}}(f) \coloneqq (y_1 f(x_1), \dots, y_n f(x_n))$.

Les codes de Reed-Solomon généralisés peuvent également être présentés comme des codes géométriques. On note $\mathbb{P}^1(\mathbb{F}_q)$ l'ensemble des points de la droite projective. Étant donné un ensemble de points $\mathcal{P} = \{P_1, \dots, P_n\} \subset \mathbb{P}^1(\mathbb{F}_q)$ et un diviseur D de $\mathbb{P}^1$, le code de Reed-Solomon généralisé est :

$$\mathcal{C}_{\mathbb{P}^1}(\mathcal{P}, D) = \{\mathrm{ev}_{\mathcal{P}}(f) \mid f \in L(D)\} .$$

On note $\mathcal{H}$ le groupe des homographies, autrement dit des automorphismes de la droite projective :

$$\mathcal{H} = \left\{ \sigma : \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} ax + by \\ cx + dy \end{pmatrix} \;\middle|\; ad - bc \neq 0 \right\}$$

Question 1.– Soit $\sigma \in \mathcal{H}$. Démontrer que $\mathcal{C}_{\mathbb{P}^1}(\mathcal{P}, D) = \mathcal{C}_{\mathbb{P}^1}(\sigma(\mathcal{P}), \sigma(D))$.

Dans une variation du schéma de McEliece employant des codes de Reed–Solomon, Alice choisit aléatoirement $\boldsymbol{x}$ et $\boldsymbol{y}$, et publie $\boldsymbol{G} \in \mathbb{F}_q^{k \times n}$, une matrice génératrice de $\mathcal{C} = \mathrm{GRS}_k(\boldsymbol{x}, \boldsymbol{y}) = \mathcal{C}_{\mathbb{P}^1}(\mathcal{P}, D)$. Les vecteurs $\boldsymbol{x}$ et $\boldsymbol{y}$ sont gardés secrets par Alice.

On rappelle que la connaissance de $(\boldsymbol{x}', \boldsymbol{y}')$ tels que $\mathcal{C} = \mathrm{GRS}_k(\boldsymbol{x}', \boldsymbol{y}')$ permet de décoder n'importe quel message chiffré à l'aide de la clé publique $\boldsymbol{G}$. Une attaque sur la clé correspond donc à retrouver, à partir de $\boldsymbol{G}$ uniquement, des vecteurs $\boldsymbol{x}'$, $\boldsymbol{y}'$ (ou de manière équivalente, $\mathcal{P}'$ et D'), tels que

$$\mathrm{GRS}_k(\boldsymbol{x}', \boldsymbol{y}') = \mathcal{C} .$$

Notons qu'il n'est pas nécessaire retrouver exactement le couple $(\boldsymbol{x}, \boldsymbol{y})$ engendré pae Alice.

Question 2.– Déduire de la **Question 1** que l'on peut supposer que l'on connaît déjà trois des points de $\boldsymbol{x}$ choisis par Alice.

Question 3.– Si $n - k \leq 3$, donner une borne sur la capacité de correction du code $\mathcal{C}$. En déduire que le cryptosystème de McEliece n'est pas sûr dans ce cas.

Question 4.– Si $k = 1$, démontrer que l'on peut retrouver aisément $(\boldsymbol{x}', \boldsymbol{y}')$ tels que $\mathrm{GRS}_k(\boldsymbol{x}', \boldsymbol{y}') = \mathcal{C}$.

Dorénavant, on suppose donc que $k \geq 2$, $n - k \geq 4$ et $x'_1 = 0$, $x'_2 = 1$ et $x'_3 = \infty$. On rappelle également que pour tout code MDS, tout ensemble $I \subset [1, n]$ de cardinal k est un ensemble d'information du code.

On peut maintenant passer à l'attaque. Pour cela, on rappelle que l'on suppose que l'on connaît uniquement $\boldsymbol{G}$.

Question 5.– Comment peut-on calculer deux mots non-collinéaires $\boldsymbol{u}$ et $\boldsymbol{v}$ de $\mathcal{C}$ qui partagent $k - 2$ zéros en commun sur leurs $k - 2$ dernières coordonnées ? Démontrer que l'on peut également imposer $u_{n-k+1} = 0$ et $v_{n-k+2} = 0$.

Question 6.– Soient f et g tels que $\boldsymbol{u} = \mathrm{ev}_{\boldsymbol{x}',\boldsymbol{y}'}(f)$ et $\boldsymbol{v} = \mathrm{ev}_{\boldsymbol{x}',\boldsymbol{y}'}(g)$ où $\boldsymbol{u}$ et $\boldsymbol{v}$ sont déterminés à la question précédente. Que dire du degré de la fraction rationnelle f/g ? Connaît-on ses évaluations sur certains points x'_i ?

Question 7.– Démontrer que l'on peut déterminer exactement f/g par la résolution d'un système linéaire de trois équations à quatre inconnues, dont on extrait une solution non-nulle.

Question 8.– En déduire que l'on peut maintenant retrouver des valeurs de x'_i sur $n - k - 3$ nouveaux points que l'on déterminera. En déduire que l'on peut retrouver tous les points $\boldsymbol{x}'$.

Question 9.– On suppose maintenant qu'un vecteur $\boldsymbol{x}'$ a été déterminé. Comment peut-on trouver $\boldsymbol{y}'$ tel que $\mathrm{GRS}_k(\boldsymbol{x}, \boldsymbol{y}) = \mathrm{GRS}_k(\boldsymbol{x}', \boldsymbol{y}')$?

Question 10.– Conclure avec un algorithme qui attaque la variante du cryptosystème de McEliece employant des codes de Reed-Solomon. On donnera sa complexité.

Solutions de l'Exercice 1.

Solution Q1. On a

$$\begin{aligned}\mathcal{C}_{\mathbb{P}^1}(\mathcal{P}, D) &= \{\mathrm{ev}_{\mathcal{P}}(f) \mid f \in L(D)\} = \{\mathrm{ev}_{\sigma(\mathcal{P})}(f \circ \sigma^{-1}) \mid f \in L(D)\} \\ &= \{\mathrm{ev}_{\sigma(\mathcal{P})}(g) \mid g \in L(\sigma(D))\} = \mathcal{C}_{\mathbb{P}^1}(\sigma(\mathcal{P}), \sigma(D))\,.\end{aligned}$$

Solution Q2. Le groupe des homographies de la droite projective est 3-transitif : pour tous points de la droite projective P_1, P_2, P_3, Q_1, Q_2, Q_3 (où les P_i d'une part, et les Q_j d'autre part, sont deux-à-deux distincts), il existe $\sigma \in \mathcal{H}$ tel que $\sigma(P_i) = Q_i$ pour tout $i \in \{1, 2, 3\}$.

Supposons que $\mathcal{P} = \{P_1, \ldots, P_n\}$ soit l'ensemble des points d'évaluation du code recherché. Alors, il existe une homographie σ telle que $\sigma(P_1) = 1$, $\sigma(P_2) = 0$ et $\sigma(P_3) = \infty$. D'après la question précédente, on a alors $\mathcal{C}_{\mathbb{P}^1}(\sigma(\mathcal{P}), \sigma(D)) = \mathcal{C}$, et au lieu de chercher à retrouver $\mathcal{P}$ et D, on cherchera à retrouver $\sigma(\mathcal{P})$ et $\sigma(D)$. En d'autres termes, on peut supposer que $P_1 = 1$, $P_2 = 0$ et $P_3 = \infty$.

Solution Q3. Si $n - k \leq 3$, alors Bob peut seulement introduire des erreurs de poids ≤ 1 (car $\lfloor \frac{d-1}{2} \rfloor = \lfloor \frac{d-1}{2} \rfloor \leq 1$). Il existe seulement n supports d'erreur de poids 1, donc la recherche exhaustive de l'erreur est de complexité polynomiale.

Solution Q4. Si $k = 1$, alors $\mathcal{C} = \{\lambda(y_1, \ldots, y_n) \mid \lambda \in \mathbb{F}_q\}$. En choisissant un mot non-nul quelconque $\boldsymbol{y}' \in \mathcal{C}$ et des points $\boldsymbol{x}' \in \mathbb{F}_q^n$ deux à deux distincts quelconques, on a donc $\mathcal{C} = \mathrm{GRS}_1(\boldsymbol{x}', \boldsymbol{y}')$.

Solution Q5. Soit $\boldsymbol{G}$ la matrice génératrice publique de $\mathcal{C}$. On réalise une élimination gaussienne sur $\boldsymbol{G}$, en utilisant comme pivots les k dernières colonnes de $\boldsymbol{G}$. On peut alors obtenir une matrice de la forme

$$\boldsymbol{G}' = (\boldsymbol{A} \mid \boldsymbol{B}) = \left(\begin{array}{ccc|ccc} (*) & & (*) & 0 & & 1 \\ & & & & \iddots & \\ & & & & \iddots & \\ (*) & & (*) & 1 & & 0 \end{array}\right).$$

Les deux dernières lignes sont donc de la forme convenue :

$$\begin{aligned}\boldsymbol{u} &= \big((*) \quad \dots \quad (*) \quad | \quad 0 \quad 1 \quad 0 \quad \cdots \quad 0\big)\,,\\ \boldsymbol{v} &= \big((*) \quad \dots \quad (*) \quad | \quad 1 \quad 0 \quad 0 \quad \cdots \quad 0\big)\,.\end{aligned}$$

Solution Q6. Comme $\boldsymbol{u}$ et $\boldsymbol{v}$ partagent $k-2$ coordonnées nulles, leurs polynômes f et g associés sont divisibles par $\prod_{i=n-k+3}^{n}(X-x'_i)$ de degré $k-2$. La fraction rationnelle f/g est donc de degré 1 : on peut écrire

$$(\frac{f}{g})(x) = \frac{ax+b}{cx+d} \text{ avec } ad-bc \neq 0$$

Par ailleurs, on connaît les évaluations de f/g sur les $n-k+2$ premiers x'_i : c'est le vecteur $\boldsymbol{w} \in \mathbb{F}_q^{n-k+2}$ défini par $w_i = u_i/v_i$, avec comme convention $1/0 = \infty$. On peut également éviter les cas spéciaux x'_{n-k+1} et x'_{n-k+2} et se restreindre aux $n-k$ premières coordonnées.

Solution Q7. On a à notre disposition les évaluations de f/g en $x'_1 = 1$, $x'_2 = 0$ et $x'_3 = \infty$. Par ailleurs, ces évaluations sont différentes de 0 et ∞, car f et g ne peuvent s'annuler qu'en $k-1$ points (ces points sont nécessairement parmi $x'_{n-k+1}, \dots, x'_n$). Si $\boldsymbol{w}$ est le vecteur défini à la question précédente, on obtient le système d'équations

$$\begin{cases} a+b &= w_1(c+d)\\ b &= w_2 d\\ a &= w_3 c\end{cases}$$

Le système est linéaire et admet un espace de solutions de dimension 1. On peut en prendre une solution non-nulle, qui définit f/g car a, b, c et d sont définis à une constante multilicative près.

Solution Q8. Une fois que f est déterminé, on peut trouver x'_i en résolvant l'équation de degré 1 donnée par $f(x'_i) = w_i$, pour tous les $i \in \{4, \dots, n-k\}$.

Pour obtenir les x'_i où $i \geq n-k+1$, on peut réitéré en procédé des questions 5 à 8 avec un autre ensemble d'information que $\{n-k+1, \dots, n\}$. Si on choisit un ensemble d'information I tel que $I \cap \{1,2,3\} = \varnothing$, alors on peut retrouver les valeurs x'_j pour $j \in \{1, \dots, n\} \setminus (I \cup \{1,2,3\})$.

Solution Q9. On peut mettre en équation $\boldsymbol{y}'$ de la manière suivante. Pour tout $j \in \{0, \dots, k-1\}$, le vecteur $\mathrm{ev}_{\boldsymbol{x}',\boldsymbol{y}'}(X^j)$ doit appartenir à $\mathcal{C}$. Si $\boldsymbol{H}$ est une matrice de contrôle de $\mathcal{C}$ (que l'on peut aisément construire à partir de la matrice publique $\boldsymbol{G}$), on a alors $\boldsymbol{H}\,\mathrm{ev}_{\boldsymbol{x}',\boldsymbol{y}'}(X^j)^\top = \boldsymbol{0}$, autrement dit

$$\boldsymbol{H} \cdot (\mathrm{Diag}(\boldsymbol{x}')^j \boldsymbol{y}')^\top = \boldsymbol{0}\,,$$

pour tous $j \in \{0, \dots, k-1\}$. On peut donc créer un système d'équations

$$\begin{pmatrix} \boldsymbol{H} \cdot \mathrm{Diag}(\boldsymbol{x}')^0 \\ \boldsymbol{H} \cdot \mathrm{Diag}(\boldsymbol{x}')^1 \\ \vdots \\ \boldsymbol{H} \cdot \mathrm{Diag}(\boldsymbol{x}')^{k-1} \end{pmatrix} \cdot \boldsymbol{y}' = \boldsymbol{0}$$

qui donne $\boldsymbol{y}'$ à une constante multiplicative près. Ce n'est pas pas problématique car $\mathrm{GRS}_k(\boldsymbol{x}', \boldsymbol{y}') = \mathrm{GRS}_k(\boldsymbol{x}', \lambda\boldsymbol{y}')$.

Solution Q10. Voir Algorithme 1. La complexité de l'algorithme est en $O(n^4)$.

Références

[1] V.M. Sidelnikov and S.O. Shestakov. On cryptosystems based on generalized Reed-Solomon codes. *Discrete Mathematics*, 4(3) :57–63, 1991.

Algorithme 1 : Description de l'algorithme de Sidelnikov–Shestakov

Entrée : une matrice génératrice $\boldsymbol{G}$ de $\mathrm{GRS}_k(\boldsymbol{x}, \boldsymbol{y})$
Sortie : des vecteurs $\boldsymbol{x}'$ et $\boldsymbol{y}'$ tels que $\mathrm{GRS}_k(\boldsymbol{x}', \boldsymbol{y}')$ est engendré par $\boldsymbol{G}$

$X_{\text{found}} = \{1, 2, 3\}$
Tant que $X_{\text{found}} \neq [1, n]$ **faire**
  Choisir $I \subset [1, n] \setminus X_{\text{found}}$ de cardinal $n - k - 3$.
  Choisir $J \subset [1, n] \setminus I$ de cardinal k.
  Faire une élimination gaussienne de $\boldsymbol{G}$ avec les colonnes indexées par J comme pivots.
  Prendre deux lignes $\boldsymbol{u}$ et $\boldsymbol{v}$ de la matrice, et calculer $\boldsymbol{w}$ défini par $w_i = u_i / v_i$ sur les coordonnées où $\boldsymbol{u}$ et $\boldsymbol{v}$ sont non-nuls.
  Déterminer l'unique fraction rationnelle $\frac{f}{g}$ de degré 1 qui s'évalue à w_1, w_2, w_3 sur $P_1 = 1$, $P_2 = 0$ et $P_3 = \infty$.
  Pour chaque $i \in I$, en déduire les P_i tels que $\frac{f}{g}(P_i) = w_i$.
  Remplacer X_{found} par $X_{\text{found}} \cup I$.
Définir $\boldsymbol{x}' = (P_1, \ldots, P_n)$.
Exclure ∞ de $\boldsymbol{x}'$ en lui appliquant une homographie bien choisie.
Calculer une matrice de parité $\boldsymbol{H}$ à partir de $\boldsymbol{G}$.
Calculer $\boldsymbol{y}'$ solution su système

$$\begin{pmatrix} \boldsymbol{H} \cdot \mathrm{Diag}(\boldsymbol{x}')^0 \\ \boldsymbol{H} \cdot \mathrm{Diag}(\boldsymbol{x}')^1 \\ \vdots \\ \boldsymbol{H} \cdot \mathrm{Diag}(\boldsymbol{x}')^{k-1} \end{pmatrix} \cdot \boldsymbol{y}' = \boldsymbol{0}$$

Retourner $\boldsymbol{x}'$ et $\boldsymbol{y}'$.