Agrégation externe de mathématiques, session 2010
Épreuve de modélisation, option C : algèbre et calcul formel

(*public 2010*)

Résumé : On étudie des problèmes de mise en commun de secrets.

Mots clefs : polynômes, interpolation, corps finis

► *Il est rappelé que le jury n'exige pas une compréhension exhaustive du texte. Vous êtes laissé(e) libre d'organiser votre discussion comme vous l'entendez. Des suggestions de développement, largement indépendantes les unes des autres, vous sont proposées en fin de texte. Vous n'êtes pas tenu(e) de les suivre. Il vous est conseillé de mettre en lumière vos connaissances à partir du fil conducteur constitué par le texte. Le jury appréciera que la discussion soit accompagnée d'exemples traités sur ordinateur.*

1. Introduction

Longtemps cantonnée à la problématique du chiffrement, la cryptologie a, depuis les années 1970 et l'explosion de l'informatique, su s'inventer nombre de nouvelles applications. Une bonne partie de ces applications relève de techniques mathématiques de nature algébrique et arithmétique. Le présent texte propose d'étudier le problème du *partage de secret*.

Une personne détient un secret qu'elle souhaite léguer à un groupe de n individus. La solution la plus évidente consiste à donner le secret à chacun d'entre eux. Néanmoins ceci fait courir le risque qu'un individu peu scrupuleux tire seul profit du secret. Il faut donc découper le secret en n parts, qui ne donnent accès au secret qu'une fois rassemblées. Une façon concrète de procéder est la suivante, utilisée dans bien des romans d'aventure : on écrit le secret sur une feuille de papier, que l'on découpe en autant de morceaux qu'il y a de protagonistes. La réunion des différents morceaux est nécessaire pour reconstruire le secret.

Nous allons formaliser ce problème et modéliser différents niveaux de sécurité pouvant être souhaités, puis proposer des solutions correspondant à chaque niveau de sécurité. Nous entreprendrons ensuite l'étude d'un problème plus général, et autoriserons enfin certains des protagonistes à mentir lors de la reconstruction du secret.

2. Modélisation du problème

2.1. *Notations*

Nous supposerons que les secrets sont pris dans un ensemble S. D'un point de vue informatique, on peut toujours voir un secret comme une suite finie de bits, donc comme un nombre entier ; il est souvent commode de considérer que le secret est de taille fixée t, quitte à découper

le secret en plusieurs sous-secrets. On considère donc que S est l'intervalle $[0,2^t[$ ou encore le corps fini à 2^t éléments.

Un (système de) partage de secret en n parties est la donnée d'un ensemble A non vide et d'un couple d'applications $i : A \times S \to \mathbb{N}^n$ et $\sigma : \mathbb{N}^n \to S$, telles que $\sigma \circ i(a,s) = s$. Dans ce formalisme, i réalise le partage de secret en n morceaux, et σ réalise la reconstruction. Le rôle de l'ensemble A sera clarifié dans la partie §2.3 : nous verrons qu'un partage de secret fort doit utiliser une partie aléatoire. Cette modélisation contient le cas naïf où il n'y a pas d'ensemble A, en prenant A réduit à un élément.

Pour réaliser le partage d'un secret s, on tire un élément a au hasard dans A, et on distribue au participant j la j-ème composante de $i(a,s)$. La reconstruction du secret est indépendante de la valeur de a choisie. On pourra noter qu'à a fixé, l'application $s \mapsto i(a,s)$ est nécessairement injective. Dans la suite, nous noterons $(i_k)_{1\leqslant k\leqslant n}$ les composantes de i, c'est-à-dire que

$$i(a,s) = (i_1(a,s),\dots,i_n(a,s)).$$

2.2. *Un exemple*

La découpage naïf de l'introduction correspond à $S = [0,2^t[$, $A = \{0\}$. Tout entier s de S s'écrit de façon unique sous la forme $\sum_{i\geq 0} b_i(s)2^i$, avec $b_i(s) \in \{0,1\}$ et $b_i(s) = 0$ pour $s \geq t$. Soit τ la partie entière de t/n ; on définit alors

$$i_\ell(a,s) = \sum_{j=0}^{\tau-1} b_{\tau(\ell-1)+j}(s)2^j, \quad \text{et} \quad \sigma(x_1,\dots,x_n) = \sum_{\ell=1}^{n} x_\ell 2^{\tau(\ell-1)}.$$

2.3. *Étude de la sécurité*

Il est possible de définir de nombreuses notions de sécurité, plus ou moins fortes. Quelle serait une exigence de sécurité minimale ? Il paraît naturel que l'on ne puisse pas, à l'aide de $k < n$ parties de secret, reconstruire de façon unique le secret initial. Cette résistance faible est insuffisante : dans le cas de notre exemple naïf, si l'on connaît $n-1$ parties de secret, le nombre de secrets possibles n'est plus 2^t mais essentiellement 2^τ. Cela peut signifier qu'il devient possible d'énumérer tous les secrets compatibles avec l'information disponible.

Une façon de formaliser cette remarque est de demander que la connaissance de $n-1$ parties du secret n'apporte *aucune* information sur le secret :

Définition 1. *Un partage de secret est* résistant *si pour tout vecteur $x = (x_1,\dots,x_n) \in i(A\times S)$, la fonction*

$$F_x(\ell,s) = |\{\alpha \in A\colon \forall k \neq \ell,\ i_k(\alpha,s) = x_k\}|$$

est constante (ne dépend pas de $\ell \leq n$ ni de $s \in S$).

En particulier, $n-1$ participants associés ne disposent d'aucune information sur s, même en sachant que a a été tiré uniformément dans A.

Proposition 1. *Si $|A| < |S|^{n-1}$, aucun partage de secrets en n parties sur $A \times S$ n'est résistant.*

Démonstration. Soit a, s arbitraires, et $(b_1, \ldots, b_n) = i(a,s)$. Supposons le système résistant : pour tout $s' \in S$, il existe $a' \in A$ tel que

$$i(a', s') = (b_1, \ldots, b_{n-1}, i_n(a', s')).$$

Comme $i(a', s') \neq i(a'', s'')$ pour $s' \neq s''$, la n-ème composante prend au moins $|S|$ valeurs distinctes. En itérant ce raisonnement $|i(A \times S)| \geq |S|^n$, et on obtient la minoration de $|A|$. □

Corollaire 1. *Le système naïf n'est pas résistant si $n > 1$.*

3. Un système résistant

Soit $\mathbb{K}$ un corps fini ; quitte à prendre une extension algébrique de $\mathbb{K}$, on peut supposer $|\mathbb{K}| > n$ et fixer des éléments $t_1, \ldots, t_n$ non nuls et deux à deux distincts de $\mathbb{K}$. On pose $S = \mathbb{K}$, $A = \mathbb{K}^{n-1}$, et on définit un partage de secrets par

$$i_\ell(a,s) = s + \sum_{j=1}^{n-1} a_j t_\ell^j.$$

Alors $\sigma(x_1, \ldots, x_n)$ est la valeur en 0 de l'unique polynôme de degré $\leqslant n-1$ passant par les points $(t_k, x_k)_{1 \leqslant k \leqslant n}$.

Proposition 2. *Ce système est résistant.*

4. Partage de secrets avec seuil

4.1. *Problème et solution*

Une version plus élaborée du partage de secret demande que, pour un seuil $k \leq n$ fixé, la donnée de k parties de secret suffise à retrouver le secret s. La méthode du paragraphe précédent s'adapte naturellement : il suffit de choisir un polynôme de degré $k-1$ à la place d'un polynôme de degré $n-1$, de sorte que k valeurs suffisent à reconstruire le polynôme.

4.2. *Menteurs*

La redondance introduite permet toutefois de résoudre un problème différent : que faire si certains protagonistes « mentent », ou ne disposent que d'une information erronée suite à un problème de transmission ? On retrouve alors une problématique de codes correcteurs d'erreur, où, à partir d'une information altérée on s'efforce de reconstituer la donnée d'origine.

Supposons que l'on dispose de $m > k$ parties de secrets. Nous allons voir que si un nombre suffisant d'entre elles sont fiables, on sait reconstituer le secret d'origine de façon unique ; et démasquer les menteurs.

4.3. *Décodage, première méthode*

Une première solution consiste à supposer d'abord qu'il y a un seul menteur et à l'éliminer pour le calcul du polynôme d'interpolation. Si l'on obtient un polynôme P de degré $\leqslant k-1$, alors $P(0)$ est la bonne réponse, pourvu que $m-2>k-1$. Comme on ne sait pas qui est le menteur, il faut essayer en éliminant les différents morceaux un par un. On recommence ensuite la même procédure dans le cas de deux menteurs en éliminant tous les couples possibles de valeurs à leur tour, etc. Si le nombre de menteurs n'est pas trop élevé, cette méthode produit une solution unique :

Proposition 3. *Si le nombre de menteurs est $\leqslant (m-k)/2$, la méthode ci-dessus appliquée à un nombre de menteurs compris entre 0 et $(m-k)/2$ produit une solution unique.*

4.4. *Décodage, seconde méthode*

Réalisant de nombreuses interpolations, cette méthode est coûteuse. On peut faire mieux :

Théorème 1. *Supposons donnés m couples (x_i,y_i), et soit $\ell \leqslant (m-k)/2$.*

(1) *Il existe Q_0,Q_1 dans $\mathbb{K}[X]$, $Q_1\neq 0$, tels que* $\deg Q_0 \leqslant m-1-\ell$, $\deg Q_1 \leqslant m-\ell-k$, *et*
$$Q_0(x_i)+y_iQ_1(x_i)=0, \quad \forall 1\leqslant i\leqslant m.$$

(2) *S'il existe P de degré $\leqslant k-1$ tel que $P(x_i)=y_i$ pour au moins $m-\ell$ valeurs de i, alors $Q_1 \mid Q_0$ et $P=-Q_0/Q_1$.*

4.5. *Plus de menteurs*

Si le nombre de menteurs est plus élevé, on généralise l'approche précédente en cherchant un polynôme $Q(X,Y)\in\mathbb{K}[X,Y]$ tel que $Q(x_i,y_i)=0$ pour tout $1\leqslant i\leqslant m$. Dès que le nombre de valeurs correctes est strictement plus grand que $\deg Q(X,P(X))$, on obtient $Q(X,P(X))=0$. Donc P est racine (en Y) de l'équation $Q(X,Y)=0$. Les différentes racines s'obtiennent à l'aide du lemme suivant :

Lemme 1. *Soit $p(X)\in\mathbb{K}[X]$ tel que $Q(X,p(X))=0 \bmod X^r$ et $\frac{\partial Q}{\partial Y}(X,p(X))\neq 0 \bmod X$. Alors, en posant*
$$p_2(X)=p(X)-\frac{Q(X,p(X))}{\frac{\partial Q}{\partial Y}(X,p(X))} \bmod X^{2r},$$
on a $Q(X,p_2(X))=0 \bmod X^{2r}$, et p_2 est la seule racine en Y de $Q(X,Y)=0$ dans $\mathbb{K}[X]/(X^{2r})$ qui soit congrue à p modulo X^r.

Il reste à décrire comment trouver $Q\in\mathbb{K}[X,Y]$, de sorte que $\deg Q(X,P(X))$ soit le plus petit possible. Notons δ le degré en Y de Q et supposons que $\deg Q(X,P(X))\leqslant m-\ell-1$, on cherche $Q(X,Y)$ sous la forme
$$Q(X,Y)=\sum_{0\leq j\leq \delta}\ \sum_{i=0}^{m-\ell-1-(k-1)j} q_{ij}X^iY^j.$$

Un polynôme de ce type et tel que $Q(x_i, y_i) = 0$ pour tout $1 \leqslant i \leqslant m$ existe dès que

$$\left(m - \ell - \frac{(k-1)}{2}\delta\right)(\delta+1) \geqslant m+1,$$

ou, de façon équivalente,

$$\ell \leqslant \frac{\delta m - 1}{\delta + 1} - \frac{k-1}{2}\delta.$$

Suggestions pour le développement

► *Soulignons qu'il s'agit d'un menu à la carte et que vous pouvez choisir d'étudier certains points, pas tous, pas nécessairement dans l'ordre, et de façon plus ou moins fouillée. Vous pouvez aussi vous poser d'autres questions que celles indiquées plus bas. Il est très vivement souhaité que vos investigations comportent une partie traitée sur ordinateur et, si possible, des représentations graphiques de vos résultats.*

- Détaillez les preuves des différentes propositions énoncées dans le texte.
- Illustrer sur ordinateur des méthodes de partage de secret au choix (éventuellement de votre cru, éventuellement naïves), et les tester avec le jury.
- Si le système n'est pas résistant, on peut se poser la question de l'équité entre participants : chaque portion du secret contient-elle la même quantité d'information sur s ? Discuter dans le cas de la méthode naïve. On pourra proposer une définition de la notion d'équité.
- Considérer la méthode alternative suivante : on encode le secret dans un grand entier s, on choisit des premiers p_i tels $N := \prod_{i=1}^{k} p_i > s$, et on pose $i_k(s) = s \bmod p_i$. Étudier la reconstruction du secret et la sécurité de ce schéma. Discuter l'équité entre les participants. Comment modifier la méthode de façon à l'améliorer ?
- Pourquoi le système résistant du §3 suppose t'il que les t_i sont non nuls ? Pourquoi travailler sur un corps $\mathbb{K}$ *fini*, et pas dans $\mathbb{Q}$ par exemple ?
- Le texte affirme que la seconde méthode de décodage est préférable ; commenter cette affirmation.
- Étudier le coût de calcul des différentes méthodes proposées, à la fois du point de vue du partage de secret (évaluation de i) que de la reconstruction (évaluation de σ).
- Détailler la limite de la méthode du §4.5 ; comment retrouve-t-on le §4.4 comme cas particulier ? Quelle est la valeur optimale de δ en fonction de k et m, et quel est le nombre maximum de menteurs supporté par la méthode ?
- On pose $S = \mathbb{F}_{2^t}$, $A = S^{n-1}$, et i est un élément de $GL(A \times S)$, où $A \times S$ est vu comme un $\mathbb{F}_{2^t}$ espace vectoriel. En quoi la méthode de l'énoncé en est-elle un cas particulier ? Est-ce une méthode de partage de secret dans tous les cas ? Est-elle résistante ?
- Discuter l'extension de la méthode précédente à un schéma avec seuil.