Décodage Statistique

Thomas Debris

Inria - SECRET, Paris, France

27 avril 2017
Journées C2

Code-based Cryptography

Cryptographie à clef publique : McEliece 1978

$\rightarrow$ Repose sur le problème difficile du décodage générique pour les codes correcteurs linéaires

Codes Correcteurs

Code correcteur : sous-espace vectoriel de $\mathbb{F}_2^n$

On les voit comme une matrice $G \in \mathbb{F}_2^{Rn \times n}$

Ses éléments : mG pour $m \in \mathbb{F}_2^{Rn}$

Problème du décodage générique

- Donnée : $\left(G \in \mathbb{F}_2^{nR \times n}, y \in \mathbb{F}_2^n, t \in \mathbb{N}\right)$ où G quelconque
- Recherche : $\exists (m, e)$ où e de poids t, $y = mG + e$

$\rightarrow$ Problème de décision NP-complet

Problème équivalent : Trouver e de poids t tel que $He^T = s$ où $GH^T = 0$ et $H \in \mathbb{F}_2^{n(1-R) \times n}$

Décodage par ensemble d'information

On cherche à résoudre $He^T = s$:

$$\begin{cases} s_1 & = & h_{1,1}e_1 + h_{1,2}e_2 + \cdots + h_{1,n}e_n \\ & \vdots & \\ s_{n(1-R)} & = & h_{n(1-R),1}e_1 + h_{n(1-R),2}e_2 + \cdots + h_{n(1-R),n}e_n \end{cases}$$

$\rightarrow$ Système de $n(1-R)$ équations à n inconnues.

Décodage par ensemble d'information

- Si $e_i = 0$ sur un ensemble de nR positions i :

$$\begin{cases} s_1 = h_{1,J_1} e_{J_1} + h_{1,J_2} e_{J_2} + \cdots + h_{1,J_{n(1-R)}} e_{J_{n(1-R)}} \\ \vdots \\ s_{n(1-R)} = h_{n(1-R),J_1} e_{J_1} + h_{n(1-R),J_2} e_{J_2} + \cdots + h_{n(1-R),J_{n(1-R)}} e_{J_{n(1-R)}} \end{cases}$$

$\rightarrow$ Système de $n(1-R)$ équations à $n(1-R)$ inconnues.

Complexité exponentielle car probabilité exponentiellement faible de "tomber" sur un ensemble de positions qui convient

Décodage par ensemble d'information

Algorithmes de décodage génériques les plus étudiés issus de celui de Prange (1962) :

Lee-Brickell (1988) - Leon (1988) - Stern (1988) - CC (1998) - - MMT (2011)- BLP (2011) - BJMM (2012) - MO (2015)

Si $t = o(n)$, ces algorithmes ont tous le même exposant de complexité asymptotique (Canto-Torres&Sendrier 2016) :

$$\widetilde{O}\left(2^{-\ln(1-R)\cdot t}\right)$$

Décodage Statistique

Il existe un algorithme n'entrant pas dans cette famille :
le décodage statistique de Al. Jabri (2001)

Etudié par R.Overbeck en 2006

Pas d'étude de sa complexité asymptotique !

Résultats

- Exposant asymptotique pour la complexité du décodage statistique donné par une formule simple ;
- Le décodage statistique jamais meilleur que Prange dans certaines zones d'erreur.

Intuition décodage statistique

$$y = c + e \text{ où } c \in \mathcal{C} = \{mG : m \in \mathbb{F}_2^k\}$$

$$h \in \mathcal{C}^\perp \Rightarrow \langle y, h \rangle = \langle e, h \rangle$$

$$\mathcal{C}^\perp = \{h \in \mathbb{F}_2^n : \forall c \in \mathcal{C},\ \langle h, c \rangle = 0\}$$

- Si $e_i = 1$ et $h_i = 1$,

$$\langle y, h \rangle = 1 \iff \#(Supp(e) \cap Supp(h) - \{i\}) \text{ pair}$$

- Si $e_i = 0$ et $h_i = 1$

$$\langle y, h \rangle = 1 \iff \#(Supp(e) \cap Supp(h) - \{i\}) \text{ impair}$$

$\rightarrow$ Biais des $\langle y, h \rangle$ selon que $e_i = 1$ ou 0

Notations

- $\mathcal{H}_w \subseteq \{h \in \mathcal{C}^\perp : \ w_H(h) = w\}$ où w_H poids de Hamming
- $\mathcal{H}_{w,i} \subseteq \mathcal{H}_w \cap \{m \in \mathbb{F}_2^n \ : \ m_i = 1\}$

On fixe un poids w, et un bruité $y = xG + e$ où $w_H(e) = t$.

Deux distributions

$$e_i = 1 \ : \ q_1(e,w,i) \stackrel{\triangle}{=} \mathbb{P}_{h \sim \mathcal{H}_{w,i}} \left(\langle y,h \rangle = \langle e,h \rangle = 1 \right)$$

$$e_i = 0 \ : \ q_0(e,w,i) \stackrel{\triangle}{=} \mathbb{P}_{h \sim \mathcal{H}_{w,i}} \left(\langle y,h \rangle = \langle e,h \rangle = 1 \right)$$

Deux distributions

$$q_1(e,w,i) = \frac{\sum_{j \text{ pair}}^{w-1} \binom{t-1}{j}\binom{n-t}{w-1-j}}{\binom{n-1}{w-1}} = \frac{1}{2} + \varepsilon_1$$

$$q_0(e,w,i) = \frac{\sum_{j \text{ impair}}^{w-1} \binom{t}{j}\binom{n-t-1}{w-1-j}}{\binom{n-1}{w-1}} = \frac{1}{2} + \varepsilon_0$$

Discerner deux distributions

Objectif : discerner deux distributions distantes de $|\varepsilon_1 - \varepsilon_0|$

$\rightarrow$ Chernoff + Neymann-Pearson : échantillon de taille minimale

$$P_w \triangleq \frac{\log_2(n)}{(\varepsilon_0 - \varepsilon_1)^2}$$

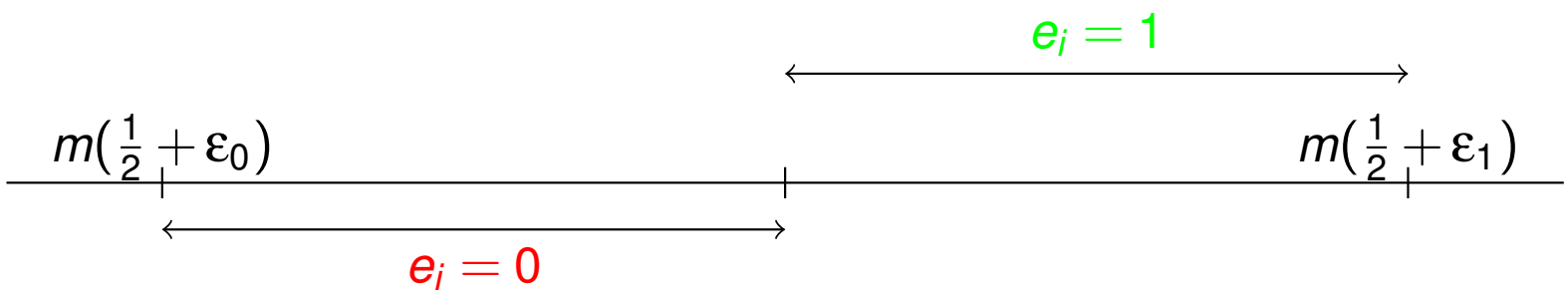
Un distingueur

$$V_m = \sum_{k=1}^{m} \mathrm{sgn}(\varepsilon_1 - \varepsilon_0)\langle y, h^k \rangle \in \mathbb{Z}$$

Proposition (Borne de Chernoff)

Selon $e_i = l$ on a :

$$\mathbb{P}\left(|V_m - m\,\mathrm{sgn}(\varepsilon_1 - \varepsilon_0)(1/2 + \varepsilon_l)| \geq m\frac{|\varepsilon_1 - \varepsilon_0|}{2} \right) \leq 2^{-2m\frac{(\varepsilon_1 - \varepsilon_0)^2}{2\ln(2)}}$$



Décodage Statistique

$\rightarrow$ Difficulté : trouver assez de vecteurs $h \in \mathcal{H}_w$ avec un algorithme $\texttt{CalculParité}_w$

$\rightarrow$ Il en faut : $O(P_w)$

Proposition

La complexité du décodage statistique est à un facteur polynomial près :

- *Si les équations de parité déja calculés :* $O(P_w)$
- *Sinon :* $O(P_w) + O(|\texttt{CalculParité}_w|)$

$$|\texttt{CalculParité}_w| \geq P_w$$

Notations

$$\pi(\omega,\tau) \stackrel{\triangle}{=} \varliminf_{n\to+\infty} \frac{1}{n} \log_2 P_w$$

$$\pi^{complete}(\omega,\tau) \stackrel{\triangle}{=} \varliminf_{n\to+\infty} \frac{1}{n} \max\left(\log_2 P_w, \log_2 \left|\texttt{CalculParité}_w\right| \right)$$

Exposant asymptotique

Théorème

On pose $\omega \stackrel{\triangle}{=} \frac{w}{n}$, $\tau \stackrel{\triangle}{=} \frac{t}{n}$ *et* $\gamma \stackrel{\triangle}{=} \frac{1}{\omega}$,

- *If* $\tau \in \left(0, \frac{1}{2} - \sqrt{\omega - \omega^2}\right)$:
  $\pi(\omega,\tau) = 2\omega\log_2(r) - 2\tau\log_2(1-r) - 2(1-\tau)\log_2(1+r) + 2H(\omega)$
  où r plus petite racine de $(1-\omega)X^2 - (1-2\tau)X + \omega = 0$.
- *If* $\tau \in \left(\frac{1}{2} - \sqrt{\omega - \omega^2}, \frac{1}{2}\right)$: $\pi(\omega,\tau) = H(\omega) + H(\tau) - 1$.

Polynômes de Krawtchouk

Polynôme de degré v, d'ordre m, p_v^m défini comme :

$$p_v^m(X) = \frac{(-1)^v}{2^v} \sum_{j=0}^{v} (-1)^j \binom{X}{j} \binom{m-X}{v-j}$$

Soit x un mot de poids i : $p_v^m(i) = \sum_{y \,:\, w_H(y)=v} (-1)^{\langle x,y \rangle}$

Polynômes de Krawtchouk et distribution de poids

$$W_i^{\mathcal{C}} \stackrel{\triangle}{=} \{x \in \mathcal{C} : w_H(x) = i\}$$

On a l'identité pour un code linéaire $\mathcal{C} \subset \mathbb{F}_2^n$:

$$W_u^{\mathcal{C}^\perp} = \frac{1}{|\mathcal{C}|} \sum_{j=0}^{n} W_i^{\mathcal{C}} p_u^n(i)$$

Biais et polynômes de Krawtchouk

$$\frac{(-2)^{w-2}}{\binom{n-1}{w-1}} p_{w-1}^{n-1}(t) = \varepsilon_0$$

$$-\frac{(-2)^{w-2}}{\binom{n-1}{w-1}} p_{w-1}^{n-1}(t-1) = \varepsilon_1$$

Équations de poids $\frac{Rn}{2}$

On calcule la matrice de parité H du code $\mathcal{C}$

Pivot de Gauss sur H : $[I_{n(1-R)}|H']$

Les lignes ont un poids de $\frac{Rn}{2}(1+o(1))$

$\rightarrow$ Coût polynomial par solution

Exposant asymptotique

Théorème

Avec l'algorithme qui précède

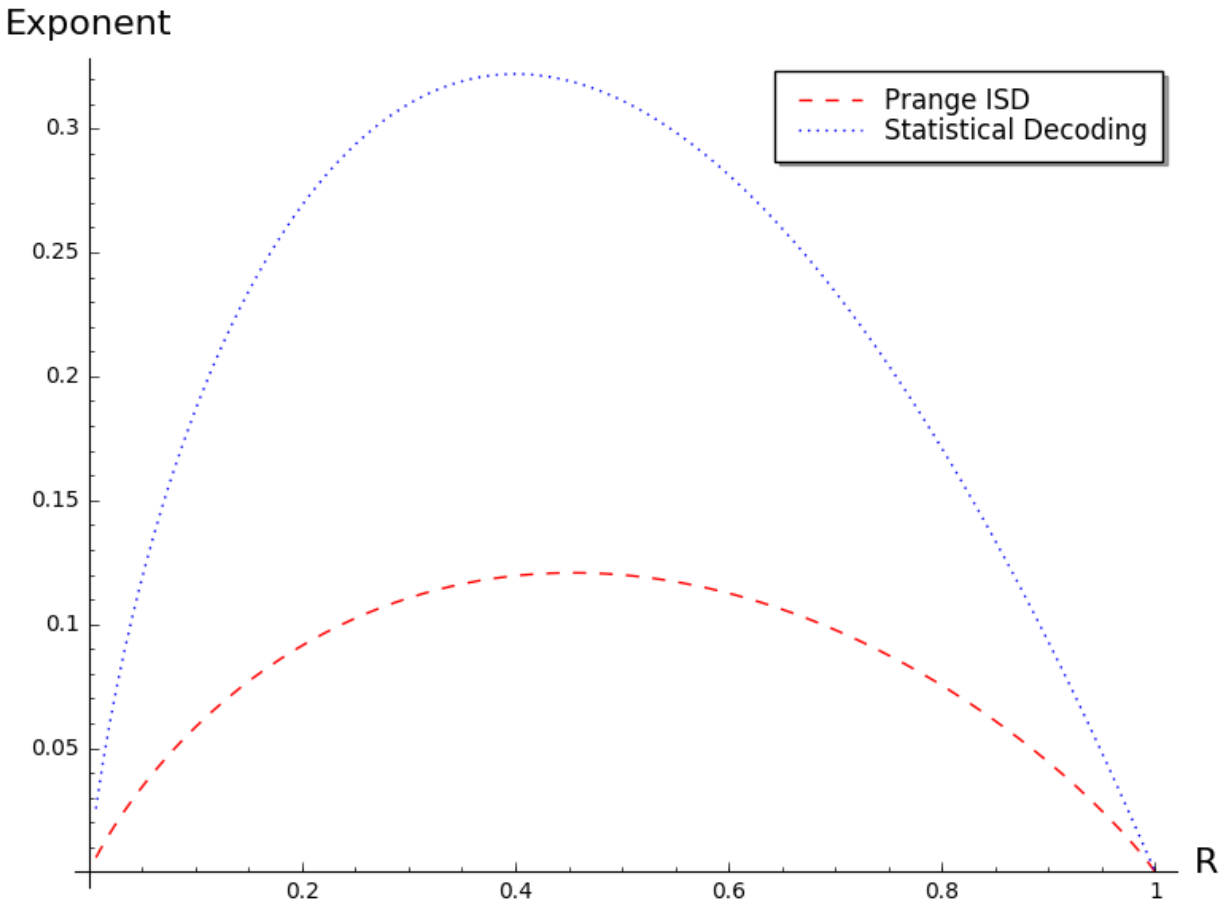
- *Si* $\tau = H^{-1}(1-R)$ *:*

$$\pi(R/2,\tau) = \pi(R/2,\tau)^{complete} = H(R/2) - R$$

- *Si* $\tau = o(n)$ *:*

$$\pi(R/2,\tau) = \pi^{complete}(R/2,\tau) = -2\tau\log_2(1-R)$$

Comparaison des exposants



Stratégie

On cherche un nombre P_w équations de $\mathcal{C}^\perp$ d'un poids w

$P_w \searrow$ si $w \searrow$

Trouver un algorithme pour w faible

Equations de parité

Dans un code aléatoire il y a $H_w \stackrel{\triangle}{=} \frac{\binom{n}{w}}{2^{nR}}$ équations de parité

$\rightarrow$ On cherche le w_0 le plus faible tel que :

$$P_{w_0} \leq H_{w_0}$$

La complexité du décodage statistique ne peut jamais être $< P_{w_0}$.

Fait suprenant

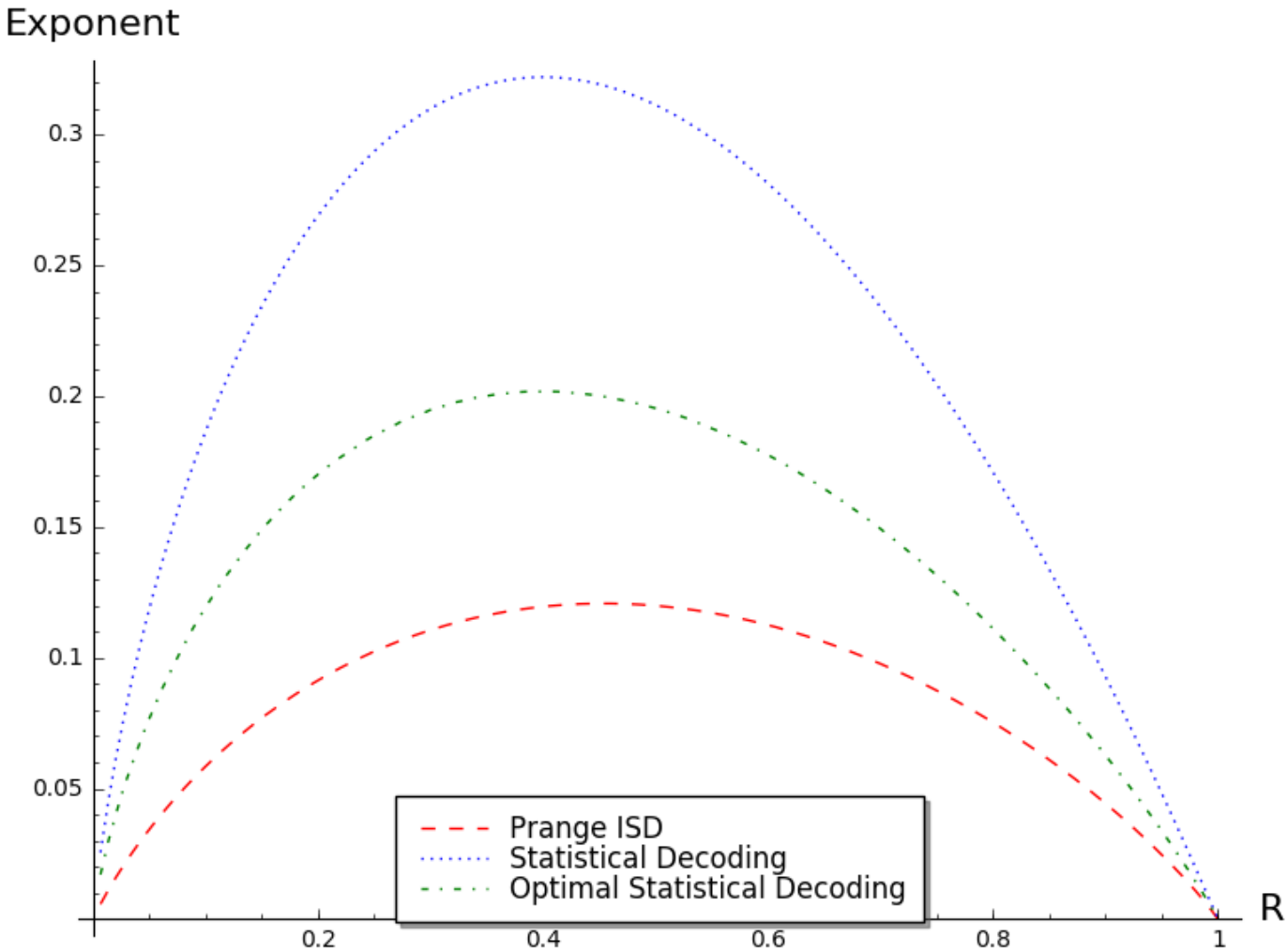
$t = nH^{-1}(1-R)$: nombre d'erreurs le plus difficile à corriger

Pour $\tau = H^{-1}(1-R) : \forall w \geq w_0 \quad : \quad P_w = H_w$

où

$$w_0 = n\left(\frac{1}{2} - \sqrt{\tau - \tau^2}\right)$$

Exposant optimal à Gilbert-Varshamov



Question Ouverte

Dans le cas où $t = o(n)$:

Nombre sous-exponentiel d'équations de poids $< Rn/2$ en temps sous-exponentiel ?