SORBONNE UNIVERSITÉ
CRÉATEURS DE FUTURS DEPUIS 1257

Inria

Cryptographie fondée sur les codes : nouvelles approches pour constructions et preuves ; contribution en cryptanalyse

Soutenance de thèse

17 Décembre 2019

Presenté par Thomas Debris-Alazard

Cryptographie

Objectif: permettre des échanges sécurisés

Cryptographie

Objectif: permettre des échanges sécurisés

Un chat apparait!

Il veut intercepter les messages!

Panorama de la cryptographie

Cryptographie à clef secrète

- Chiffrement symétrique,
- Fonction de hachage, authentification

Cryptographie à clef publique

- Chiffrement asymétrique, échange de clef
- Signature

Cryptographie à clef publique traditionnelle

La cryptographie à clef publique déployée repose sur deux problèmes:

Problème (Factorisation)

Entrée: $n = pq$ *où* p *et* q *sont premiers,*
Sortie: p *et* q*.*

Problème (Logarithme discret)

Entrée: g^x *où* g *générateur d'un groupe* $\mathbb{G}$,
Sortie: x*.*

$\rightarrow$ Résolution par l'algorithme quantique de Shor!

Pour une sécurité à long terme nous avons besoin de nouveau(x) problème(s)!

Cryptographie post-quantique

Il existe des domaines mathématiques dont les problèmes associés semblent difficile même avec un ordinateur quantique :

- Réseaux euclidiens,
- Codes correcteurs d'erreurs (dans cette thèse)
- Fonctions de hachage,
- Multi-variés,
- Isogénies.

Processus de standardisation de cryptographie à clef publique du gouvernement américain (NIST)

NIST standardisation post-quantique

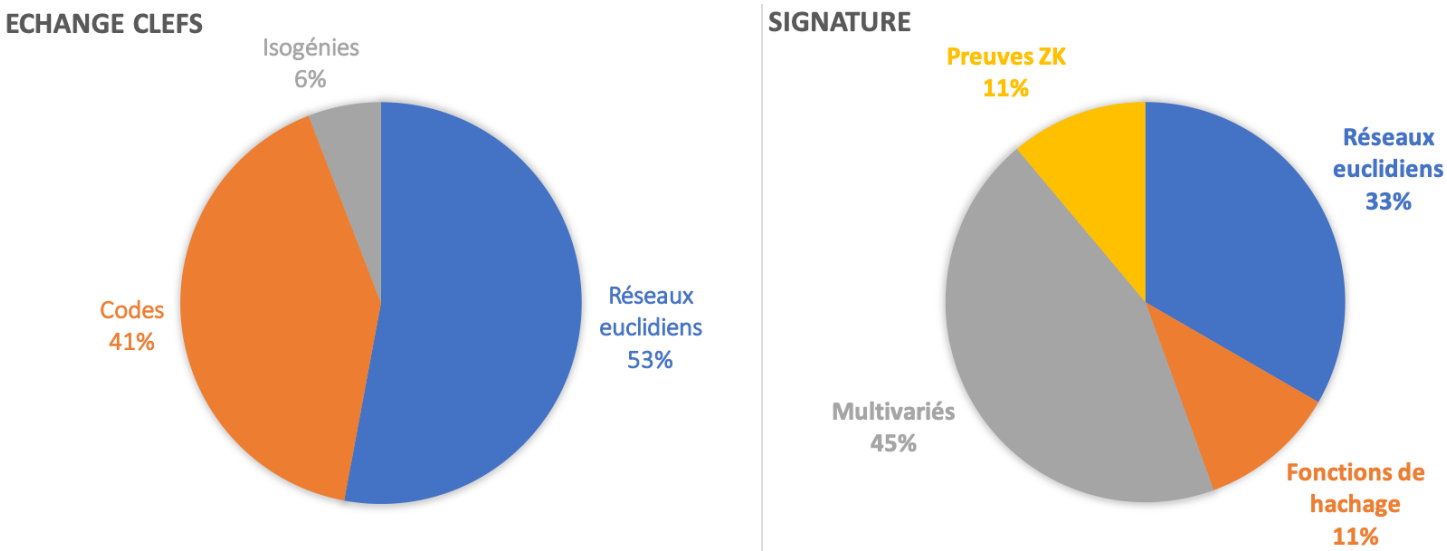


Figure: Propositions au second tour du NIST

→ Il n'y a pas de signature utilisant des codes...

Codes : problème du décodage générique

- Poids de Hamming:

$$|\mathbf{x}| \stackrel{\triangle}{=} \#\{1 \leq i \leq n : x_i \neq 0\}$$

Problème (Décodage générique)

Donnée: $\mathbf{H} \in \mathbb{F}_q^{(n-k)\times n}$ *et* $\mathbf{s} \in \mathbb{F}_q^{n-k}$
Sortie: $\mathbf{e}$ *tel que:* $|\mathbf{e}| = w$ *et* $\mathbf{e}\mathbf{H}^\intercal = \mathbf{s}$.

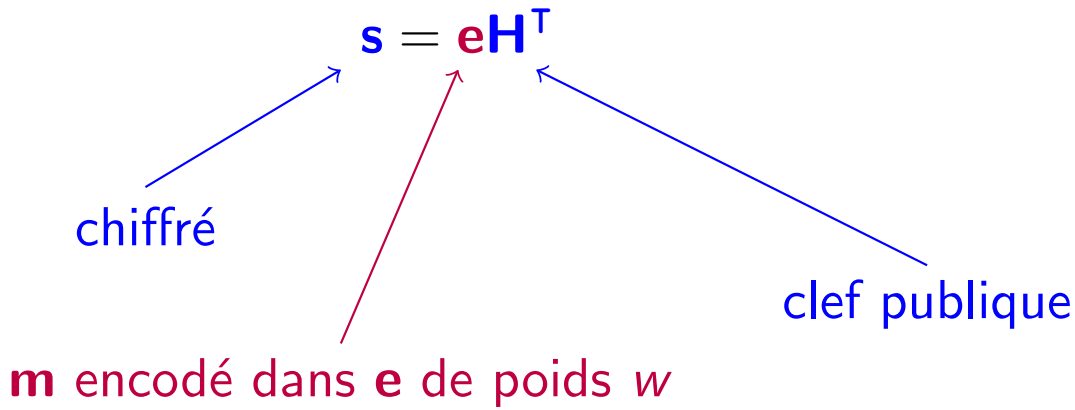
Après 60 années de recherche...

- NP-complet,
- Difficile en moyenne, meilleurs algorithmes pour le résoudre de complexité :

$$2^{cw(1+o(1))}$$

Chiffrement de McEliece/Niederreiter

- Chiffrement de $\mathbf{m}$:



- Déchiffrement de $\mathbf{s}$: utiliser une trappe.

Codes avec un décodage

- $\mathcal{C}$ code de longueur n, dimension k sur $\mathbb{F}_q$:

$$\mathcal{C} = \left\{ \mathbf{c} : \mathbf{c}\mathbf{H}^\intercal = \mathbf{0} \right\} \text{ où } \mathbf{H} \in \mathbb{F}_q^{(n-k)\times n} \text{ de rang } n-k.$$

$\rightarrow$ On sait décoder $\mathcal{C}$ avec une structure particulière!

Secret (McEliece/Niederreiter) : structure sur $\mathbf{H}$.

Problème : attaquer n'implique pas résoudre le décodage générique.

Sécurité de McEliece/Niederreiter

Attaquer McEliece/Niederreiter avec un chiffré implique:

1. résoudre le décodage générique à distance w,

Ou

2. distinguer $\mathbf{H}$ avec "structure" d'une matrice aléatoire

$\rightarrow$ Des codes tels que **2.** est difficile?

Des codes cryptographiques

Décodeur algébrique

- Codes de Goppa, décodage à distance $\approx \frac{n-k}{\log_2(n)}$.

Décodeur probabiliste

- Codes MDPC, décodage à distance $\approx \sqrt{n}$

Les signatures avec des codes?

- Protocole ZK Stern 93' + transformation de Fiat-Shamir 87'

  Longues signatures $\approx \Theta(\lambda^2)$ bits 

- KKS [Kabatianskii, Krouk, Smeets] 97', $\approx$ signature de Schnorr

  Fuite d'information, signature unique 

- CFS [Courtois, Finiasz, Sendrier] 01', hache et signe,

  Des Gigaoctets de taille de clef (128 bits de sécurité) 

- RankSign [GRSZ] 14', hache et signe (poids rang),

  Cassée par une attaque polynomiale 

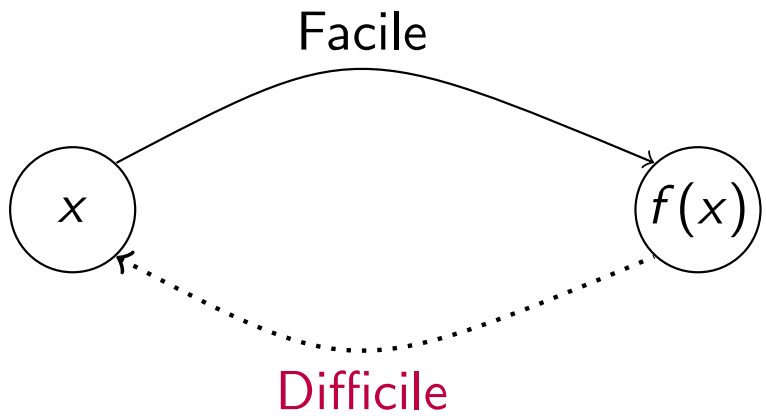
- Pas de signature au 2*nd* tour du NIST...

-  : Durandal [ABGHZ] 19', (poids rang), Schnorr-Lyubashevsky

  Fuite d'information? Absence de preuve 

CFS: une signature de type hache et signe

- $\mathcal{H}(\cdot)$ fonction de hachage,
- f fonction à sens unique à trappe



- Signature de $\mathbf{m}$:

$$\text{Calculer } \sigma \in f^{-1}(\mathcal{H}(\mathbf{m})).$$

CFS

- $\mathbf{H}_{\text{Goppa}} \in \mathbb{F}_2^{(n-k)\times n} \rightarrow$ code de Goppa,
- $w \triangleq (n-k)/\log_2(n)$ distance de décodage d'un code de Goppa.

Fonction à sens unique injective :

$$\begin{array}{rccc} f_{w,\mathbf{H}_{\text{Goppa}}} : & \{\mathbf{e} \in \mathbb{F}_2^n : |\mathbf{e}| = w\} & \longrightarrow & \mathbb{F}_2^{n-k} \\ & \mathbf{e} & \longmapsto & \mathbf{e}\mathbf{H}_{\text{Goppa}}^{\mathsf{T}} \end{array}$$

Ici,

$$\binom{n}{(n-k)/\log_2(n)} \approx 2^{n-k} \iff k \approx n$$

Le problème de CFS

- $k \nearrow n$ implique $\searrow$ coût inversion,
- $k \nearrow n$ implique $\searrow$ coût attaques.

$\rightarrow$ Mauvaise mise à l'échelle des paramètres...

Résultats thèse (I) : Cryptanalyse

- Codes LRPC RankSign [GRSZ14] : mots de petits poids rang

  $\rightarrow$ contraintes sur le décodeur par effacement imposées par le hache et signe

- Attaque RankSign en retrouvant (par base de Gröbner) ces mots

  $\longrightarrow$ Ne peut être évité en changeant les paramètres

- Attaque du chiffrement fondé sur l'identité (IBE) [GHPT17] avec des codes (métrique rang);

  $\longrightarrow$ Paramètres possibles pour éviter l'attaque

- IBE : rang $\rightarrow$ Hamming à ne pas faire.

Résultats thèse (II) : Wave

- Une signature avec des codes dans un régime surjectif

  $\rightarrow$ Régime avec possible fuite d'information...

- Approche des réseaux euclidiens, Gentry-Peikert-Vaikuntanathan (GPV)

  $\rightarrow$ Aucune fuite d'information, méthode de rejet

- Nouvelle trappe en cryptographie : codes $(U, U+V)$-généralisés

  $\rightarrow$ Distinguer ces codes de codes aléatoires : NP-complet

- Réduction de sécurité aux problèmes :
  - Distinguer codes $(U, U+V)$-généralisés de codes aléatoires,
  - Décodage générique en grande distance.

Résultats thèse (III) : décodage en grande distance

- Problème de décodage en grande distance : $n/2 < w$
- Cruciale pour paramètres de Wave.
- Adaptation algorithmes par ensemble d'information

  $\rightarrow$ Plus difficile dans $\mathbb{F}_q$ $(q \geq 3)$ que de décoder à petite distance dans $\mathbb{F}_2$!

Résultats thèse (IV) : décodage statistique

- Décodage statistique : $\mathbf{c} + \mathbf{e}$ où $\mathbf{c} \in \mathcal{C}$, retrouver $\mathbf{e}$...

  $\rightarrow \mathbf{h} \in \mathcal{C}^{\perp}$, calculer $\langle \mathbf{c} + \mathbf{e}, \mathbf{h} \rangle = \langle \mathbf{e}, \mathbf{h} \rangle$: donne des stat. sur $\mathbf{e}$

  $\rightarrow$ Complexité asymptotique inconnue

- Obtention de cette complexité (nouveaux résultats sur les polynômes de Krawtchouk)

- Amélioration du calcul d'équations de parité $\mathbf{h} \in \mathcal{C}^{\perp}$

- Décodage statistique non compétitif avec le plus simple des algos. par ensemble d'information.

La fonction à sens unique

- $|\cdot|$ poids de Hamming
- $\mathbf{H}$ matrice sur $\mathbb{F}_q$ avec $n-k$ lignes et n colonnes
- w entier (poids)

Fonction à sens unique en code :

$$\begin{array}{rccc} f_{w,\mathbf{H}} : & \{\mathbf{e} \in \mathbb{F}_q^n : |\mathbf{e}| = w\} & \longrightarrow & \mathbb{F}_q^{n-k} \\ & \mathbf{e} & \longmapsto & \mathbf{H}\mathbf{e}^\intercal \end{array}$$

Pour espérer $f_{w,\mathbf{H}}$ surjective, choisir w assez gros

$$w \geq (1+\varepsilon) w_{\mathrm{GV}} \text{ où } q^{n-k} \approx \binom{n}{w_{\mathrm{GV}}} (q-1)^{w_{\mathrm{GV}}}$$

Typiquement nombre exponentiel de pré-images...

Gentry-Peikert-Vaikuntanathan (GPV) approche

Ajouter des propriétés à $f_{w,\mathbf{H}}$: fonction "preimage sampleable" !

- $\overset{\$}{\leftarrow}$ tirage uniforme,
- S_w mots de poids de Hamming w.

1. Trap. algo: $\forall \mathbf{s}$, $\mathbf{e} \leftarrow f^{-1}_{w,\mathbf{H}}(\mathbf{s})$ distribué comme $\mathbf{e} \overset{\$}{\leftarrow} S_w \cap f^{-1}_{w,\mathbf{H}}(\mathbf{s})$.

Gentry-Peikert-Vaikuntanathan (GPV) approche

Ajouter des propriétés à $f_{w,\mathbf{H}}$: fonction "preimage sampleable"!

- $\xleftarrow{\$}$ tirage uniforme,
- S_w mots de poids de Hamming w.

1. ~~Trap. algo: $\forall \mathbf{s}, \mathbf{e} \leftarrow f^{-1}_{w,\mathbf{H}}(\mathbf{s})$ distribué comme $\mathbf{e} \xleftarrow{\$} S_w \cap f^{-1}_{w,\mathbf{H}}(\mathbf{s})$.~~

Nous relaxons : $f^{-1}_{w,\mathbf{H}}(\mathbf{s}^{\text{unif}}) \xleftarrow{\$} S_w$ pour $\mathbf{s}^{\text{unif}}$ uniformément distribué.

$\rightarrow$ Suffisant pour la réduction de sécurité dans le ROM

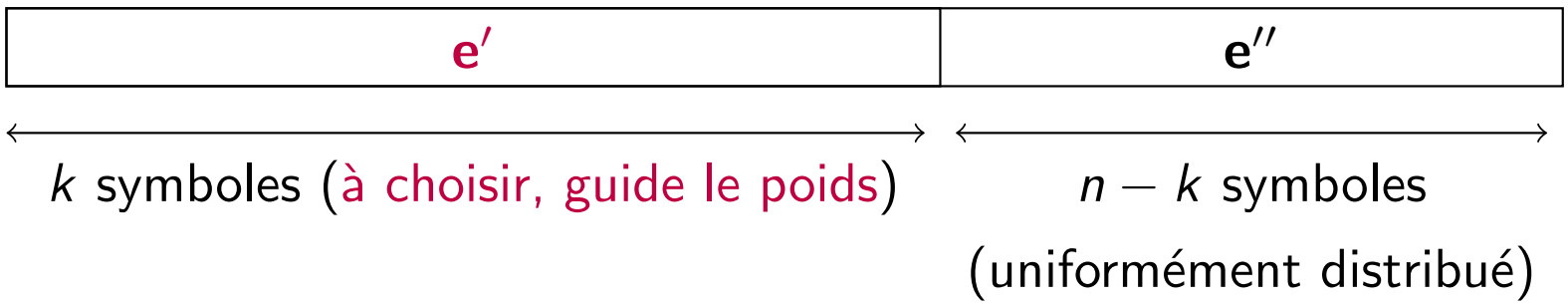
2. $f_{w,\mathbf{H}}(\mathbf{e})$ uniformément distribué quand $\mathbf{e} \xleftarrow{\$} S_w$,

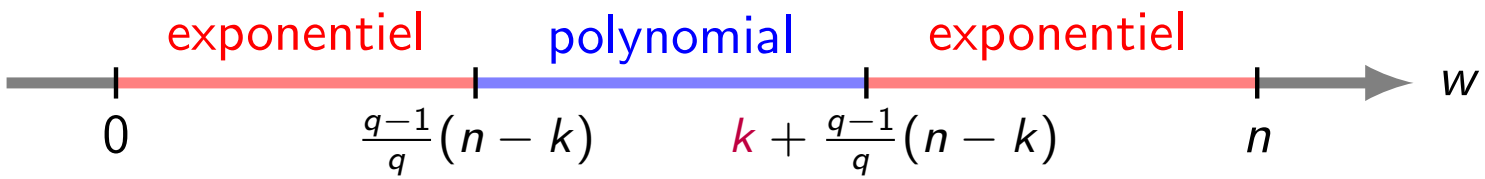
Algorithme de Prange

Donnée : $\mathbf{H} \in \mathbb{F}_q^{(n-k)\times n}$ et $\mathbf{s}$ uniformément distribué sur $\mathbb{F}_q^{n-k}$;

Sortie : $\mathbf{e} \in \mathbb{F}_q^n$ tel que (i) $|\mathbf{e}| = w$ et (ii) $\mathbf{H}\mathbf{e}^\intercal = \mathbf{s}^\intercal$.

$\rightarrow$ système linéaire avec n inconnues $> n-k$ équations

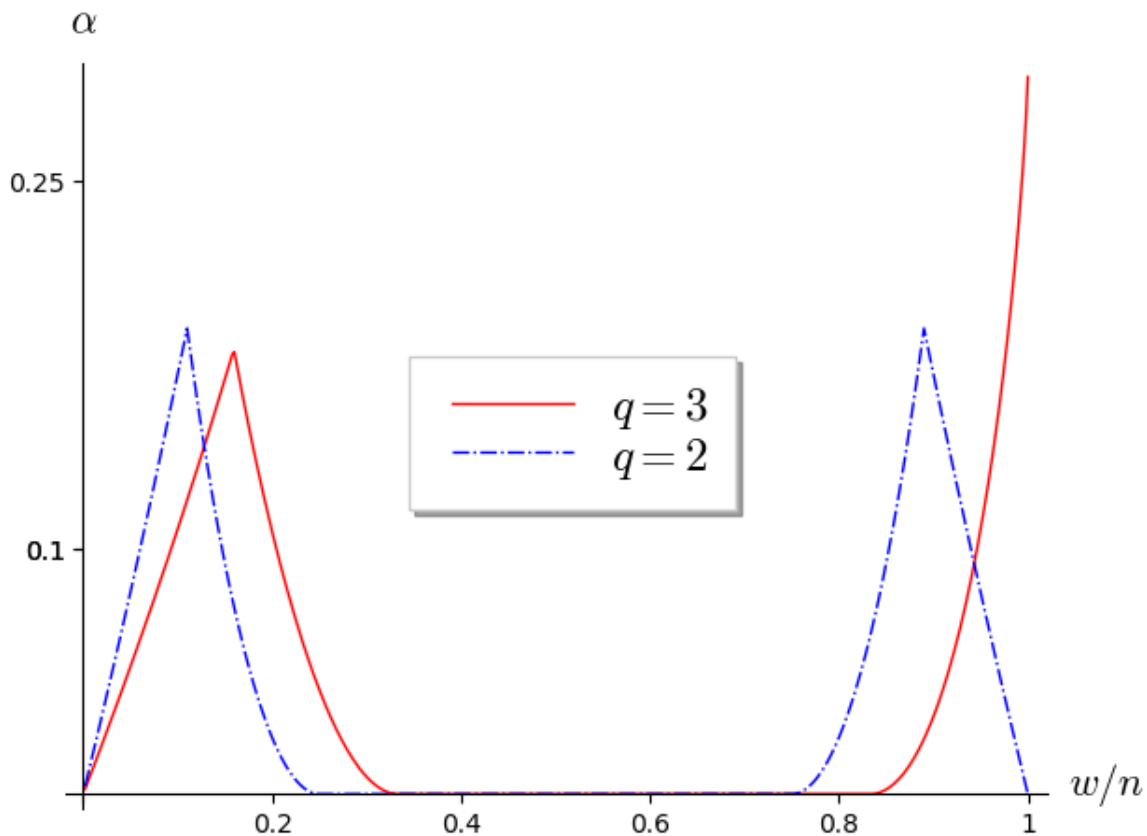




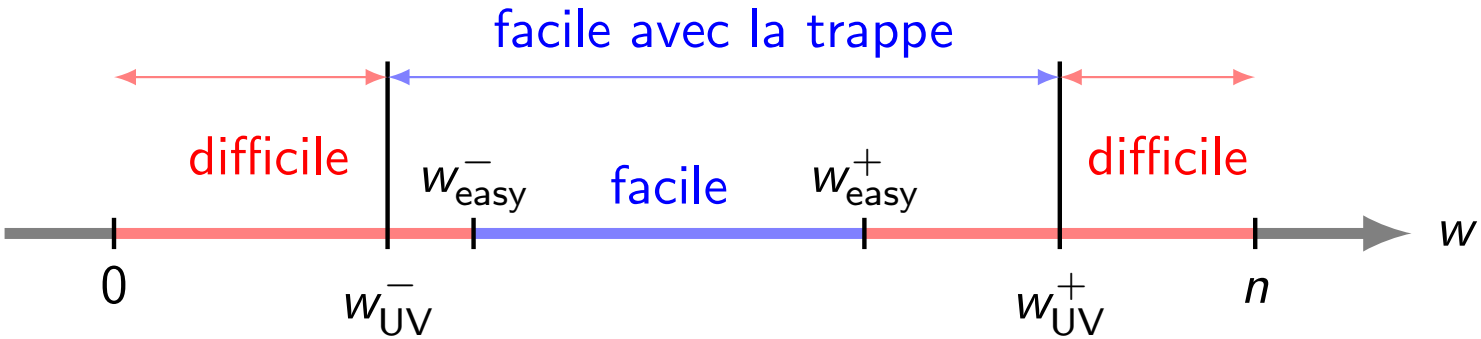
Exposant de Prange pour $q = 2$ et $q = 3$

Complexité : $2^{\alpha n}$ où α fonction de w/n.

Figure: Exposant vs poids relatif pour $k/n = 1/2$



Décoder avec la trappe



Notre trappe (I)

Matrice de parité d'un $(U, U+V)$: $\mathbf{H}_{\text{sec}} \triangleq \begin{pmatrix} \mathbf{H}_U & \mathbf{0} \\ -\mathbf{H}_V & \mathbf{H}_V \end{pmatrix}$ (n colonnes ; $n/2$ colonnes ; $n/2 - k_U$ lignes ; $n/2 - k_V$ lignes)

où $\mathbf{H}_U$ et $\mathbf{H}_V$ sont aléatoires!

Cacher la trappe: $\mathbf{P}$ permutation, $\mathbf{S}$ inversible et

$$\mathbf{H}_{\text{pub}} \triangleq \mathbf{S}\mathbf{H}_{\text{sec}}\mathbf{P} \text{ : publique}$$

Hypothèse de sécurité : Distinguer $\mathbf{H}_{\text{pub}}$/matrice aléatoire (même taille) est difficile.

Proposition

Le problème de décision sous-jacent est NP-complet.

Notre trappe (II)

Soit,

$$\mathbf{e} = (\mathbf{e}_U, \mathbf{e}_U + \mathbf{e}_V) \quad ; \quad \mathbf{s} = (\mathbf{s}_U, \mathbf{s}_V)$$

$$\boxed{\mathbf{H}_{\text{sec}}\mathbf{e}^\intercal = \mathbf{s}^\intercal \iff \begin{cases} \mathbf{H}_U\mathbf{e}_U^\intercal = \mathbf{s}_U^\intercal \\ \mathbf{H}_V\mathbf{e}_V^\intercal = \mathbf{s}_V^\intercal \end{cases}}$$

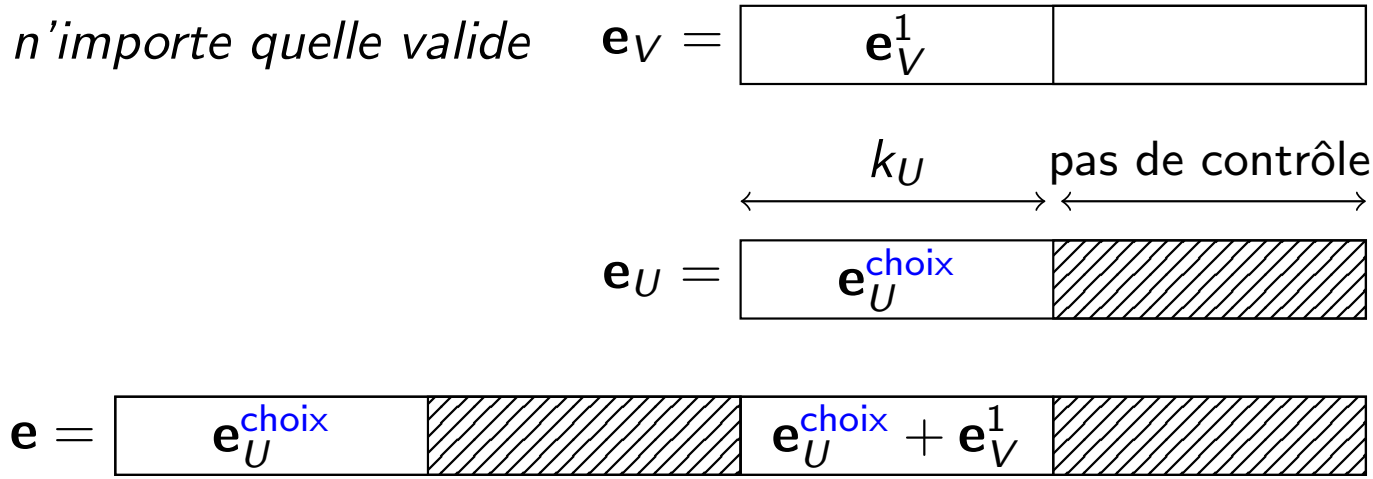
$$k_U + k_V = \text{Ncols}(\mathbf{H}_{\text{sec}}) - \text{Nlgns}(\mathbf{H}_{\text{sec}})$$

$$k_U = \text{Ncols}(\mathbf{H}_U) - \text{Nlgns}(\mathbf{H}_U) \quad \text{and} \quad k_V = \text{Ncols}(\mathbf{H}_V) - \text{Nlgns}(\mathbf{H}_V)$$

$\rightarrow$ Prange directement sur $\mathbf{H}_{\text{sec}}$ choisit $k_U + k_V$ symboles de $\mathbf{e}$ mais $\mathbf{e}_U$ apparait deux fois ($k_U > k_V$)...

Notre décodeur

Solution $\mathbf{e} = (\mathbf{e}_U, \mathbf{e}_U + \mathbf{e}_V) \in \mathbb{F}_q^n$ de forme :



Pour obtenir une solution de poids maximal

- choisir k_U symboles $\mathbf{e}_U^{\text{choix}}(i)$ t.q : $\begin{cases} \mathbf{e}_U^{\text{choix}}(i) \neq 0 \\ \mathbf{e}_U^{\text{choix}}(i) + \mathbf{e}_V^1(i) \neq 0 \end{cases}$

$\rightarrow$ Possible car travail dans $\mathbb{F}_q$ avec $q \geq 3$

$\rightarrow$ Gain en choisissant $2k_U > k_U + k_V$

1 Introduction

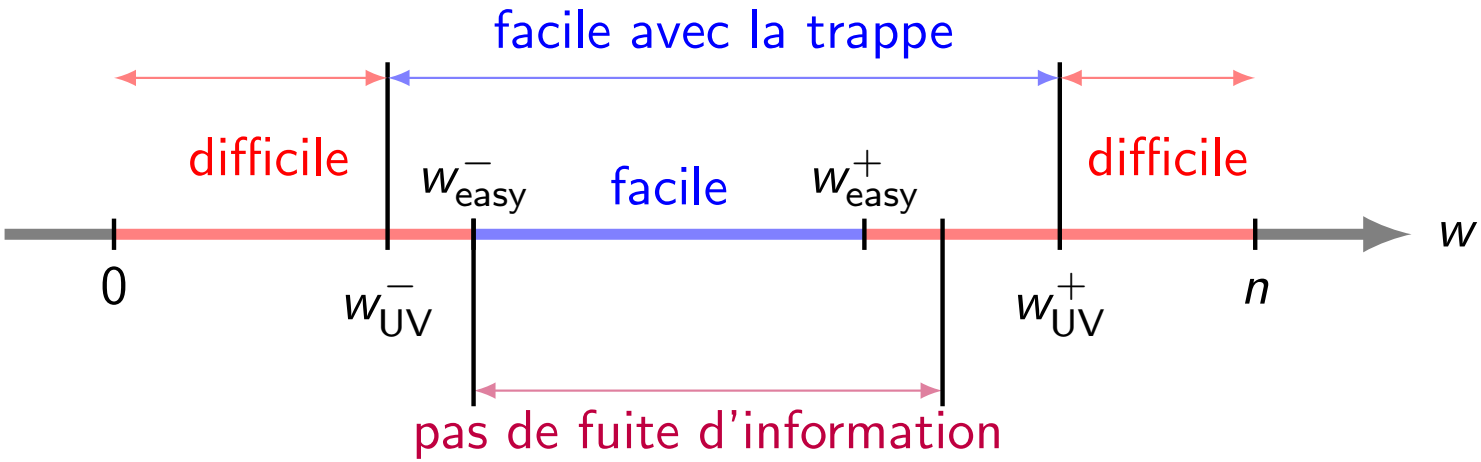
2 Résultats de la thèse

3 Wave

- Contexte
- Décoder avec la trappe
- Signatures sans fuite d'information
- Conclusion sur Wave

4 Conclusion et perspective

Signatures sans fuite d'information



Travail maintenant avec $q = 3$.

Signatures sans fuite d'information

$\mathbf{e}^{\text{sgn}} \triangleq (\mathbf{e}_1^{\text{sgn}}, \mathbf{e}_2^{\text{sgn}})$ signature, $\mathbf{e}^{\text{unif}} \triangleq (\mathbf{e}_1, \mathbf{e}_2)$ mot unif de poids w.

Objectif,

$$\mathbf{e}^{\text{sgn}} \sim \mathbf{e}^{\text{unif}}$$

$$\begin{cases} \mathbf{e}_1^{\text{sgn}} = \mathbf{e}_U \\ \mathbf{e}_2^{\text{sgn}} = \mathbf{e}_U + \mathbf{e}_V \end{cases} \iff \begin{cases} \mathbf{e}_1^{\text{sgn}} = \mathbf{e}_U \\ \mathbf{e}_2^{\text{sgn}} - \mathbf{e}_1^{\text{sgn}} = \mathbf{e}_V \end{cases}$$

Première étape,

$$\mathbf{e}_V \sim \mathbf{e}_2 - \mathbf{e}_1 \quad \text{où} \quad \mathbf{e}_V = \text{Prange}(\mathbf{H}_V, \mathbf{s}_V)$$

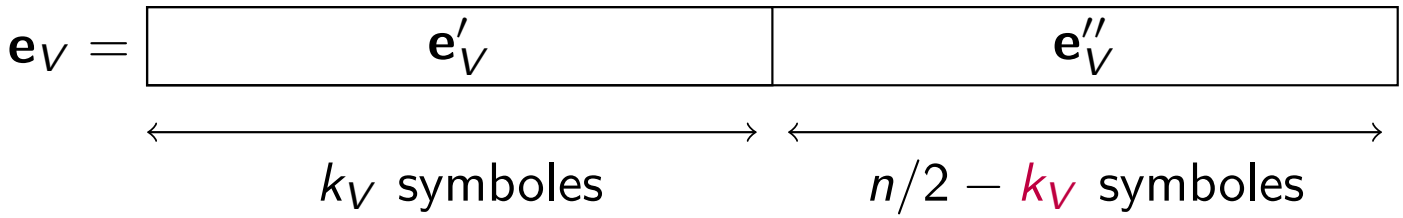
Première approximation, distribution de Prange seulement fonction du poids :

$$\mathbb{P}(\text{Prange}(\cdot) = \mathbf{x} \mid |\text{Prange}(\cdot)| = |\mathbf{x}|) = \frac{1}{\#\{\mathbf{y} : |\mathbf{y}| = |\mathbf{x}|\}}$$

$\rightarrow$ Avec uniformité de poids, suffisant que $|\mathbf{e}_V| \sim |\mathbf{e}_2 - \mathbf{e}_1|$

Guider le poids de $\mathbf{e}_V$

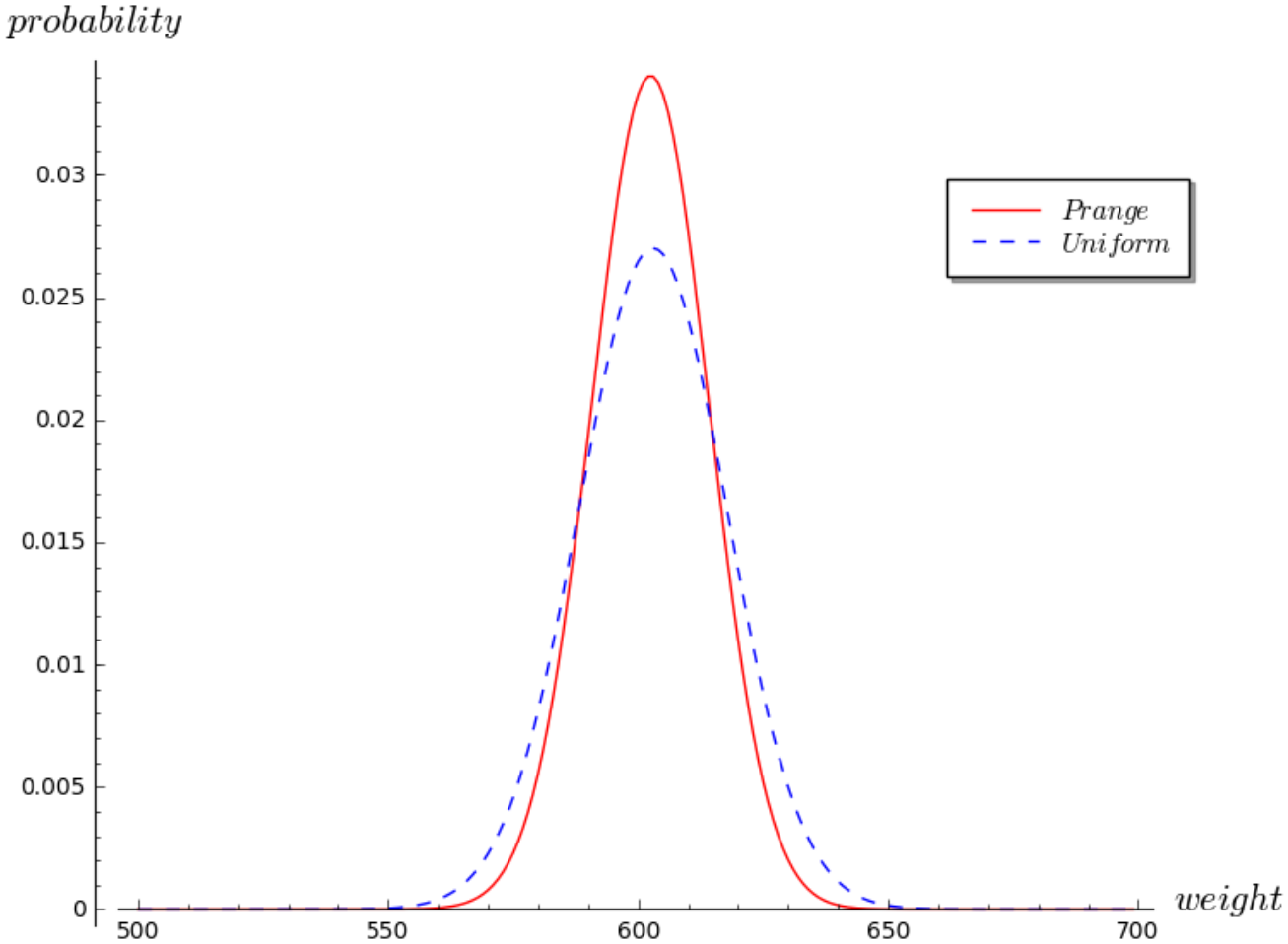
- Première condition : $\mathbb{E}(|\mathbf{e}_V|) = \mathbb{E}(|\mathbf{e}_2 - \mathbf{e}_1|)$ où $\mathbf{e}^{\text{unif}} \triangleq (\mathbf{e}_1, \mathbf{e}_2)$



- $\mathbf{e}''_V$ distribuée uniformément sur $\mathbb{F}_3^{n/2-k_V}$: $\mathbb{E}(|\mathbf{e}''_V|) = \frac{2}{3}(n/2 - k_V)$
- $\mathbf{e}'_V$ peut être choisi

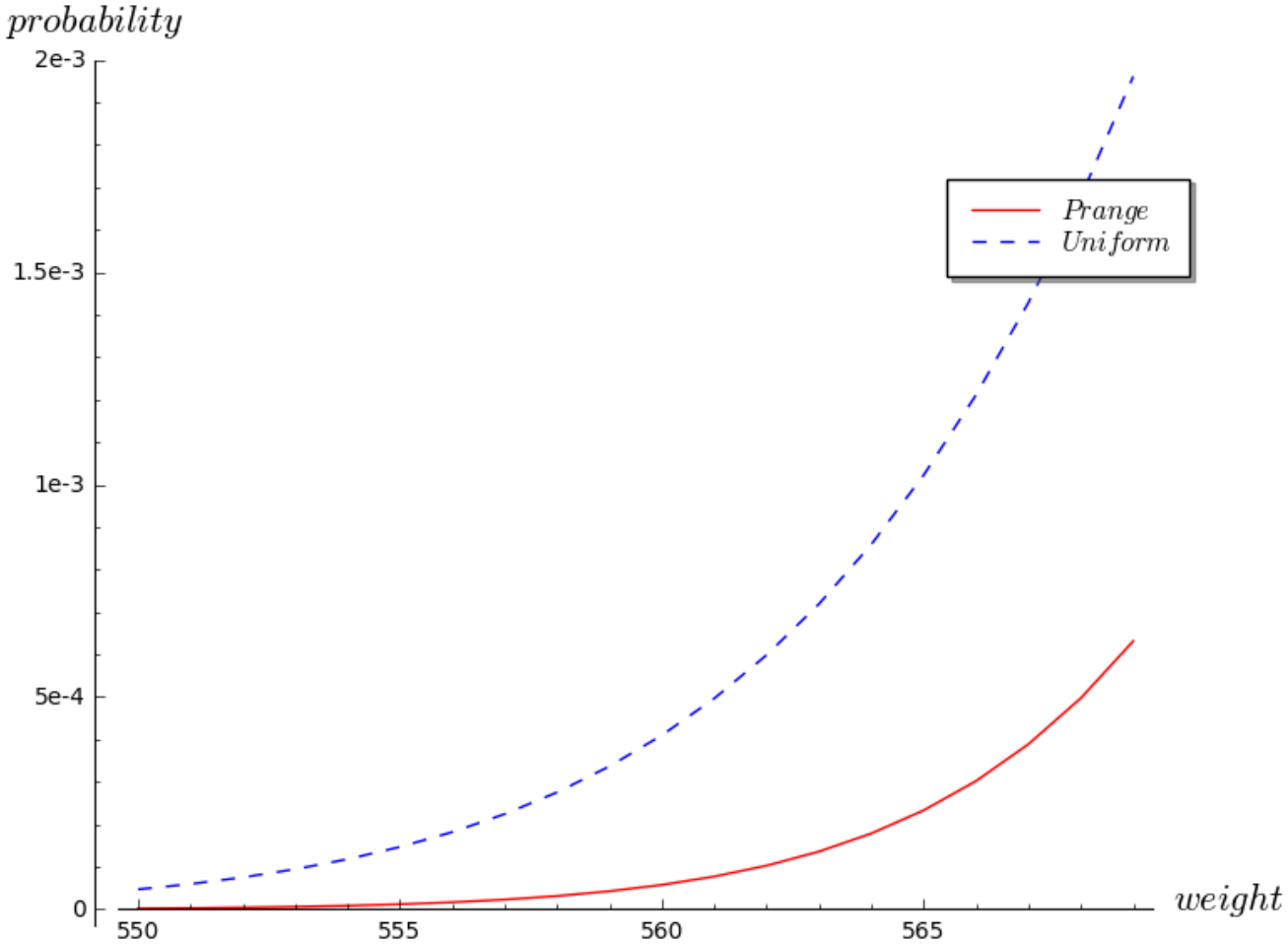
$\rightarrow k_V$ est choisi comme : $|\mathbf{e}'_V| + \frac{2}{3}(n/2 - k_V) = \mathbb{E}(|\mathbf{e}_2 - \mathbf{e}_1|)$

Méthode de rejet



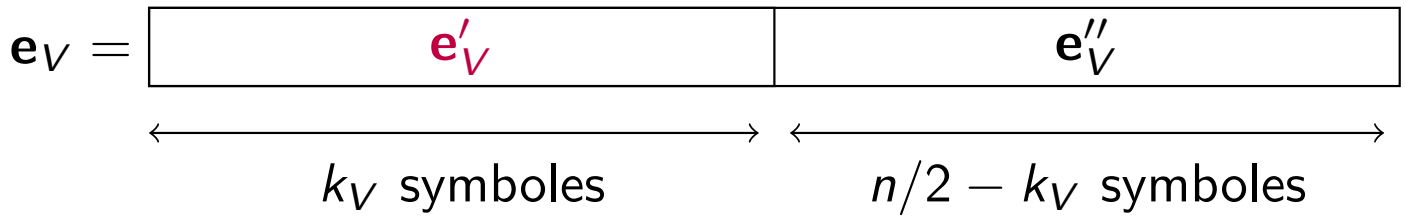
$$\mathbb{P}(\text{accepte}) = \min_j \frac{\mathbb{P}(|\mathbf{e}_V| = j)}{\mathbb{P}(|\mathbf{e}_2 - \mathbf{e}_1| = j)}$$

Méthode de rejet : queue



$$\mathbb{P}(\text{accepte}) = \min_j \frac{\mathbb{P}(|\mathbf{e}_V| = j)}{\mathbb{P}(|\mathbf{e}_2 - \mathbf{e}_1| = j)}$$

Choix probabiliste de $\mathbf{e}'_V$

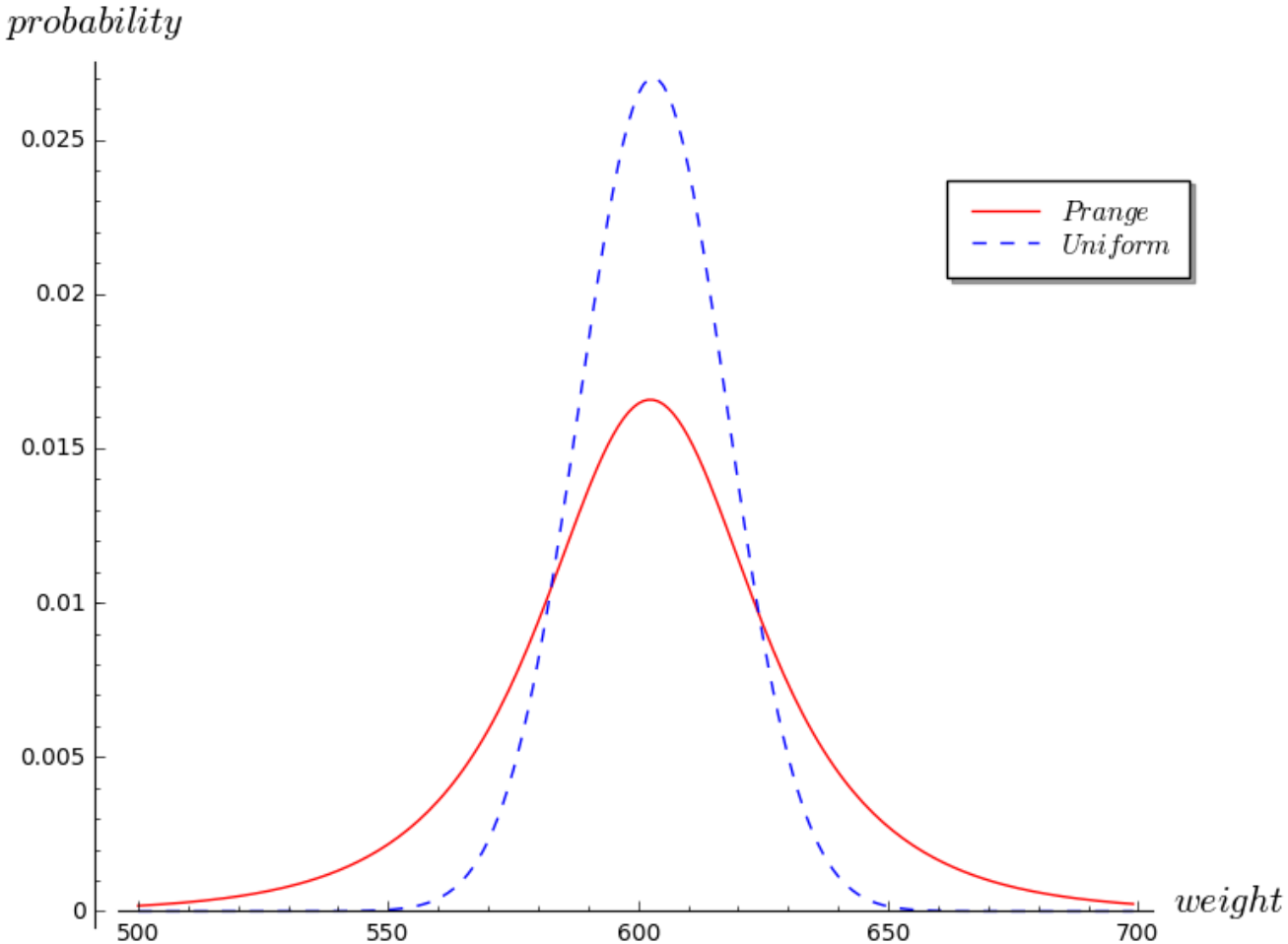


- $\mathbf{e}''_V$: variance fixée,

Choisir le poids de $\mathbf{e}'_V$ comme une variable aléatoire!

- $|\mathbf{e}'_V|$ tel que : $\begin{cases} \mathbb{E}(|\mathbf{e}'_V|) + \frac{2}{3}(n/2 - k_V) = \mathbb{E}(|\mathbf{e}_2 - \mathbf{e}_1|) \\ |\mathbf{e}'_V| \text{ grande variance!} \end{cases}$

Méthode de rejet



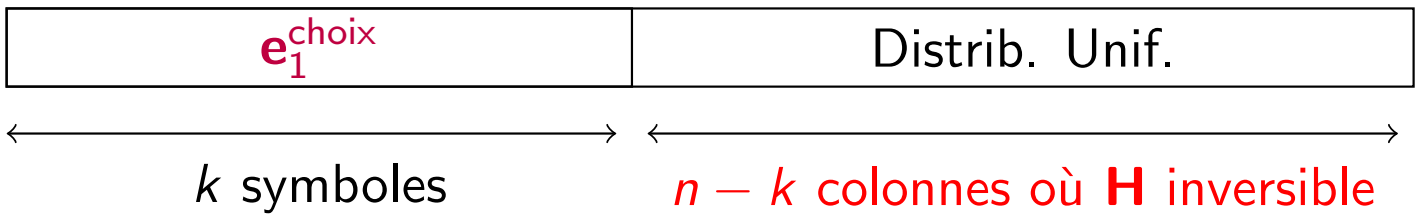
$$\mathbb{P}(\text{accepte}) = \min_j \frac{\mathbb{P}(|\mathbf{e}_V| = j)}{\mathbb{P}(|\mathbf{e}_2 - \mathbf{e}_1| = j)}$$

Propriété non-uniforme Prange

$$\mathbb{P}(\text{Prange}(\cdot) = \mathbf{x} \mid |\text{Prange}(\cdot)| = |\mathbf{x}|) = \frac{1}{\#\{\mathbf{y} : |\mathbf{y}| = |\mathbf{x}|\}} \quad : \text{ seulement } \approx.$$

Étant donné $\mathbf{H} \in \mathbb{F}_3^{(n-k)\times n}$, $\mathbf{s} \in \mathbb{F}_3^{n-k}$, trouver $\mathbf{e} \in \mathbb{F}_3^n$ t.q $\mathbf{H}\mathbf{e}^\intercal = \mathbf{s}^\intercal$.

$\rightarrow$ Système linéaire n inconnues $> n-k$ équations

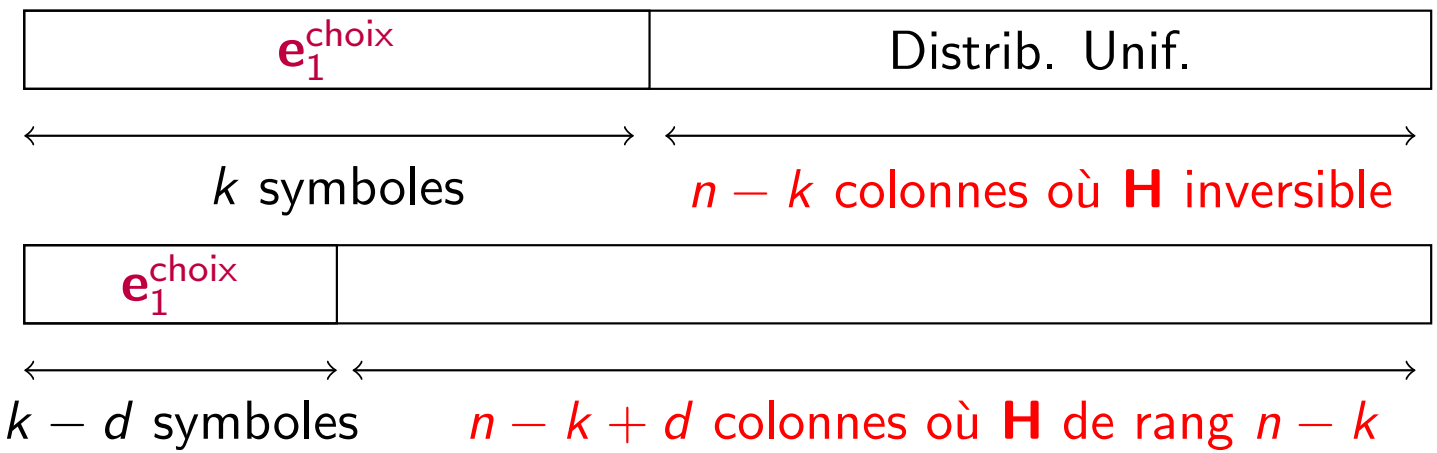


Propriété non-uniforme Prange

$$\mathbb{P}(\text{Prange}(\cdot) = \mathbf{x} \mid |\text{Prange}(\cdot)| = |\mathbf{x}|) = \frac{1}{\#\{\mathbf{y} : |\mathbf{y}| = |\mathbf{x}|\}} \quad : \text{ seulement } \approx.$$

Étant donné $\mathbf{H} \in \mathbb{F}_3^{(n-k)\times n}$, $\mathbf{s} \in \mathbb{F}_3^{n-k}$, trouver $\mathbf{e} \in \mathbb{F}_3^n$ t.q $\mathbf{H}\mathbf{e}^\intercal = \mathbf{s}^\intercal$.

$\rightarrow$ Système linéaire n inconnues $> n-k$ équations



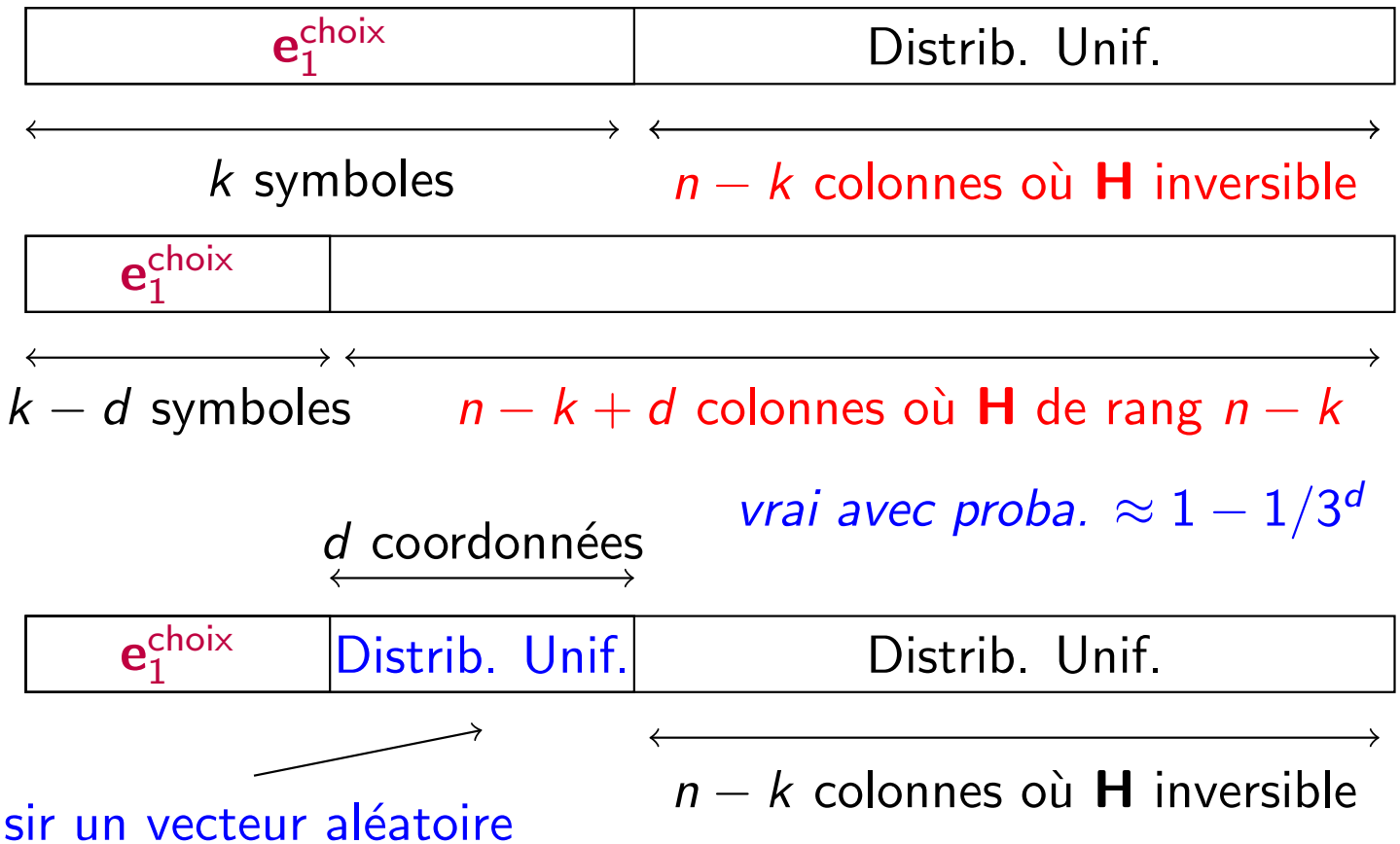
vrai avec proba. $\approx 1 - 1/3^d$

Propriété non-uniforme Prange

$$\mathbb{P}(\text{Prange}(\cdot) = \mathbf{x} \mid |\text{Prange}(\cdot)| = |\mathbf{x}|) = \frac{1}{\#\{\mathbf{y} : |\mathbf{y}| = |\mathbf{x}|\}} \quad : \text{ seulement } \approx.$$

Étant donné $\mathbf{H} \in \mathbb{F}_3^{(n-k)\times n}$, $\mathbf{s} \in \mathbb{F}_3^{n-k}$, trouver $\mathbf{e} \in \mathbb{F}_3^n$ t.q $\mathbf{H}\mathbf{e}^\intercal = \mathbf{s}^\intercal$.

$\rightarrow$ Système linéaire n inconnues $> n-k$ équations



Obtenir des signatures uniformes

Théorème

Soient $\mathbf{e}^{\text{sgn}}$ *signature,* $\mathbf{e}^{\text{unif}}$ *uniformément distribuée de poids* w*. Nous avons pour* P, Q *des polynômes et* ρ *distance statistique*

$$\mathbb{P}_{\mathbf{H}_{\text{pub}}}\left(\rho(\mathbf{e}^{\text{sgn}}, \mathbf{e}^{\text{unif}}) > Q(d)3^{-d}\right) \leq P(d)3^{-d}.$$

Nous prouvons aussi :

$$\rho(\mathbf{H}_{\text{pub}}\mathbf{e}^{\mathsf{T}}, \mathbf{s}^{\text{unif}}) \text{ négligeable} \quad \text{où} \quad \mathbf{e} \xleftarrow{\$} S_w \quad \text{et} \quad \mathbf{s}^{\text{unif}} \xleftarrow{\$} \mathbb{F}_3^{n-k}$$

$\rightarrow$ Dans les deux cas, l'aléa sur les matrices $\mathbf{H}_U$ et $\mathbf{H}_V$ est crucial...

Conclusion sur Wave

- La première signature avec des codes de type "hache et signe" qui suit la stratégie GPV

- Mise à l'échelle du schéma (en bits):

$$\text{signature de longueur} = 105\lambda \quad \text{et} \quad \text{taille de clef} = 1565\lambda^2$$

Perspective :

- Algorithmes pour des distinguer des code $(U, U+V)$-généralisés de codes aléatoires: actuellement utilise des algorithmes de décodage générique,

- Une meilleure réduction que la NP-complétude

- Espoir de supprimer la méthode du rejet

  $\rightarrow$ De nombreux degrés de liberté dans l'algorithme de Prange!

1 Introduction

2 Résultats de la thèse

3 Wave

Contexte

Décoder avec la trappe

Signatures sans fuite d'information

Conclusion sur Wave

4 Conclusion et perspective

Conclusion et perspective

Conclusion:

- Wave, une signature utilisant des codes,
- Étude du décodage générique,
  1. pour des nouveaux régimes (grande distance),
  2. avec un algorithme non-usuel (décodage statistique).
- Deux cryptanalyses, RankSign et un IBE.

Perspectives:

- Primitives cryptographiques avec le décodage en grande distance?
- Étendre techniques réseaux euclidiens aux codes?
  - Une signature de type Schnorr-Lyubashevsky avec des codes,
  - Un paramètre de lissage avec des codes.

Conclusion et perspective

Conclusion:

- Wave, une signature utilisant des codes,
- Étude du décodage générique,
  1. pour des nouveaux régimes (grande distance),
  2. avec un algorithme non-usuel (décodage statistique).
- Deux cryptanalyses, RankSign et un IBE.

Perspectives:

- Primitives cryptographiques avec le décodage en grande distance?
- Étendre techniques réseaux euclidiens aux codes?
  - Une signature de type Schnorr-Lyubashevsky avec des codes,
  - Un paramètre de lissage avec des codes.

Merci!