Decoding One Out of Many

Nicolas Sendrier

INRIA Paris-Rocquencourt, équipe-projet SECRET

Code-based Cryptography Workshop

11-12 May 2011, Eindhoven, The Netherlands

Computational Syndrome Decoding

Problem: Syndrome Decoding
Instance: $H \in \{0,1\}^{r \times n}$, $s \in \{0,1\}^r$ and $w > 0$
Question: is there a word e of Hamming weight w such that $He^T = s$?

Problem: Computational Syndrome Decoding (CSD)
Given $H \in \{0,1\}^{r \times n}$, $s \in \{0,1\}^r$ and $w > 0$
Find a word e of Hamming weight w such that $He^T = s$

NP-hard, conjectured hard in the average case

We will denote CSD(H, s, w) this problem as well as the set of its solutions

Typically $n = 2048$, $r = 352$ and $w = 32$

Computational Syndrome Decoding – Multiple instances

Problem: Syndrome Decoding One Out of Many

Instance: $H \in \{0,1\}^{r \times n}$, $\mathcal{S} \subset \{0,1\}^r$ and $w > 0$

Question: is there a word e of Hamming weight w such that $He^T \in \mathcal{S}$?

Problem: Computational Syndrome Decoding One Out of Many

Given $H \in \{0,1\}^{r \times n}$, $\mathcal{S} \subset \{0,1\}^r$ and $w > 0$

Find a word e of Hamming weight w such that $He^T \in \mathcal{S}$

For convenience, we will also denote CSD$(H, \mathcal{S}, w)$ this problem and the set of its solutions

Message Security of Code-Based Public-Key Cryptosystems

The public key is a parity check matrix $H_0 \in \{0,1\}^{r \times n}$ (or a generator matrix) of some binary (n,k) error correcting code ($r = n - k$)

Solving CSD(H_0, y, w) for a cryptogram y and some prescribed value of w breaks the system

- In McEliece system the cryptogram is a noisy codeword x; we have $y = H_0 x^T$ and $w = t = r / \lfloor \log_2 n \rfloor$ is the error correcting capability of the (secret) Goppa code
- In Niederreiter system the cryptogram is the syndrome y and $w = t$ as above
- In CFS signature y is the hash of the message and either $w = t$ and we decode one out of $t!$ instances, or $w = t + \delta = d_{\mathrm{GV}}$ (the Gilbert-Varshamov distance)

Best Decoding Algorithms

Fixed binary (n, k) code, solve CSD for growing w

codimension $r = n - k$, Gilbert-Varshamov distance $\binom{n}{d_{GV}} > 2^r$

ISD: Information Set Decoding

GBA: Generalized Birthday Algorithm

In the present study we will consider $w \leq d_{GV}$ and the impact of multiple instances on the complexity of GBA and ISD

Problem Statement

The size of the problem (i.e. r and n) is fixed

Three facts:

- Decoding one out of N is easier when N grows
- One cannot gain more than a factor N
- It is useless to let N grow indefinitely

Two questions:

- How easier is it to solve CSD$(H, \mathcal{S}, w)$ rather than CSD(H, s, w) when $|\mathcal{S}| = N$ grows ?
- What is the largest useful value of N ?

Generalized Birthday Algorithm for Decoding

Generalized Birthday Algorithm for Decoding – Bibliography

- Order 2 GBA

  Camion and Patarin, EUROCRYPT'91

- GBA

  Wagner, CRYPTO 2005

- GBA for decoding

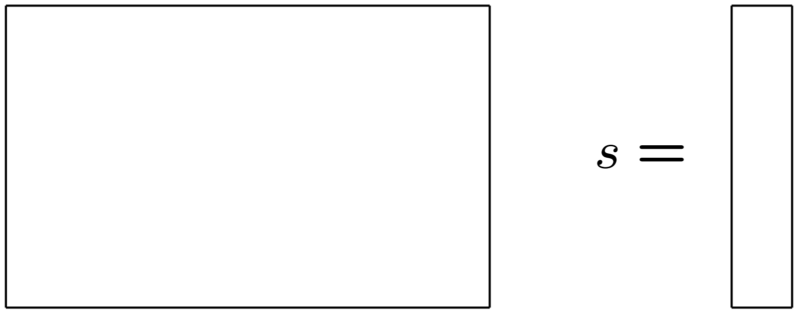
  Coron and Joux, 2004 (IACR eprint), attack against FSB

- GBA for decoding one out of many

  Bleichenbacher, 200? (unpublished), attack against CFS

Generalized Birthday Algorithm for Decoding – Order 2

$H =$ $s =$



CSD(H, s, w)

find w columns of H adding to s

Order 2

Build 4 subsets of $\{0,1\}^r$, $i \in \{1,2,3,4\}$ (ℓ is optimized later)

$$W_i \subset s_i + \{He^T \mid \text{wt}(e) = w_i\}$$

with $s = \sum_i s_i$, $w_i \approx w/4$, $w = \sum_i w_i$ and $|W_i| = 2^\ell$

Next build $W_{1,2}$ and $W_{3,4}$ as

$$W_{i,j} = \{x + y \mid x \in W_i \text{ and } y \in W_j \text{ match on their first } \ell \text{ bits}\}$$

Any element of $W_{1,2} \cap W_{3,4}$ provides a solution to CSD(H, s, w)

Generalized Birthday Algorithm for Decoding – Complexity

$H =$ $s =$

CSD(H, s, w)

find w columns of H adding to s

Order 2

If $\sqrt[4]{\binom{n}{w}} \geq 2^{r/3}$ then one may choose $\ell = r/3$ and $W_{1,2} \cap W_{3,4} \neq \emptyset$ with probability $> 1/2 \rightarrow$ complexity $O(r2^{r/3})$

Else $|W_i| = 2^\ell = \sqrt[4]{\binom{n}{w}}$ and $W_{1,2} \cap W_{3,4} \neq \emptyset$ with probability $\approx 2^{r-3\ell}$
$\rightarrow$ complexity $O\left(r2^{r-2\ell}\right) = O\left(\frac{r2^r}{\sqrt{\binom{n}{w}}}\right)$

When $w = d_{GV}$ then $\binom{n}{w} \approx 2^r$ and the complexity is $O(r2^{r/2})$

$H =$ $s =$

CSD(H, s, w)

find w columns of H adding to s

Order a

The best value for ℓ is

$$\ell = \min\left(\frac{r}{a+1}, \log_2 \sqrt[2^a]{\binom{n}{w}}\right)$$

$\longrightarrow$ complexity $O(r2^{r-a\ell})$

When $\sqrt[2^a]{\binom{n}{w}} \geq 2^{\frac{r}{a+1}}$ the complexity is $O\left(r2^{\frac{r}{a+1}}\right)$ else it is $O\left(\frac{r2^r}{\binom{n}{w}^{\frac{a}{2^a}}}\right)$

Only interesting for very large values of w

GBA for Decoding One Out of Many

Order 2 GBA with Multiple Instances

$H =$ $\qquad$ $s =$

CSD$(H, \mathcal{S}, w)$

find w columns of H
adding to $s \in \mathcal{S}$, $N = |\mathcal{S}|$

Order 2

Build 3 subsets of $\{0,1\}^r$, $i \in \{1,2,3\}$

$$W_i \subset s_i + \{He^T \mid \mathrm{wt}(e) = w_i\}$$

with $s_1 + s_2 + s_3 = 0$, $w_1 + w_2 + w_3 \leq w$ and a fourth set

$$W_4 \subset \mathcal{S} + \{He^T \mid \mathrm{wt}(e) = w_4\}$$

where $w_4 = w - w_1 - w_2 - w_3$ (possibly $w_4 = 0$) and all $|W_i| = 2^\ell \geq N$

Next build $W_{1,2}$ and $W_{3,4}$ as

$$W_{i,j} = \{x + y \mid x \in W_i \text{ and } y \in W_j \text{ match on their first } \ell \text{ bits}\}$$

Any element of $W_{1,2} \cap W_{3,4}$ provides a solution to CSD$(H, \mathcal{S}, w)$

Order 2 GBA with Multiple Instances – Complexity

$H =$ $s =$

CSD$(H, \mathcal{S}, w)$

find w columns of H
adding to $s \in \mathcal{S}$, $N = |\mathcal{S}|$

Order 2

If $\sqrt[4]{N\binom{n}{w}} \geq 2^{r/3}$ then we may choose $\ell = r/3$ and $W_{1,2} \cap W_{3,4} \neq \emptyset$ with probability $> 1/2$ $\rightarrow$ complexity $O(r2^{r/3})$

Else $|W_i| = 2^\ell = \sqrt[4]{\binom{n}{w}}$ and $W_{1,2} \cap W_{3,4} \neq \emptyset$ with probability $\approx 2^{r-3\ell}$

$\rightarrow$ complexity $O\left(r2^{r-2\ell}\right) = O\left(\frac{r2^r}{\sqrt{N\binom{n}{w}}}\right)$

There is a gain of a factor $\sqrt{N}$ as long as $N \leq 2^{4r/3}/\binom{n}{w}$

When $w = d_{GV}$ then $\binom{n}{w} \approx 2^r$ and $N = 2^{r/3}$ $\Rightarrow$ complexity $O(r2^{r/3})$

Bleichenbacher's Attack

For CFS (original counter version) one can build as many syndromes as needed by hashing many variants of a favorable message

We need to decode $w = t$ errors in a code of length $n = 2^m$ and codimension $r = tm$

For those value, $\binom{n}{t} \approx 2^r/t!$ and the largest value for N is $\sqrt[3]{\binom{n}{t}}$ (common size of the 4 lists) the complexity of CSD becomes

$$O\left(r 2^{r/3} (t!)^{2/3}\right)$$

with $t = 9$ and $m = 16$ we get $\approx 2^{67.5}$ with 2^{42} instances which can be improved a bit (around $2^{63.3}$) because we can use slightly larger lists ($\sqrt{\binom{n}{2w/3}}$ instead of $\sqrt[3]{\binom{n}{w}}$)

Finally there is a small multiplicative constant (2 to 6) which seems difficult to avoid

Bleichenbacher's Attack

For CFS counterless version, the attacker needs to perform a complete decoding. As many variants as needed of a favorable message are hashed to produce the syndromes

We need to decode $w = d_{GV} > t$ errors in a code of length $n = 2^m$ and codimension $r = tm$

For those value, $\binom{n}{w} \geq 2^r$ and the good choice for N and the list size is $2^{r/3}$ the complexity of CSD becomes

$$O\left(r2^{r/3}\right)$$

with $w = 11$ and $m = 16$ we get $\approx 2^{53.6}$ with 2^{48} instances

However because w is not a multiple of 3, some ajustement are required and the cost is $2^{54.9}$ with $2^{45.4}$ instances

GBA with Multiple Instances – General Case

$H =$ $s =$

CSD$(H, \mathcal{S}, w)$

find w columns of H
adding to $s \in \mathcal{S}$

Order a

The best value for ℓ is

$$\ell = \min\left(\frac{r}{a+1}, \log_2 \sqrt[2^a]{N\binom{n}{w}}\right)$$

$\longrightarrow$ complexity $O(r2^{r-a\ell})$

When $\sqrt[2^a]{N\binom{n}{w}} \geq 2^{\frac{r}{a+1}}$ the complexity is $O\left(r2^{\frac{r}{a+1}}\right)$

Else the complexity is $O\left(\frac{r2^r}{\left(N\binom{n}{w}\right)^{\frac{a}{2^a}}}\right)$ and we only gain a factor $N^{\frac{a}{2^a}}$

Information Set Decoding

Information Set Decoding – Bibliography

- ISD

  Folklore, $\leq$ 1978

- Collision decoding

  Stern, 1989

  Canteaut and Chabaud, IEEE-IT 1998 (1995)

  Bernstein, Lange, and Peters, PQCrypto 2008

- One out of many

  Johansson and Jönsson, IEEE-IT 2002

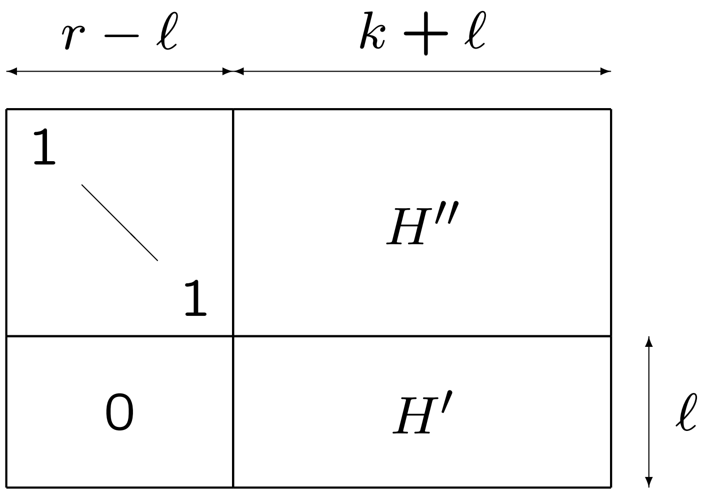
Information Set Decoding – First Step

Problem: Solve CSD(H_0, y, w)

The algorithm involves two parameters p and ℓ which will be chosen to minimize the cost

Step 1: Column permutation and Gaussian elimination

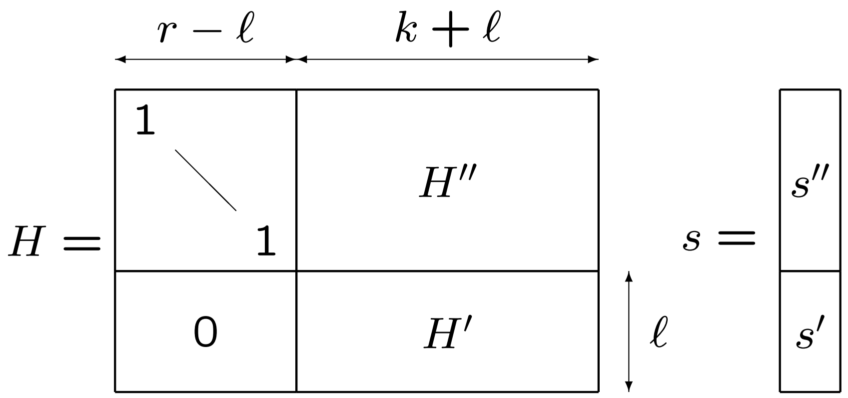
- Pick a random permutation matrix P
- Compute $H = UH_0P =$



with $U \in \{0,1\}^{r \times r}$ non singular and $s = Uy$

$$e \in \text{CSD}(H, s, w) \Leftrightarrow eP^T \in \text{CSD}(H_0, y, w)$$

Information Set Decoding – Second Step



Problem:
Solve CSD(H, s, w)

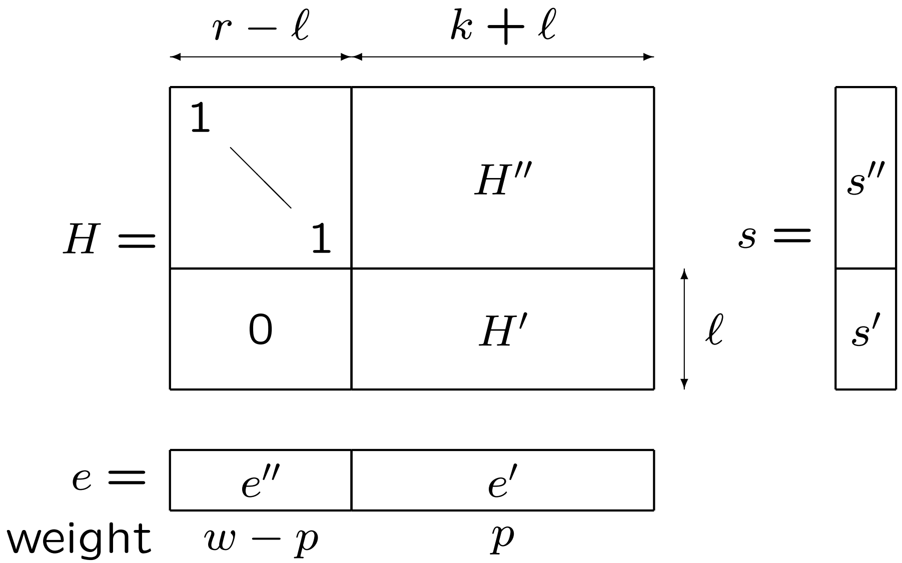
Step 2: Find (all) solutions of CSD(H', s', p)

Build two subsets of $\{0,1\}^\ell$: $\begin{cases} W_1 \subset \{H'e^T \mid \mathrm{wt}(e) = \lfloor p/2 \rfloor\} \\ W_2 \subset \{H'e^T \mid \mathrm{wt}(e) = \lceil p/2 \rceil\} \end{cases}$

Any element of $W_1 \cap (s' + W_2)$ corresponds to a pair $(e_1, e_2) \in W_1 \times W_2$ such that $e_1 + e_2 \in$ CSD(H', s', p)

Birthday attack with a search space of size $\binom{k+\ell}{p}$, we expect that it is optimal for $L = |W_1| = |W_2| = \sqrt{\binom{k+\ell}{p}}$

Information Set Decoding – Third Step



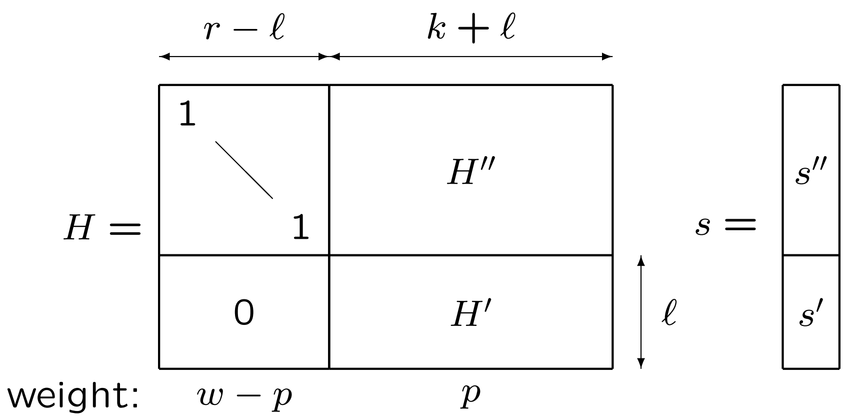
Problem:
Solve $\mathsf{CSD}(H, s, w)$

Step 3: For all $e' \in \mathsf{CSD}(H', s', p)$ found in **Step 2.**

Let $e'' = s'' + H''e'^T \in \{0,1\}^{r-\ell}$ and $e = (e'', e')$

If $\mathrm{wt}(e'') = w - p$ then $e = (e'', e') \in \mathsf{CSD}(H, s, w)$ ($\rightarrow$ SUCCESS)

Information Set Decoding – Algorithm



Subset size in **Step 2.**

$$L = \sqrt{\binom{k+\ell}{p}}$$

(could be less)

Iteration success probability

$$\mathcal{P} = \frac{L^2 \binom{r-\ell}{w-p}}{\binom{n}{w}}$$

Repeat:

1. Permutation + elimination
   Cost polynomial in n

2. Solve CSD(H', s', p)
   Birthday attack
   Total cost is $\geq 2\ell L$ for $\approx L^2/2^\ell$ solutions

3. For each e' found in step 2, test the weight of $H''e'^T + s''$
   One test costs
   $K_{w-p} \geq 2(1 + w - p)$
   $(\approx 2p(1 + w - p))$ in practice)
   Total cost is $\approx K_{w-p} L^2 / 2^\ell$

All costs in binary operations

ISD – Lower Bound on the Binary Work Factor

We neglect the cost of step 1

$$\text{WF}_{\text{ISD}} \geq \min_{p,\ell} \frac{1}{\mathcal{P}_p(\ell)} \left(2\ell L_p(\ell) + \frac{L_p(\ell)^2 K_{w-p}}{2^\ell} \right)$$

nb iter. step 2 step 3

where $\begin{cases} \mathcal{P}_p(\ell) \text{ is the success probability of one iteration} \\ L_p(\ell) \text{ is the optimal subset size in step 2} \end{cases}$

In practice we have $\mathcal{P}_p(\ell) = \frac{\binom{k+\ell}{p}\binom{r-\ell}{w-p}}{\binom{n}{w}}$ and $L_p(\ell) = \sqrt{\binom{k+\ell}{p}}$, but the general formula is

$$\mathcal{P}_p(\ell) = 1 - (1-\varepsilon)^{\binom{k+\ell}{p}} \text{ and } L_p(\ell) = \sqrt{\frac{\mathcal{P}_p(\ell)}{\varepsilon}} \text{ where } \varepsilon = \frac{\binom{r-\ell}{w-p}}{\min\left(\binom{n}{w}, 2^r\right)}.$$

ISD – Lower Bound on the Binary Work Factor

Assuming $L_p(\ell)/\mathcal{P}_p(\ell)$ varies slowly with ℓ, for a given p the optimal value of the parameter ℓ is

$$\ell_p \approx \log_2\left(\frac{\ln(2) K_{w-p} L_p(\ell_p)}{2}\right)$$

Taking into account the variation of $L_p(\ell)/\mathcal{P}_p(\ell)$ leads to a marginaly smaller value of ℓ_p with no easy closed expression

For convenience, we will use below the notations ℓ, L and $\mathcal{P}$ (instead of ℓ_p, $L_p(\ell_p)$ and $\mathcal{P}_p(\ell_p)$) to denote the optimal values

Claim. Provided there are solutions to CSD(H_0, y, w), the cost for finding one with ISD is not smaller than

$$\mathrm{WF}_{\mathrm{ISD}} \geq \min_p \frac{2\ell L}{\mathcal{P}}$$

ISD

One Out of Many

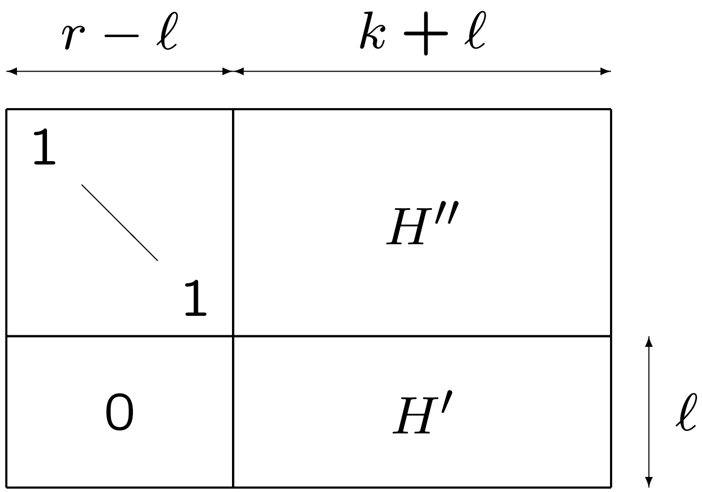
Information Set Decoding One Out of Many – First Step

Problem: Solve CSD$(H_0, \mathcal{Y}, w)$

The algorithm involves two parameters p and ℓ which will be chosen to minimize the cost

Step 1: Column permutation and Gaussian elimination

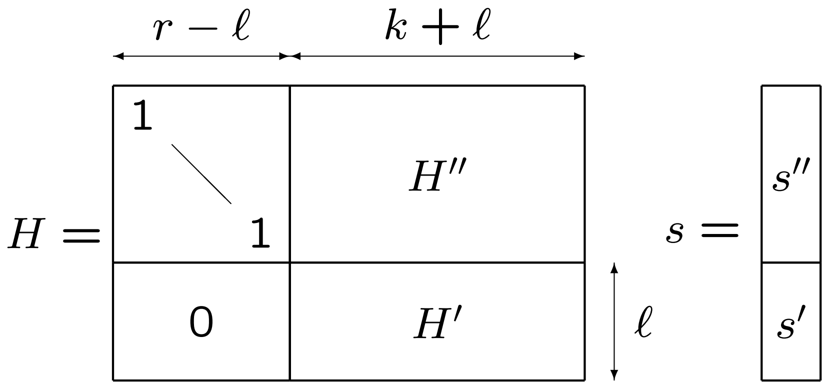
- Pick a random permutation matrix P
- Compute $H = UH_0P =$



with $U \in \{0,1\}^{r \times r}$ non singular and $\mathcal{S} = \{Uy \mid y \in \mathcal{Y}\}$

$$e \in \text{CSD}(H, \mathcal{S}, w) \Leftrightarrow eP^T \in \text{CSD}(H_0, \mathcal{Y}, w)$$

Information Set Decoding One Out of Many – Second Step



Problem:
Solve CSD$(H, \mathcal{S}, w)$

$\mathcal{S}'$ the set of all s'

Step 2: Find (all) solutions of CSD$(H', \mathcal{S}', p)$

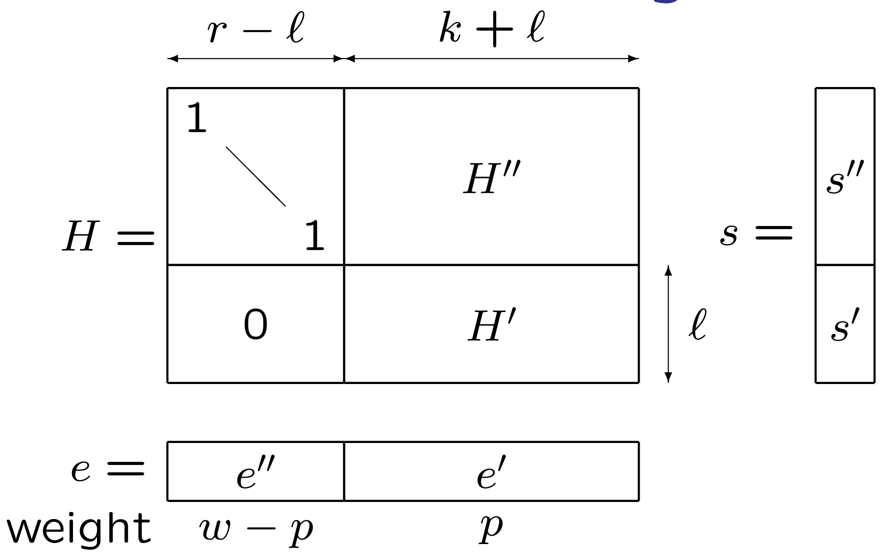
Build two subsets of $\{0,1\}^\ell$: $\begin{cases} W_1 \subset \{H'e^T \mid \text{wt}(e) = a\} \\ W_2 \subset \{H'e^T \mid \text{wt}(e) = b\} \end{cases}$ $(a + b = p)$

Any element of $W_1 \cap (\mathcal{S}' + W_2)$ corresponds to a pair $(e_1, e_2) \in W_1 \times W_2$ such that $e_1 + e_2 \in$ CSD$(H', \mathcal{S}', p)$

In fact the solutions are triples $(e_1, e_2, s = (s'', s')) \in W_1 \times W_2 \times \mathcal{S}$

Birthday attack with a search space of size $N\binom{k+\ell}{p}$, we expect that it is optimal for $L = |W_1| = N|W_2| = \sqrt{N\binom{k+\ell}{p}}$ $(\Rightarrow N \leq L \leq \binom{k+\ell}{p})$

Information Set Decoding One Out of Many – Third Step



Problem:
Solve $\mathrm{CSD}(H, \mathcal{S}, w)$

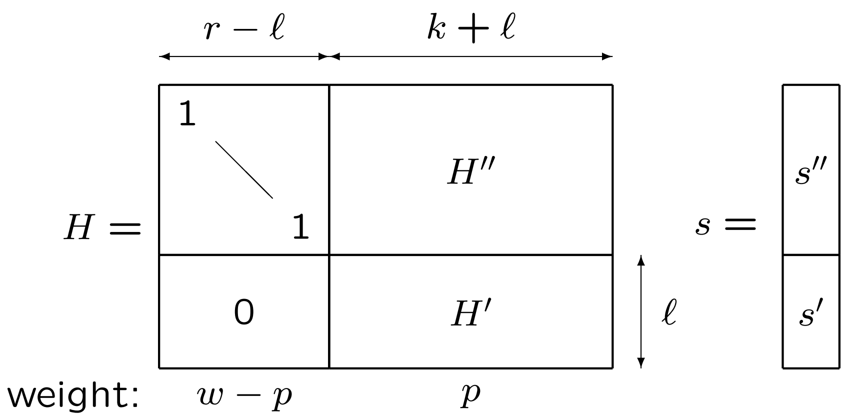
Step 3: For all e' found in **Step 2.**

(e' is associated to some $s = (s'', s') \in \mathcal{S}$)

Let $e'' = s'' + H''e'^T \in \{0,1\}^{r-\ell}$ and $e = (e'', e')$

If $\mathrm{wt}(e'') = w - p$ then $e = (e'', e') \in \mathrm{CSD}(H, s, w) \subset \mathrm{CSD}(H, \mathcal{S}, w)$
($\longrightarrow$ SUCCESS)

Information Set Decoding One Out of Many – Algorithm



Subset size in **Step 2.**

$$L = \sqrt{N\binom{k+\ell}{p}}$$

(could be less)

Iteration success probability

$$\mathcal{P} = \frac{L^2\binom{r-\ell}{w-p}}{\binom{n}{w}} \quad \text{if } \mathcal{P} \ll 1$$

Repeat:

1. Permutation + elimination
   Cost polynomial in n + ?

2. Solve CSD$(H', \mathcal{S}', p)$
   Birthday attack
   Total cost is $\geq 2\ell L$ for $\approx L^2/2^\ell$ solutions

3. For each e' found in step 2, test the weight of $H''e'^T + s''$
   One test costs
   $K_{w-p} \geq 2(1+w-p)$
   ($\approx 2p(1+w-p)$) in practice)
   Total cost is $\approx K_{w-p}L^2/2^\ell$

All costs in binary operations

ISDOOM – Lower Bound on the Binary Work Factor

We neglect the cost of step 1

$$\mathrm{WF}_{\mathrm{ISD}}^{(N)} \geq \min_{p,\ell} \frac{1}{\mathcal{P}_p^{(N)}(\ell)} \left(2\ell L_p^{(N)}(\ell) + \frac{L_p^{(N)}(\ell)^2 K_{w-p}}{2^\ell} \right)$$

nb iter. step 2 step 3

where $\begin{cases} \mathcal{P}_p^{(N)}(\ell) \text{ is the success probability of one iteration} \\ L_p^{(N)}(\ell) \text{ is the optimal subset size in step 2} \end{cases}$

In practice we have $\mathcal{P}_p^{(N)}(\ell) = \frac{N\binom{k+\ell}{p}\binom{r-\ell}{w-p}}{\binom{n}{w}}$ and $L_p^{(N)}(\ell) = \sqrt{N\binom{k+\ell}{p}}$,

but the general formula is

$$\mathcal{P}_p^{(N)}(\ell) = 1-(1-\varepsilon)^{N\binom{k+\ell}{p}} \text{ and } L_p^{(N)}(\ell) = \sqrt{\frac{\mathcal{P}_p^{(N)}(\ell)}{\varepsilon}} \text{ where } \varepsilon = \frac{\binom{r-\ell}{w-p}}{\min\left(\binom{n}{w}, 2^r\right)}.$$

ISDOOM – Lower Bound on the Binary Work Factor

For a given p the optimal value of the parameter ℓ is

$$\ell_p^{(N)} \approx \log_2 \left(\frac{\ln(2) K_{w-p} L_p^{(N)}(\ell_p^{(N)})}{2} \right)$$

For convenience, we will use below the notations ℓ', L' and $\mathcal{P}'$ instead of $\ell_p^{(N)}$, $L_p^{(N)}(\ell_p^{(N)})$ and $\mathcal{P}_p^{(N)}(\ell_p^{(N)})$ to denote the optimal values

Claim. Provided there are solutions to CSD(H_0, y, w) **for all** $y \in \mathcal{Y}$, the cost for finding one solution of CSD$(H_0, \mathcal{Y}, w)$ with ISD is not smaller than

$$\mathrm{WF}_{\mathrm{ISD}}^{(N)} \geq \min_p \frac{2\ell' L'}{\mathcal{P}'}$$

For fixed p and ℓ we have $L' \approx \sqrt{N} L$ and $\mathcal{P}' \approx \sqrt{N} \mathcal{P}$ so we expect a gain of a factor $\approx \sqrt{N}$

ISDOOM – Complexity gain

More precisely, as long as N is not too large

$$
\begin{aligned}
\ell' &\approx \ell + x & &\approx \ell + \log_2 \sqrt{N} \\
L' &\approx \sqrt{N}\sqrt{\binom{k+\ell+x}{p}} & &\approx \sqrt{N} L \exp\left(\tfrac{c_1}{2} x\right) \\
\mathcal{P}' &\approx N \frac{\binom{k+\ell+x}{p}\binom{r-\ell-x}{w-p}}{\binom{n}{w}} & &\approx N\mathcal{P} \exp\left(c_1 x - c_2 x\right)
\end{aligned}
$$

where $c_1 \approx \dfrac{p}{k+\ell-\frac{p-1}{2}}$ and $c_2 \approx \dfrac{w-p}{r-\ell-\frac{w-p-1}{2}}$ (both $\ll 1$)

$$
\frac{2\ell' L'}{\mathcal{P}'} \approx \frac{2\ell L}{\mathcal{P}} \left(1 + \frac{\log_2 \sqrt{N}}{\ell}\right) \frac{1}{\sqrt{N^{1-c}}}
$$

where $c \approx (c_2 - c_1/2)/\ln 2$ is a small (usually positive) constant

About tightness

I've been cheating you !

It is not possible to claim a computational gain from lower bounds !!!

We need tight bounds to do that and so **we must make sure it was legitimate to neglect the cost of the first step**

Computing the set $\mathcal{S} = \{Uy \mid y \in \mathcal{Y}\}$ will cost something like

$$\frac{2r(K_{w-p} + \ell)N}{\log_2 N}$$

possibly less because there are ways to reduce the impact of **Step 1.** [Bernstein, Lange, Peters, PQCrypto 2008]

This has to be compared with $2\ell L$, the cost of an iteration

Consequence: if $\dfrac{r(K_{w-p} + \ell)N}{\log_2 N} \geq \ell L$ the gain is smaller than expected

Some Numbers

McEliece or Niederreiter					
$n = 2^{11}$, $w = 32$, $r = 352$					
	single		multiple		
p	ℓ	WF	ℓ	N	WF′
4	22	85.9	40	2^{38}	74.2
6	30	**85.9**	55	2^{52}	66.2
8	37	86.3	61	2^{49}	**66.1**
10	45	87.0	65	2^{41}	69.9

CFS - counterless version					
$n = 2^{16}$, $w = 11$, $r = 144$					
	single		multiple		
p	ℓ	WF	ℓ	N	WF′
4	31	85.2	56	2^{57}	**63.0**
6	44	81.1	60	2^{38}	66.6
8	56	77.8	64	2^{20}	70.9
10	68	**76.2**	69	2^{5}	76.0

Conclusion – Further work

DOOM is a threat to code-based crypto

Its impact can be cancelled

- Against the signature scheme

  Repared by Finiasz (SAC 2010) $\rightarrow$ decode several (3 or 4) related syndromes

- Against McEliece (or Niederreiter)

  If you are going to encrypt many messages you may chain them

- Security of FSB: what about $w > d_{GV}$ or regular words?

- Are there other ways to use multiple instances?

Thank you