Ternary Syndrome Decoding with Large Weight

Rémi Bricout, André Chailloux, Thomas Debris-Alazard, Matthieu Lequesne

Ternary Syndrome Decoding with Large Weight

Rémi Bricout [1,2] André Chailloux [2] Thomas Debris-Alazard[1,2] Matthieu Lequesne [1,2]

[1] Sorbonne Université, UPMC Univ Paris 06 [2] Inria, Paris

The Syndrome Decoding Problem

Syndrome Decoding - SD(q, R, W)
Instance: $\mathbf{H} \in \mathbb{F}_q^{(n-k)\times n}$ of full rank,
$\mathbf{s} \in \mathbb{F}_q^{n-k}$ (usually called the *syndrome*).
Output: $\mathbf{e} \in \mathbb{F}_q^n$ such that $|\mathbf{e}| = w$ and $\mathbf{e}\mathbf{H}^\intercal = \mathbf{s}$,
where $k \triangleq \lceil Rn \rceil$, $w \triangleq \lceil Wn \rceil$ and $|\mathbf{e}| \triangleq |\{i : \mathbf{e}_i \neq 0\}|$.

Binary vs. Ternary case

Depending on R and W, the complexity of the SD problem can greatly vary. Let us fix a value R, and let W_{GV} denote the Gilbert-Varshamov bound. For $W \in \left[0, \frac{1}{2}\right]$, there exist three regimes.

1. $W \approx W_{\mathrm{GV}}$. There is a small number of solutions. This is the regime where the problem is the hardest and where it is the most studied.
2. $W \gg W_{\mathrm{GV}}$. There are exponentially many solutions and this makes the problem simpler. When W reaches $\frac{1-R}{2}$, the problem can be solved in average polynomial time.
3. $W \ll W_{\mathrm{GV}}$. In this regime, we have with high probability a unique solution. However, the search space, *i.e.* the set of vectors $\mathbf{e}$ st. $|\mathbf{e}| = \lceil Wn \rceil$ is much smaller.

Decoding in Large Weight

Symetry. SD$(2, R, W)$ and SD$(2, R, 1-W)$ are equivalent.

However, **this argument is no longer valid for $\mathbf{q \geq 3}$.** The problem has a quite different behavior in small and large weights. **The goal of this work is to study the complexity of the SD problem for $\mathbf{q = 3}$ and $\mathbf{W > 0.5}$.**

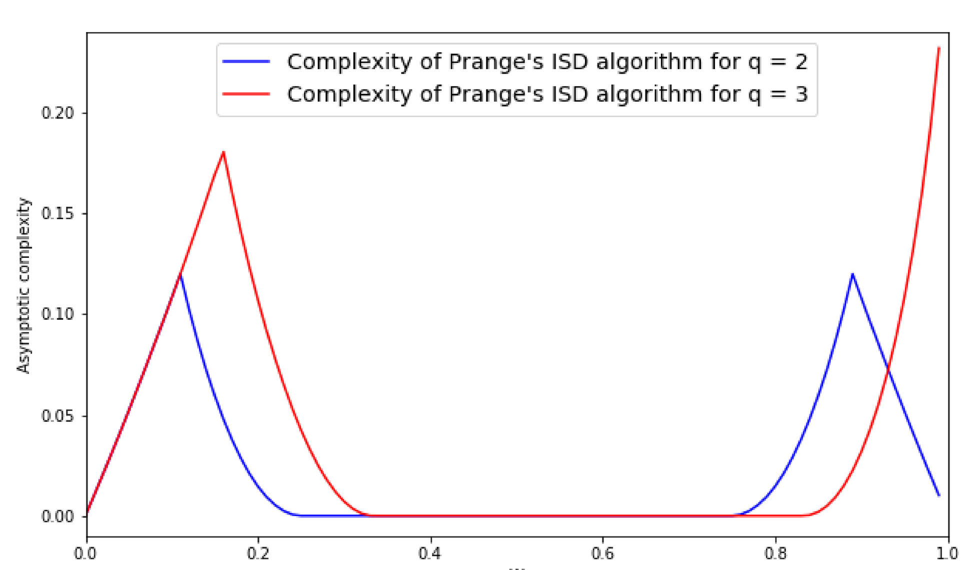

Figure 1: Asymptotic complexity of Prange's ISD algorithm for $R = 0.5$.

From SD to Subset Sum

Most algorithms designed to solve the SD problem follow the same framework, which can be generalized for the non-binary case. Let $(\mathbf{H}, \mathbf{s})$ be the instance that we want to solve. Let us introduce ℓ and p, two parameters of the system.

1. Apply a **random permutation** π on the columns of $\mathbf{H}$.
2. Perform a **partial Gaussian elimination** on the rows of $\mathbf{H}_\pi$ using the first $n-k-\ell$ columns. Let $\mathbf{S} \in \mathbb{F}_q^{(n-k)\times(n-k)}$ be the matrix corresponding to this operation and introduce two matrices $\mathbf{H}' \in \mathbb{F}_q^{(n-k-\ell)\times(k+\ell)}$ and $\mathbf{H}'' \in \mathbb{F}_q^{\ell\times(k+\ell)}$ such that
$$\mathbf{S}\mathbf{H}_\pi = \begin{pmatrix} \mathbf{1}_{n-k-\ell} & \mathbf{H}' \\ \mathbf{0} & \mathbf{H}'' \end{pmatrix}.$$
The problem can be rewritten as follows.
$$\begin{aligned} \mathbf{H}_\pi \mathbf{e}^\intercal = \mathbf{s}^\intercal &\iff \mathbf{S}\mathbf{H}_\pi\mathbf{e}^\intercal = \mathbf{S}\mathbf{s}^\intercal \\ &\iff \begin{pmatrix} \mathbf{1}_{n-k-\ell} & \mathbf{H}' \\ \mathbf{0} & \mathbf{H}'' \end{pmatrix} \begin{pmatrix} \mathbf{e}'^\intercal \\ \mathbf{e}''^\intercal \end{pmatrix} = \begin{pmatrix} \mathbf{s}'^\intercal \\ \mathbf{s}''^\intercal \end{pmatrix} \\ &\iff \begin{cases} \mathbf{e}'^\intercal + \mathbf{H}'\mathbf{e}''^\intercal = \mathbf{s}'^\intercal \\ \mathbf{H}''\mathbf{e}''^\intercal = \mathbf{s}''^\intercal \end{cases} \end{aligned}$$
To solve the problem, we will try to find a solution $(\mathbf{e}', \mathbf{e}'')$ to the above system such that $|\mathbf{e}''| = p$ and $|\mathbf{e}'| = w - p$.
3. Compute a set $\mathcal{S}$ of solutions of $\mathbf{H}''\mathbf{e}''^\intercal = \mathbf{s}''^\intercal$ such that $|\mathbf{e}''| = p$. This is an instance of the **Subset Sum problem**.
4. Take a vector $\mathbf{e}'' \in \mathcal{S}$ and let $\mathbf{e}'^\intercal = \mathbf{s}'^\intercal - \mathbf{H}'\mathbf{e}''^\intercal$. If $|\mathbf{e}'| = w - p$, $\mathbf{e} = (\mathbf{e}', \mathbf{e}'')$ is a solution for inputs $\mathbf{H}_\pi$ and $\mathbf{s}$, which can be turned into a solution of the initial problem.

Reduction Lemma

If we have an algorithm that solves SS$(3, k+\ell, \ell, L, \emptyset)$ then we have an algorithm that solves SSNZC$(3, k+\ell, \ell, L, k+\ell)$ with the same complexity.

The Subset Sum Problem

Subset Sum problem - SS(q, n, m, L, p)
Instance: n vectors $\mathbf{x}_i \in \mathbb{F}_q^m$ for $1 \leq i \leq n$, a target $\mathbf{s} \in \mathbb{F}_q^m$.
Output: L solutions $(b_1^{(j)}, \ldots, b_n^{(j)}) \in \{0,1\}^n$ for $1 \leq j \leq L$, such that for all j, $\sum_{i=1}^n b_i^{(j)}\mathbf{x}_i = \mathbf{s}$ and $|\mathbf{b}^{(j)}| = p$.

We denote SSNZC(q, n, m, L, p) the problem when we look for solutions with $b_i^{(j)} \in \mathbb{F}_q$. We denote SS$(q, n, m, L, \emptyset)$ the problem without any constraint on the weight.

Wagner's algorithm

Wagner's algorithm [1] is an algorithm to solve SS$(2, n, \ell, L, \emptyset)$. For some parameters, it finds L solution in time $O(L)$. It can easily be adapted to solve the SS problem in the ternary case. The algorithm works as follows:

- divide the vectors $\mathbf{x}_1, \ldots, \mathbf{x}_n$ in 2^a stacks of size $n/2^a$;
- for each stack, compute a list $\mathcal{L}_p$ of L random linear combinations of the vectors in the stack ;
- merge the lists two by two: from $\mathcal{L}_p$ and $\mathcal{L}_{p+1}$ create the list $\{\mathbf{y}_p + \mathbf{y}_{p+1} : \mathbf{y}_i \in \mathcal{L}_i$ and the last ℓ/a bits of $\mathbf{y}_p + \mathbf{y}_{p+1}$ are 0s.$\}$;
- repeat $a-1$ times.

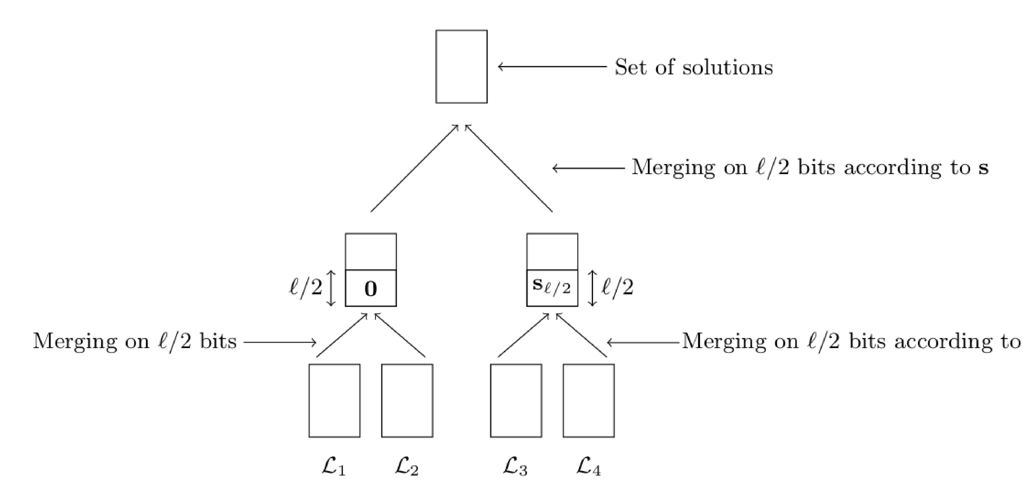

Figure 2: Wagner's algorithm with $a = 2$.

Theorem. Fix $k, \ell \in \mathbb{N}^*$ and $a \in \mathbb{N}$ such that $3^{\ell/a} \leq 2^{(k+\ell)/2^a}$. The associated SS$(3, k+\ell, \ell, 3^{\ell/a}, \emptyset)$ problem can be solved in average time and space $O(3^{\ell/a})$.

Smoothing Wagner's Algorithm

Wagner's algorithm shows how to find L solutions in time $O(L)$ for $L = 3^{\ell/a}$. The smaller L is, the better the algorithm performs. So the idea is to take the largest integer a such that $3^{\ell/a} < 2^{(k+\ell)/2^a}$. But this induces a discontinuity in the complexity. We propose a refinement of the theorem that **reduces the discontinuity**.

Using Representations

When looking for vectors $\mathbf{b}$, Wagner's algorithm looks for $\mathbf{b}$ in the form $\mathbf{b} = \mathbf{b}_1 + \mathbf{b}_2$, where the second half of $\mathbf{b}_1$ and the first half of $\mathbf{b}_2$ are only zeros, as on Figure 3 (1). As in the BJMM algorithm [2], the idea of representations is to remove this constraint and replace it by a less restrictive one. Here, we fix the number of 0s, 1s and 2s in $\mathbf{b}_1$ and $\mathbf{b}_2$. This allows **more possibilities to write b as the sum of two vectors** (2).

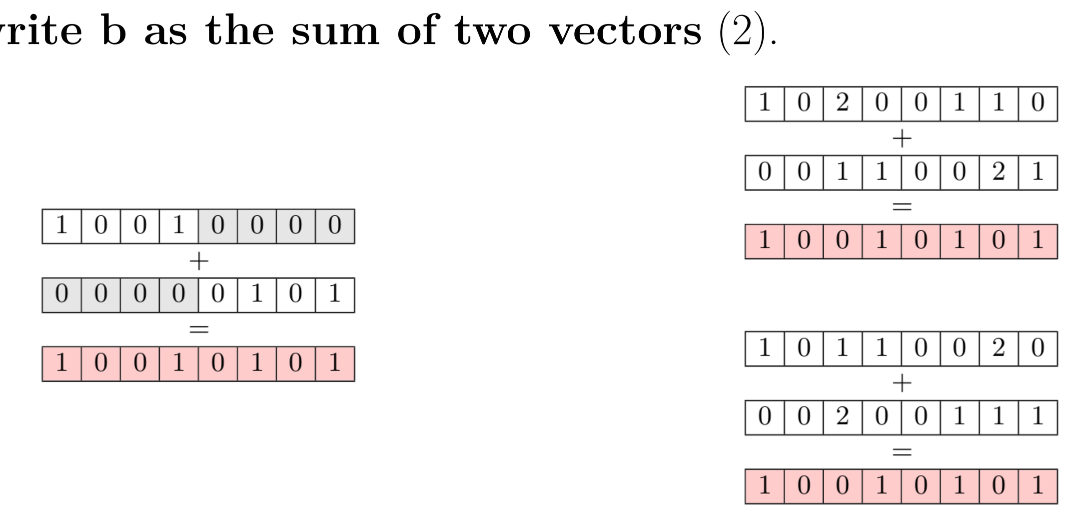

Figure 3: Same vector (1) using left-right split and (2) using representations.

Our best algorithm uses a tree on seven levels. It uses Wagner's left-right splits at the top and at the bottom of the tree and representations on the intermediate levels.

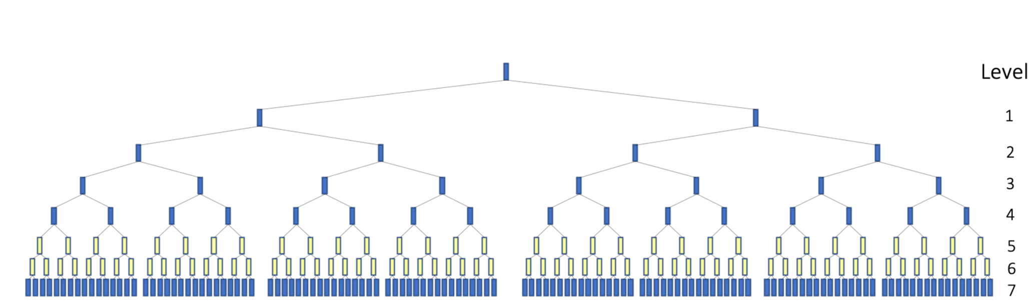

Figure 4: Wagner tree for $a = 7$.
Yellow lists correspond to representations and blue lists to left-right splits.

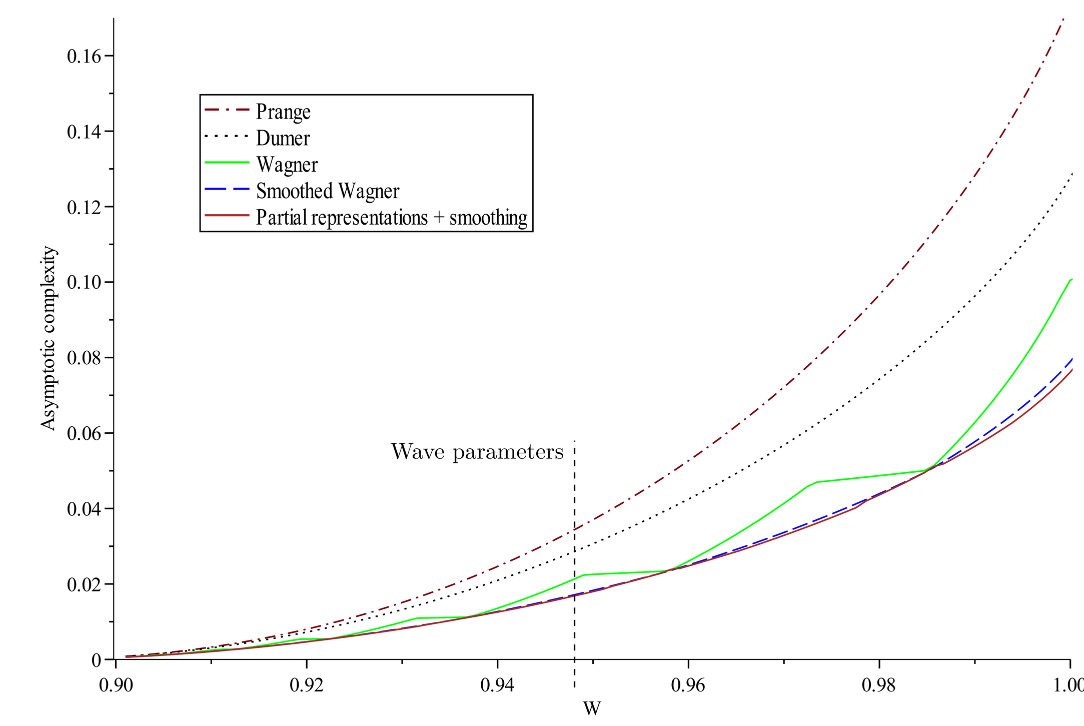

Figure 5: Comparison of the exponent complexities for $R = 0.676$

New Parameters for Wave

Wave [3] is a new code-based signature scheme. It uses a *hash-and-sign* approach and follows the GPV paradigm. Wave is the firts cryptographic scheme relying on the ternary SD problem with large weight. Forging a signature in the Wave scheme amounts to solving the ternary SD problem (more exactly the *Decoding One Out of Many* version of the problem). We provide new parameters for Wave so that the scheme resists attacks using our algorithms.

Hardest Instances

We now focus on the parameters where the problem is the hardest. We compared the performance of 3 standard algorithms: Prange's algorithm, Dumer's algorithm and the BJMM algorithm on binary SD problem with our equivalents for the ternary SD problem. We performed a case study and showed that the hardest case is reached for $R \approx 0.36907$ and $W = 1$.

Algorithm	$q = 2$	$q = 3$ and $W > 0.5$
Prange	0.121 ($R = 0.454$)	0.369 ($R = 0.369$)
Dumer/Wagner	0.116 ($R = 0.447$)	0.269 ($R = 0.369$)
BJMM/our algorithm	0.102 ($R = 0.427$)	0.247 ($R = 0.369$)

Table 1: Best exponents with associated rates.

The ternary SD problem appears significantly harder than the binary one. But the input matrices have elements in $\mathbb{F}_3$ and not $\mathbb{F}_2$, so matrices of the same dimension contain more information.

To get rid of this bias, we define the following metric. **What is the smallest input size for which the algorithms need at least $\mathbf{2^{128}}$ operations to decode?** The input matrix $\mathbf{H} \in \mathbb{F}_3^{n(1-R)\times n}$ is represented in systematic form.

Algorithm	$q = 2$	$q = 3$ and $W > 0.5$
Prange	275 ($R = 0.384$)	44 ($R = 0.369$)
Dumer/Wagner	295 ($R = 0.369$)	83 ($R = 0.369$)
BJMM/our algorithm	374 ($R = 0.326$)	99 ($R = 0.369$)

Table 2: Minimum input sizes (in Kbits) for a time complexity of 2^{128}.

Conclusion

- Strong difference between the cases $q = 2$ and $q \geq 3$.
- Two algorithms to solve the Syndrome Decoding problem in this new regime : a q-ary version of Wagner's approach and a second algorithm making use of representations.
- Application: new parameters for the Wave signature scheme.
- Study of ths hardest case: complexity of SD in large weight is higher than in small weight.

This work opens many new perspectives. It seems there are many cases in code-based cryptography, from encryption schemes to signatures, where **this problem could replace the binary SD problem to get smaller key sizes**.

[1] David Wagner.
A generalized birthday problem.
In Moti Yung, editor, *Advances in Cryptology - CRYPTO 2002*, volume 2442 of *LNCS*, pages 288–303. Springer, 2002.

[2] Anja Becker, Antoine Joux, Alexander May, and Alexander Meurer.
Decoding random binary linear codes in $2^{n/20}$: How $1 + 1 = 0$ improves information set decoding.
In *Advances in Cryptology - EUROCRYPT 2012*, LNCS. Springer, 2012.

[3] Thomas Debris-Alazard, Nicolas Sendrier, and Jean-Pierre Tillich.
Wave: A new code-based signature scheme.
Cryptology ePrint Archive, Report 2018/996, October 2018.
https://eprint.iacr.org/2018/996.