Integer Syndrome Decoding Problem in the Presence of Noise

Pierre-Louis Cayrel   Brice Colombier   Vlad-Florin Drăgoi   Vincent Grosso

A variant of the syndrome decoding problem

The security argument of many post-quantum code-based cryptographic schemes relies on a well-known NP-complete problem, *i.e.* the Syndrome Decoding Problem (SDP). Recent results show that, when the syndrome is computed, for instance during the encryption step in the Niederreiter scheme, an adversary may obtain extra information by means of side-channel analysis. Therefore, depending on the nature of this information, new theoretical problems could be defined and solved.

Here, we target one such problem, where the side-channel information retrieved is the integer value of each syndrome entry. This problem is referred to as the integer syndrome decoding problem ($\mathbb{N}-\mathsf{SDP}$). An adversary is given a parity check matrix of a linear code $\boldsymbol{H} \in \{0,1\}^{(n-k)\times n}$, a weight parameter $t \in \mathbb{N}^*$, and an integer-valued syndrome vector $\boldsymbol{s} \in \mathbb{Z}^{n-k}$. The challenge is to retrieve a binary solution $\boldsymbol{x} \in \{0,1\}^n$ to the linear system $\boldsymbol{Hx} = \boldsymbol{s}$, with a given weight $\mathrm{HW}(\boldsymbol{x}) = t$. $\mathbb{N}-\mathsf{SDP}$ was first defined in [DCC$^+$20] and analyzed further in [CCD$^+$21, CDCG22]. Remark that compared to SDP, where the computations are performed over the binary finite field, in the $\mathbb{N}-\mathsf{SDP}$ the matrix-vector multiplication is performed over the ring of integers. As a consequence, it might be easier to find solutions for this new problem. This is exactly the purpose of this document, *i.e.* to propose an algorithm that efficiently retrieves a solution to the $\mathbb{N}-\mathsf{SDP}$.

However, due to experimental factors involved in the side-channel analysis process, the exact integer value of the syndrome entry might be difficult to obtain. Therefore, the retrieved integer values can be seen and modeled as noisy versions of the exact integer syndrome. As a consequence, we will analyze a more realistic, and also more general, problem, called integer syndrome decoding in presence of noise. Instead of having access to an instance of the $\mathbb{N}-\mathsf{SDP}$, *i.e.* $(\boldsymbol{H}, \boldsymbol{s}, t)$ we are given a noisy syndrome $\widetilde{\boldsymbol{s}} = \boldsymbol{s} + \epsilon$, where $\epsilon_i \sim \mathcal{D}$ a discrete probability distribution, and the value $\boldsymbol{s}^* = \boldsymbol{s} \pmod 2$ (component-wise). Equivalently, we are dealing with the SDP $(\boldsymbol{H}, \boldsymbol{s}^*, t)$ with additional information $(\widetilde{\boldsymbol{s}})$.

Under a different name, and in a completely different context, $\mathbb{N}-\mathsf{SDP}$ was studied in [FL20] and is known as *Quantitative group testing*. Here, we combine techniques from [FL20] and Information-Set Decoding in order to adapt them to this new problem, *i.e.* $\mathbb{N}-\mathsf{SDP}$ the in presence of noise. In [CDCG22] the main idea was to assigning a score to each column of the parity-check matrix. The objective is to distinguish columns of $\boldsymbol{H}$ in the support of the solution vector from columns which are outside the support. The following score decoder, based on the dot product and proposed in [FL20], proved to be particularly discriminant in the context of $\mathbb{N}-\mathsf{SDP}$:

$$\forall i \in \mathbb{Z}_n^* \quad \psi_i(\widetilde{\boldsymbol{s}}) = \sum_{\ell=1}^{n-k} \left(h_{\ell,i}\widetilde{\boldsymbol{s}}_\ell + (1-h_{\ell,i})(t-\widetilde{\boldsymbol{s}}_\ell)\right). \tag{1}$$

A column permutation is derived from this score, by sorting the indices in decreasing order with respect to $\psi_i(\widetilde{\boldsymbol{s}})$. This can be seen as a "good" permutation in an Information-Set Decoding-type algorithm on $\boldsymbol{H}$ with a binary syndrome $\boldsymbol{s}^*$. Hence, once this permutation is found, linear algebra in the Prange style could be performed in order to retrieve the solution. We refer to this algorithm as `Rank-Threshold Score Decoder`. Other optimizations techniques could be imagined, *e.g.* Lee-Brickell or Stern variants.

Experimental results

We simulated the score decoder in a noiseless setting, *i.e* with a perfect integer syndrome, to establish the general behavior of the algorithm. `Rank-Threshold Score Decoder` from [CDCG22] presents the following features, when the length of the code is fixed: i) for fixed t the probability of success increases with $n-k$, which is somehow expected as more information is added when $n-k$ increases; ii) for fixed $n-k$, the probability of success decreases when t increases.

In Figure 1, we plot the number of ones found in the first $n-k$ positions for all the parameter sets of the *Classic McEliece* proposal. The colored horizontal stripes show the $[t-3, t]$ interval. We estimate this interval to be within reach of enumeration when following the Information-Set Decoding approach.

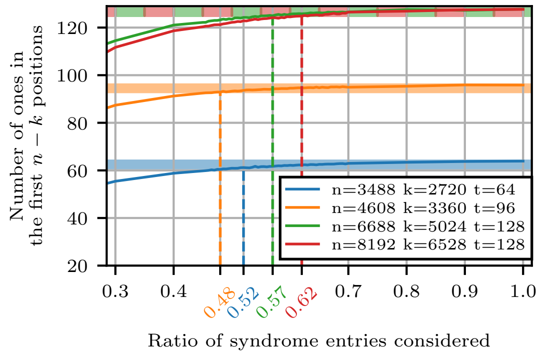

Figure 1: Number of ones in the first $n-k$ positions for the four *Classic McEliece* sets of parameters and different ratio of syndrome entries considered

One of the important features observed in these simulations is that, for almost all selected parameters, a ratio of less than 62 % of the syndrome entries, and thus of the matrix rows, are sufficient to solve the problem. This aspect is of great interest in the context of side-channel analysis, where the integer syndrome might not be perfectly recovered due to experimental variations.

In the second set of simulations we illustrate how the score decoder performs in the presence of noise, *i.e.* when the integer syndrome is not perfect. For this purpose, we generated random vectors ϵ following a binomial distribution centered around zero of parameters $\mathcal{B}(m, 1/2)$ where $0 \leq m \leq t$. For each matrix $\boldsymbol{H}$, we computed the number of 1s in the first $n-k$ positions.

The results are shown in Figure 2. The numerical results are supporting the claim that if the noise is centered around zero and symmetric, the score decoder can tolerate such levels of noise. For all *Classic McEliece* parameters, a binomial noise as high as $\mathcal{B}(t/4, 1/2)$ is manageable by the score decoder when considering all the integer syndrome entries.

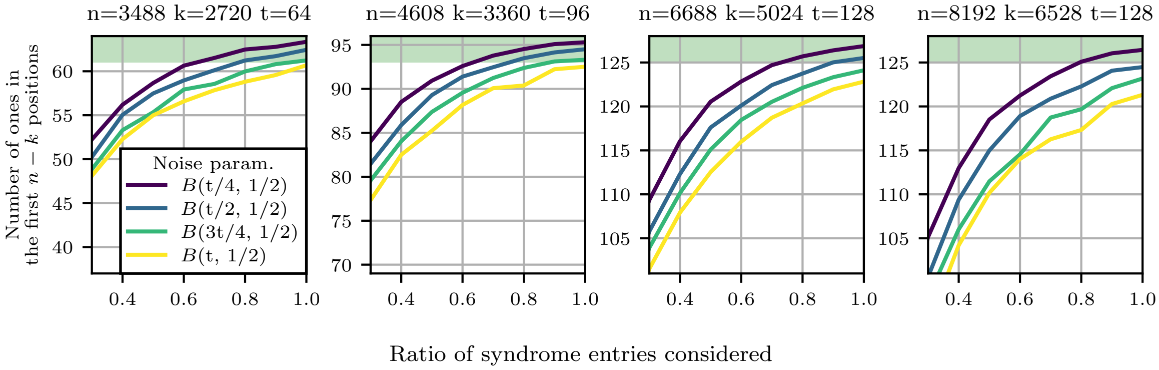

Figure 2: Number of ones in the first $n-k$ positions for the four *Classic McEliece* sets of parameters and a symmetric binomial noise centered around zero

References

[CCD+21] Pierre-Louis Cayrel, Brice Colombier, Vlad-Florin Dragoi, Alexandre Menu, and Lilian Bossuet. Message-Recovery Laser Fault Injection Attack on the Classic McEliece Cryptosystem. In Anne Canteaut and François-Xavier Standaert, editors, *Advances in Cryptology - EUROCRYPT 2021 - 40th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Zagreb, Croatia, October 17-21, 2021, Proceedings, Part II*, volume 12697 of *Lecture Notes in Computer Science*, pages 438–467. Springer, 2021.

[CDCG22] Brice Colombier, Vlad-Florin Dragoi, Pierre-Louis Cayrel, and Vincent Grosso. Message-recovery profiled side-channel attack on the classic mceliece cryptosystem. Cryptology ePrint Archive, Report 2022/125, 2022.

[DCC+20] Vlad-Florin Dragoi, Pierre-Louis Cayrel, Brice Colombier, Dominic Bucerzan, and Sorin Hoara. Solving a modified syndrome decoding problem using integer programming. *INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS and CONTROL*, 15(5), 2020.

[FL20] Uriel Feige and Amir Lellouche. Quantitative group testing and the rank of random matrices. *CoRR*, abs/2006.09074, 2020.