Syndrome Decoding in the Non-Standard Cases

Matthieu Finiasz

1 Introduction

In the late 70's the McEliece cryptosystem was invented and the syndrome decoding problem was proven to be NP-complete. The proof of NP-completeness shows that among some instances (those which can be derived from a 3D mapping problem), some are difficult to solve. The fact that no attack has yet been found on the McEliece cryptosystem tends to show that for standard parameters (like those used in instances of the cryptosystem), finding an attack is not easy. However, other code based cryptographic constructions rely on the hardness of syndrome decoding, some of them using non-standard parameters. Through the review of a few of these alternate constructions I will introduce some techniques used to solve the problem of syndrome decoding.

2 The Problem of Syndrome Decoding

The problem of syndrome decoding is probably the most famous hard problem of coding theory. It is basically a re-writing of the general problem of decoding in terms of syndrome: given a linear code (defined by its parity check matrix) and a random vector, compute the syndrome of this vector and try to find an error pattern of low weight having the same syndrome. More formally it can be written as:

Syndrome Decoding: (SD)

Input: a binary n-$k \times n$ matrix $\mathcal{H}$, an n-k bit vector $\mathcal{S}$ and a weight w.
Output: an n bit vector e of Hamming weight $\leq w$ such that $\mathcal{H} \cdot e = \mathcal{S}$.

This problem was proven to be NP-complete in 1978 [2] but of course, depending on the input, some instances can be solved in polynomial time. For instance, when w is larger than $\frac{n}{2}$, solving SD becomes easy, as computing a pseudo-inverse $\mathcal{H}^{-1}$ of $\mathcal{H}$ and computing $\mathcal{H}^{-1} \cdot \mathcal{S}$ will return a valid solution with large probability. However, for smaller values of w, when a single solution exists, finding it becomes much harder. For instance, assuming

$\mathcal{H}$ and $\mathcal{S}$ are perfectly random, a solution exists if:

$$\binom{n}{w} \geq 2^{n-k}. \tag{1}$$

When w is close to the value defined by this bound, the problem of SD is expected to be hard.

The McEliece and Niederreiter Cryptosystems

In the cryptosystems of McEliece [7] and Niederreiter [8], illegitimate decryption is possible if an instance of SD can be solved. The input of this instance is a scrambled Goppa code parity check matrix of size $mt \times 2^m$, a syndrome of length mt and the target weight is t. In general, for such parameters, no solution exists as $\binom{2^m}{t} < 2^{mt}$, but for instances taken from the McEliece (or Niederreiter) cryptosystems, one knows that a solution exists and that it is unique. Most algorithms designed to solve SD are focused on this type of parameters, but will also work well for any parameters close to the bound given by equation (1).

In this case, the best known algorithms are based on information set decoding and consist in finding a set of k columns of $\mathcal{H}$ such that a solution e to the problem of SD is all zero on these k positions. Computing a pseudo-inverse $\mathcal{H}^{-1}$ of the remaining $n-k$ columns of $\mathcal{H}$ makes it possible to invert $\mathcal{S}$ and obtain the solution e. The first algorithm to use this technique was due to Lee and Brickell [6] and then improved by Stern [9]. The latest refinement is due to Canteaut and Chabaud [3] and greatly improves the polynomial part of the complexity of the algorithm. However, this complexity remains exponential of the form $\mathcal{O}\left(\mathcal{P}oly\,(n)\left(\frac{n}{n-k}\right)^w\right)$.

Typical parameters are $m = 11$ and $t = 30$, and thus $n = 2^{11}$, $n - k = 330$, and $w = 30$, leading to an attack complexity just above 2^{80}.

3 McEliece-Based Signatures

In 2001, Courtois, Finiasz and Sendrier [4] presented the first code-based signature scheme to rely on the problem of SD. In this construction a counter is appended to the message to sign before hashing it into a syndrome. The signer then tries to decode the syndrome and increments the counter until a decodable syndrome is found. The signature then consists of the counter and the error pattern e corresponding to the syndrome.

As in the McEliece cryptosystem, the matrix is a scrambled Goppa code matrix, but here the value of t has to be very small: the average number of decoding attempts before obtaining a decodable syndrome is $t!$. Typical parameters are $m = 16$ and $t = 9$, and thus $n = 2^{16}$, $n - k = 144$ and

$w = 9$, once again leading to an attack complexity just above 2^{80} using the Canteaut-Chabaud algorithm.

However, these parameters have been much less studied than previous ones: can some attacks take advantage of the very small value of t? Moreover, in this construction the attacker does not have a single instance of SD to solve, but has to solve one among many: each value of the counter corresponds to a different instance of SD involving the same matrix. The attacker can now try to solve multiple instances of SD in parallel, some of which might be easier than others. This does not seem to improve the complexity of the existing attacks, but might be a weakness against new attacks.

4 Fast Syndrome Based Hash Function

Relying on a variation of the SD problem, a family of fast syndrome based cryptographic hash functions [1] was proposed. In this construction a binary matrix $\mathcal{H}$ is used to build the compression function of a Merkle-Damgård hash function. The input (chaining or initial vector and message block) is converted into a low weight word using a one to one mapping and the output is the product of this word by $\mathcal{H}$. For efficiency reasons, instead of converting the input into any low weight word, it is converted in a *regular* word, that is, the word is split in w block of $\frac{n}{w}$ bits and each block has a Hamming weight of 1.

Inverting or finding collisions on the compression function can be reduced to solving an instance of a problem very close to SD, but using regular words, which is also proven to be NP-complete. However, as we need the compression function to compress, it is clear that when trying to invert it or to find collisions many solutions exist, and finding a single one is enough to break the construction. In this case the best algorithm is no longer based on information set decoding but uses the generalized birthday algorithm of Wagner [10].

This generalized birthday algorithm improves the standard birthday paradox by only looking for specific solutions and discarding other solutions. Therefore, it does not apply to the standard McEliece parameters of SD as a single solution exists, but is suitable for values of w yielding many solutions. It applies for any values of w such that:

$$\binom{n}{\frac{w}{4}} \geq 2^{\frac{n-k}{3}}.$$

The larger w, the more efficient the algorithm will be.

5 The Multiple of Low Weight Problem

In stream cipher analysis and in the trapdoor stream cipher construction of Finiasz and Vaudenay [5], a key problem is that of finding a multiple of low weight of a given polynomial on $\mathbf{F}_2$.

Multiple of Low Weight Problem:

Input: a polynomial P of degree d_P on $\mathbf{F}_2$, a weight w and a degree d.
Output: a polynomial K of degree $\leq d$ and weight $\leq w$, multiple of P.

This problem can be translated in terms of SD by simply computing the matrix $\mathcal{H}$ which columns are defined by $\mathcal{H}_i = X^i \mod P(X)$, for $i \in [1; d]$. Then, finding a multiple of P of weight $\leq w$ simply consists in solving an instance of SD with input $\mathcal{H}$, $\mathcal{S} = (1, 0, \ldots, 0)$ and weight $w - 1$.

The best known techniques for solving the multiple of low weight problem use algorithms designed for SD but it might be possible to improve them based on the specific structure of the matrix $\mathcal{H}$. Also, for cryptanalysis, d and w can be chosen by the adversary: a larger d will require to capture more output bits of the stream cipher, a larger w will also require more output bits and will increase the complexity of the attack. Each algorithm will thus lead to different optimal parameters for the attack and the best attack will depend on the way the multiple polynomial K is used afterwards.

6 Conclusion

Algorithms based on information set decoding techniques seem to perform best for most instances of syndrome decoding derived from the cryptosystems of McEliece and Niederreiter. However, other cryptographic constructions lead to completely different inputs for which deciding which is the best algorithm is not always obvious. New algorithms do not have to target the cryptosystems of McEliece and Niederreiter, but can also target these newer constructions which might offer more suitable inputs.

References

[1] D. Augot, M. Finiasz, and N. Sendrier. A family of fast syndrome based cryptographic hash functions. In E. Dawson and S. Vaudenay, editors, *Mycrypt 2005*, volume 3715 of *LNCS*, pages 64–83. Springer, 2005.

[2] E. R. Berlekamp, R. J. McEliece, and H. C. van Tilborg. On the inherent intractability of certain coding problems. *IEEE Transactions on Information Theory*, 24(3), May 1978.

[3] A. Canteaut and F. Chabaud. A new algorithm for finding minimum-weight words in a linear code: Application to McEliece's cryptosystem and to narrow-sense BCH codes of length 511. *IEEE Transactions on Information Theory*, 44(1):367–378, January 1998.

[4] N. Courtois, M. Finiasz, and N. Sendrier. How to achieve a McEliece-based digital signature scheme. In C. Boyd, editor, *Asiacrypt 2001*, volume 2248 of *LNCS*, pages 157–174. Springer, 2001.

[5] M. Finiasz and S. Vaudenay. When stream cipher analysis meets public-key cryptography. In E. Biham and A.M. Youssef, editors, *SAC 2006*, volume to appear of *LNCS*. Springer, 2006.

[6] P. J. Lee and E. F. Brickell. An observation on the security of McEliece's public-key cryptosystem. In C. G. Günther, editor, *Eurocrypt 88*, volume 330 of *LNCS*, pages 275–280. Springer, 1988.

[7] R. J. McEliece. A public-key cryptosystem based on algebraic coding theory. *DSN Prog. Rep.,* Jet Prop. Lab., California Inst. Technol., Pasadena, CA, pages 114–116, January 1978.

[8] H. Niederreiter. Knapsack-type crytosystems and algebraic coding theory. *Prob. Contr. Inform. Theory*, 15(2):157–166, 1986.

[9] J. Stern. A method for finding codewords of small weight. In G. Cohen and J. Wolfmann, editors, *Coding theory and applications*, volume 388 of *LNCS*, pages 106–113. Springer, 1989.

[10] D. Wagner. A generalized birthday problem. In M. Yung, editor, *Crypto 2002*, volume 2442 of *Lecture Notes in Computer Science*, pages 288–304. Springer, 2002.