
A FAULT ATTACK ON THE NIEDERREITER CRYPTOSYSTEM USING BINARY IRREDUCIBLE GOPPA CODES

JULIAN DANNER AND MARTIN KREUZER

Fakultät für Informatik und Mathematik, Universität Passau, D-94030 Passau, Germany
e-mail address: Julian.Danner@uni-passau.de

Fakultät für Informatik und Mathematik, Universität Passau, D-94030 Passau, Germany
e-mail address: Martin.Kreuzer@uni-passau.de

ABSTRACT. A fault injection framework for the decryption algorithm of the Niederreiter public-key cryptosystem using binary irreducible Goppa codes and classical decoding techniques is described. In particular, we obtain low-degree polynomial equations in parts of the secret key. For the resulting system of polynomial equations, we present an efficient solving strategy and show how to extend certain solutions to alternative secret keys. We also provide estimates for the expected number of required fault injections, apply the framework to state-of-the-art security levels, and propose countermeasures against this type of fault attack.

1. INTRODUCTION

Many established public-key cryptosystems rely on the hardness of the factorization problem or the discrete logarithm problem. However, their long-term security is not guaranteed, because they can be broken in polynomial time using sufficiently large quantum-computers [22]. This motivates the need for post-quantum cryptosystems. One of the oldest public-key cryptosystems is the McEliece cryptosystem [17] which was designed to be used by NASA. However, partly due to its large public-key sizes, it was never standardized. The security of the McEliece cryptosystem relies on the hardness of the decoding problem for random linear codes [19].

The Niederreiter cryptosystem is a variant of the McEliece cryptosystem which offers some improvements as to the costs of encryption and decryption and requires smaller public-key sizes [18]. Particularly promising variants are based on the problem of decoding binary Goppa codes [7].

For some special subclasses of Goppa codes, namely quasi-dyadic and quasi-cyclic Goppa codes, there exist successful algebraic attacks which take advantage of the particular structure of the code [9]. But apart from these subclasses, binary (irreducible) Goppa codes still appear to resist structural attacks. The best known algorithms for decoding arbitrary linear codes use information set decoding, and their running time is still exponential [16].

Key words and phrases: Niederreiter cryptosystem, binary Goppa code, side-channel analysis, fault attack.

Structural attacks on the Niederreiter and McEliece cryptosystems have been widely researched. However, their side-channel analysis is not quite as advanced. In general, any additional source of information that can be derived from a specific hardware or software implementation of a cryptosystem, or even its execution, may be considered a side-channel. Clearly, side-channel attacks have to be measured by their practical feasibility. Passive side-channel analysis is frequently based on power-analysis of hardware devices or timings of the execution of certain parts of the encryption or decryption algorithms.

In this paper we consider a type of active side-channel attacks called fault attacks. In particular, we assume that we are able to inject an error into the usual flow of the algorithms of the cryptosystem by corrupting the contents of specific memory cells at a particular moment. Common methods of achieving such a fault injection are manipulations of the power supply and the usage of pulsed laser or ion beams. The usability of a fault attack depends chiefly on the fault model which describes the required physical capabilities of the attacker.

For the Niederreiter and McEliece cryptosystems, there exist successful passive side-channel attacks [2, 21, 23, 24] which exploit traditional side-channels, such as timing and power consumption attacks. For active side-channel attacks such as fault attacks, much less seems to be known. We found only the article [6] which analyses the sensitivity of these cryptosystems to fault injections in the encryption and key generation algorithms.

However, the most natural target for a fault attack at a public-key cryptosystem is the decryption algorithm, because it uses the secret key whose knowledge would allow us to completely break the system. In fact, it suffices to find an *alternative key*, i.e., a key which also allows us to decrypt ciphertexts. The current paper is a first attempt at constructing such fault attacks. Our target class of cryptosystems are Niederreiter cryptosystems based on binary irreducible Goppa codes. We call them briefly *BIG-N cryptosystems*. The attack is based on the following fault model:

- (1) The decryption algorithm follows the standard pattern: first the error locator polynomial is computed, and then it is evaluated at the support elements to deduce the plaintext.
- (2) The decryption algorithm not only reconstructs plaintext units of weight t , the designed distance, but also *illegal* plaintexts p , i.e., plaintexts of weight $1 \leq \text{wt}(p) < t$.
- (3) After the error-locator polynomial has been computed, we are able to inject a uniformly random fault into a chosen coefficient of this polynomial, i.e., we are able to replace it by a random bit-tuple.

Moreover, we assume that we are able to repeat these injections hundreds or even several thousand times. The resulting BIG-N fault attack breaks all state-of-the-art security levels [4], even “long term” secure ones, within minutes. Consequently, we suggest explicit countermeasures for hardware- or software-implementations of BIG-N cryptosystems.

Feasibility and Applicability. Let us briefly discuss the practical applicability and relevance of the new fault attack. As we shall see in Section 7, for carrying out the actual fault injections, we need to hit the register holding a certain 10-13 bit wide coefficient at a specified point in time. Using modern equipment, this is a mild requirement (see for instance [12] and [5]).

Thus it seems more pertinent to discuss the vulnerability of current cryptographic BIG-N schemes to the proposed attacks. Looking through the NIST Post-Quantum Standardization candidates, there are three original submissions based on Goppa codes: “BIG Quake” which is vulnerable to our attack, “Classic McEliece”, and “NTS-KEM”. The latter two are actually

key exchange systems where the key cannot be chosen freely, and for the random vector a hash value is transmitted. This is clearly a setting in which the current attack does not apply directly. However, the reference implementation [27] for the BIG-N part of the schemes is not protected against the attack. Furthermore, the other published reference implementations [10] and [11] of the BIG-N cryptosystem use constant weight encoders and decoders whose standard implementation is also vulnerable. Thus, although it may not be very difficult to defend against, apparently all current implementations and applications of the BIG-N cryptosystem are vulnerable to the BIG-N fault attack.

Contents. Let us describe the structure of the paper in more detail. After recalling some basic facts about binary Goppa codes and BIG-N cryptosystems in Sections 2 and 3, we present a BIG-N fault attack framework in Section 4. In particular, we analyze the assumptions underlying the attack carefully and propose countermeasures. Then we introduce the general framework for the fault attack: we assume that we are able to replace the error locator polynomial $\sigma_e(x)$ by an erroneous one of the form $\tilde{\sigma}_e(x) = \varepsilon x^d + \sigma_e(x)$ where $\varepsilon \in \mathbb{F}_{2^m}$ is distributed uniformly at random and d is the chosen degree under attack.

In Section 5 we analyze the resulting equations for the components of the support vector $\alpha = (\alpha_1, \dots, \alpha_n)$ of the binary Goppa code in two particular cases: we attack the constant and the quadratic coefficient of $\sigma_e(x)$. From a successful constant injection we derive a linear equation for the components of α , and from a successful quadratic injection we get a linear or a quadratic equation. However, this typically requires a sequence of injections until we succeed in obtaining an erroneously deciphered word of weight two.

The next steps are taken in Section 6 where we combine the acquired linear and quadratic equations into a fault equation system and then carry out the actual BIG-N fault attack in three steps: Firstly, we solve the fault equation system and get a set of support candidates. Secondly, using the fact that it suffices to find the support and the Goppa polynomial of a larger binary Goppa code containing the publicly known one, we determine a support candidate which can be extended. Finally, we use this alternative secret key to break the given BIG-N cryptosystem.

In the final section we offer some experiments and timings for the BIG-N fault attack. In particular, we provide estimates for the average numbers of constant and quadratic fault injections needed to succeed. Moreover, we collect the timings for breaking various security levels, ranging from one minute for “short term” 60-bit security to about 25 minutes for “long term” 266-bit security.

2. BINARY GOPPA CODES

For starters, let us recall the definition of a binary Goppa code. By \mathbb{F}_{2^m} we denote the finite field having 2^m elements. We also fix the following notation: given a tuple $c = (c_1, \dots, c_n) \in \mathbb{F}_2^n$, we let $\mathbb{I}_c := \{i \in \{1, \dots, n\} \mid c_i = 1\}$.

Definition 2.1. Let $m, t, n \in \mathbb{N}_+$ such that $mt < n \leq 2^m$.

- (a) A tuple $\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{F}_{2^m}^n$ such that $\alpha_i \neq \alpha_j$ for $i \neq j$ will be called a **support tuple**.
- (b) A polynomial $g \in \mathbb{F}_{2^m}[x]$ with $\deg(g) = t$ and $g(\alpha_i) \neq 0$ for $i \in \{1, \dots, n\}$ is called a **Goppa polynomial** for the support tuple α .

(c) The **(binary) Goppa code** with the generating pair (α, g) is given by

$$\Gamma(\alpha, g) = \{c \in \mathbb{F}_2^n \mid \sum_{i \in \mathbb{I}_c} (x - \alpha_i)^{-1} = 0 \text{ in } \mathbb{F}_{2^m}[x]/\langle g \rangle\}$$

In particular, if g is an irreducible polynomial, then $\Gamma(\alpha, g)$ is called an **irreducible Goppa code**.

Remark 2.2. For a Goppa code $C = \Gamma(\alpha, g)$ and for $c = (c_1, \dots, c_n) \in \mathbb{F}_2^n$, we have

$$\begin{aligned} c \in C &\Leftrightarrow \sum_{i \in \mathbb{I}_c} (x - \alpha_i)^{-1} = 0 && \text{(in } \mathbb{F}_{2^m}[x]/\langle g \rangle) \\ &\Leftrightarrow g \mid \sum_{i \in \mathbb{I}_c} \prod_{j \in \mathbb{I}_c \setminus \{i\}} (x - \alpha_j) && \text{(in } \mathbb{F}_{2^m}[x]) \end{aligned}$$

Remark 2.3. Given a Goppa code $C = \Gamma(\alpha, g)$ it is well-known that a parity-check matrix $H \in \text{Mat}_{mt \times n}(\mathbb{F}_2)$ of C can be obtained from the matrix

$$H' = \begin{pmatrix} \beta_1 & \cdots & \beta_n \\ \alpha_1 \beta_1 & \cdots & \alpha_n \beta_n \\ \vdots & & \vdots \\ \alpha_1^{t-1} \beta_1 & \cdots & \alpha_n^{t-1} \beta_n \end{pmatrix} \in \text{Mat}_{t,n}(\mathbb{F}_{2^m})$$

where $\beta_i = g(\alpha_i)^{-1}$ for $i = 1, \dots, n$, by replacing each entry of H' by a column of m bits that arise by fixing an \mathbb{F}_2 -basis of \mathbb{F}_{2^m} .

From here on, let $C = \Gamma(\alpha, g)$ be a binary Goppa code with parameters $m, t, n \in \mathbb{N}_+$ as described above, and let $H \in \text{Mat}_{mt,n}(\mathbb{F}_2)$ be a parity-check matrix of C . For $\tilde{c} \in \mathbb{F}_2^n$, we call $s_{\tilde{c}} = \tilde{c}H^{tr} \in \mathbb{F}_2^{mt}$ the **syndrome** of \tilde{c} with respect to H . Then we have $\tilde{c} \in C$ if and only if $\tilde{c}H^{tr} = 0$.

It is known that $\dim C \geq n - mt$ and that the minimal distance of C satisfies $d_{\min}(C) \geq t$. If C is irreducible, we even have $C = \Gamma(\alpha, g^2)$ and $d_{\min}(C) \geq 2t$. Hence, in general, up to $\frac{t}{2}$ errors can be corrected, and if C is irreducible, even up to t errors can be corrected uniquely.

Since we are going to use it extensively, let us briefly recall the classical syndrome decoding method for binary Goppa codes.

Remark 2.4. (Syndrome Decoding for Goppa Codes)

Consider a received word $\tilde{c} = c + e \in \mathbb{F}_2^n$ with $c \in C$ and $e \in \mathbb{F}_2^n$. Then we define the **error-locator polynomial** by

$$\sigma_e(x) = \prod_{i \in \mathbb{I}_e} (x - \alpha_i) \in \mathbb{F}_{2^m}[x]$$

and the **syndrome polynomial** of $\tilde{c} = (\tilde{c}_1, \dots, \tilde{c}_n) \in \mathbb{F}_2^n$ by

$$s_{\tilde{c}}(x) = \sum_{i=1}^n \frac{\tilde{c}_i}{g(\alpha_i)} \frac{g(x) - g(\alpha_i)}{x - \alpha_i} \in \mathbb{F}_{2^m}[x]$$

Then we have $s_e(x) \equiv s_{\tilde{c}}(x) \pmod{g(x)}$, and we obtain the **key equation**

$$\sigma_e(x) \cdot s_{\tilde{c}}(x) \equiv \sigma'_e(x) \pmod{g(x)}$$

Let us write $g(x) = g_t x^t + \cdots + g_1 x + g_0$ with $g_0, \dots, g_t \in \mathbb{F}_{2^m}$. Using the same \mathbb{F}_2 -basis of \mathbb{F}_{2^m} as in Remark 2.3, we combine sequences of m entries in the syndrome

$s_{\tilde{c}} = \tilde{c}H^{tr} = eH^{tr} \in \mathbb{F}_2^{mt}$ and get $\hat{s} \in \mathbb{F}_2^t$. Now the coefficients of the syndrome polynomial $s_{\tilde{c}}(x)$ can be computed by a simple multiplication of $\hat{s} \in \mathbb{F}_2^t$ with the matrix

$$S_g = \begin{pmatrix} g_t & g_{t-1} & \cdots & g_1 \\ & \ddots & \ddots & \vdots \\ & & g_t & g_{t-1} \\ & & & g_t \end{pmatrix} \in \text{GL}_t(\mathbb{F}_2^m)$$

Hence it is sufficient to solve the key equation for the error-locator polynomial $\sigma_e(x)$. Then the zeros of $\sigma_e(x)$ determine the error vector $e \in \mathbb{F}_2^n$ via the observation that, for $i \in \{1, \dots, n\}$, we have $e_i = 1$ if and only if $\sigma_e(\alpha_i) = 0$. Finally, we decode \tilde{c} to $c = \tilde{c} + e$.

The main task in this method is to solve the key equation. This can be done in several ways, for instance explicitly using the Sugiyama-Algorithm [25], or implicitly using the Berlekamp-Massey Algorithm [3, 15]. As shown in [8], one may consider these two algorithms as essentially equivalent. For a binary irreducible Goppa code C , up to t errors can be corrected using the Patterson Algorithm [20] which also uses the key equation to obtain the error-locator polynomial. Independently of the chosen method, decoding consists of the following three basic steps.

Algorithm 2.5. (The Syndrome Decoding Algorithm)

Let $C = \Gamma(\alpha, g) \subseteq \mathbb{F}_2^n$ be a binary Goppa code, where $\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{F}_2^m$ and $g \in \mathbb{F}_2^m[x]$, and let $t = \deg(g)$. Let $H \in \text{Mat}_{mt \times n}(\mathbb{F}_2)$ be a parity check matrix of C obtained as in Remark 2.3, and let $s_{\tilde{c}} = \tilde{c}H^{tr} \neq 0$ be the syndrome of a given received word $\tilde{c} \in \mathbb{F}_2^n \setminus C$. We assume that \tilde{c} is of the form $\tilde{c} = c + e$ with $c \in C$ and $e \in \mathbb{F}_2^n$ such that $\#\mathbb{I}_e \leq \frac{t}{2}$ (or $\#\mathbb{I}_e \leq t$, if C is irreducible). Then we compute $e \in \mathbb{F}_2^n$ from $s_{\tilde{c}}$ using the following steps.

- (1) Compute the syndrome polynomial $s_{\tilde{c}}(x) \in \mathbb{F}_2^m[x]$ from $s_{\tilde{c}}$.
- (2) Compute the error-locator polynomial $\sigma_e(x)$ by solving the key equation (explicitly or implicitly), e.g., via one of the cited algorithms.
- (3) For $i = 1, \dots, n$, let $e_i = 1$ if $\sigma_e(\alpha_i) = 0$ and $e_i = 0$ otherwise. Return $e = (e_1, \dots, e_n) \in \mathbb{F}_2^n$ and stop.

3. BIG-N CRYPTOSYSTEMS

In this section we introduce Niederreiter cryptosystems using binary irreducible Goppa codes (see [18]) and recall some of their basic properties. Subsequently, we write $W_{n,t}$ for the set of all elements $v \in \mathbb{F}_2^n$ of **Hamming weight** t , i.e., such that $\text{wt}(v) = \#\mathbb{I}_v = t$.

Definition 3.1. (BIG-N Cryptosystems)

A **Niederreiter cryptosystem using a binary irreducible Goppa code C (BIG-N cryptosystem)** with parameters $m, t, n \in \mathbb{N}_+$ such that $mt < n \leq 2^m$ is represented by a tuple $(K_{\text{sec}}, K_{\text{pub}}, \mathcal{P}, \mathcal{C}, \text{encr}, \text{decr})$ consisting of the following parts.

- (a) The tuple $K_{\text{sec}} = (S, H, P, \alpha, g)$, where $H \in \text{Mat}_{mt \times n}(\mathbb{F}_2)$ is a parity check matrix of the irreducible Goppa code $C = \Gamma(\alpha, g)$ with $\deg(g) = t$ and dimension $n - mt$, where $P \in \text{Mat}_n(\mathbb{F}_2)$ is a permutation matrix, and where $S \in \text{GL}_{mt}(\mathbb{F}_2)$, is called the **secret key**.

- (b) The tuple $K_{\text{pub}} = (m, t, n, H_{\text{pub}})$, where m, t, n are the parameters of C , and where $H_{\text{pub}} = S \cdot H \cdot P \in \text{Mat}_{mt,n}(\mathbb{F}_2)$, is called the **public key**. The matrix H_{pub} is also called the **public parity check matrix**.
- (c) The set $\mathcal{P} = W_{n,t}$ is called the **plaintext space**, and an element $p \in \mathcal{P}$ is called a **plaintext unit**.
- (d) The set $\mathcal{C} = \{pH_{\text{pub}}^{tr} \mid p \in \mathcal{P}\}$ is called the **ciphertext space**, and an element $c \in \mathcal{C}$ is called a **ciphertext unit**.
- (e) The map $\text{encr}: \mathcal{P} \rightarrow \mathcal{C}$ given by $\text{encr}(p) = pH_{\text{pub}}^{tr}$ is called the **encryption map**.
- (f) The map $\text{decr}: \mathcal{C} \rightarrow \mathcal{P}$ given by $\text{decr}(c) = \gamma(c \cdot (S^{tr})^{-1}) \cdot (P^{tr})^{-1}$ is called the **decryption map**. Here the map $\gamma: \mathbb{F}_2^{mt} \rightarrow \mathbb{F}_2^n$ satisfies $\gamma(eH^{tr}) = e$ for all $e \in W_{n,t}$ and is a syndrome decoding algorithm which corrects up to t errors.

In particular, notice that we have $\text{decr}(\text{encr}(p)) = p$ for all $p \in \mathcal{P}$. Indeed, for a plaintext unit $p \in \mathcal{P}$ and its ciphertext $c = \text{encr}(p) = pH_{\text{pub}}^{tr}$, we have $\text{wt}(pP^{tr}) = \text{wt}(p) = t$, and therefore $\gamma(c(S^{tr})^{-1}) = \gamma((pP^{tr})H^{tr}) = pP^{tr}$. Hence we get $\text{decr}(c) = p$, as claimed.

A BIG-N cryptosystem is a public-key cryptosystem, i.e., the encryption map encr can be computed solely using the public key K_{pub} , but the application of the decryption map decr requires the knowledge of the secret key K_{sec} . The cryptosystem $(K_{\text{sec}}, K_{\text{pub}}, \mathcal{P}, \mathcal{C}, \text{encr}, \text{decr})$ is considered broken if one can efficiently compute a fast **alternative decryption map** $\text{decr}' : \mathcal{C} \rightarrow \mathcal{P}$ which satisfies $\text{decr}'(\text{encr}(p)) = p$ for all $p \in \mathcal{P}$. If, for an alternative decryption map decr' , there exists a secret key K'_{sec} such that $(K'_{\text{sec}}, K_{\text{pub}}, \mathcal{P}, \mathcal{C}, \text{encr}, \text{decr}')$ is a BIG-N cryptosystem, then K'_{sec} is called an **alternative secret key**.

In the following we let $m, t, n \in \mathbb{N}_+$ such that $mt < n \leq 2^m$, we let $C = \Gamma(\alpha, g)$ be an irreducible Goppa code with these parameters, and we let $(K_{\text{pub}}, K_{\text{sec}}, \mathcal{P}, \mathcal{C}, \text{encr}, \text{decr})$ be a BIG-N cryptosystem using C . Moreover, we fix an \mathbb{F}_2 -basis of \mathbb{F}_{2^m} and use it to convert elements of \mathbb{F}_{2^m} to m -tuples of bits and vice versa.

Remark 3.2. (Simplified BIG-N Cryptosystems)

In the above setting, let $H \in \text{Mat}_{mt,n}(\mathbb{F}_2)$ be the parity check matrix for C obtained as in Remark 2.3, and let $P \in \text{Mat}_n(\mathbb{F}_2)$ be the permutation matrix contained in the secret key. Then we define $\tilde{\alpha} := \alpha \cdot P \in \mathbb{F}_{2^m}^n$ and $\tilde{C} = \Gamma(\tilde{\alpha}, g)$. Clearly, also \tilde{C} is a binary irreducible Goppa code, and the matrix $\tilde{H} := H \cdot P$ is a parity check matrix for \tilde{C} which has the shape described in Remark 2.3.

Since the public parity check matrix satisfies $H_{\text{pub}} = S \cdot \tilde{H}$, an alternative secret key for the original cryptosystem is given by $(S, \tilde{H}, I_n, \tilde{\alpha}, g)$, where I_n is the identity matrix of size n . Consequently, we may use the Goppa code \tilde{C} instead of C and get rid of the permutation matrix P as part of the secret key.

From here on we use secret keys of the form $K_{\text{sec}} = (S, H, \alpha, g)$ and simplify all BIG-N cryptosystems accordingly. Notice that the code defined by the public parity check matrix H_{pub} of the simplified cryptosystem is now equal to C and continues to be publicly known.

In view of the preceding section, we know that a BIG-N cryptosystem is broken if an efficient t -error correcting algorithm is found for the Goppa code C . The next algorithm shows that this requirement can be weakened even further.

Algorithm 3.3. (An Alternative Decryption Algorithm)

Consider a BIG-N cryptosystem $(K_{\text{sec}}, K_{\text{pub}}, \mathcal{P}, \mathcal{C}, \text{encr}, \text{decr})$ using the Goppa code $C = \Gamma(\alpha, g)$ and having the public key $K_{\text{pub}} = (m, t, n, H_{\text{pub}})$. Assume that there exist a support

tuple $\tilde{\alpha} \in \mathbb{F}_2^n$ and a Goppa polynomial $\tilde{g} \in \mathbb{F}_2[x]$ for $\tilde{\alpha}$ such that $\tilde{t} := \deg(\tilde{g}) \geq 2t$ and $C \subseteq \tilde{C} := \Gamma(\tilde{\alpha}, \tilde{g})$. Given a ciphertext unit $c \in \mathcal{C}$, consider the following sequence of instructions.

- (1) Compute a parity check matrix $\tilde{H} \in \text{Mat}_{m\tilde{t},n}(\mathbb{F}_2)$ for \tilde{C} as in Remark 2.3.
- (2) Compute the matrix $\tilde{S} \in \text{Mat}_{m\tilde{t},mt}(\mathbb{F}_2)$ such that $\tilde{S} \cdot H_{\text{pub}} = \tilde{H}$.
- (3) Compute $\tilde{c} = c\tilde{S}^{\text{tr}}$ and apply the syndrome decoding algorithm of \tilde{C} to \tilde{c} . Return the resulting tuple p .

This is an algorithm which computes a plaintext unit $p \in \mathcal{P}$ such that we have $\text{encr}(p) = c$.

Proof. First notice that $C \subseteq \tilde{C}^\perp$ implies $\tilde{C}^\perp \subseteq C^\perp$. Since the rows of \tilde{H} generate \tilde{C}^\perp and the rows of H_{pub} generate C^\perp , it follows that there exists a matrix $\tilde{S} \in \text{Mat}_{m\tilde{t},mt}(\mathbb{F}_2)$ such that $\tilde{S} \cdot H_{\text{pub}} = \tilde{H}$, and this matrix can be computed in step (2). Furthermore, the input for Algorithm 2.5 in step (3) is correct, because $c = \text{encr}(p)$ for some $p \in \mathcal{P}$, and therefore $\tilde{c} = c\tilde{S}^{\text{tr}} = p H_{\text{pub}}^{\text{tr}} \tilde{S}^{\text{tr}} = p \tilde{H}^{\text{tr}}$ is a syndrome of weight $\text{wt}(p) = t$ with respect to \tilde{H} . Consequently, the algorithm can be executed.

Its finiteness is clear. Its correctness follows from the fact that Algorithm 2.5 corrects up to $\frac{\deg(\tilde{g})}{2} = \frac{\tilde{t}}{2} \geq t = \text{wt}(p)$ errors, and hence determines $p \in \mathcal{P}$ correctly. \square

In view of this algorithm, it is clear that a BIG-N cryptosystem using the code $C = \Gamma(\alpha, g)$ is broken if a generating pair $(\tilde{\alpha}, \tilde{g})$ for a binary Goppa code is found such that $\deg(\tilde{g}) \geq 2t$ and such that $C \subseteq \tilde{C} = \Gamma(\tilde{\alpha}, \tilde{g})$. Therefore such a pair $(\tilde{\alpha}, \tilde{g})$ will be called an **alternative secret pair**.

4. THE BIG-N FAULT INJECTION FRAMEWORK

In this section we let $m, t, n \in \mathbb{N}_+$ such that $mt < n \leq 2^m$, we let $C = \Gamma(\alpha, g)$ be a binary irreducible Goppa code with these parameters, and we let $(K_{\text{pub}}, K_{\text{sec}}, \mathcal{P}, \mathcal{C}, \text{encr}, \text{decr})$ be a BIG-N cryptosystem using C . Moreover, we write $K_{\text{pub}} = (m, t, n, H_{\text{pub}})$ for the public key and $K_{\text{sec}} = (S, H, \alpha, g)$ for the secret key. Using Remark 3.2, we assume that $H_{\text{pub}} = S \cdot H$. Recall that we let $\mathbb{I}_p = \{i \in \{1, \dots, n\} \mid p_i = 1\}$ for $p = (p_1, \dots, p_n) \in \mathbb{F}_2^n$.

In order to mount the proposed fault attack, we require that the implementation of the decryption map decr satisfies three assumptions. They are motivated by the following *usual* implementation which is based on the classical syndrome decoding method of Algorithm 2.5.

Algorithm 4.1. (Implementing the Decryption Map decr)

Input: a ciphertext unit $c = \text{encr}(p) \in \mathcal{C}$, the secret key $K_{\text{sec}} = (S, H, \alpha, g)$

Output: a plaintext unit $p \in \mathcal{P}$

- 1: Compute the syndrome $s = c(S^{\text{tr}})^{-1} \in \mathbb{F}_2^{mt}$ with respect to H .
- 2: Compute the syndrome polynomial $s_p(x)$.
- 3: Compute the error-locator polynomial $\sigma_p(x)$.
- 4: Determine $p = (p_1, \dots, p_n) \in \mathcal{P}$ by setting $p_i = 1$ if $\sigma_p(\alpha_i) = 0$ and $p_i = 0$ otherwise.
- 5: **return** p

In view of this algorithm, the following assumption seems natural.

Assumption 4.2. *The implementation of the decryption map decr makes use of the error-locator polynomial in such a way that it is first computed explicitly, then evaluated at the support elements, and finally the resulting plaintext unit is returned.*

Next we let $p \in \mathbb{F}_2^n$ with $0 < \text{wt}(p) < t$ and $c = pH_{\text{pub}}^{\text{tr}}$. Since Algorithm 4.1 is based on the Syndrome Decoding Algorithm 2.5, and since syndrome decoding corrects *up to* t errors, applying this algorithm to c will correctly return p . Hence we make the following assumption.

Assumption 4.3. *Let $p \in \mathbb{F}_2^n$ with $0 < \text{wt}(p) \leq t$, and let $c = pH_{\text{pub}}^{\text{tr}}$. If we apply the decryption map decr to c , it returns p .*

In order to be able to inject a fault into the decryption process, we require one final assumption.

Assumption 4.4. *Let $d \in \mathbb{N}$ with $d < t$, let $p \in \mathbb{F}_2^n$ with $0 < \text{wt}(p) \leq t$, and let $c = pH_{\text{pub}}^{\text{tr}}$. After the error-locator polynomial $\sigma_p(x)$ has been computed during the decryption process of c , we assume that we can inject a uniformly random fault into the d -th coefficient of $\sigma_p(x)$. In other words, we assume that we may replace $\sigma_p(x)$ by a polynomial $\tilde{\sigma}_p(x) = \varepsilon x^d + \sigma_p(x)$ where $\varepsilon \in \mathbb{F}_{2^m}$ is chosen uniformly at random.*

As a testimony to the applicability of these assumptions, consider the hardware implementation described in [27]. It follows the steps of Algorithm 4.1, and after $\sigma_p(x)$ has been computed, this polynomial is transferred to the evaluation module by sending $m(t+1)$ bits, where each m -bit subtuple represents one of coefficients of $\sigma_p(x)$. This transfer process is suited for an injection of a (uniformly distributed) random error $\varepsilon \in \mathbb{F}_{2^m}$ into the d -th coefficient of $\sigma_p(x)$. From a hardware point of view, it corresponds to randomly changing the state of m chosen consecutive bits.

Every implementation of the decryption map that satisfies these assumptions is vulnerable to the following fault injection framework.

Algorithm 4.5. (The BIG-N Fault Injection Framework)

For a BIG-N cryptosystem $(K_{\text{sec}}, K_{\text{pub}}, \mathcal{C}, \mathcal{P}, \text{encr}, \text{decr})$ as above, assume that the implementation of the decryption map decr satisfies Assumptions 4.2, 4.3, and 4.4. Choose a word $p \in \mathbb{F}_2^n$ with $0 < \text{wt}(p) \leq t$ and a number $d \in \mathbb{N}$ with $d < t$. Then consider the following sequence of instructions.

- (1) Compute $c = pH_{\text{pub}}^{\text{tr}} \in \mathbb{F}_2^{mt}$.
- (2) Start the decryption algorithm decr with input $c \in \mathbb{F}_2^{mt}$ and inject a uniformly random fault $\varepsilon \in \mathbb{F}_{2^m}$ in the d -th coefficient of $\sigma_p(x)$ such that $\tilde{\sigma}_p(x) = \varepsilon x^d + \sigma_p(x)$ is evaluated instead of $\sigma_p(x)$.
- (3) Return $\tilde{p} \in \mathbb{F}_2^n$, the output of the faulty decryption of step (2).

This is an algorithm which returns a tuple $\tilde{p} = (\tilde{p}_1, \dots, \tilde{p}_n) \in \mathbb{F}_2^n$ such that, for $i \in \{1, \dots, n\}$, we have $\tilde{p}_i = 1$ if and only if $\varepsilon \alpha_i^d + \prod_{j \in \mathbb{I}_p} (\alpha_i - \alpha_j) = 0$.

Consequently, every component $\tilde{p}_i = 1$ yields a polynomial equation in $\mathbb{F}_{2^m}[x_0, \dots, x_n]$ which is satisfied for $(\varepsilon, \alpha_1, \dots, \alpha_n)$

Proof. By Assumption 4.3, the decryption algorithm is correct for all syndromes $c = pH_{\text{pub}}^{\text{tr}}$ with $p \in \mathbb{F}_2^n$ and $0 < \text{wt}(p) \leq t$. In combination with Assumption 4.2, this means that in the course of the decryption algorithm, the error-locator polynomial $\sigma_p(x)$ is computed correctly. By Assumption 4.2, the output $\tilde{p} = (\tilde{p}_1, \dots, \tilde{p}_n) \in \mathbb{F}_2^n$ satisfies $\tilde{p}_i = 1$ if and only if $\tilde{\sigma}_p(\alpha_i) = 0$. Since we have $\tilde{\sigma}_p(x) = \varepsilon x^d + \sigma_p(x)$ and $\sigma_p(x) = \prod_{i \in \mathbb{I}_p} (x - \alpha_i)$, the claim follows. \square

In the setting of this framework, we call the triple (p, d, \tilde{p}) a **BIG-N fault injection** in degree d . We also say that this injection **uses the fault** ε . The possibility to perform BIG-N fault injections can be prevented as follows.

Remark 4.6. (Countermeasures)

Let (p, d, \tilde{p}) be a BIG-N fault injection.

- (a) The output of the decryption map of a BIG-N cryptosystem is an n -bit tuple of weight t . In general, for a BIG-N fault injection (p, d, \tilde{p}) , the output \tilde{p} will have weight $\leq t$. Thus checking the weight of \tilde{p} discovers most fault injections.
- (b) A further way to detect fault injections is to re-encrypt the output \tilde{p} . If $\tilde{p} \neq p$, we will get $\tilde{p} H_{\text{pub}}^{tr} \neq c$.

The next proposition collects some observations on BIG-N fault injections.

Proposition 4.7. *Let (p, d, \tilde{p}) be a BIG-N fault injection which uses the fault $\varepsilon \in \mathbb{F}_{2^m}$, and let $\tilde{\sigma}_p(x) = \varepsilon x^d + \sigma_p(x)$. Then the following claims hold.*

- (a) *If $\text{wt}(\tilde{p}) > \text{wt}(p)$ then $\varepsilon \neq 0$.*
- (b) *We have either $p = \tilde{p}$ or $\#(\mathbb{I}_p \cap \mathbb{I}_{\tilde{p}}) = 1$.*
- (c) *If $\varepsilon \neq 0$ then $\mathbb{I}_p \cap \mathbb{I}_{\tilde{p}} = \{i\}$ is equivalent to $i \in \mathbb{I}_p$, $\alpha_i = 0$, and $d > 0$.*

Proof. First we prove (a). Assuming that $\text{wt}(\tilde{p}) > \text{wt}(p)$, we have to show $\varepsilon \neq 0$. Considering the way in which \tilde{p} is determined by the zeros of $\tilde{\sigma}_p(x)$, we see that $\deg(\tilde{\sigma}_p(x)) \geq \text{wt}(\tilde{p})$. Since $\text{wt}(p) = \deg(\sigma_p(x))$, we get $\deg(\tilde{\sigma}_p(x)) > \deg(\sigma_p(x))$, and hence $\varepsilon \neq 0$.

To show (b), we consider two cases. In the case $\varepsilon = 0$, we clearly have $\mathbb{I}_p = \mathbb{I}_{\tilde{p}}$ and thus $p = \tilde{p}$. It remains to examine the case $\varepsilon \neq 0$. For a contradiction, assume that $\#(\mathbb{I}_p \cap \mathbb{I}_{\tilde{p}}) \geq 2$. Let $j_1, j_2 \in \mathbb{I}_p \cap \mathbb{I}_{\tilde{p}}$ with $j_1 \neq j_2$. Then $\varepsilon x^d = \tilde{\sigma}_p(x) - \sigma_p(x)$ has the distinct zeros α_{j_1} and α_{j_2} . This is a contradiction to the fact that εx^d has no two distinct zeros. Consequently, we get $\#(\mathbb{I}_p \cap \mathbb{I}_{\tilde{p}}) \leq 1$.

Finally, we prove (c). Let $\varepsilon \neq 0$. To show the implication “ \Rightarrow ”, we note that α_i is a zero of $\tilde{\sigma}_p(x)$ and of $\sigma_p(x)$. Hence α_i is a zero of $\tilde{\sigma}_p(x) - \sigma_p(x) = \varepsilon x^d$. As $\varepsilon \neq 0$, this implies $\alpha_i = 0$ and $d > 0$.

Now we show the reverse implication “ \Leftarrow ”. Since $\alpha_i = 0$ is a zero of $\sigma_p(x)$, we have $x \mid \sigma_p(x)$, and thus $x \mid \varepsilon x^d + \sigma_p(x) = \tilde{\sigma}_p(x)$. Consequently, α_i is a zero of $\tilde{\sigma}_p(x)$, and we get $i \in \mathbb{I}_{\tilde{p}}$. Thus we have $i \in \mathbb{I}_p \cap \mathbb{I}_{\tilde{p}}$, and (b) says that either $p = \tilde{p}$ or $\mathbb{I}_p \cap \mathbb{I}_{\tilde{p}} = \{i\}$. In the second case, we are already done. So, assume that $p = \tilde{p}$. Then we deduce that, for all $j \in \mathbb{I}_p = \mathbb{I}_{\tilde{p}}$, the element α_j is a zero of both $\sigma_p(x)$ and $\tilde{\sigma}_p(x)$. Hence α_j is a zero of $\tilde{\sigma}_p(x) - \sigma_p(x) = \varepsilon x^d$ for all $j \in \mathbb{I}_p = \mathbb{I}_{\tilde{p}}$. Thus we have $\alpha_j = 0$ for all $j \in \mathbb{I}_p$, and since the elements $\alpha_1, \dots, \alpha_n$ are pairwise distinct, we must have $\#(\mathbb{I}_p \cap \mathbb{I}_{\tilde{p}}) = 1$, as claimed. \square

From a BIG-N fault injection one can derive polynomial equations in the unknown support α , as the next remark explains.

Remark 4.8. Let (p, d, \tilde{p}) be a BIG-N fault injection using the fault $\varepsilon \in \mathbb{F}_{2^m}$, and assume that $\text{wt}(\tilde{p}) \geq 2$. Choosing $i, j \in \mathbb{I}_{\tilde{p}}$ with $i \neq j$, we obtain

$$0 = \varepsilon \alpha_i^d \alpha_j^d + \prod_{k \in \mathbb{I}_p} (\alpha_i - \alpha_k) \alpha_j^d = \varepsilon \alpha_i^d \alpha_j^d + \prod_{k \in \mathbb{I}_p} (\alpha_j - \alpha_k) \alpha_i^d$$

Therefore the tuple $(\alpha_1, \dots, \alpha_n)$ is a zero of the polynomial

$$x_i^d \prod_{k \in \mathbb{I}_p} (x_j - x_k) - x_j^d \prod_{k \in \mathbb{I}_p} (x_i - x_k)$$

in $\mathbb{F}_{2^m}[x_1, \dots, x_n]$. Of course, using multiple BIG-N fault injections, we can generate a polynomial system which has $(\alpha_1, \dots, \alpha_n)$ as one of its \mathbb{F}_{2^m} -rational solutions.

Notice that each polynomial has degree $d + \text{wt}(p)$ and involves either $\text{wt}(p) + 1$ or $\text{wt}(p) + 2$ indeterminates. Moreover, from one fault injection we obtain $\binom{\text{wt}(\tilde{p})}{2}$ polynomials, and we have $\text{wt}(\tilde{p}) \leq \deg(\tilde{\sigma}_p) \leq \max(\text{wt}(p), d)$. Therefore we should choose both d and $\text{wt}(p)$ small, so that the polynomial system contains only small degree polynomials in relatively few indeterminates. But note that then only a few equations can be obtained from each injection, and hence we have to perform a large number of fault injections in order to obtain a polynomial system that involves all indeterminates.

In the next section we present two specific BIG-N fault injection classes which allow us to obtain equations of degree even lower than $d + \text{wt}(p)$. In particular, we will see how to generate linear and quadratic equations in merely a few indeterminates.

5. CONSTANT AND QUADRATIC FAULT INJECTION SEQUENCES

In this section we construct algorithms which repeatedly perform BIG-N fault injections until we obtain a linear or quadratic polynomial satisfied by the support tuple $(\alpha_1, \dots, \alpha_n)$. We continue to use the setting of the preceding section: let $(K_{\text{sec}}, K_{\text{pub}}, \mathcal{C}, \mathcal{P}, \text{encr}, \text{decr})$ be a BIG-N cryptosystem, and assume that the decryption map decr satisfies Assumptions 4.2, 4.3, and 4.4. For $i = 1, \dots, n$, let $e^{(i)}$ denote the i -th standard basis vector of \mathbb{F}_2^n . Subsequently, we are mainly interested in the following types of fault injections.

Definition 5.1. Let (p, d, \tilde{p}) be a BIG-N fault injection.

- (a) The fault injection (p, d, \tilde{p}) is called a **constant injection** if $d = 0$ and we have $p = e^{(i_1)} + e^{(i_2)} \in \mathbb{F}_2^n$ for some $i_1, i_2 \in \{1, \dots, n\}$ such that $i_1 \neq i_2$.
- (d) The fault injection (p, d, \tilde{p}) is called a **quadratic injection** if $d = 2$ and we have $p = e^{(i)} \in \mathbb{F}_2^n$ for some $i \in \{1, \dots, n\}$.
- (c) A constant or quadratic injection (p, d, \tilde{p}) is called **successful** if we have $\text{wt}(\tilde{p}) = 2$.

The term *successful* is adequately chosen, as the following proposition shows.

Proposition 5.2. Let (p, d, \tilde{p}) be a BIG-N fault injection.

- (a) Suppose that a constant injection with $p = e^{(i_1)} + e^{(i_2)}$, where $i_1, i_2 \in \{1, \dots, n\}$ and $i_1 \neq i_2$, is successful. Let $\mathbb{I}_{\tilde{p}} = \{j_1, j_2\}$. Then we have

$$\alpha_{i_1} + \alpha_{i_2} = \alpha_{j_1} + \alpha_{j_2}$$

- (b) Suppose that a quadratic fault injection with $p = e^{(i)}$ for some $i \in \{1, \dots, n\}$ is successful. Let $\mathbb{I}_{\tilde{p}} = \{j_1, j_2\}$. Then we have

$$\alpha_i \alpha_{j_1} + \alpha_i \alpha_{j_2} + \alpha_{j_1} \alpha_{j_2} = 0$$

If, additionally, $\alpha_i \neq 0$, then we also have $\alpha_{j_1} \neq 0$ and $\alpha_{j_2} \neq 0$.

Proof. First we show (a). Let $\varepsilon \in \mathbb{F}_{2^m}$ be the fault that is used by the fault injection (p, d, \tilde{p}) . By definition of the error-locator polynomial, we have $\sigma_p(x) = (x - \alpha_{i_1})(x - \alpha_{i_2})$. Then $\tilde{\sigma}_p(x) = \sigma_p(x) + \varepsilon$, and for $i \in \{1, \dots, n\}$ we have $\tilde{p}_i = 1$ if and only if $\tilde{\sigma}_p(\alpha_i) = 0$. Using $j_1, j_2 \in \mathbb{I}_{\tilde{p}}$, we get $\tilde{\sigma}_p(\alpha_{j_1}) = \tilde{\sigma}_p(\alpha_{j_2}) = 0$. Then $\deg(\tilde{\sigma}_p) = 2$ and $j_1 \neq j_2$ yield $\tilde{\sigma}_p(x) = (x - \alpha_{j_1})(x - \alpha_{j_2})$. Thus $\tilde{\sigma}_p(x) = \sigma_p(x) + \varepsilon$ implies

$$x^2 + (\alpha_{i_1} + \alpha_{i_2})x + (\alpha_{i_1}\alpha_{i_2} + \varepsilon) = x^2 + (\alpha_{j_1} + \alpha_{j_2})x + \alpha_{j_1}\alpha_{j_2}$$

Comparing coefficients yields $\alpha_{i_1} + \alpha_{i_2} = \alpha_{j_1} + \alpha_{j_2}$, as claimed.

Next we prove (b). Let $i \in \{1, \dots, n\}$ and $p = e^{(i)}$. By definition, the tuple $(p, 2, \tilde{p})$ is a BIG-N fault injection. Let $\varepsilon \in \mathbb{F}_{2^m}$ be the fault used by this fault injection. By definition of the error-locator polynomial, we have $\sigma_p(x) = x - \alpha_i$. Hence we have $\tilde{\sigma}_p(x) = \varepsilon x^2 + x - \alpha_i$. Notice that, for $i \in \{1, \dots, n\}$, we have $\tilde{p}_i = 1$ if and only if $\tilde{\sigma}_p(\alpha_i) = 0$. Since $j_1, j_2 \in \mathbb{I}_{\tilde{p}}$, we get $\tilde{\sigma}_p(\alpha_{j_1}) = \tilde{\sigma}_p(\alpha_{j_2}) = 0$, and therefore $j_1 \neq j_2$ implies $\varepsilon x^2 + x - \alpha_i = \varepsilon(x - \alpha_{j_1})(x - \alpha_{j_2})$. Comparing coefficients yields $1 = -\varepsilon(\alpha_{j_1} + \alpha_{j_2})$ and $-\alpha_i = \varepsilon\alpha_{j_1}\alpha_{j_2}$. By multiplying the second equation with $-(\alpha_{j_1} + \alpha_{j_2})$, we get

$$(\alpha_{j_1} + \alpha_{j_2}) \cdot \alpha_i = (-\varepsilon(\alpha_{j_1} + \alpha_{j_2}))\alpha_{j_1}\alpha_{j_2} = -\alpha_{j_1}\alpha_{j_2}$$

and the first claim of (b) follows.

To show the second claim, let $\alpha_i \neq 0$. We want to prove that $\alpha_{j_1} \neq 0$ and $\alpha_{j_2} \neq 0$. For a contradiction, assume that $\alpha_{j_1} = 0$ or $\alpha_{j_2} = 0$. In both cases we have $\alpha_{j_1}\alpha_{j_2} = 0$, and therefore the above equation yields $-\alpha_i = \varepsilon\alpha_{j_1}\alpha_{j_2} = 0$, in contradiction to the fact that $\alpha_i \neq 0$. Hence we have $\alpha_{j_1} \neq 0$ and $\alpha_{j_2} \neq 0$. \square

Part (a) of this proposition can now be exploited for a fault injection sequence algorithm which finds a linear equation for $(\alpha_1, \dots, \alpha_n)$, if it terminates.

Algorithm 5.3. (A Constant Fault Injection Sequence)

Given a BIG-N cryptosystem $(K_{\text{sec}}, K_{\text{pub}}, \mathcal{C}, \mathcal{P}, \text{encr}, \text{decr})$ as above and two distinct indices $i_1, i_2 \in \{1, \dots, n\}$, consider the following instructions.

- (1) Perform a BIG-N fault injection for the word $p = e^{(i_1)} + e^{(i_2)}$ in degree zero, and let $\tilde{p} \in \mathbb{F}_2^n$ be its output.
- (2) If $\text{wt}(\tilde{p}) = 2$ and $\tilde{p} \neq p$, then write $\mathbb{I}_{\tilde{p}} = \{j_1, j_2\}$, return the polynomial

$$x_{i_1} + x_{i_2} + x_{j_1} + x_{j_2} \in \mathbb{F}_{2^m}[x_1, \dots, x_n]$$

and stop. Otherwise, continue with (1).

This is a Las Vegas algorithm, i.e., it may not terminate, but if it does terminate, then it returns a linear polynomial $f \in \mathbb{F}_{2^m}[x_1, \dots, x_n]$ such that $f(\alpha_1, \dots, \alpha_n) = 0$.

Proof. To prove correctness, it suffices to note that, if the algorithm stops in step (2), Proposition 5.2.a can be applied and yields $\alpha_{i_1} + \alpha_{i_2} = \alpha_{j_1} + \alpha_{j_2}$. \square

Naturally, the question arises how many faults have to be injected on average until this algorithm stops. It turns out that in the case $n = 2^m$ the precise number is given by $\frac{2^m}{2^{m-1}-1} \approx 2$. For the general case $n \leq 2^m$, the probability of a successful constant fault injection is estimated in Table 2 for a selection of parameters (see Section 7).

Similarly, also Proposition 5.2.b can be used via repeated fault injections to gain a quadratic equation satisfied by $(\alpha_1, \dots, \alpha_n)$. Moreover, it allows us to check whether $\alpha_i = 0$ for a given $i \in \{1, \dots, n\}$.

Algorithm 5.4. (A Quadratic Fault Injection Sequence)

Given a BIG-N cryptosystem $(K_{\text{sec}}, K_{\text{pub}}, \mathcal{C}, \mathcal{P}, \text{encr}, \text{decr})$ as above and $i \in \{1, \dots, n\}$, consider the following sequence of instructions.

- (1) Perform a BIG-N fault injection for the word $p = e^{(i)}$ in degree 2, and let $\tilde{p} \in \mathbb{F}_2^n$ be its output.
- (2) If $\text{wt}(\tilde{p}) > 1$ and $i \in \mathbb{I}_{\tilde{p}}$ then return $x_i \in \mathbb{F}_{2^m}[x_1, \dots, x_n]$ and stop.

(3) If $\text{wt}(\tilde{p}) = 2$ then write $\mathbb{I}_{\tilde{p}} = \{j_1, j_2\}$, return the polynomial

$$x_i x_{j_1} + x_i x_{j_2} + x_{j_1} x_{j_2} \in \mathbb{F}_{2^m}[x_1, \dots, x_n]$$

and stop. Otherwise, continue with (1).

This is a Las-Vegas algorithm. If it terminates, it returns a linear or quadratic polynomial $f \in \mathbb{F}_{2^m}[x_1, \dots, x_n]$ such that $f(\alpha_1, \dots, \alpha_n) = 0$. Moreover, if the algorithm stops in step (3) then we have $\alpha_i, \alpha_{j_1}, \alpha_{j_2} \in \mathbb{F}_{2^m} \setminus \{0\}$.

Proof. To show correctness, we distinguish two cases. If the algorithm terminates in step (2), it suffices to prove $\alpha_i = 0$. Let $(e^{(i)}, 2, \tilde{p})$ be the quadratic fault injection of step (1), and let $\varepsilon \in \mathbb{F}_{2^m}$ be the fault that it uses. Since $\text{wt}(\tilde{p}) > 1$, we have $\text{wt}(e^{(i)}) = 1 < \text{wt}(\tilde{p})$, and hence Proposition 4.7.a yields $\varepsilon \neq 0$. Now $i \in \mathbb{I}_{\tilde{p}}$ and Proposition 4.7.c imply $\alpha_i = 0$, as claimed. This also proves that, if the algorithm terminates in step (3), we must have $\alpha_i \neq 0$.

Next, assume that the algorithm terminates in step (3). We just saw that this forces α_i to be non-zero. Let $(e^{(i)}, 2, \tilde{p})$ be the quadratic fault injection of step (1). Since $\text{wt}(\tilde{p}) = 2$, we write $\mathbb{I}_{\tilde{p}} = \{j_1, j_2\}$ and note that Proposition 5.2.b yields $\alpha_i \alpha_{j_1} + \alpha_i \alpha_{j_2} + \alpha_{j_1} \alpha_{j_2} = 0$. Thus $(\alpha_1, \dots, \alpha_n)$ is a zero of the given polynomial f . Moreover, as $\alpha_i \neq 0$, we also get $\alpha_{j_1} \neq 0$ and $\alpha_{j_2} \neq 0$ by the same proposition. \square

Again, it is not clear how many fault injections are typically required for one execution of the algorithm. In the case $n = 2^m$, it turns out that on average $\frac{2^m}{2^m-1-1} \approx 2$ faults need to be injected if $\alpha_i \neq 0$, and otherwise $\frac{2^m}{2^m-1} \approx 1$ faults are required. For the general setting $n \leq 2^m$, estimates for this number can be found in Table 2 (see Section 7).

6. THE BIG-N FAULT ATTACK

In this section we first derive some simplifications of the systems of polynomial equations we obtain by performing constant and quadratic fault injection sequences (see Algorithms 5.3 and 5.4). After that we present a strategy to determine all solutions of such a system. Finally, we explain how one can check whether they can be extended to an alternative secret pair. To begin with, consider the following property of Goppa codes.

Proposition 6.1. *Let $\alpha \in \mathbb{F}_{2^m}^n$ be a support tuple, let $g \in \mathbb{F}_{2^m}[x]$ be a Goppa polynomial for α , and let $a \in \mathbb{F}_{2^m} \setminus \{0\}$. Then we have*

$$\Gamma(\alpha, g(x)) = \Gamma(a \cdot \alpha, g(a^{-1} \cdot x))$$

Proof. For $c \in \Gamma(\alpha, g)$, we let $\eta_{c,\alpha}(x) = \sum_{i \in \mathbb{I}_c} \prod_{j \in \mathbb{I}_c \setminus \{i\}} (x - \alpha_j)$. By Remark 2.2, we have $g \mid \eta_{c,\alpha}$. By applying the ring homomorphism $\Psi_a : \mathbb{F}_{2^m}[x] \rightarrow \mathbb{F}_{2^m}[x]$ defined by $\Psi_a(x) = a^{-1} \cdot x$, we get $\Psi_a(g(x)) \mid \Psi_a(\eta_{c,\alpha}(x))$, where

$$\Psi_a(\eta_{c,\alpha}(x)) = \eta_{c,\alpha}(a^{-1} \cdot x) = a^{-(\text{wt}(c)-1)} \eta_{c,a \cdot \alpha}(x)$$

Thus $\Psi_a(g) \mid \eta_{c,a \cdot \alpha}$, and using Remark 2.2 we see that $c \in \Gamma(a \cdot \alpha, \Psi_a(g))$. Therefore we have $\Gamma(\alpha, g) \subseteq \Gamma(a \cdot \alpha, g(a^{-1} \cdot x))$.

Conversely, we apply this inclusion to the Goppa code $\Gamma(a\alpha, g(a^{-1}x))$ with the factor a^{-1} and obtain $\Gamma(a \cdot \alpha, g(a^{-1} \cdot x)) \subseteq \Gamma(a^{-1} \cdot a \cdot \alpha, g(a \cdot a^{-1} \cdot x)) = \Gamma(\alpha, g)$. This finishes the proof. \square

In the following we fix a BIG-N cryptosystem $(K_{\text{sec}}, K_{\text{pub}}, \mathcal{C}, \mathcal{P}, \text{encr}, \text{decr})$ and use the notation introduced in the preceding sections.

Remark 6.2. Let $L \subseteq \mathbb{F}_{2^m}[x_1, \dots, x_n]$ be a set of polynomials obtained by constant and quadratic fault injection sequences. By

$$\mathcal{Z}_{\mathbb{F}_{2^m}}(L) = \{(a_1, \dots, a_n) \in \mathbb{F}_{2^m}^n \mid f(a_1, \dots, a_n) = 0 \text{ for all } f \in L\}$$

we denote the zero set of L in \mathbb{F}_{2^m} . Now we consider the set

$$S_L = \{(a_1, \dots, a_n) \in \mathcal{Z}_{\mathbb{F}_{2^m}}(L) \mid a_i \neq a_j \text{ for } i \neq j\}$$

Since every $f \in L$ results from Algorithms 5.3 or 5.4, we have $f(\alpha_1, \dots, \alpha_n) = 0$, and since $\alpha = (\alpha_1, \dots, \alpha_n)$ is a support tuple, we even get $\alpha \in S_L$.

Notice that the polynomials in L are homogeneous. Hence the set L generates a homogeneous ideal in $\mathbb{F}_{2^m}[x_1, \dots, x_n]$. Consequently, for $(a_1, \dots, a_n) \in S_L$ and $b \in \mathbb{F}_{2^m} \setminus \{0\}$ we also have $(ba_1, \dots, ba_n) \in S_L$. In view of Proposition 6.1 and the fact that $g(x)$ is irreducible if and only if $g(b^{-1} \cdot x)$ is irreducible, we can therefore assume without loss of generality that one non-zero support element α_i with $i \in \{1, \dots, n\}$ is chosen arbitrarily to be 1.

This means that we may choose an index $i \in \{1, \dots, n\}$ for which we know that $\alpha_i \neq 0$. Then we consider the dehomogenization of L with respect to the indeterminate x_i , i.e., we add the polynomial $x_i - 1$ to L . To determine such an index we can use a quadratic fault injection sequence. Clearly, it is best to choose the indeterminate that occurs most frequently in the polynomials of L , in order to simplify the fault equations system as much as possible.

In view of this remark, we are led to the following definitions.

Definition 6.3. Let $C = \Gamma(\alpha, g)$ be a binary irreducible Goppa code, and let $(K_{\text{sec}}, K_{\text{pub}}, \mathcal{P}, \mathcal{C}, \text{encr}, \text{decr})$ be a BIG-N cryptosystem which uses C .

(a) Let $L_1 \subseteq \mathbb{F}_{2^m}[x_1, \dots, x_n]$ be a set of linear polynomials obtained from constant fault injection sequences (see Algorithm 5.3). Let $L_2 \subseteq \mathbb{F}_{2^m}[x_1, \dots, x_n]$ be a set of linear and quadratic polynomials obtained from quadratic fault injection sequences (see Algorithm 5.4). Let $i \in \{1, \dots, n\}$ be such that the indeterminate x_i occurs in some quadratic polynomial of L_2 . Then the set $L_1 \cup L_2 \cup \{x_i - 1\}$ is called a **fault equation system**.

(b) Given a set of polynomials $L \subseteq \mathbb{F}_{2^m}[x_1, \dots, x_n]$, we call

$$S_L = \{(a_1, \dots, a_n) \in \mathcal{Z}_{\mathbb{F}_{2^m}}(L) \mid a_i \neq a_j \text{ for } i \neq j\} \subseteq \mathbb{F}_{2^m}^n$$

the **support candidate set** of L , and every element $(a_1, \dots, a_n) \in S_L$ is called a **support candidate** of L .

Given a fault equation system L , Remark 6.2 shows that we may assume $\alpha \in S_L$. Hence the task of finding α is reduced to solving a suitable fault equation system. In view of Algorithm 3.3, we may reduce this task further.

Definition 6.4. Let $C = \Gamma(\alpha, g)$ be a binary Goppa code, let $\tilde{\alpha} \in \mathbb{F}_{2^m}^n$ be a support tuple, and let $u \in \mathbb{N}_+$.

(a) A Goppa polynomial $\tilde{g} \in \mathbb{F}_{2^m}[x]$ for $\tilde{\alpha}$ with $C \subseteq \Gamma(\tilde{\alpha}, \tilde{g})$ and $\deg(\tilde{g}) = u$ is called a **degree- u extension** of $\tilde{\alpha}$ with respect to C .

(b) A support tuple $\tilde{\alpha}$ is called **degree- u extendable** to C if there exists a degree- u extension of $\tilde{\alpha}$ with respect to C .

If the code C is clear from the context, it may also be omitted.

Remark 6.5. As explained at the end of Section 3, one can apply Algorithm 3.3 to break a BIG-N cryptosystem which uses a binary Goppa code $C = \Gamma(\alpha, g)$ with $\deg(g) = t$ as follows: Find a support tuple $\tilde{\alpha} \in \mathbb{F}_{2^m}^n$ and a Goppa polynomial $\tilde{g} \in \mathbb{F}_{2^m}[x]$ with $\deg(\tilde{g}) = u \geq 2t$ such that \tilde{g} is a degree- u extension of $\tilde{\alpha}$ with respect to C . This is the reason why $(\tilde{\alpha}, \tilde{g})$ is called an alternative secret pair.

In order to find such a support tuple $\tilde{\alpha} \in \mathbb{F}_{2^m}^n$, we generate a fault equation system $L \subseteq \mathbb{F}_{2^m}[x_1, \dots, x_n]$ using constant and quadratic fault injection sequences. Note that $C = \Gamma(\alpha, g^2)$ and $\alpha \in S_L$ imply that at least one degree- $2t$ extendable support candidate can be found in S_L .

Based on these observations, we now perform a BIG-N fault attack which is based on the following three steps: Firstly, we compute the support candidate set S_L of a fault equation system L . Secondly, we determine a degree- u extendable support candidate $\tilde{\alpha} \in S_L$ with $u \geq 2t$ and its corresponding extension. Finally, we combine everything and compute an alternative secret pair, thereby breaking the system.

6.1. Finding Support Candidates. Let $L \subseteq \mathbb{F}_{2^m}[x_1, \dots, x_n]$ be a fault equation system. Then the problem of computing the support candidate set S_L can be reduced to finding the zero set $\mathcal{Z}_{\mathbb{F}_{2^m}}(L)$. Computing this zero set is a classic problem of computer algebra which can be solved, for instance, using Gröbner basis techniques (see [13, Sec. 3.7] or [14, Sec. 6.3]). However, in order to improve the efficiency of these methods, it is important that we first use the linear equations in L to eliminate some indeterminates and reduce the complexity of the quadratic equations. The following algorithm aids this task.

Algorithm 6.6. (Solving a Fault Equation System)

Let \mathcal{N} be a BIG-N cryptosystem, and let $L \subseteq \mathbb{F}_{2^m}[x_1, \dots, x_n]$ be a fault equation system obtained by applying constant and quadratic fault injection sequences to \mathcal{N} . Moreover, let σ be a term ordering. Consider the following sequence of instructions.

- (1) Let $L^{\text{lin}} = \{f \in L \mid \deg(f) = 1\}$.
- (2) Interreduce the set L^{lin} linearly and get a set $L^{\text{irlin}} = \{\ell_1, \dots, \ell_r\}$ such that, for $i = 1, \dots, r$, the indeterminate $\text{LT}_{\sigma}(\ell_i)$ does not occur in the other polynomials of L^{irlin} . Renumber the indeterminates such that $\text{LT}_{\sigma}(\ell_i) = x_i$ for $i \in \{1, \dots, r\}$.
- (3) Define a ring homomorphism $\Psi : \mathbb{F}_{2^m}[x_1, \dots, x_n] \rightarrow \mathbb{F}_{2^m}[x_{r+1}, \dots, x_n]$ by $\Psi(x_i) = \ell_i + x_i$ for $i \in \{1, \dots, r\}$, and by $\Psi(x_i) = x_i$ for $i \in \{r+1, \dots, n\}$.
- (4) Let $L^{\text{red}} = \Psi(L) \setminus \{0\}$. Compute $S^{\text{red}} = \mathcal{Z}_{\mathbb{F}_{2^m}}(L^{\text{red}})$ using Gröbner basis techniques.
- (5) Define a map $\psi : \mathbb{F}_{2^m}^{n-r} \rightarrow \mathbb{F}_{2^m}^n$ by $\psi(\gamma) = (\Psi(x_1)(\gamma), \dots, \Psi(x_n)(\gamma))$.
- (6) Return $S = \{(\tilde{\alpha}_1, \dots, \tilde{\alpha}_n) \in \psi(S^{\text{red}}) \mid \tilde{\alpha}_i \neq \tilde{\alpha}_j \text{ for } i \neq j\}$.

This is an algorithm which computes the support candidate set S_L of L .

Proof. Since finiteness is clear, we have to show that the set S returned in step (6) is indeed equal to S_L .

First we show the inclusion $S \subseteq S_L$. Let $\gamma \in S^{\text{red}}$ with $\psi(\gamma) \in S$. By construction, we have $\Psi(f) \in L^{\text{red}}$ for all $f \in L$. Hence we get $0 = \Psi(f)(\gamma) = f(\Psi(x_1)(\gamma), \dots, \Psi(x_n)(\gamma)) = f(\psi(\gamma))$. Since this equality holds for all $f \in L$, we conclude that $\psi(\gamma) \in \mathcal{Z}_{\mathbb{F}_{2^m}}(L)$. Therefore we have $\psi(S^{\text{red}}) \subseteq \mathcal{Z}_{\mathbb{F}_{2^m}}(L)$, and by the construction of S also $S \subseteq S_L$.

Conversely, let $\tilde{\alpha} = (\tilde{\alpha}_1, \dots, \tilde{\alpha}_n) \in S_L$. To prove $\tilde{\alpha} \in S$, it suffices to show that $\tilde{\alpha} = \psi(\gamma)$ and $\gamma \in S^{\text{red}}$ for $\gamma = (\tilde{\alpha}_{r+1}, \dots, \tilde{\alpha}_n)$. By the definition of Ψ and the fact

that $\ell_i(\tilde{\alpha}) = 0$ for $i \in \{1, \dots, r\}$, we have $\Psi(x_i)(\gamma) = \tilde{\alpha}_i$ for $i \in \{1, \dots, n\}$. This yields $\psi(\gamma) = (\Psi(x_1)(\gamma), \dots, \Psi(x_n)(\gamma)) = (\tilde{\alpha}_1, \dots, \tilde{\alpha}_n) = \tilde{\alpha}$.

Thus it remains to show that $\gamma \in S^{\text{red}} = \mathcal{Z}_{\mathbb{F}_{2^m}}(L^{\text{red}})$. Note that, by construction, we have $L^{\text{red}} = \{\Psi(f) \mid f \in L\}$. Using $\Psi(x_i)(\gamma) = \tilde{\alpha}_i$ for $i \in \{1, \dots, n\}$ and $\tilde{\alpha} \in S_L$, we have

$$\Psi(f)(\gamma) = f(\Psi(x_1)(\gamma), \dots, \Psi(x_n)(\gamma)) = f(\tilde{\alpha}_1, \dots, \tilde{\alpha}_n) = f(\tilde{\alpha}) = 0$$

for all $\Psi(f) \in L^{\text{red}}$. This proves $\gamma \in \mathcal{Z}_{\mathbb{F}_{2^m}}(L^{\text{red}})$, and hence $\tilde{\alpha} \in S$. \square

6.2. Finding Suitable Goppa Polynomials. Recall that the generating pair (α, g) of the irreducible Goppa code $C = \Gamma(\alpha, g)$ is part of the secret key of any BIG-N cryptosystem using C . The code C itself is published via the public parity check matrix H_{pub} . Let $L \subseteq \mathbb{F}_{2^m}[x_1, \dots, x_n]$ be a fault equation system, let S_L be its support candidate set, let $\tilde{\alpha} \in S_L$ be a support candidate which is degree- u extendable with respect to C for some $u \geq 2t$, and let $\tilde{g} \in \mathbb{F}_{2^m}[x]$ be a degree- u extension of $\tilde{\alpha}$. This means that $C \subseteq \Gamma(\tilde{\alpha}, \tilde{g})$. By Remark 2.2, it follows for all $c \in \Gamma(\tilde{\alpha}, \tilde{g})$ that we have

$$\tilde{g} \mid \sum_{i \in \mathbb{I}_c} \prod_{j \in \mathbb{I}_c \setminus \{i\}} (x - \tilde{\alpha}_j)$$

in $\mathbb{F}_{2^m}[x]$. In particular, this divisibility then holds for all $c \in C$. Hence, knowing only C and $\tilde{\alpha}$, we can compute multiples of the desired polynomial \tilde{g} . The following algorithm uses these observations and computes an extension \tilde{g} of a given support tuple $\tilde{\alpha}$ if and only if $\tilde{\alpha}$ is extendable. Note that its core idea is based on [19, p. 125].

Algorithm 6.7 (GoppaGCD).

Input: A support tuple $\tilde{\alpha} \in \mathbb{F}_{2^m}^n$, $t \in \mathbb{N}_+$, and a binary Goppa code $C \subseteq \mathbb{F}_2^n$.

Output: Fail, or a degree- $2t$ extension of $\tilde{\alpha}$ with respect to C .

- 1: Let $\tilde{g} := 0$, and let B be an \mathbb{F}_2 -basis of C .
- 2: **for** $c \in B$ **do**
- 3: $\eta_{c, \tilde{\alpha}} := \sum_{i \in \mathbb{I}_c} \prod_{j \in \mathbb{I}_c \setminus \{i\}} (x - \tilde{\alpha}_j)$
- 4: $\tilde{g} := \text{gcd}(\tilde{g}, \eta_{c, \tilde{\alpha}})$
- 5: **if** $\deg(\tilde{g}) < 2t$ **then**
- 6: **return** Fail
- 7: **end if**
- 8: **end for**
- 9: **for** $i = 1, \dots, n$ **do**
- 10: **while** $\tilde{g}(\tilde{\alpha}_i) = 0$ **do**
- 11: $\tilde{g} := \frac{\tilde{g}}{x - \tilde{\alpha}_i}$
- 12: **end while**
- 13: **end for**
- 14: **if** $\deg(\tilde{g}) < 2t$ **then**
- 15: **return** Fail
- 16: **else**
- 17: **return** \tilde{g}
- 18: **end if**

Proposition 6.8. *Let $\tilde{\alpha} \in \mathbb{F}_{2^m}^n$ be a support tuple, let $t \in \mathbb{N}_+$, and let $C \subseteq \mathbb{F}_2^n$ be a binary Goppa code. Then Algorithm 6.7 is finite and the following conditions are equivalent.*

- (a) The function $\text{GoppaGCD}(\tilde{\alpha}, t, C)$ returns a degree- u extension of $\tilde{\alpha}$ with respect to C for some $u \geq 2t$.
- (b) The tuple $\tilde{\alpha}$ is degree- u extendable with respect to C for some $u \geq 2t$.

Otherwise, the algorithm returns **Fail**.

Proof. Since finiteness is clear, it suffices to prove the equivalence of (a) and (b) and the additional claim. To begin with, we show that the output of the function $\text{GoppaGCD}(\tilde{\alpha}, t, C)$ is either **Fail** or a degree- u extension of $\tilde{\alpha}$ for some $u \geq 2t$.

Clearly, the algorithm ends either in step (6), (15), or (17). Hence it terminates with **Fail** or a polynomial $\tilde{g} \in \mathbb{F}_{2^m}[x]$. Suppose that it ends in step (17) with a polynomial \tilde{g} . Notice that this entails $\deg(\tilde{g}) \geq 2t$, as otherwise the algorithm would have terminated at the latest in step (15).

Thus it remains to prove $C \subseteq \Gamma(\tilde{\alpha}, \tilde{g})$. By the loop in steps (9)-(13), we have $\tilde{g}(\tilde{\alpha}_i) \neq 0$ for all $i \in \{1, \dots, n\}$. Therefore \tilde{g} is a Goppa polynomial for the support tuple $\tilde{\alpha}$ and $\Gamma(\tilde{\alpha}, \tilde{g})$ is a binary Goppa code. Let $B = \{c_1, \dots, c_k\}$ be the \mathbb{F}_2 -basis of C chosen in step (1). By the construction of \tilde{g} in steps (2)-(8), we have $\tilde{g} \mid \eta_{c_i, \tilde{\alpha}}$ for all $i \in \{1, \dots, k\}$. Since in steps (9)-(13) only factors of \tilde{g} are removed, these divisibilities still hold true in step (17). Therefore the polynomial \tilde{g} returned in step (17) satisfies $\tilde{g} \mid \eta_{c_i, \tilde{\alpha}}$ for all $i \in \{1, \dots, k\}$. By Remark 2.2, this implies that $c_i \in \Gamma(\tilde{\alpha}, \tilde{g})$ for all $i \in \{1, \dots, k\}$. Using the fact that B is an \mathbb{F}_2 -basis of C , we deduce $C \subseteq \Gamma(\tilde{\alpha}, \tilde{g})$.

Since the implication (a) \Rightarrow (b) is trivially true, it remains to prove that (b) implies (a). Let $\hat{g} \in \mathbb{F}_{2^m}[x]$ be a degree- u extension of $\tilde{\alpha}$, where $u \geq 2t$. Let $B = \{c_1, \dots, c_k\}$ be the \mathbb{F}_2 -basis of C chosen in step (1). From the hypothesis and Remark 2.2, we get that $\hat{g} \mid \eta_{c_i, \tilde{\alpha}}$ for $i \in \{1, \dots, k\}$. Since we compute \tilde{g} in steps (2)-(8) as the greatest common divisor of the polynomials $\eta_{c_i, \tilde{\alpha}}$ for $i \in \{1, \dots, k\}$, we have $\hat{g} \mid \tilde{g}$ in every iteration of the loop. In particular, this implies that we have $\deg(\tilde{g}) \geq \deg(\hat{g}) \geq 2t$ and thus the algorithm does not terminate in this loop. Now, in steps (9)-(13), all linear factors of the form $(x - \tilde{\alpha}_i)$ for $i \in \{1, \dots, n\}$ are removed, and thus we get $\tilde{g}(\tilde{\alpha}_i) \neq 0$ for all $i \in \{1, \dots, n\}$. As \hat{g} is a Goppa polynomial for $\tilde{\alpha}$, we also have $\hat{g}(\tilde{\alpha}_i) \neq 0$ for $i \in \{1, \dots, n\}$. Consequently, we have $\hat{g} \mid \tilde{g}$, and therefore $\deg(\tilde{g}) \geq \deg(\hat{g}) \geq 2t$. Finally, the algorithm terminates in step (17) by returning the polynomial $\tilde{g} \in \mathbb{F}_{2^m}[x]$. This finishes the proof. \square

Note that an \mathbb{F}_2 -basis B of C , as required in step (1), can be deduced from the rows of a generator matrix of C . Moreover, when the degree of the polynomial \tilde{g} is equal to $2t$ in the course of the loop (2)-(8), the polynomial will either stay the same for the remaining execution of the algorithm, or the result is **Fail**. Therefore we can also stop the algorithm at this point and use any other method to check if \tilde{g} is a Goppa polynomial for $\tilde{\alpha}$ and if we have $C \subseteq \Gamma(\tilde{\alpha}, \tilde{g})$. For instance, an efficient method is to compare a parity check matrix of $\Gamma(\tilde{\alpha}, \tilde{g})$ with H_{pub} .

6.3. Computing an Alternative Secret Pair. In this subsection we combine the algorithms of the previous two subsections and obtain the following fault attack algorithm which returns an alternative secret pair.

Algorithm 6.9. (BIG-N Fault Attack)

Let $m, t, n \in \mathbb{N}_+$ be such that $mt < n \leq 2^m$, let $C = \Gamma(\alpha, g)$ be a binary irreducible Goppa code with parameters (m, t, n) , and let \mathcal{N} be a BIG-N cryptosystem which uses C . Suppose that the implementation of the decryption map satisfies Assumptions 4.2, 4.3, and 4.4.

Let $L \subseteq \mathbb{F}_{2^m}[x_1, \dots, x_n]$ be a fault equation system which has been constructed using constant and quadratic fault injection sequences applied to \mathcal{N} .

- (1) Compute the support candidate set S_L of L using Algorithm 6.6.
- (2) Choose $\tilde{\alpha} \in S_L$ and remove it from S_L .
- (3) Use Algorithm 6.7 to compute $\text{GoppaGCD}(\tilde{\alpha}, t, C)$. If the output is `Fail`, then go to step (2). Otherwise, the output is a polynomial \tilde{g} . Return the pair $(\tilde{\alpha}, \tilde{g})$ and stop.

This is an algorithm which computes an alternative secret pair $(\tilde{\alpha}, \tilde{g})$.

Proof. First we prove finiteness. For the support candidate set S_L of step (1), we may assume $\alpha \in S_L$ by Remark 6.2. Since $S_L \subseteq \mathbb{F}_{2^m}^n$ is a finite set, after finitely many iterations of steps (2) and (3) the support tuple $\tilde{\alpha} = \alpha \in S_L$ is chosen. Since α is degree- $2t$ extendable using the extension g^2 , Proposition 6.8 shows that the call to $\text{GoppaGCD}(\tilde{\alpha}, t, C)$ in step (3) returns a polynomial $\tilde{g} \in \mathbb{F}_{2^m}[x]$. Hence the algorithm terminates.

It remains to prove that the output $(\tilde{\alpha}, \tilde{g})$ is an alternative secret pair. By Proposition 6.8, the algorithm terminates in step (3) if and only if $\tilde{\alpha} \in S_L$ is degree- u extendable for some $u \geq 2t$. In this case the polynomial \tilde{g} is a degree- u extension of $\tilde{\alpha}$ for some $u \geq 2t$, and Algorithm 3.3 implies that $(\tilde{\alpha}, \tilde{g})$ is an alternative secret pair. \square

7. EXPERIMENTS AND TIMINGS

In this section we apply the BIG-N fault attack (Algorithm 6.9) to a selection of state-of-the-art security levels. Table 1 contains a list of recommended parameter choices for the Goppa codes to be used in the BIG-N cryptosystem along with their claimed security.

Security Level		n	m	t
insec	(60bit)	1024	10	38
short I	(80bit)	2048	11	27
short II	(80bit)	1632	11	33
mid I	(128bit)	2960	12	56
mid II	(147bit)	3408	12	67
long I	(191bit)	4624	13	95
long II	(256bit)	6624	13	115
long III	(266bit)	6960	13	119

Table 1: Security parameters for BIG-N cryptosystems proposed in [4].

Recall that the only input of the BIG-N fault attack is a fault equation system. It can be computed, for instance, using the following algorithm. Our experiments show that fault equation systems obtained with this method can be solved efficiently with Algorithm 6.6.

Algorithm 7.1. (Computing a Fault Equation System)

Let $m, t, n \in \mathbb{N}^+$ be such that $mt < n \leq 2^m$, let $C = \Gamma(\alpha, g)$ be a binary irreducible Goppa code with parameters (m, t, n) , and let \mathcal{N} be a BIG-N cryptosystem which uses C . Suppose that the implementation of the decryption map satisfies Assumptions 4.2, 4.3, and 4.4. Consider the following sequence of instructions.

- (1) Let $I := \{(n, 1)\} \cup \{(i, i+1) \mid i \in \{1, \dots, n-1\}\} \subseteq \{1, \dots, n\}^2$.

- (2) For each pair $(i_1, i_2) \in I$, perform a constant fault injection sequence and collect the resulting linear polynomials in $L_1 \subseteq \mathbb{F}_{2^m}[x_1, \dots, x_n]$.
- (3) Let $I \subseteq \{1, \dots, n\}$ with $\#I = \lfloor \frac{n}{10} \rfloor$ be chosen uniformly at random.
- (4) For each index $i \in I$, perform a quadratic fault injection sequence and collect the resulting polynomials in $L_2 \subseteq \mathbb{F}_{2^m}[x_1, \dots, x_n]$.
- (5) Let x_i be an indeterminate which occurs most often in the terms of the polynomials of L_1 and which appears in $L_2 \setminus \{x_1, \dots, x_n\}$.
- (6) Return $L = L_1 \cup L_2 \cup \{x_i - 1\}$ and stop.

This is a Las-Vegas algorithm which performs n constant fault injection sequences and $\lfloor \frac{n}{10} \rfloor$ quadratic fault injection sequences. It returns a fault equation system $L \subseteq \mathbb{F}_{2^m}[x_1, \dots, x_n]$.

In order to determine the expected number of BIG-N fault injections needed during the computation of a fault equation system using this algorithm, we denote the probability of a successful constant fault injection by p_0 and the probability of a successful quadratic fault injection by p_2 . Then we expect that $\frac{n}{p_0} + \lfloor \frac{n}{10} \rfloor \frac{1}{p_2}$ faults have to be injected in total to generate the fault equation system using Algorithm 7.1.

Since there is no obvious formula to calculate these probabilities, they were estimated in the following way: For each security level, we considered three random irreducible Goppa codes. For each Goppa code, we chose uniformly at random 200 distinct words p for the fault injection sequences in step (2) and (4), respectively. For each word p , we computed the exact number of faults which lead to a successful fault injection. Dividing the average of those three numbers by 2^m yields our estimates \hat{p}_0 and \hat{p}_2 respectively.

Table 2 contains, both for the constant and the quadratic fault injection sequences, the average number of faults which lead to a successful fault injection and the estimated standard deviation. The estimates \hat{p}_0 and \hat{p}_2 for p_0 and p_2 are also given.

Sec Lvl	Parameters			succ const inj		\hat{p}_0	succ quad inj		\hat{p}_2
	n	m	t	avg	std dev	(%)	avg	std dev	(%)
insec	1024	10	38	511.0	0.0	49.9	511.0	0.0	49.9
short I	2048	11	27	1023.0	0.0	50.0	1023.0	0.0	50.0
short II	1632	11	33	649.5	5.3	31.7	649.5	5.2	31.7
mid I	2960	12	56	1067.9	9.3	26.1	1069.2	8.8	26.1
mid II	3408	12	67	1416.4	6.6	34.6	1417.0	6.5	34.6
long I	4624	13	95	1304.5	15.4	15.9	1303.7	15.3	15.9
long II	6624	13	115	2677.5	10.0	32.7	2676.8	9.9	32.7
long III	6960	13	119	2955.5	8.3	36.1	2955.8	8.8	36.2

Table 2: Average numbers of faults and success probabilities.

Observe that the success probabilities drop significantly when the ratio $\frac{n}{2^m}$ gets smaller. This can be attributed to the fact that a constant or quadratic fault injection is successful if and only if the *faulty* error-locator polynomial $\tilde{\sigma}_p(x)$ has two zeros among the support elements $\{\alpha_1, \dots, \alpha_n\}$, and it seems natural that there are more such faults when n is larger.

An implementation of a BIG-N cryptosystem following the FPGA-based implementation of [26, 27] is provided in the computer algebra system CoCoA-5 [1] along with all algorithms

of this paper. For the computation of the zero set of the *reduced* fault equation system L^{red} in step (4) of Algorithm 6.6, we make use the CoCoA-5 function `RationalSolve` which uses Gröbner basis computations.

Timings of the BIG-N fault attack (Algorithm 6.9) for all security parameters of Table 1 are given in Table 3. We also list the size of the reduced fault equation system L^{red} , as computed in step (4) of Algorithm 6.6, the time for computing its zero set $\mathcal{Z}_{\mathbb{F}_{2^m}}(L^{\text{red}})$ using the CoCoA-5 function `RationalSolve`, and the time for extending a support candidate to an alternative secret pair using Algorithm 6.7. The timings represent the average of three runs of the algorithm applied to distinct randomly generated BIG-N cryptosystems.

sec lvl	interreduced L^{red}		RatSol (s)	Alg 6.7 (s)	total (s)	exp no req fault inj
	ind	eq				
insec	10	46	4.5	2.6	13.1	2258.41
short I	10-11	55-56	28.4	4.1	19.0	4510.40
short II	11	56	15.7	4.2	19.0	5563.51
mid I	12	67	88.0	7.1	71.2	12474.52
mid II	11-12	66-67	119.9	6.3	101.7	10840.84
long I	13	79	304.2	13.1	240.9	31946.16
long II	13	79	866.1	13.1	438.2	22295.50
long III	13	79	938.4	13.2	481.0	21220.49

Table 3: Timings of the BIG-N fault attack (Algorithm 6.9).

This table shows that a straightforward implementation of the BIG-N cryptosystem using classical decoding methods is quite susceptible to the BIG-N fault attack. Even state-of-the-art security parameters were broken in about 20 minutes. Therefore we recommended to implement the countermeasures proposed in Remark 4.6.

Acknowledgements. This research was supported by DFG (German Research Foundation) project “Algebraische Fehlerangriffe” grant KR 1907/6-2.

REFERENCES

- [1] John Abbott, Anna M. Bigatti, and Lorenzo Robbiano. CoCoA: a system for doing Computations in Commutative Algebra. Available at <http://cocoa.dima.unige.it>.
- [2] Roberto Avanzi, Simon Hoerder, Dan Page, and Michael Tunstall. Side-channel attacks on the McEliece and Niederreiter public-key cryptosystems. *J. Cryptogr. Eng.*, 1(4):271–281, 2011.
- [3] Elwyn Berlekamp. Nonbinary BCH decoding (abstr.). *IEEE Trans. Inf. Theory*, 14(2):242–242, 1968.
- [4] Daniel J. Bernstein, Tanja Lange, and Christiane Peters. Attacking and defending the McEliece cryptosystem. In *International Workshop on Post-Quantum Cryptography*, volume 5299 of *LNCS*, pages 31–46. Springer-Verlag, 2008.
- [5] Jakub Breier, Xiaolu Hou, and S. Bhasin. *Automated Methods in Cryptographic Fault Analysis*. Springer Int. Publishing, Cham, 2019.
- [6] Pierre-Louis Cayrel and Pierre Dusart. McEliece/Niederreiter PKC: sensitivity to fault injection. In *5th International Conference on Future Information Technology*, pages 1–6. IEEE, 2010.
- [7] Hang Dinh, Cristopher Moore, and Alexander Russell. McEliece and Niederreiter cryptosystems that resist quantum Fourier sampling attacks. In *Advances in Cryptology – CRYPTO 2011*, volume 6841 of *LNCS*, pages 761–779. Springer-Verlag, 2011.

- [8] Jean L. Dornstetter. On the equivalence between Berlekamp's and Euclid's algorithms (corresp.). *IEEE Trans. Inf. Theory*, 33(3):428–431, 1987.
- [9] Jean-Charles Faugere, Ayoub Otmani, Ludovic Perret, and Jean-Pierre Tillich. Algebraic cryptanalysis of McEliece variants with compact keys. In *Advances in Cryptology – EUROCRYPT 2010*, volume 6110 of *LNCS*, pages 279–298. Springer-Verlag, 2010.
- [10] Stefan Heyse and Tim Güneysu. Code-based cryptography on reconfigurable hardware: tweaking Niederreiter encryption for performance. *J. Cryptogr. Eng.*, 3(1):29–43, 2013.
- [11] Jinhwei Hu, Wangchen Dai, Liu Yao, and Ray C. C. Cehung. An application specific instruction set processor (ASIC) for the Niederreiter cryptosystem. In Asaf Varol, Murat Karabatak, and Cihat Varol, editors, *6th Int. Symp. on Digital Forensic and Security (ISDFS 2018)*, Piscataway, 2018. IEEE.
- [12] Marc Joye and Michael Tunstall. *Fault Analysis in Cryptography*. Springer-Verlag, Berlin Heidelberg, 2012.
- [13] Martin Kreuzer and Lorenzo Robbiano. *Computational Commutative Algebra 1*. Springer-Verlag, Heidelberg, 2000.
- [14] Martin Kreuzer and Lorenzo Robbiano. *Computational Linear and Commutative Algebra*. Springer Int. Publ., Cham, 2016.
- [15] James L. Massey. Shift-register synthesis and BCH decoding. *IEEE Transactions on Information Theory*, 15(1):122–127, 1969.
- [16] Alexander May, Alexander Meurer, and Enrico Thomae. Decoding random linear codes in $\mathcal{O}(2^{0.054n})$. In *Advances in Cryptology – ASIACRYPT'11*, volume 7073 of *LNCS*, pages 107–124. Springer-Verlag, 2011.
- [17] Robert J. McEliece. A public-key cryptosystem based on algebraic coding theory. *Deep Space Network Progress Report*, 44:114–116, 1978.
- [18] Harald Niederreiter. Knapsack-type cryptosystems and algebraic coding theory. *Prob. Control and Inf. Theory*, 15(2):159–166, 1986.
- [19] Raphael Overbeck and Nicolas Sendrier. Code-based cryptography. In Daniel J. Bernstein, Johannes Buchmann, and Erik Dahmen, editors, *Post-Quantum Cryptography*, pages 95–145. Springer-Verlag, Berlin, Heidelberg, 2009.
- [20] Nicholas Patterson. The algebraic decoding of Goppa codes. *IEEE Trans. Inf. Theory*, 21(2):203–207, 1975.
- [21] Mélissa Rossi, Mike Hamburg, Michael Hutter, and Mark E. Marson. A side-channel assisted cryptanalytic attack against QcBits. In Wieland Fischer and Naofumi Homma, editors, *Cryptographic Hardware and Embedded Systems – CHES 2017*, volume 10529 of *LNCS*, pages 3–23. Springer-Verlag, 2017.
- [22] Peter W. Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM J. Comput.*, 26(5):1484–1509, 1997.
- [23] Falko Strenzke. Fast and secure root finding for code-based cryptosystems. In Josef Pieprzyk, Ahmad-Reza Sadeghi, and Mark Manulis, editors, *Cryptology and Network Security – CANS 2012*, volume 7712 of *LNCS*, pages 232–246. Springer-Verlag, 2012.
- [24] Falko Strenzke, Erik Tews, H. Gregor Molter, Raphael Overbeck, and Abdulhadi Shoufan. Side channels in the McEliece PKC. In Johannes Buchmann and Jintai Ding, editors, *Post-Quantum Cryptography – PQCrypto 2008*, volume 5299 of *LNCS*, pages 216–229. Springer-Verlag, 2008.
- [25] Yasuo Sugiyama, Masao Kasahara, Shigeichi Hirasawa, and Toshihiko Namekawa. A method for solving key equation for decoding Goppa codes. *Inform. Control*, 27(1):87–99, 1975.
- [26] Wen Wang, Jakub Szefer, and Ruben Niederhagen. FPGA-based key generator for the Niederreiter cryptosystem using binary Goppa codes. In Wieland Fischer and Naofumi Homma, editors, *Cryptographic Hardware and Embedded Systems – CHES 2017*, volume 10529 of *LNCS*, pages 253–274. Springer-Verlag, 2017.
- [27] Wen Wang, Jakub Szefer, and Ruben Niederhagen. FPGA-based Niederreiter cryptosystem using binary Goppa codes. In Tanja Lange and Rainer Steinwandt, editors, *Post-Quantum Cryptography*, volume 10786 of *LNCS*, pages 77–98. Springer-Verlag, 2018.