IET Journals
The Institution of Engineering and Technology

On the Security of Generalized Reed-Solomon Code-Based Cryptosystems

ISSN 1751-8644
doi: 0000000000
www.ietdl.org

Marco Baldi[1] *Franco Chiaraluce*[1] *Joachim Rosenthal*[2] *Paolo Santini*[1] *Davide Schipani*[2]

[1] *DII, Università Politecnica delle Marche, Ancona, Italy*
[2] *Department of Mathematics, University of Zurich, Zurich, Switzerland*
✉ *E-mail: {m.baldi, f.chiaraluce}@univpm.it, p.santini@pm.univpm.it,{rosenthal, davide.schipani}@math.uzh.ch*

Abstract: In this paper, we elaborate on a recently proposed variant of the public-key McEliece and Niederreiter cryptosystems using Generalized Reed-Solomon (GRS) codes as private codes. The use of these codes brings known advantages in terms of public key size, but particular care is needed in the choice of parameters not to endanger the system security. In fact, the considered system exploits a strong disguising technique of the private code within the public code. However, it has recently been pointed out that some new attacks exist which may threaten some instances of such a system, therefore the choice of parameters needs to consider some further constraints compared to the original version. After outlining these constraints, we propose a new modification of the system achieving a greater flexibility in the parameter choice. Moreover, the new system exhibits lower complexity than the original GRS code-based system. We highlight its very competitive features like key size and encryption rate in comparison with classical systems.

1 Introduction

In recent years, research on code-based public-key cryptography has been boosted by the fact that code-based cryptosystems may offer shelter to attacks exploiting quantum computers. In fact, it is known that quantum algorithms like Shor's algorithm [1] are able to turn into polynomial-time solvable some hard mathematical problems that are at the basis of many widespread public-key cryposystems, like RSA, ElGamal, DSA, ECDSA, Diffie-Hellman and others. While quantum computers represented a mostly theoretical threat until a few years, they are rapidly becoming real nowadays [2, 3]. Therefore, research efforts devoted to the definition of efficient post-quantum replacements of quantum vulnerable systems have dramatically increased in last years. Also NIST's post-quantum crypto project [4] goes in this direction, by encouraging the community to select secure and efficient post-quantum cryptosystems to be used in commercial applications.

According to NIST [5], code-based cryptosystems are among the most promising solutions to replace quantum vulnerable public-key cryptosystems. Code-based public key cryptography has been initiated by Robert McEliece, who proposed the first public-key cryptosystem based on codes in 1978 [6]. The hard problem exploited by the McEliece cryptosystem is that of decoding a random-like linear block code with no visible structure, which is known to be a non-polynomial-time solvable problem [7]. Instead, many families of non-random or quasi-random codes exist which are provided with efficient and fast decoding algorithms working on suitable representations of the codes. McEliece chose to use Goppa codes of length n able to correct t errors to form the secret key. By knowing the secret representation of these codes, they can be decoded through Patterson's algorithm running in time $O(n \cdot t)$. Then, McEliece proposed to use a linear transformation to disguise the secret code having a random-like representation, which prevents the use of efficient decoders and creates the trapdoor. While McEliece used generator matrices to represent private and public codes, Niederreiter later introduced an equivalent cryptosystem exploiting parity-check matrices [8], which is still secure when Goppa codes are used as secret codes.

However, when Goppa codes are used, these systems have the drawback of requiring rather large public keys. This is mainly due to the random-like nature of Goppa codes characteristic matrices and of their disguised versions, which does not facilitate their storage. A common approach to face this problem is to replace Goppa codes with other families of codes characterized by more structured representations allowing a more compact storage of the public keys. In fact, several families of codes have been considered for this purpose, but often with security drawbacks [9–11]. Also Low-Density Parity-Check (LDPC) codes have a long history in the framework of these cryptosystems [12–15], and some secure LDPC code-based instances have been found [16], [17].

The use of Generalized Reed-Solomon (GRS) codes as replacements of Goppa codes in the McEliece and Niederreiter cryptosystems is a challenging matter as well. On the one hand, GRS codes would allow to reduce the public key size thanks to their optimum error correction capability, since they are maximum distance separable codes. In addition, GRS codes are widespread in many applications, and much research has been done on their encoding and decoding, with the aim to reduce complexity and facilitate implementation. On the other hand, GRS codes have more structure than Goppa codes (and wild Goppa codes), therefore it is harder to avoid attacks aimed at recovering the secret code from the public one. The first successful attack against GRS code-based cryptosystems was the Sidelnikov-Shestakov attack [18], which was addressed to the Niederreiter system. This highlighted the fact that, differently from Goppa codes, exposing a public code which is permutation equivalent to the secret GRS code does not provide security.

A GRS code-based variant of the McEliece and Niederreiter cryptosystems has been proposed in [19], with the purpose of providing a strong protection of the secret GRS code. The rationale of the proposal is to replace, in the transformation from the private to the public code, the classical permutation matrix with a more general matrix obtained as the sum of a low rank matrix ($\mathbf{R}$) and a sparse matrix ($\mathbf{T}$). Recently, an attack procedure already devised in [20] has been improved in [21], resulting in a polynomial-time attack against the system parameters proposed in [19]. This induces to increase the rank of $\mathbf{R}$, which entails a higher decryption complexity, and/or the density of $\mathbf{T}$, which can weaken the advantage brought by the high correction capability of GRS codes. Nevertheless, some choices of the system parameters still avoid the mentioned attack (e.g., using codes with rate $R \approx 1/2$, rank of $\mathbf{R}$ equal to 1 and density of $\mathbf{T}$ larger than or equal to $1 + R$), so they can be proposed and combine security with an acceptable decryption complexity. A few instances will be given in Table 6, where the key size can be seen

© The Institution of Engineering and Technology 2015

to be a fraction (0.67 and 0.61, respectively) of that required by the Goppa code-based solution to achieve the same security level.

In this paper, we introduce a modification of the scheme proposed in [19], which allows to increase the rank of $\mathbf{R}$ without any effect on the decryption complexity, which is even considerably reduced. This is paid in terms of the ciphertext and public key lengths, which may be somewhat increased with respect to [19]. However, in several applications of public-key cryptosystems, having long ciphertexts is less detrimental than having large public keys, and this solution can still allow to achieve considerably shorter keys than the classical binary Goppa code-based counterparts. Moreover, the new system incurs a very small penalty (or even no penalty in some cases) in terms of encryption rate with respect to the binary Goppa code-based counterparts. On the other hand, the secret GRS code disguising technique is weakened by the new modification, in that a subcode of the public code is now exposed. In some instances of the system (that is, when $m = 1$) this subcode is permutation equivalent to a subcode of the secret code, which introduces some new constraints and guides the choice of parameters that avoid attacks on the subcode. In fact, in this case the security of the system can be reduced to the security of the system in [22]. Choosing $m > 1$ instead provides a higher level of protection and at the same time allows achieving more competitive parameters in terms of key size.

The paper is organized as follows. In Section 2, we describe the original GRS code-based cryptosystem that we consider as a starting point. In Section 3, we describe the new variant we propose. In Section 4, we compare the original system and the new variant with the classical binary Goppa code-based systems in terms of key size and complexity. In Section 5, we provide some final remarks and hints for future works.

2 Original system

In the following we focus mainly on the Niederreiter version of the system in [19], that is reminded next, but the same arguments can be applied to the McEliece version, with minor modifications. A few differences, e.g., in encryption rate, between the two versions are anyway pointed out later in Section 4.

2.1 Key generation

Bob's secret key is formed by three matrices, $\{\mathbf{H}, \mathbf{S}, \mathbf{Q}\}$. The first one is the $r \times n$ parity-check matrix of a GRS code with length n and dimension $k = n - r$, defined over $\mathbb{F}_q$, able to correct $t = \frac{r}{2}$ errors. The matrix $\mathbf{S}$ is a non-singular $r \times r$ scrambling matrix and $\mathbf{Q}$ is a non-singular $n \times n$ transformation matrix. The latter is obtained as $\mathbf{Q} = \mathbf{R} + \mathbf{T}$, where $\mathbf{R}$ is a dense $n \times n$ matrix with rank $z \ll n$ and $\mathbf{T}$ is a sparse $n \times n$ matrix with average row and column weight $m \ll n$. Both $\mathbf{R}$ and $\mathbf{T}$ take values in $\mathbb{F}_q$. The matrix $\mathbf{R}$ is obtained as $\mathbf{R} = \mathbf{a}^T \cdot \mathbf{b}$, where $\mathbf{a}$ and $\mathbf{b}$ are two matrices defined over $\mathbb{F}_q$, having size $z \times n$ and rank z. Therefore, $\mathbf{R}$ has rank z as well. The matrix $\mathbf{T}$ instead is an $n \times n$ non-singular sparse matrix having average row and column weight equal to m, where m is not necessarily an integer value. More details on its construction are provided in [19, Sect. 2.4]. Bob's public key $\mathbf{H}'$ is then obtained as

$$\mathbf{H}' = \mathbf{S}^{-1} \cdot \mathbf{H} \cdot \mathbf{Q}^T. \tag{1}$$

2.2 Encryption

In order to perform encryption, Alice maps the cleartext vector into an error vector $\mathbf{e} = [e_1, e_2, \ldots, e_n]$, having weight $t_p \leq t$. Differently from the original McEliece and Niederreiter systems, where $t_p = t$, this system uses $t_p = \left\lfloor \frac{t}{m} \right\rfloor$. Alice then computes the ciphertext as the syndrome

$$\mathbf{x} = \mathbf{H}' \cdot \mathbf{e}^T. \tag{2}$$

2.3 Decryption

In order to decrypt $\mathbf{x}$, Bob first computes

$$\begin{aligned} \mathbf{x}' &= \mathbf{S} \cdot \mathbf{x} \\ &= \mathbf{H} \cdot \mathbf{Q}^T \cdot \mathbf{e}^T \\ &= \mathbf{H} \cdot (\mathbf{e} \cdot \mathbf{Q})^T \\ &= \mathbf{H} \cdot [\mathbf{e} \cdot (\mathbf{R} + \mathbf{T})]^T \\ &= \mathbf{H} \cdot \mathbf{b}^T \cdot \gamma + \mathbf{H} \cdot \mathbf{T}^T \cdot \mathbf{e}^T, \end{aligned} \tag{3}$$

where $\gamma = \mathbf{a} \cdot \mathbf{e}^T$. Bob then has to guess the entries of γ, which are unknown to him. This means discovering the values of z elements of $\mathbb{F}_q$, which can be done through a sequential search over the whole set of q^z possible solutions, requiring $q^z/2$ attempts on average. As shown in [19, Sect. 3.2], Bob is able to tell if each attempt is successful or not. After having correctly guessed the entries of γ, Bob computes

$$\begin{aligned} \mathbf{x}'' &= \mathbf{x}' - \mathbf{H} \cdot \mathbf{b}^T \cdot \gamma \\ &= \mathbf{H} \cdot \mathbf{T}^T \cdot \mathbf{e}^T \\ &= \mathbf{H} \cdot \mathbf{e}_T^T. \end{aligned}$$

Since $\mathbf{e}_T = \mathbf{e} \cdot \mathbf{T}$ has weight $\leq m \cdot t_p \leq t$, $\mathbf{x}''$ is a correctable syndrome through the secret GRS code, and he can recover $\mathbf{e}_T$, having weight $\leq t$, by performing syndrome decoding. Then, he computes

$$\mathbf{e} = \mathbf{e}_T \cdot \mathbf{T}^{-1} \tag{4}$$

and finally demaps $\mathbf{e}$ into its associated cleartext vector to get the secret message.

2.4 Parameters design

In this system, the two values m and z must be small, since:

- for a given t_p, determined by the target security level against decoding attacks, increasing m requires to increase t and, hence, the code size and the public key size;
- in order to guess the value of γ, Bob needs to perform $q^z/2$ attempts on average, therefore increasing z increases the decryption complexity.

Keeping both z and m too small exposes the system to polynomial-time attacks [21], therefore a security / performance trade-off arises. In fact, the attack in [21] can be applied only if m and z are chosen in such a way as to verify both the following conditions:

$$\begin{cases} 1 \leq m \leq 1 + R - \frac{1}{n} - \sqrt{\frac{8}{n}R + \frac{1}{n^2}} < 2, \\ z = 1. \end{cases} \tag{5}$$

The special choice $m = z = 1$ exposes the system to the original version of the attack, introduced in [20]. The improved version of the attack proposed in [21] is built upon a distinguisher of GRS codes based on computing the dimension of the square of shortenings of the public code. The squares of some of these shortenings have a smaller dimension than the squares of shortened random codes of the same size, due to the structure of the hidden private code. The core of the attack in [21] is an algorithm to distinguish between rows of $\mathbf{T}$ with Hamming weight 1 and rows of $\mathbf{T}$ with Hamming weight 2. In fact, for $1 < m < 2$ the rows of $\mathbf{T}$ have Hamming weight 1 or 2. Then, the effect of weight-2 columns of $\mathbf{T}$ is reverted to that of weight-1 columns through linear combinations of columns of the public parity-check matrix. Through these steps, the public key of an alternative system with the same private code but with $m = 1$ is

© The Institution of Engineering and Technology 2015

recovered by the opponent, who can then mount the attack in [20] against such an alternative system to recover the private key.

In order to counter these attacks, the system parameters must be chosen such that conditions (5) on m and z are not verified. This, however, has a detrimental effect on the public key size and complexity when the original system is used. In the following, a new variant of this system is presented that aims at preventing known vulnerabilities by increasing z while, at the same time, avoiding to increase the decryption complexity by publishing $\mathbf{a}$. This leaks some information on $\mathbf{R}$, which affects the computation of the work factor of decoding attacks and in some cases leads us to choose set of parameters that thwart some known algebraic attacks. However, as stated above, choosing high values of z allows avoiding the attack presented in [21], which can be applied successfully only if conditions (5) on m and z are verified, thus resulting in some instances of the new system that bring significant advantages over Goppa code-based systems and other GRS code-based systems.

3 New system

In the variant we propose, $\mathbf{a}$ is made public together with $\mathbf{H}'$, so Bob's public key is now the pair $(\mathbf{H}', \mathbf{a})$. Moreover, during encryption, Alice computes $\gamma = \mathbf{a} \cdot \mathbf{e}^T$ and sends it along with the ciphertext. In this case, Bob no longer needs to guess the value of γ, which significantly reduces the decryption complexity. Basically the new system works as follows.

3.1 Encryption

Alice maps the cleartext vector into an error vector $\mathbf{e} = [e_1, e_2, \ldots, e_n]$, having weight $t_p = \lfloor \frac{t}{m} \rfloor$. Alice then sends the pair $(\mathbf{x}, \gamma)$, where $\mathbf{x}$ is given by (2) and

$$\gamma = \mathbf{a} \cdot \mathbf{e}^T. \tag{6}$$

3.2 Decryption

As in the original system (see Section 2.3), in order to decrypt $\mathbf{x}$, Bob first computes

$$\mathbf{x}' = \mathbf{H} \cdot \mathbf{b}^T \cdot \gamma + \mathbf{H} \cdot \mathbf{T}^T \cdot \mathbf{e}^T.$$

Now, knowing γ, he can directly compute

$$\begin{aligned} \mathbf{x}'' &= \mathbf{x}' - \mathbf{H} \cdot \mathbf{b}^T \cdot \gamma \\ &= \mathbf{H} \cdot \mathbf{e}_T^T, \end{aligned}$$

and then recover $\mathbf{e}_T$ by performing syndrome decoding in order to get the secret message.

3.3 Parameters design

The parameters of the new system variant are chosen as follows to avoid potential vulnerabilities. In fact, when $\mathbf{a}$ is public, the system is exposed to the subcode vulnerability described in [19, Sect. 3.1]. In fact, from (1) it follows that

$$\begin{aligned} \mathbf{H}' &= \mathbf{S}^{-1} \cdot \mathbf{H} \cdot \mathbf{R}^T + \mathbf{S}^{-1} \cdot \mathbf{H} \cdot \mathbf{T}^T \\ &= \mathbf{S}^{-1} \cdot \mathbf{H} \cdot \mathbf{b}^T \cdot \mathbf{a} + \mathbf{S}^{-1} \cdot \mathbf{H} \cdot \mathbf{T}^T. \end{aligned} \tag{7}$$

An attacker could consider the following alternative parity-check matrix

$$\mathbf{H}_S = \begin{bmatrix} \mathbf{H}' \\ \mathbf{a} \end{bmatrix} = \begin{bmatrix} \mathbf{S}^{-1} \cdot \mathbf{H} \cdot \mathbf{b}^T \cdot \mathbf{a} + \mathbf{S}^{-1} \cdot \mathbf{H} \cdot \mathbf{T}^T \\ \mathbf{a} \end{bmatrix}. \tag{8}$$

Compared to $\mathbf{H}'$, $\mathbf{H}_S$ includes an additional set of parity-check equations, defined by $\mathbf{a}$, which imply that any codeword $\mathbf{c}$ belonging to the code defined by $\mathbf{H}_S$ satisfies $\mathbf{a} \cdot \mathbf{c}^T = \mathbf{0}$. This in turn implies that $\mathbf{S}^{-1} \cdot \mathbf{H} \cdot \mathbf{b}^T \cdot \mathbf{a} \cdot \mathbf{c}^T = \mathbf{0}$. Therefore, the constraint imposed by the set of parity-check equations appearing in (8) due to $\mathbf{H}'$ becomes $\mathbf{S}^{-1} \cdot \mathbf{H} \cdot \mathbf{T}^T \cdot \mathbf{c}^T = \mathbf{0}$, for any codeword $\mathbf{c}$ belonging to the code defined by $\mathbf{H}_S$. In summary, $\mathbf{H}_S$ as in (8) defines a subcode of the public code where any codeword $\mathbf{c}$ satisfies $\mathbf{S}^{-1} \cdot \mathbf{H} \cdot \mathbf{T}^T \cdot \mathbf{c}^T = \mathbf{0}$. Hence, the effect of the dense matrix $\mathbf{R}$ is removed and, when $\mathbf{T}$ is a permutation matrix (that is, when $m = 1$), the subcode defined by $\mathbf{H}_S$ is permutation-equivalent to a subcode of the secret code. A cryptosystem exposing a subcode of a private GRS code is the Berger-Loidreau (BL) scheme [22]. Based on the above considerations, for the case $m = 1$, the security of the new system is equivalent to that of a BL scheme with the same subcode parameters. This also makes the attack presented in [10] applicable, and designing parameters for the BL scheme that are secure against known attacks is related to designing secure parameters for the new variant we propose in the case $m = 1$.

However, the dimension of such a subcode is equal to $n - \mathrm{rank}\{\mathbf{H}_S\}$ and we can choose parameters which avoid known attacks on the subcode (which apply in case $m = 1$): if we look at [10, Eq. 8], for dimension $k \geq n/2$ we can avoid the attack by requiring $k - z - 1 < 2k - n + 1$. In particular, we will show examples for $m = 1$ with high rate R and $z \geq n - k$, while for $m > 1$ we can consider lower values of z. For this case, we will choose a few values of $m > 1 + \frac{r-3}{n}$ as in [19] to avoid distinguishers, and see how the key size changes for a few values of z in the range of interest.

Another point to take into account is that, although this type of subcode attack may be avoided, knowing $\mathbf{a}$ and γ facilitates decoding attacks. In fact, $\begin{bmatrix} \mathbf{x} \\ \gamma \end{bmatrix} = \mathbf{H}_S \cdot \mathbf{e}^T$ and an attacker could perform syndrome decoding on the code defined by $\mathbf{H}_S$, rather than the public code. Such a code has rate $\frac{k-z}{n} < \frac{k}{n}$, and this facilitates attacks exploiting general decoding algorithms like information set decoding (ISD). The attack complexity decreases as long as z increases (when $z \geq k$ the attack becomes very simple, since Eve could find a full rank $\mathbf{H}_S$ with size $n \times n$ and just invert it). Indeed, we will choose parameters with large dimension k and relatively small z.

It is interesting to remark that a first attempt in using a disguising matrix like $\mathbf{R}$ together with publishing $\mathbf{a}$ was reported in [23], but this involved choosing $\mathbf{e}$ such that $\mathbf{a} \cdot \mathbf{e}^T = 0$. This constraint on the choice of $\mathbf{e}$ forced to choose very small values of z, otherwise the work factor of decoding attacks would be significantly reduced. In the case of the present proposal, there is no constraint on $\mathbf{e}$, which allows to choose higher values of z. The price to pay is some loss in encryption rate due to the transmission of γ along with the ciphertext.

4 Comparison with classical systems

Let us compare the new system with related previous systems. As shown in Section 3.3, designing secure parameters for the BL scheme is related to designing secure parameters for the new variant we propose in the case $m = 1$. However, differently from our scheme, the BL system uses the subcode to encrypt messages. As we will see next, this implies some difference in encryption rate when compared to our system instances with $m = 1$ and the same security level.

In the following, we first compare the new variant with $m = 1$ (and $z = n - k$) with the BL system based on GRS subcodes and the classical cryptosystem based on binary Goppa codes, for some fixed security levels. We then take m slightly larger than 1 as already suggested in [19]. In this case the exposed subcode is no longer permutation equivalent to a subcode of the secret code, therefore security is higher than security of BL. The reduction in error correction capability is counterbalanced by the possibility of decreasing z, so that overall the effect on key size is favourable. We therefore add new instances (choosing a few different values for m and z) in the comparison.

© The Institution of Engineering and Technology 2015

Table 1 System performance comparison for $SL = 2^{180}$: (a) Goppa code-based system, (b) new system with $m = 1$ and $z = n - k$, (c) Berger-Loidreau.

Variant	n	k	$t = t_p$	WF	KS (KiB)	R_e (Niederreiter)	R_e (McEliece)
(a)	4096	3004	91	180.06	400.44	0.5724	0.7334
(b)	1282	1146	68	180.01	392.89	0.3850	0.8082
(c)	1212	1062	75	180.05	342.15	0.3806	0.7525

4.1 Public key size

In the Niederreiter cryptosystem, we can use a systematic version of the public key, that is, $\mathbf{H}'' = \left(\mathbf{H}'_l\right)^{-1} \cdot \mathbf{H}' = \left[\mathbf{I}|\left(\mathbf{H}'_l\right)^{-1} \cdot \mathbf{H}'_r\right] = \left[\mathbf{I}|\mathbf{H}''_r\right]$, where $\mathbf{H}'_l$ and $\mathbf{H}'_r$ are the left $r \times r$ and right $r \times k$ submatrices of $\mathbf{H}'$, that is, $\mathbf{H}' = \left[\mathbf{H}'_l|\mathbf{H}'_r\right]$. By considering public parity-check matrices in such a form, the public key size is kr bits for the binary Goppa code-based system. In the new system, $kr + z(n - z)$ values over $\mathbb{F}_q$, i.e $(kr + zn - z^2)\log_2 q$ bits are required to store the public key $\left\{\mathbf{H}', \mathbf{a}\right\}$, with $\mathbf{a}$ in systematic form, while in the BL scheme using a subcode of dimension $k - z$, the public key size is $(k - z)(r + z)\log_2 q$ bits. It follows that, for a given z, when k increases the difference in key size between this new system and BL decreases, although BL always has a smaller key size. When $m = 1$, the new system requires the same vale of z used in BL to achieve the same security level. Therefore, in such a case the BL system exhibits some advantage over the new system in terms of public key size. However, choosing $m > 1$ allows to reduce the value of z in the new system with respect to BL for the same security level, and this yields significant reductions in the public key size of the new system with respect to BL.

The values of the key size are the same for the McEliece versions of the above systems using matrices in systematic form.

4.2 Encryption rate

Due to increased storage capabilities in recent years, the encryption rate R_e is currently often considered more critical than the public key size. In the Niederreiter versions, the information is encrypted through the intentional error vector, so the encryption rates are

$$R_e = \frac{\log_2 \binom{n}{t}}{(n-k)} \tag{9}$$

for the Goppa code-based system,

$$R_e = \frac{\log_2 \binom{n}{t_p} + t_p \log_2(q-1)}{(n-k+z)\log_2 q} \tag{10}$$

for the new system (considering γ of length z as part of the ciphertext), and

$$R_e = \frac{\log_2 \binom{n}{t} + t \log_2(q-1)}{(n-k+z)\log_2 q} \tag{11}$$

for the BL system. In the McEliece versions, the encryption rates are $\frac{k}{n}$ for the Goppa code-based system, $\frac{k-z}{n}$ for BL and $\frac{k}{n+z}$ for the new system. From (10) and (11) we observe that, with the Niederreiter version, when the same set of parameters is chosen and $m = 1$ (hence, $t_p = t$), the encryption rate of the new system equals that of BL with the same security level, while with the McEliece version it is higher.

4.3 Security analysis

Cryptanalysis of the new system has to study two possible scenarios. First, the system should be secure against the best decoding algorithm capable of decoding a general linear code. As we will explain next, information set decoding (ISD) is the best known algorithm in the literature for decoding a general linear code over a finite field $\mathbb{F}_q$. Second, we should consider possible structural attacks involving some distinguishers.

ISD was first introduced by Prange [24], then improved by Lee-Brickell [25], Leon [26] and Stern [27], and in recent years has benefited from a number of further improvements [28–30]. In the general area of code based cryptography, ISD represents the most dangerous non-polynomial time attack against general code based systems. In this paper we estimate the cost of the algorithm as it can be found in [28], where the extension to non-binary fields (that are of interest here) is also provided, and we use it for both the binary and the non-binary cases in order to perform a fair comparison. In fact, there are no simple generalizations of the most recent algorithms [30] to work over non-binary fields. However, as shown in [19], we can suppose that their possible application could result in a work factor reduction in the order of 2^9 or less with respect to the algorithm in [28]. Hence, such an advance would result in a general, slight reduction of the security level of the systems we consider and, more importantly, would not significantly affect the results of our comparative assessment.

Let us now discuss the possibility of a structural attack. Starting with the work of Sidelnikov and Shestakov [18], it has been recognized that McEliece type systems having as an underlying (disguised) structure a GRS code are insecure. The main reason for this is that the Schur square code of a GRS code has in general a very small dimension compared to a random code of the same dimension and the dimension of the Schur square stays invariant under monomial transformations. In other words the Schur square can serve as a distinguisher. Couvreur et al. [21] used the Schur square to attack the system in [19], and Marquez-Corbella et al. explained in [31] how the Berger-Loidreau system [22] can be attacked using the Schur square. More recently Couvreur et al. [32] came up with general polynomial time attacks against a large class of algebraic geometric codes and their subcodes (and GRS codes are algebraic geometric).

The question therefore is if the Schur square of the proposed code could be dimension deficient. For this, let us consider once more the relevant equations

$$\mathbf{x} = \mathbf{H}' \cdot \mathbf{e}^T \text{ and } \gamma = \mathbf{a} \cdot \mathbf{e}^T.$$

Combining these equations one can study a related parity-check matrix, that is

$$\tilde{\mathbf{H}} = \begin{bmatrix} \mathbf{0} & \mathbf{H}' \\ -\gamma & \mathbf{a} \end{bmatrix} \tag{12}$$

where $\mathbf{0}$ is an all-zero matrix with size $r \times z$. Clearly if one has efficient decoding for this parity-check matrix the system becomes insecure. For this, we wish to comment on two extreme cases.

1. Assume that $m = 1$, i.e., $\mathbf{H}$ simply represents a disguised GRS code. Then, it is fairly clear that the Schur square of the defined code is dimension deficient unless the size of z is chosen sufficiently large. So, it is important that $m > 1$, what we will do in all our examples.
2. Consider the other extreme case where $z = 0$. Here Couvreur et al. [21] derived a polynomial time attack in case that $1 < m < 2$. Bolkema et al. [33] studied the case of $m = 2$ and they called this situation "the weight two masking of the GRS system". Using extensive simulations, they conjectured that over large fields the Schur square of the public code has the expected dimension of a random code. More recently, Weger [34] has shown that the probability that

© The Institution of Engineering and Technology 2015

Table 2 System performance comparison for $SL = 2^{260}$: (a) Goppa code-based system, (b) new system with $m = 1$ and $z = n - k$, (c) Berger-Loidreau.

Variant	n	k	$t = t_p$	WF	KS (KiB)	R_e (Niederreiter)	R_e (McEliece)
(a)	8192	6957	95	260.57	1048.82	0.6012	0.8492
(b)	1950	1754	98	260.10	917.37	0.3798	0.8173
(c)	1788	1560	114	260.11	801.13	0.3732	0.7450

Table 3 New system performance for $m = 1.2$, $SL = 2^{180}$ and $SL = 2^{260}$.

SL	z	n	k	t	t_p	WF	KS (KiB)	R_e (Niederreiter)	R_e (McEliece)
180	10	796	634	81	67	181.03	130.09	0.5870	0.7866
180	30	886	722	82	68	180.50	172.24	0.5304	0.7882
180	50	918	740	89	74	180.19	210.43	0.4879	0.7644
260	10	1162	928	117	97	260.64	284.27	0.5894	0.7918
260	30	1222	976	123	102	260.21	435.37	0.5467	0.7796
260	50	1276	1018	129	107	260.03	408.04	0.5128	0.7677

the Schur square has maximal dimension approaches 1 as the field size q goes to infinity.

These remarks should make it clear that a distinguisher attack using the Schur square computation becomes out of reach when m and z are chosen sufficiently large. When $z = 0$ it seems that m should be chosen at least two in order to avoid distinguisher attacks. When $z > 0$ we do not know what the minimal value for m should be to avoid the Schur square distinguisher attack and it will be a question of future research to come up with a precise value for this.

4.4 Examples

On the basis of the above performance and security metrics, let us compare the new system based on GRS codes with the classical binary Goppa code-based system and the BL system based on GRS subcodes. For such purpose, we consider some instances of these systems approximately achieving the same security level and compare their features. For GRS code-based systems, we consider full length GRS codes defined over $\mathbb{F}_q$, with $q = n + 1$ being a prime. Goppa code-based systems instead exploit irreducible binary Goppa codes with length equal to a power of two.

In Tables 1 and 2, we respectively consider codes with a security level (SL) of at least 2^{180} and 2^{260}, estimated as the work factor (WF) of attacks based on ISD, computed according to [28]. On the basis of an exhaustive search performed over the range of parameters of interest, we report the solutions achieving the smallest public key size (KS) expressed in kibibytes (KiB) for the classical binary Goppa code-based cryptosystem, for the new variant of GRS code-based cryptosystem with $m = 1$, and for the BL cryptosystem, using $z = n - k$. We can notice that the new system with $m = 1$ is able to achieve smaller key sizes than the Goppa-based solution, and it may have higher encryption rate in the McEliece version. The new system has always higher encryption rate than BL, particularly for the McEliece version, but larger key size.

As mentioned, $m = 1$ was considered as in this case the new system is comparable to BL, but an m slightly larger than 1 is certainly preferable, to better protect the secret code. In Tables 3, 4 and 5 a few values of m and z are tested and the instances with smallest key sizes are presented. It is evident how it is possible to achieve very interesting parameters, namely high encryption rates and compact public keys. Considering the instances in the tables, the reduction in public key size with respect to the Goppa code-based solution with the same security level can reach 73%.

To conclude the comparison, in Table 6 we report two instances of the original GRS code-based cryptosystem in [19], achieving smallest key sizes for a security level of 2^{180} and 2^{260}, respectively, and satisfying $m = 1.1 \cdot (1 + R) > 1 + R$ to avoid the attack in [21]. Through the comparison with the previous tables, we observe that, for the same security levels, the new system is able to achieve a reduction in public key size by more than 50%.

4.5 Complexity

In this section, we assess the encryption and decryption complexity of the new system, and compare it with that of the original version of GRS code-based system and the classical binary Goppa code-based system. We restrict our comparison to the Niederreiter versions, as looking at other instances will not change significantly the sense of this comparison.

As done in [35], [19], let us assume that one addition between elements of $\mathbb{F}_q$ costs $\mathcal{S} = \lceil \log_2(q) \rceil$ binary operations, while one multiplication costs $\mathcal{M} = 2\mathcal{S}^2$ binary operations. We also consider that an inversion over $\mathbb{F}_q$ has the same cost as a multiplication. The right (or left, respectively) multiplication of an $x \times y$ matrix by a vector having w non-null elements requires to sum w columns (or rows, respectively) of the matrix, that is, $(w-1)x\mathcal{S}$ (or $(w-1)y\mathcal{S}$, respectively) binary operations, plus $wx\mathcal{M}$ (or $wy\mathcal{M}$, respectively) binary operations to multiply each element of the vector by the corresponding matrix column (or row, respectively). The latter term must be considered only when working over $\mathbb{F}_q$ with $q > 2$. Computing the element-wise sum or product of two vectors with length x requires, on average, $x\frac{q-1}{q}$ sums or multiplications. Since in GRS code-based systems q is in the order of some hundreds or more, the term $\frac{q-1}{q}$ can be set equal to 1 with a negligible impact on the numerical results.

Concerning encryption, if we consider the systematic version of the public parity-check matrix and split the vector $\mathbf{e}$ into its left and right parts, $\mathbf{e} = [\mathbf{e}_l|\mathbf{e}_r]$, the encryption function becomes $\mathbf{x} = \mathbf{e}_l^T + \mathbf{H''}_r \cdot \mathbf{e}_r^T$, where $\mathbf{e}_r$ has weight equal to $w = \frac{k}{n}t_p$ on average. Hence, if we neglect the transformation of the information vector into a constant weight vector and back, we obtain an encryption cost equal to:

- $[(w-1)r + t_p - w]$ binary operations for the classical binary code-based system (with codes over $\mathbb{F}_2$),
- $[(w-1)r + t_p - w]\mathcal{S} + wr\mathcal{M}$ binary operations for the original GRS code-based system (with codes over $\mathbb{F}_q$, $q > 2$).

© The Institution of Engineering and Technology 2015

Table 4 New system performance for $m = 1.3$, $SL = 2^{180}$ and $SL = 2^{260}$.

SL	z	n	k	t	t_p	WF	KS (KiB)	R_e (Niederreiter)	R_e (McEliece)
180	10	760	544	108	83	180.04	146.06	0.5399	0.7065
180	30	810	576	117	90	180.16	186.60	0.4989	0.6857
180	50	852	594	129	99	180.11	229.80	0.4671	0.6585
260	10	1122	810	156	120	260.56	326.36	0.5399	0.7155
260	30	1200	888	156	120	260.14	389.81	0.5104	0.7220
260	50	1236	898	169	130	260.30	454.98	0.4843	0.6983

Table 5 New system performance for $m = 1.8$, $SL = 2^{180}$ and $SL = 2^{260}$.

SL	z	n	k	t	t_p	WF	KS (KiB)	R_e (Niederreiter)	R_e (McEliece)
180	10	1092	804	144	80	180.19	298.65	0.4042	0.7296
180	30	1116	792	162	90	180.05	357.44	0.3789	0.6911
180	50	1180	852	164	91	180.00	418.54	0.3594	0.6927
260	10	1600	1168	216	120	260.12	676.31	0.4012	0.7255
260	30	1626	1154	236	131	260.03	771.67	0.3828	0.7054
260	50	1722	1268	227	126	260.04	865.19	0.3691	0.7156

Concerning the proposed variant, we must consider the number of binary operations needed to compute $\gamma = \mathbf{a} \cdot \mathbf{e}^T$, that is, $(t_p - 1)\, z\mathcal{S} + t_p z\mathcal{M}$ binary operations.

Concerning decryption, as done in [19], we consider the classical GRS syndrome decoding algorithm [35], requiring:

1. $4t(2t+2)\mathcal{M} + 2t(2t+1)\mathcal{S}$ binary operations for the key equation solver,
2. $n(t-1)\mathcal{M} + nt\mathcal{S}$ binary operations for the Chien search,
3. $(2t^2+t)\mathcal{M} + t(2t-1)\mathcal{S}$ binary operations for Forney's formula.

After decoding, Bob needs to compute $\mathbf{x}' = \mathbf{S} \cdot \mathbf{x}$, requiring further $(r-1)r\mathcal{S} + r^2\mathcal{M}$ binary operations, and $\mathbf{x}'' = \mathbf{x}' - \mathbf{H} \cdot \mathbf{b}^T \cdot \gamma$, requiring further $r\mathcal{S} + (z-1)r\mathcal{S} + zr\mathcal{M}$ binary operations (taking into account that $\mathbf{H} \cdot \mathbf{b}^T$ can be precomputed only once). The original GRS code-based system also requires Bob to guess the value of γ. Each guessing attempt requires to repeat the key equation solver and the computation of $\mathbf{x}''$ [19]. Therefore, for the GRS code-based systems, we can estimate the overall decryption complexity as

$$D_{\mathrm{GRS}} = \left\{[4t(2t+2) + zr]g + 2t^2 + (2n+1)t + r^2 - n\right\}\mathcal{M} + \left\{[2t(2t+1) + zr]g + 2t^2 + (2n-1)t + (r-1)r - n\right\}\mathcal{S}, \quad (13)$$

where g is the number of times the key equation solver needs to be executed.

For the original system, g equals the average number of attempts needed for a correct guessing by Bob, that is, $g = \frac{q^z}{2}$. For the proposed variants, instead, we have $g = 1$, since no repeated attempts are needed to obtain γ. Finally, for the classical binary Goppa code-based Niederreiter cryptosystem, the decryption complexity is [36]

$$D_{\mathrm{Goppa}} = n + 2\left[\log_2(n)\right]^2 t(2t+1) + \log_2(n)n(2t+1) + \frac{r^2}{2}. \quad (14)$$

In Table 7 we provide the values of the encryption (E) and decryption (D) complexity per information bit for the classical binary Goppa code-based system and instances of the old (Table 6) and new variant (Table 3, $z = 10$) of GRS code-based system considered in this paper for achieving 180-bit security.

By looking at the table, we observe that the gain in public key size with respect to the classical binary Goppa code-based system is paid in terms of complexity. This already occurred for the original GRS code-based system, and still holds for the new variant. However, it has to be noted that the new system achieves a considerable reduction in the decryption complexity with respect to the original GRS code-based system. In fact, the decryption complexity is reduced by more than two orders of magnitude. This makes the new system more balanced from the complexity standpoint, since the encryption and decryption complexities are now more similar.

5 Conclusion

A new type of GRS code-based cryptosystem has been presented which improves over previous solutions by allowing a larger set of instances featuring competitive parameters, such as key size and encryption rate, as well as resilience with respect to any known attack. This is also achieved without affecting complexity, which is even reduced with respect to the original GRS code-based scheme.

Acknowledgements

We would like to thank Alain Couvreur and Ludovic Perret for fruitful discussion on subcode attacks and the relations with the BL scheme. The Swiss National Science Foundation partially supported the work of Joachim Rosenthal under the grant no. 169510 and the work of Davide Schipani under the project IZK0Z2_173546.

6 References

1 Shor P. W.: Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. SIAM J. Comput. 26(5), 1484–1509 (1997).
2 Mitchell R.: IBM to sell use of its new 17-qubit quantum computer over the cloud. [Online]. Available: https://www.allaboutcircuits.com/news/ibm-to-sell-use-of-its-new-17-qubit-quantum-computer-over-the-cloud/, May 2017.

© The Institution of Engineering and Technology 2015

Table 6 Original GRS code-based system performance for $SL = 2^{180}$ and $SL = 2^{260}$.

SL	n	k	m	t	t_p	WF	KS (KiB)	R_e (Niederreiter)	R_e (McEliece)
180	946	504	1.686	221	131	180.24	268.87	0.4209	0.5328
260	1422	786	1.708	318	186	260.26	639.19	0.4111	0.5527

Table 7 Complexity comparison between the binary Goppa code-based Niederreiter cryptosystem, the original GRS code-based cryptosystem and the new variant.

	Binary Goppa code-based	GRS code-based original	GRS code-based new variant 2
n	4096	946	796
k	3004	504	634
q	2	947	797
t	91	221	81
t_p	91	131	67
E	115	3520	2009
D	23024	20818690	47710

3 Hardesty L.: Toward mass-producible quantum computers. [Online]. Available: http://news.mit.edu/2017/toward-mass-producible-quantum-computers-0526, May 2017.
4 National Institute of Standards and Technology: Post-quantum crypto project. [Online]. Available: https://csrc.nist.gov/projects/post-quantum-cryptography, Dec. 2016.
5 Chen, L., Liu, Y.-K., Jordan, S., Moody, D., Peralta, R., Perlner, R., Smith-Tone, D.: Report on post-quantum cryptography. National Institute of Standards and Technology, Tech. Rep. NISTIR 8105, 2016.
6 McEliece R. J.: A public-key cryptosystem based on algebraic coding theory. DSN Progress Report, 114–116, 1978.
7 Berlekamp, E., McEliece, R., van Tilborg, H.: On the inherent intractability of certain coding problems. IEEE Trans. Inform. Theory 24(3), 384–386 (1978).
8 Niederreiter, H.: Knapsack-type cryptosystems and algebraic coding theory. Probl. Contr. and Inform. Theory 15(2), 159–166 (1986).
9 Overbeck, R.: Structural attacks for public key cryptosystems based on Gabidulin codes. Journal of Cryptology 21(2), 280–301 (2008).
10 Wieschebrink, C.: Cryptanalysis of the Niederreiter public key scheme based on GRS subcodes. In: Sendrier, N. (ed.) Post-Quantum Cryptography (PQCrypto 2010), Lecture Notes in Computer Science 6061, 61–72. Springer (2010).
11 Umana, V.G., Leander, G.: Practical key recovery attacks on two McEliece variants. In: Cid, C., Faugère, J.C. (eds.) Proc. 2nd Int. Conf. on Symbolic Computation and Cryptography, 27–44. Egham, UK (Jun. 2010).
12 Monico, C., Rosenthal, J., Shokrollahi, A.: Using low density parity check codes in the McEliece cryptosystem. In: Proc. IEEE International Symposium on Information Theory (ISIT 2000), 215. Sorrento, Italy (Jun. 2000).
13 Baldi, M., Chiaraluce, F.: Cryptanalysis of a new instance of McEliece cryptosystem based on QC-LDPC codes. In: Proc. IEEE International Symposium on Information Theory (ISIT 2007), 2591–2595. Nice, France (Mar. 2007).
14 Baldi, M., Chiaraluce, F., Garello R., Mininni, F.: Quasi-cyclic low-density parity-check codes in the McEliece cryptosystem. In: Proc. IEEE International Conference on Communications (ICC 2007), 951–956. Glasgow, Scotland (Jun. 2007).
15 Otmani, A., Tillich, J.P., Dallot, L.: Cryptanalysis of two McEliece cryptosystems based on quasi-cyclic codes. In: Proc. First International Conference on Symbolic Computation and Cryptography (SCC 2008), 129–140. Beijing, China (Apr. 2008).
16 Baldi, M., Bianchi, M., Chiaraluce, F.: Security and complexity of the McEliece cryptosystem based on QC-LDPC codes. IET Information Security 7(3), 212–220 (2013).
17 Misoczki, R, Tillich, J.-P., Sendrier, N., Barreto, P.: MDPC-McEliece: New McEliece variants from moderate density parity-check codes. In: Proc. IEEE International Symposium on Information Theory (ISIT 2013), 2069–2073, Istanbul, Turkey (Jul. 2013).
18 Sidelnikov, V. M., Shestakov, S. O.: On insecurity of cryptosystems based on generalized Reed-Solomon codes, Discrete Mathematics and Applications 2(4), 439–444 (1992).
19 Baldi, M., Bianchi, M., Chiaraluce, F., Rosenthal, J., Schipani, D. (2016): Enhanced Public Key Security for the McEliece Cryptosystem. Journal of Cryptology 29(1), 1–27 (2016).
20 Couvreur, A., Gaborit, P., Gauthier-Umaña, V., Otmani, A., Tillich, J.-P.: Distinguisher-based attacks on public-key cryptosystems using Reed-Solomon codes. Designs, Codes and Cryptography 73(2), 641–666 (2014).
21 Couvreur, A., Otmani, A., Tillich, J.-P., Gauthier-Umaña, V.: A polynomial-time attack on the BBCRS scheme. In: Katz, J. (ed.) Public-Key Cryptography (PKC 2015), Lecture Notes in Computer Science 9020, 175–193. Springer (2015).
22 Berger, T., Loidreau, P.: How to mask the structure of codes for a cryptographic use. Des. Codes Cryptogr. 35(1), 63–79 (2005).
23 Baldi, M., Bianchi, M., Chiaraluce, F., Rosenthal, J., Schipani, D.: A variant of the McEliece cryptosystem with increased public key security. In: Proc. 19th International Workshop on Coding and Cryptography (WCC), 173–182. Paris, France (Apr. 2011).
24 Prange, E.: The use of information sets in decoding cyclic codes. IRE Trans. Inform. Theory 8(5), 5–9 (1962).
25 Lee, P., Brickell, E.: An observation on the security of McEliece's public-key cryptosystem. In: Barstow D. et al. (eds.) Advances in Cryptology (EUROCRYPT 88), Lecture Notes in Computer Science 330, 275–280. Springer (1988).
26 J. Leon: A probabilistic algorithm for computing minimum weights of large error-correcting codes. IEEE Trans. Inform. Theory 34(5), 1354–1359 (1988).
27 Stern, J.: A method for finding codewords of small weight. In: Cohen G. and Wolfmann J. (eds.) Coding Theory and Applications, Lecture Notes in Computer Science 388, 106–113. Springer (1989).
28 Peters, C.: Information-set decoding for linear codes over $\mathbb{F}_q$. In: Sendrier, N. (ed.) Post-Quantum Cryptography, Lecture Notes in Computer Science 6061, 81–94. Springer (2010).
29 Bernstein, D. J., Lange, T., Peters, C.: Smaller decoding exponents: ball-collision decoding. In: Rogaway P. (ed.) Advances in Cryptology (CRYPTO 2011), Lecture Notes in Computer Science 6841, 743–760. Springer (2011).
30 Becker, A., Joux, A., May, A., Meurer, A.: Decoding random binary linear codes in $2^{n/20}$: How 1 + 1 = 0 improves information set decoding. In: Pointcheval D., Johansson T. (eds) Advances in Cryptology (EUROCRYPT 2012), Lecture Notes in Computer Science 7237, 520–536. Springer (2012).
31 Márquez-Corbella, I., Martinez-Moro, E., Pellikaan, R.: The non-gap sequence of a subcode of a generalized Reed-Solomon code. Des. Codes Cryptogr., 66(1-3), 317–333 (2013).
32 Couvreur, A., Márquez-Corbella, I., Pellikaan, R.: Cryptanalysis of McEliece cryptosystem based on algebraic geometry codes and their subcodes. IEEE Trans. Inform. Theory, 63(8), 5404–5418 (2017).
33 Bolkema, J., Gluesing-Luerssen, H., Kelley, C. A., Lauter, K. E., Malmskog, B., Rosenthal, J.: Variations of the McEliece Cryptosystem. In: Algebraic Geometry for Coding Theory and Cryptography, 129–150. Springer (2017).
34 Weger, V.: A Code-Based Cryptosystem using GRS Codes. Master Thesis at the University of Zürich (Switzerland) (2016).
35 Chen, N., Yan, Z.: Complexity analysis of Reed-Solomon decoding over $GF(2^m)$ without using syndromes. EURASIP Journal on Wireless Communication and Networking, Article ID 843634 (2008).
36 Canteaut, A.: Attaques de cryptosystemes à mots de poids faible et construction de fonction t-résilentes. PhD Thesis, Université Paris (Oct. 1996).

© The Institution of Engineering and Technology 2015