Improved Attacks for Characteristic-2 Parameters of the Cubic ABC Simple Matrix Encryption Scheme

Dustin Moody[1], Ray Perlner[1], and Daniel Smith-Tone[1,2]

[1]National Institute of Standards and Technology,
Gaithersburg, Maryland, USA
[2]Department of Mathematics, University of Louisville,
Louisville, Kentucky, USA

dustin.moody@nist.gov, ray.perlner@nist.gov, daniel.smith@nist.gov

Abstract. In the last few years multivariate public key cryptography has experienced an infusion of new ideas for encryption. Among these new strategies is the ABC Simple Matrix family of encryption schemes which utilize the structure of a large matrix algebra to construct effectively invertible systems of nonlinear equations hidden by an isomorphism of polynomials. One promising approach to cryptanalyzing these schemes has been structural cryptanalysis, based on applying a strategy similar to MinRank attacks to the discrete differential. These attacks however have been significantly more expensive when applied to parameters using fields of characteristic 2, which have been the most common choice for published parameters. This disparity is especially great for the cubic version of the Simple Matrix Encryption Scheme.

In this work, we demonstrate a technique that can be used to implement a structural attack which is as efficient against parameters of characteristic 2 as are attacks against analogous parameters over higher characteristic fields. This attack demonstrates that, not only is the cubic simple matrix scheme susceptible to structural attacks, but that the published parameters claiming 80 bits of security are less secure than claimed (albeit only slightly.) Similar techniques can also be applied to improve structural attacks against the original Simple Matrix Encryption scheme, but they represent only a modest improvement over previous structural attacks. This work therefore demonstrates that choosing a field of characteristic 2 for the Simple Matrix Encryption Scheme or its cubic variant will not provide any additional security value.

Key words: multivariate public key cryptography, differential invariant, MinRank, encryption

1 Introduction

The National Institute of Standards and Technology (NIST) is currently engaged in an effort to update the public key infrastructure, providing alternatives to the classical public key schemes based on arithmetic constructions. The discovery by Peter Shor in the 1990s of efficient algorithms for factoring and computing discrete logarithms, see [1], accelerated research towards building the necessary class of computers, those that Feynman famously suggested in [2]: quantum computers. There has been growing interest among scientists in our discipline in the years since, to provide protocols and algorithms that are post-quantum, that is, secure in the quantum model of computing. The recent publication by (NIST), see [3], of a call for proposals for post-quantum standards directly addresses the challenge of migration towards a more diverse collection of tools for our public key infrastructure.

Public key schemes based on the difficulty of inverting nonlinear systems of equations provide one possibility for post-quantum security. Multivariate Public Key Cryptography (MPKC) is a reasonable option because the problem of solving systems of nonlinear equations, even if only quadratic, is known to be NP-complete; thus, the generic problem is likely beyond the reach of quantum adversaries. Furthermore, there are a variety of standard techniques to metamorphosize multivariate schemes, to introduce new properties, to enhance security, to reduce power consumption, to resist side-channel analysis, etc.

There are numerous long-lived multivariate digital signature schemes. All of UOV [4], HFE- [5], and HFEv- [6] have been studied for around two decades. Moreover, some of the above schemes have optimizations which have strong theoretical support or have stood unbroken in the literature for some time. Notable among these are UOV, which has a cyclic variant [7] that dramatically reduces the key size, and Gui [8], an HFEv- scheme, that, due to tighter bounds on the complexity of algebraically solving the underlying system of equations, see [9], has much more aggressive parameters than QUARTZ, see [6].

Multivariate public key encryption, however, has a much rockier history. Several attempts at multivariate encryption, see [10, 11] for example, have been shown to be weak based on rank or differential weaknesses. Recently, a new framework for developing secure multivariate encryption schemes has surfaces, drawing on the idea that it may impose sufficiently few restrictions on a multivariate map to be merely an injective map into a much larger codomain instead of being essentially a permutation. A few interesting attempts to achieve multivariate encryption have originated from this thought. ZHFE, see [12], the quadratic and cubic variants of the ABC Simple Matrix Scheme, see [13] and [14], and Extension Field Cancellation, see [15], all use fundamentally new structures for the derivation of an encryption system.

A few of the above schemes have already suffered some setbacks. A questionable rank property in the public key of ZHFE presented in [16] makes this scheme appear dubious, while it was shown that the quadratic Simple Matrix structure leaves the signature of a differential invariant in the public key which is exploited in [17] to effect an attack.

The case of the Cubic Simple Matrix encryption scheme is more interesting; the authors in [14] present a heuristic argument for security and suggest the possibility of provable security for the scheme. These provable security claims were undermined in [18], however, with the presentation of a key recovery attack on a full scale version of the Cubic Simple Matrix encryption scheme. The complexity of the attack was on the order of q^{s+2} for characteristic $p > 3$, q^{s+3} for characteristic 3, and q^{2s+6} for characteristic 2. Here s is the dimension of the matrices in the scheme, and q is the cardinality of the finite field used. This technique was an extension and augmentation of the technique of [17], and similarly exploited a differential invariant property of the core map to perform a key recovery attack. Nonetheless, the much higher complexity of this attack for characteristic 2 left open the possibility that there may be some security advantage to using a cubic ABC map over a field with characteristic 2.

In this paper, we present an attack whose complexity is on the order of q^{s+2} for all characteristics. Similar techniques can also improve the complexity of attacks against characteristic 2 parameters for the original quadratic version of the ABC cryptosystem, from q^{s+4} (reported in [17]) to q^{s+2}.

Specifically, our technique improves the complexity of attacking CubicABC($q = 2^8$,$s = 7$), designed for 80-bit security, from the horrendous value of 2^{177} in [18] to approximately 2^{88} operations, the same as the direct algebraic attack complexity reported in [14]. More convincing is our attack on CubicABC($q = 2^8$,$s = 8$), designed for 100-bit security. We break the scheme in approximately 2^{98} operations. Furthermore, the attack is fully parallelizable and requires very little memory; hence, our technique is asymptotically far more efficient than algebraic attacks, the basis for the original security estimation. Thus, the security claims in [14] not only fail to hold in the odd characteristic case, they fail to hold in characteristic two as well.

The paper is organized as follows. In the next section, we present the structure of the Cubic ABC Simple Matrix encryption scheme. In the following section, the fingerprint of the matrix algebra used in the construction of the ABC scheme is exposed. In the subsequent section, the effect of this structure on minrank calculations is determined. We then calculate the complexity of the full attack including the linear algebra steps required for full key recovery. Finally, we review these results and discuss the security of the Cubic ABC scheme and its quadratic counterpart moving forward.

2 The Cubic ABC Matrix Encryption Scheme

In [14], the Cubic ABC Matrix encryption scheme is proposed. The motivation behind the scheme is to use a large matrix algebra over a finite field to construct an easily invertible cubic map. The construction uses matrix multiplication to combine random linear and quadratic formulae into cubic formulae in a way that allows a user with knowledge of the structure of the matrix algebra and the polynomial isomorphism used to compose the scheme to invert the map.

Let $k = \mathbb{F}_q$ be a finite field. Linear forms and variables over k will be denoted with lower case letters. Vectors of any dimension over k will be denoted with bold font, $\mathbf{v}$. Fix $s \in \mathbb{N}$ and set $n = s^2$ and $m = 2s^2$. An element of a matrix ring $M_d(k)$ or the linear transformations they represent, will be denoted by upper case letters, such as M. When the entries of the matrix are being considered functions of a variable, the matrix will be denoted $M(\mathbf{x})$. Let $\phi : M_{s\times 2s}(k) \to k^{2s^2}$ represent the vector space isomorphism sending a matrix to the column vector consisting of the concatenation of its rows. The output of this map, being a vector, will be written with bold font; however, to indicate the relationship to its matrix preimage, it will be denoted with an upper case letter, such as $\mathbf{M}$.

The scheme utilizes an isomorphism of polynomials to hide the internal structure. Let $\mathbf{x} = \left[x_1, x_2, \ldots, x_n\right]^\top \in k^n$ denote plaintext while $\mathbf{y} = \left[y_1, \ldots, y_m\right] \in k^m$ denotes ciphertext. Fix two invertible linear transformations $T \in M_m(k)$ and $U \in M_n(k)$. (One may use affine transformations, but there is no security or performance benefit in doing so.) Denote the input and output of the central map by $\mathbf{u} = U\mathbf{x}$ and $\mathbf{v} = T^{-1}(\mathbf{y})$.

The construction of the central map is as follows. Define three $s \times s$ matrices A, B, and C in the following way:

$$A = \begin{bmatrix} p_1 & p_2 & \cdots & p_s \\ p_{s+1} & p_{s+2} & \cdots & p_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ p_{s^2-s+1} & p_{s^2-s+2} & \cdots & p_{s^2} \end{bmatrix}, B = \begin{bmatrix} b_1 & b_2 & \cdots & b_s \\ b_{s+1} & b_{s+2} & \cdots & b_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ b_{s^2-s+1} & b_{s^2-s+2} & \cdots & b_{s^2} \end{bmatrix},$$

and

$$C = \begin{bmatrix} c_1 & c_2 & \cdots & c_s \\ c_{s+1} & c_{s+2} & \cdots & c_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ c_{s^2-s+1} & c_{s^2-s+2} & \cdots & c_{s^2} \end{bmatrix}.$$

Here the p_i are quadratic forms on $\mathbf{u}$ chosen independently and uniformly at random from among all quadratic forms and the b_i and c_i are linear forms on $\mathbf{u}$ chosen independently and uniformly at random from among all linear forms.

We define two $s \times s$ matrices $E_1 = AB$ and $E_2 = AC$. Since A is quadratic and B and C are linear in u_i, E_1 and E_2 are cubic in the u_i. The central map $\mathcal{E}$ is defined by

$$\mathcal{E} = \phi \circ (E_1 || E_2).$$

Thus $\mathcal{E}$ is an m dimensional vector of cubic forms in $\mathbf{u}$. Finally, the public key is given by $\mathcal{F} = T \circ \mathcal{E} \circ U$.

Encryption with this system is standard: given a plaintext $(x_1, \ldots, x_n)$, compute $(y_1, \ldots, y_m) = \mathcal{F}(x_1, \ldots, x_n)$. Decryption is somewhat more complicated.

To decrypt, one inverts each of the private maps in turn: apply T^{-1}, invert $\mathcal{E}$, and apply U^{-1}. To "invert" $\mathcal{E}$, one assumes that $A(\mathbf{u})$ is invertible, and forms a matrix

$$A^{-1}(\mathbf{u}) = \begin{bmatrix} w_1 & w_2 & \cdots & w_s \\ w_{s+1} & w_{s+2} & \cdots & w_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ w_{s^2-s+1} & w_{s^2-s+2} & \cdots & w_{s^2} \end{bmatrix},$$

where the w_i are indeterminants. Then collectinging the relations $A^{-1}(\mathbf{u})E_1(\mathbf{u}) = B(\mathbf{u})$ and $A^{-1}(\mathbf{u})E_2(\mathbf{u}) = C(\mathbf{u})$, we have $m = 2s^2$ linear equations in $2n = 2s^2$ unknowns w_i and u_i. Using, for example, Gaussian elimination one can eliminate all of the variables w_i and most of the u_i. The resulting relations can be substituted back into $E_1(\mathbf{u})$ and $E_2(\mathbf{u})$ to obtain a large system of equations in very few variables which can be solved efficiently in a variety of ways.

3 The Structure of the Cubic ABC scheme

3.1 Column Band Spaces

Each component of the central $\mathcal{E}(\mathbf{u}) = E_1(\mathbf{u})||E_2(\mathbf{u})$ map may be written as:

$$\mathcal{E}_{(i-1)s+j} = \sum_{l=1}^{s} p_{(i-1)s+l} b_{(l-1)s+j},$$

for the E_1 equations, and likewise, for the E_2 equations:

$$\mathcal{E}_{s^2+(i-1)s+j} = \sum_{l=1}^{s} p_{(i-1)s+l} c_{(l-1)s+j}$$

where i and j run from 1 to s.

Consider the s sets of s polynomials that form the columns of E_1, i.e. for each $j \in \{1, \ldots, s\}$ consider $(\mathcal{E}_j, \mathcal{E}_{s+j}, \ldots, \mathcal{E}_{s^2-s+j})$. With high probability, the linear forms $b_j, b_{s+j}, \ldots, b_{s^2-s+j}$ are linearly independent, and if so the polynomials may be re-expressed, using a linear change of variables to $(u'_1, \ldots u'_{s^2})$ where $u'_i = b_{(i-1)s+j}$ for $i = 1, \ldots, s$. After the change of variables, the only cubic monomials contained in $(\mathcal{E}_j, \mathcal{E}_{s+j}, \ldots, \mathcal{E}_{s^2-s+j})$ will be those containing at least one factor of $u'_1, \ldots, u'_s$. We can make a similar change of variables to reveal structure in the s sets of s polynomials that form the columns of E_2: Setting $u'_i = c_{(i-1)s+j}$ for $i = 1, \ldots, s$ and a fixed j, the only cubic monomials contained in $(\mathcal{E}_{s^2+j}, \mathcal{E}_{s^2+s+j}, \ldots, \mathcal{E}_{2s^2-s+j})$ will be those containing at least one factor of $u'_1, \ldots, u'_s$.

More generally, we can make a similar change of variables to reveal structure in any of a large family of s dimensional subspaces of the span of the component polynomials of E_1 and E_2, which we will call column band spaces in analogy to the band spaces used to analyze the quadratic ABC cryptosystem in [17]. Each family is defined by a fixed linear combination, (β, γ), of the columns of E_1 and E_2:

Definition 1 *The column band space defined by the $2s$-dimensional linear form (β, γ) is the space of cubic maps, $\mathcal{B}_{\beta,\gamma}$, given by:*

$$\mathcal{B}_{\beta,\gamma} = Span(\mathcal{E}_{\beta,\gamma,1}, \ldots, \mathcal{E}_{\beta,\gamma,s}),$$

where

$$\mathcal{E}_{\beta,\gamma,i} = \sum_{j=1}^{s} (\beta_j \mathcal{E}_{(i-1)s+j} + \gamma_j \mathcal{E}_{s^2+(i-1)s+j})$$

$$= \sum_{l=1}^{s} \left(p_{(i-1)s+l} \sum_{j=1}^{s} \left(\beta_j b_{(l-1)s+j} + \gamma_j c_{(l-1)s+j} \right) \right).$$

Note that under a change of variables

$$(x_1, \ldots, x_{s^2}) \xmapsto{M} (u'_1, \ldots u'_{s^2}), \text{ where } u'_i = \sum_{j=1}^{s} \left(\beta_j b_{(i-1)s+j} + \gamma_j c_{(i-1)s+j} \right) \text{ for } i = 1, \ldots, s,$$

the only cubic monomials contained in the elements of $\mathcal{B}_{\beta,\gamma}$ will be those containing at least one factor of $u'_1, \ldots, u'_s$.

In such a basis, the third formal derivative, or the 3-tensor of third partial derivatives

$$D^3\mathcal{E} = \sum_{i,j,k} \frac{\partial^3 \mathcal{E}}{\partial u'_i \partial u'_j \partial u'_k} du'_i \otimes du'_j \otimes du'_k,$$

of any map $\mathcal{E} \in \mathcal{B}_{\beta,\gamma}$ has a special block form, see Figure 1. This tensor is the same as the one used for the attack in [18], although in that case it was computed using the discrete differential. There are, however, a number of disadvantages to using this 3-tensor to represent the structural features of cubic ABC. In particular, when defined over a field of characteristic 2, the symmetry of the 3-tensor results in the loss of any information about coefficients for monomials of the form $x_i^2 x_j$, since the 3rd derivatave of such a monomial is always 0. We will therefore use a different tool to express the structure of cubic ABC.

Using the same u' basis as above, we see that the gradient $\nabla_{u'}\mathcal{E}$ produces a covector of quadratic forms, which can be though of as a quadratic map that takes any vector w of the form

$$(0, \ldots, 0, u'_{s+1}(\mathbf{w}), \ldots, u'_{s^2}(\mathbf{w}))^\top,$$

to a covector of the form

$$(y(u'_1), \ldots, y(u'_s), 0, \ldots, 0).$$

Note that, by the chain rule, we can relate $\nabla_{u'}\mathcal{E} = \left[\frac{\partial \mathcal{E}}{\partial u'_1}, \ldots, \frac{\partial \mathcal{E}}{\partial u'_{s^2}} \right]$ to the formal derivative defined over the public basis:

$$\nabla \mathcal{E} = \left[\frac{\partial \mathcal{E}}{\partial x_1}, \ldots, \frac{\partial \mathcal{E}}{\partial x_{s^2}} \right] = \nabla_{u'}\mathcal{E} \left[\frac{du'_j}{dx_i} \right]_{i,j}$$

using the nonsingular change of basis matrix whose entries are $\frac{du'_j}{dx_i}$. We can therefore conclude that even defined over the public basis, the first formal derivative of any map $\mathcal{E} \in \mathcal{B}_{\beta,\gamma}$ is a quadratic map that takes an $s^2 - s$ dimensional space of vectors to an s dimensional space of covectors.

We will define the term "band kernel" to describe this $s^2 - s$ dimensional space of vectors (including $\mathbf{w}$) which are mapped to an s dimensonal image space by the first formal derivative of $\mathcal{E}$.

Definition 2 *The band kernel of $\mathcal{B}_{\beta,\gamma}$, denoted $\mathcal{BK}_{\beta,\gamma}$, is the space of vectors x, such that*

$$u'_i = \sum_{j=1}^{s} \beta_j b_{(i-1)s+j}(x) + \gamma_j c_{(i-1)s+j}(x) \ = 0,$$

for $i = 1, \ldots, s$.

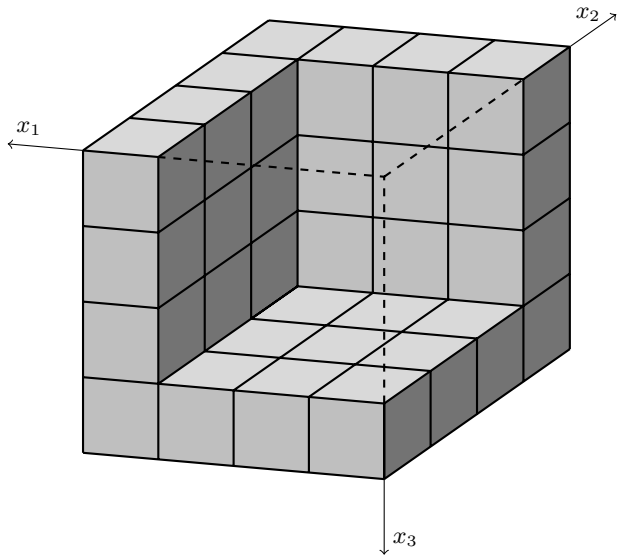

Fig. 1. 3-tensor structure of the third formal derivative of a band space map. Solid regions correspond to nonzero coefficients. Transparent regions correspond to zero coefficients.

4 A Variant of MinRank Exploiting the Column Band Space Structure

A minrank-like attack may be used to locate the column band space maps defined in the previous section. In this case, the attack proceeds by selecting s^2-dimensional vectors $\mathbf{w}_1$ and $\mathbf{w}_2$, setting

$$\begin{aligned} \sum_{i=1}^{2s^2} t_i \nabla \mathcal{E}_i(\mathbf{w}_1) = 0, \\ \sum_{i=1}^{2s^2} t_i \nabla \mathcal{E}_i(\mathbf{w}_2) = 0, \end{aligned} \tag{1}$$

and then solving for the t_i. The attack succeeds when $\sum_{i=1}^{2s^2} t_i \mathcal{E}_i \in \mathcal{B}_{\beta,\gamma}$, and $\mathbf{x}_1$ and $\mathbf{x}_2$ are within the corresponding band kernel. If these conditions are met, then the 2-tensors

$$\sum_{i=1}^{2s^2} t_i \mathbf{H}(\mathcal{E}_i)(\mathbf{w}_1) \text{ and } \sum_{i=1}^{2s^2} t_i \mathbf{H}(\mathcal{E}_i)(\mathbf{w}_2),$$

will have rank at most $2s$, and this will be easily detectable. Here $\mathbf{H}(\mathcal{E}_i)$ is the Hessian matrix

$$\mathbf{H}(\mathcal{E}_i) := \begin{bmatrix} \frac{\partial^2 \mathcal{E}_i}{\partial x_1^2} & \frac{\partial^2 \mathcal{E}_i}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 \mathcal{E}_i}{\partial x_1 \partial x_n} \\ \frac{\partial^2 \mathcal{E}_i}{\partial x_1 \partial x_2} & \frac{\partial^2 \mathcal{E}_i}{\partial x_2^2} & \cdots & \frac{\partial^2 \mathcal{E}_i}{\partial x_1 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 \mathcal{E}_i}{\partial x_n \partial x_1} & \frac{\partial^2 \mathcal{E}_i}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 \mathcal{E}_i}{\partial x_n^2} \end{bmatrix}.$$

Theorem 1 *The probability that 2 randomly chosen vectors, $\mathbf{w}_1$ and $\mathbf{w}_2$, are both in the band kernel of some band space $\mathcal{B}_{\beta,\gamma}$ is approximately $\frac{1}{q-1}$.*

Proof. The condition that the $\mathbf{w}_1$ and $\mathbf{w}_2$ are contained within a band kernel is that there be a nontrivial linear combination of the columns of the following matrix which is equal to zero (i.e. that the matrix has nonzero column corank):

$$\left[\begin{array}{cccc|cccc} b_1(\mathbf{w}_1) & b_2(\mathbf{w}_1) & \dots & b_s(\mathbf{w}_1) & c_1(\mathbf{w}_1) & c_2(\mathbf{w}_1) & \dots & c_s(\mathbf{w}_1) \\ b_{s+1}(\mathbf{w}_1) & b_{s+2}(\mathbf{w}_1) & \dots & b_{2s}(\mathbf{w}_1) & c_{s+1}(\mathbf{w}_1) & c_{s+2}(\mathbf{w}_1) & \dots & c_{2s}(\mathbf{w}_1) \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ b_{s^2-s+1}(\mathbf{w}_1) & b_{s^2-s+2}(\mathbf{w}_1) & \dots & b_{s^2}(\mathbf{w}_1) & c_{s^2-s+1}(\mathbf{w}_1) & c_{s^2-s+2}(\mathbf{w}_1) & \dots & c_{s^2}(\mathbf{w}_1) \\ \hline b_1(\mathbf{w}_2) & b_2(\mathbf{w}_2) & \dots & b_s(\mathbf{w}_2) & c_1(\mathbf{w}_2) & c_2(\mathbf{w}_2) & \dots & c_s(\mathbf{w}_2) \\ b_{s+1}(\mathbf{w}_2) & b_{s+2}(\mathbf{w}_2) & \dots & b_{2s}(\mathbf{w}_2) & c_{s+1}(\mathbf{w}_2) & c_{s+2}(\mathbf{w}_2) & \dots & c_{2s}(\mathbf{w}_2) \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ b_{s^2-s+1}(\mathbf{w}_2) & b_{s^2-s+2}(\mathbf{w}_2) & \dots & b_{s^2}(\mathbf{w}_2) & c_{s^2-s+1}(\mathbf{w}_2) & c_{s^2-s+2}(\mathbf{w}_2) & \dots & c_{s^2}(\mathbf{w}_2) \end{array}\right].$$

The matrix is a uniformly random $2s \times 2s$ matrix, which has nonzero column corank with probability approximately $\frac{1}{q-1}$. □

Theorem 2 *If $\mathbf{w}_1$ and $\mathbf{w}_2$ are chosen in such a way that they are both in the band kernel of a column band space $\mathcal{B}_{\beta,\gamma}$, and they are linearly independent from one another and statistically independent from the private quadratic forms, $p_{(i-1)s+j}$ in the matrix A, then $\mathbf{w}_1$ and $\mathbf{w}_2$ are both in the kernel of the first formal derivative of some column band space map, $\mathcal{E} = \sum_{\mathcal{E}_{\beta,\gamma,i} \in \mathcal{B}_{\beta,\gamma}} \tau_i \mathcal{E}_{\beta,\gamma,i}$ with probability approximately $\frac{1}{(q-1)q^s}$.*

Proof. An $\mathcal{E}$ meeting the above condition exists iff there is a nontrivial solution to the following system of equations

$$\begin{aligned} \sum_{\mathcal{E}_{\beta,\gamma,i} \in \mathcal{B}_{\beta,\gamma}} \tau_i \nabla \mathcal{E}_{\beta,\gamma,i}(\mathbf{w}_1) = 0, \\ \sum_{\mathcal{E}_{\beta,\gamma,i} \in \mathcal{B}_{\beta,\gamma}} \tau_i \nabla \mathcal{E}_{\beta,\gamma,i}(\mathbf{w}_2) = 0. \end{aligned} \tag{2}$$

We may express our band space maps in a basis (e.g. the u_i' basis used in Definition 2) where the first s basis vectors are chosen to be outside the band kernel, and the remaining $s^2 - s$ basis vectors are chosen from within the band kernel. Combining this with Definition 1, we see that the band space maps can be written as

$$\mathcal{E}_{\beta,\gamma,i} = \sum_{j=1}^{s} p_{(i-1)s+j} u_j'.$$

Note that $\mathbf{w}_1$ and $\mathbf{w}_2$ are band kernel vectors, and so for both vectors we have that $u'_j = 0$ for $j = 1, \ldots, s$. Therefore, in such a basis, the only formal derivatives of $\mathcal{E}$ that can be nonzero are $\frac{\partial \mathcal{E}}{\partial u'_j} = p_{(i-1)s+j}$ for $j = 1, \ldots, s$. Thus in order for there to be a nontrivial solution to Equation (2), it is necessary and sufficient that $\sum_{i=1}^{s} \tau_i p_{(i-1)s+j}(\mathbf{w}_k) = 0$ for $j = 1, \ldots, s$ and $k = 1, 2$. This condition will be satisfied if and only if the following $2s \times s$ matrix has nonzero column corank:

$$\left[\begin{array}{cccc} p_1(\mathbf{w}_1) & p_{s+1}(\mathbf{w}_1) & \cdots & p_{s^2-s+1}(\mathbf{w}_1) \\ p_2(\mathbf{w}_1) & p_{s+2}(\mathbf{w}_1) & \cdots & p_{s^2-s+2}(\mathbf{w}_1) \\ \vdots & \vdots & \ddots & \vdots \\ p_s(\mathbf{w}_1) & p_{2s}(\mathbf{w}_1) & \cdots & p_{s^2}(\mathbf{w}_1). \\ \hline p_1(\mathbf{w}_2) & p_{s+1}(\mathbf{w}_2) & \cdots & p_{s^2-s+1}(\mathbf{w}_2) \\ p_2(\mathbf{w}_2) & p_{s+2}(\mathbf{w}_2) & \cdots & p_{s^2-s+2}(\mathbf{w}_2) \\ \vdots & \vdots & \ddots & \vdots \\ p_s(\mathbf{w}_2) & p_{2s}(\mathbf{w}_2) & \cdots & p_{s^2}(\mathbf{w}_2) \end{array}\right].$$

This matrix is a random matrix over $k = \mathbb{F}_q$, which has nonzero column corank with probability approximately $\frac{1}{(q-1)q^s}$, for practical parameters. □

Combining the results of Theorems 1 and 2, we find that for a random choice of the vectors $\mathbf{w}_1$ and $\mathbf{w}_2$, there is a column band space map among the solutions of Equation (1) with probability approximately $\frac{1}{(q-1)^2 q^s}$. It may be somewhat undesirable to choose $\mathbf{w}_1$ and $\mathbf{w}_1$ completely randomly, however. The naïve algorithm for constructing the coefficients of Equation (1) for a random choice of $\mathbf{w}_1$ and $\mathbf{w}_2$ requires on the order of s^8 field operations. This can be reduced to s^6 operations if we make sure that each new choice of $\mathbf{w}_1$ and $\mathbf{w}_2$ differs from the previous choice at only a single coordinate. Then, rather than recomputing Equation (1) from scratch, we can use the previous values of the coefficients and we will only need to include corrections for the monomials that contain the variable that was changed from the previous iteration. Over a large number of iterations, the distribution of $\mathbf{w}_1$ and $\mathbf{w}_2$ should still be sufficiently close to random that the probability of success for the attack will not be meaningfully altered.

One final factor which may increase the cost of attacks is the expected dimension of the solution space of Equation (1). If this space has a high dimension, then the attack will be slowed down since the attacker much search through a large number of spurious solutions to find a real solution (i.e. one where $\sum_{i=1}^{2s^2} t_i \mathbf{H}(\mathcal{E}_i)(\mathbf{w}_l)$ has rank at most $2s$ for $l = 1, 2$). Fortunately, Equation (1) is a system of $2s^2$ equations in $2s^2$ variables and it generally has a 0-dimensional space of solutions. The lone exception occurs for characteristic 3. In this case, there are two linear dependencies among the equations, given by $\mathbf{w}_1 \left[\nabla \mathcal{E}_i(\mathbf{w}_1)\right]^\top = 0$ and $\mathbf{w}_2 \left[\nabla \mathcal{E}_i(\mathbf{w}_2)\right]^\top = 0$. In this situation we would therefore expect a 2-dimensional solution space. We can, however, recover two additional linear constraints on the t_i's by also requiring:

$$\sum_{i=1}^{2s^2} t_i \mathcal{E}_i(\mathbf{w}_l) = 0, \text{ for } l = 1, 2.$$

When these additional linear constraints are added to those given by Equation (1), the expected dimension of the solution space drops back to 0. We can therefore assess the cost of the above attack at approximately $s^6 q^{s+2}$, regardless of the characteristic.

5 Application to the Quadratic ABC Scheme

A similar technique was used to attack the original quadratic version of the ABC cryptosytem in [17]. While this technique was expressed in terms of the discrete differential, it can also be

expressed using the formal derivative. In that case, the attack proceeds by selecting two random vectors $\mathbf{w}_1$ and $\mathbf{w}_2$, and solving an equation identical to Equation (1) for t_i, where the $\mathcal{E}_i$ are quadratic rather than cubic. The attack succeeds when $\sum_{i=1}^{2s^2} t_i \mathbf{H}(\mathcal{E}_i)$ has low rank.

When this attack is applied to parameters chosen over a field with characteristic 2, it is less efficient for the same reason as the basic attack given in the previous section is less efficient for the characteristic 3 parameters: the $2s^2$ linear equations given by Equation (1) have three linear dependencies given by $\mathbf{w}_1 \left[\nabla \mathcal{E}_i(\mathbf{w}_1)\right]^\top = 0$, $\mathbf{w}_2 \left[\nabla \mathcal{E}_i(\mathbf{w}_2)\right]^\top = 0$, and $\mathbf{w}_1 \left[\nabla \mathcal{E}_i(\mathbf{w}_2)\right]^\top + \mathbf{w}_2 \left[\nabla \mathcal{E}_i(\mathbf{w}_1)\right]^\top = 0$, and the attacker must generally search through a 3-dimensional solution space of spurious solutions in order to find a 1-dimensional space of useful solutions. As a result, the complexity of the attack for characteristic 2 is $s^{2\omega} q^{s+4}$, instead of $s^{2\omega} q^{s+2}$, as it is for all other characteristics. ($\omega \approx 2.373$ is the linear algebra constant.)

However, just as with cubic ABC parameters of characteristic 3, we can add two additional linear constraints and reduce the expected dimension of the solution space to 1:

$$\sum_{i=1}^{2s^2} t_i \mathcal{E}_i(\mathbf{w}_l) = 0, \text{ for } l = 1, 2.$$

Thus, we can also reduce the attack complexity for quadratic ABC parameters with characteristic 2 to $s^{2\omega} q^{s+2}$.

6 Completing the Key Recovery

Once the MinRank instance is solved, key extraction proceeds in a similar manner to [18, Section 6] in the cubic case and [17, Section 6]. Here we discuss the cubic version.

First, note that U is not a critical element of the scheme. If A is a random matrix of quadratic forms and B and C are random matrices of linear forms, then so are $A \circ U$, $B \circ U$ and $C \circ U$ for any full rank map U. Thus, since $T \circ \phi(AB||AC) \circ U = T \circ \phi((A \circ U)(B \circ U)||(A \circ U)(C \circ U))$, we may absorb the action of U into A, B, and C, and consider the public key to be of the form

$$P(\mathbf{x}) = T \circ \phi(AB||AC)(\mathbf{x}).$$

Let $\mathcal{E} \in \mathcal{B}_{\beta,\gamma}$, and consider $\mathbf{H}(\mathcal{E})$. For $\mathbf{w}_1$ and $\mathbf{w}_2$ in the band kernel corresponding to $\mathcal{B}_{\beta,\gamma}$, there is a basis in which both $\mathbf{H}(\mathcal{E})(\mathbf{w}_1)$ and $\mathbf{H}(\mathcal{E})(\mathbf{w}_2)$ have the form illustrated in Figure 2. Thus, for $s \geq 3$, with high probability the kernels of both maps are contained in the corresponding band kernel $\mathcal{B}_{\beta,\gamma}$, and $\text{span}\{\ker(\mathbf{H}(\mathcal{E})(\mathbf{w}_1), \ker(\mathbf{H}(\mathcal{E})(\mathbf{w}_2)\} = \mathcal{B}_{\beta,\gamma}$.

Given the basis for an $s^2 - s$ dimensional band kernel $\mathcal{BK}$, we may choose a basis $\{v_1, \ldots, v_s\}$ for the subspace of the dual space vanishing on $\mathcal{BK}$. We can also find a basis $\mathcal{E}_{v_1}, \ldots, \mathcal{E}_{v_s}$ for the band space itself by solving the linear system

$$\begin{aligned}
\sum_{\mathcal{E}_i} \tau_i \mathcal{E}_i(\mathbf{w}_1) &= 0, \\
\sum_{\mathcal{E}_i} \tau_i \mathcal{E}_i(\mathbf{w}_2) &= 0, \\
\vdots &= \vdots \\
\sum_{\mathcal{E}_i} \tau_i \mathcal{E}_i(\mathbf{w}_t) &= 0,
\end{aligned}$$

where $t \approx 2s^2$ and $\mathbf{w}_i$ is in the band kernel.

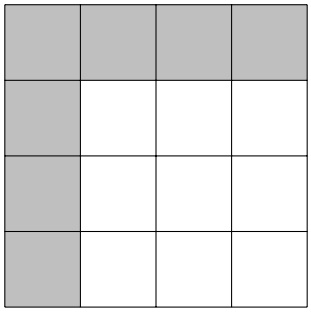

Fig. 2. Structure of $\mathbf{H}(\mathcal{E})(\mathbf{w})$ when $\mathcal{E} \in \mathcal{B}_{\beta,\gamma}$ and $\mathbf{w}$ is in the band kernel corresponding to the band space $\mathcal{B}_{\beta,\gamma}$. The shaded region corresponds to nonzero coefficients.

Since the basis $\mathcal{E}_{v_1}, \ldots, \mathcal{E}_{v_s}$ is in a single band space, there exists an element $\begin{bmatrix} b'_1 \cdots b'_s \end{bmatrix}^\top$ in ColumnSpace$(B||C)$, and two matrices Ω_1 and Ω_2 such that

$$\Omega_1 A \left(\Omega_2 \begin{bmatrix} b'_1 \\ \vdots \\ b'_s \end{bmatrix} \right) =: A' \left(\begin{bmatrix} v_1 \\ \vdots \\ v_s \end{bmatrix} \right) = \begin{bmatrix} \mathcal{E}_{v_1} \\ \vdots \\ \mathcal{E}_{v_s} \end{bmatrix}.$$

Solving the above system of equations over $\mathbb{F}_q[x_1, \ldots, x_{s^2}]$ uniquely determines A' in the quotient $\mathbb{F}_q[x_1, \ldots, x_{s^2}]/\langle v_1, \ldots, v_s \rangle$. To recover all of A', note that the above system is part of an equivalent key

$$\mathcal{F} = T' \circ A'(B'||C')$$

where $\begin{bmatrix} v_1 \cdots v_s \end{bmatrix}^\top$ is the first column of B'.

Applying T'^{-1} to both sides and inserting the information we know we may construct the system

$$A'(B'||C') = T'^{-1}\mathcal{F}. \tag{3}$$

Solving this system of equations modulo $\langle v_1, \ldots, v_s \rangle$ for B', C' and T'^{-1} we can recover a space of solutions, which we will restrict by arbitrarily fixing the value of T'^{-1}. Note that the elements of T'^{-1} are constant polynomials, and therefore $T'^{-1} (\mathrm{mod}\, \langle v_1, \ldots, v_s \rangle)$ is the same as T'^{-1}. Thus, for any choice of T'^{-1} in this space, the second column of $T'^{-1}\mathcal{F}$ is a basis for a band space. Moreover, the elements $v'_{s+1}, \ldots, v'_{2s}$ of the second column of $B' (\mathrm{mod}\, \langle v_1, \ldots, v_s \rangle)$ are the image, modulo $\langle v_1, \ldots, v_s \rangle$, of linear forms vanishing on the corresponding band kernel. Therefore, we obtain the equality

$$\left(\bigcap_{i=1}^{s} \ker(v_i) \right) \bigcap \left(\bigcap_{i=s+1}^{2s} \ker(v'_i) \right) = \mathcal{BK}_2 \cap \mathcal{BK}_1,$$

the intersection of the band kernels of our two band spaces.

We can reconstruct the full band kernel of this second band space using the same method we used to obtain our first band kernel. We take a map $\mathcal{E}_2$ from the second column of $T'^{-1}\mathcal{F}$, and two vectors $\mathbf{w}_a$ and $\mathbf{w}_b$ from $\mathcal{BK}_2 \cap \mathcal{BK}_1$, and we compute $\mathcal{BK}_2 = \mathrm{span}\{\ker(\mathbf{H}(\mathcal{E}_2)(\mathbf{w}_a) \cup \ker(\mathbf{H}(\mathcal{E}_2)(\mathbf{w}_b)\}$. We can now solve for the second column of B', $\begin{bmatrix} v_{s+1} \cdots v_{2s} \end{bmatrix}^\top$, uniquely over $\mathbb{F}_q[x_1, \ldots, x_{s^2}]$ (NOT modulo $\langle v_1, \ldots, v_s \rangle$) by solving the following system of linear equations:

$$\begin{aligned} v_i \equiv v_i' \bmod \langle v_1, \ldots, v_s \rangle, \\ v_i(\mathbf{w}_1) = 0, \\ v_i(\mathbf{w}_2) = 0, \\ \vdots = \vdots \\ v_i(\mathbf{w}_{s^2-s}) = 0, \end{aligned}$$

where $i = s+1, \ldots, 2s$, and $\{\mathbf{w}_1, \ldots, \mathbf{w}_{s^2-s}\}$ is a basis for $\mathcal{BK}_2$. We can now solve for A' (again, uniquely over $\mathbb{F}_q[x_1, \ldots, x_{s^2}]$) by solving:

$$A'\left(\begin{bmatrix} v_1 \\ \vdots \\ v_s \end{bmatrix}\right) \equiv \begin{bmatrix} \mathcal{E}_{v_1} \\ \vdots \\ \mathcal{E}_{v_s} \end{bmatrix} \bmod \langle v_1, \ldots, v_s \rangle,$$

$$A'\left(\begin{bmatrix} v_{s+1} \\ \vdots \\ v_{2s} \end{bmatrix}\right) \equiv \begin{bmatrix} \mathcal{E}_{v_{s+1}} \\ \vdots \\ \mathcal{E}_{v_{2s}} \end{bmatrix} \bmod \langle v_{s+1}, \ldots, v_{2s} \rangle,$$

where $\begin{bmatrix} \mathcal{E}_{v_{s+1}} & \cdots & \mathcal{E}_{v_{2s}} \end{bmatrix}^\top$ is the second column of $T'^{-1}\mathcal{F}$. This allows us to solve Equation (3) for the rest of B' and C', completing the attack.

The primary cost of the attack involves finding the band space map. The rest of the key recovery is additive in complexity and dominated by the band space map recovery; thus the total complexity of the attack is of the same order as the band space map recovery. Hence, the cost of private key extraction is approximately $q^{s+2}s^6$ for all characteristics.

The original parameters of Cubic ABC were designed for a security level of 80-bits and 100-bits. Since NIST has been recommending a security level of 112-bits since 2015, see [19], these figures may be a bit out of date. In fact, our attack seems more effective for larger parameter sets than small.

We note that our attack breaks CubicABC($q = 2^8, s = 7$), designed for 80-bit security, in approximately 2^{88} operations. More convincingly, our attack breaks CubicABC($q = 2^8, s = 8$), designed for 100-bit security, in approximately 2^{98} operations, indicating that for parameters as small as these, we have already crossed the threshold of algebraic attack efficiency. Furthermore, the attack is fully parallelizable and requires very little memory. Hence, this technique is asymptotically far more efficient than algebraic attacks, the basis for the original security estimation in [14].

In the case of the quadratic ABC scheme, the original 86-bit secure parameters ABC($q = 2^8, s = 8$). The attack complexity with the new methodology presented here is 2^{87}, just above the claimed level. We note, however, that the authors of [13] supplied additional parameters using odd characteristic in their presentation at PQCRYPTO 2013, see [20], with a claimed security level of 108-bits. This scheme, ABC($q = 127, s = 8$) offers resistance only to the level of 2^{77} to our slight improvement in technique over that of [17]. Thus, our attack definitively breaks these parameters.

7 Experiments

Using SAGE [21], we performed some experiments as a sanity check to confirm the efficiency of our ideas on small scale variants of the Cubic ABC scheme. The computer used has a 64 bit

quad-core Intel i7 processor, with clock cycle 2.8 GHz. Rather than considering the full attack, we were most interested in confirming our complexity estimates on the most costly step in the attack, the MinRank instance. Given as input the finite field size q, and the scheme parameter s, we computed the average number of vectors v required to be sampled in order for the rank of the 2-tensor $\mathbf{H}(\mathcal{E})(v)$ to fall to $2s$. As explained in Section 4, when the rank falls to this level, we have identified the subspace differential invariant structure of the scheme which can then be exploited to attack the scheme.

As this paper is only concerned with binary fields, we ran experiments with $q = 2, 4$ and 8. We found that for $s = 3$ and $q = 2, 4$, or 8, with high probability only a single vector was needed before the rank fell to $2s$. For $s = 4$ and $s = 5$, the computations were only feasible in SAGE for $q = 2$ and $q = 4$. The average values obtained are presented in the table below. Note that for $q = 4$ and $s = 5$ the average value is based on a small number of samples as the computation time was quite lengthy.

	$s=4$	$(q-1)^2q^s$	$s=5$	$(q-1)^2q^s$
$q=2$	24	16	35	32
$q=4$	1962	2304	7021	9216

Table 1. Average number of vectors needed for the rank to fall to $2s$ versus the predicted values.

In comparison, our previous experiments [18] were only able to obtain data for $q = 2$ and $s = 4, 5$. The average number of vectors needed in the $s = 4$ case was 244, while for $s = 5$, the average number in our experiments was 994 (with the predicted values being 256 and 1024).

8 Conclusion

The ABC schemes offer an interesting new technique for the construction of multivariate public key schemes. Previously, we have used the multiplicative structure of an extension field to generate an efficiently invertible map. Schemes built on such a construct are known as "big field" schemes. The ABC framework is essentially a "large structure" or perhaps "large algebra" scheme, depending on multiplication from a matrix algebra over the base field. Since the only simple algebras are either matrix algebras or field extensions, we seem to have exhausted the possibilities. Interestingly, MinRank techniques seem optimal in this setting, at least asymptotically in the dimension of the extension.

Also interesting to note is the fact that the authors present in [14] a heuristic security argument for the provable security of the scheme and reinforce the notion of provable security in this venue at the presentation of the scheme at [22]. Unfortunately, this analysis does not contribute a sound conclusion, as demonstrated by the methodology of [18]. With our improved attack, we rule out the possibility that the cubic variant of ABC offers any security advantage over the original quadratic scheme. Likewise, our improved attack on quadratic ABC eliminates any security benefit associated with characteristic-2 parameters in the quadratic case.

References

1. Shor, P.W.: Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. SIAM J. Sci. Stat. Comp. **26, 1484** (1997)
2. Feynman, R.P.: Simulating physics with computers. Int. J. Theor. Phys. **21** (1982) 467–488

3. Group, C.T.: Submission requirements and evaluation criteria for the post-quantum cryptography standardization process. NIST CSRC (2016) http://csrc.nist.gov/groups/ST/post-quantum-crypto/documents/call-for-proposals-final-dec-2016.pdf.
4. Kipnis, A., Patarin, J., Goubin, L.: Unbalanced oil and vinegar signature schemes. EUROCRYPT 1999. LNCS **1592** (1999) 206–222
5. Patarin, J., Goubin, L., Courtois, N.: C $^{*}_{-+}$ and HM: Variations around two schemes of T.Matsumoto and H.Imai. Asiacrypt 1998, Springer **1514** (1998) 35–49
6. Patarin, J., Courtois, N., Goubin, L.: Quartz, 128-bit long digital signatures. In Naccache, D., ed.: CT-RSA. Volume 2020 of Lecture Notes in Computer Science., Springer (2001) 282–297
7. Petzoldt, A., Bulygin, S., Buchmann, J.: Cyclicrainbow - a multivariate signature scheme with a partially cyclic public key. In Gong, G., Gupta, K.C., eds.: INDOCRYPT. Volume 6498 of Lecture Notes in Computer Science., Springer (2010) 33–48
8. Petzoldt, A., Chen, M., Yang, B., Tao, C., Ding, J.: Design principles for hfev- based multivariate signature schemes. In Iwata, T., Cheon, J.H., eds.: Advances in Cryptology - ASIACRYPT 2015 - 21st International Conference on the Theory and Application of Cryptology and Information Security, Auckland, New Zealand, November 29 - December 3, 2015, Proceedings, Part I. Volume 9452 of Lecture Notes in Computer Science., Springer (2015) 311–334
9. Ding, J., Yang, B.Y.: Degree of regularity for hfev and hfev-. [23] 52–66
10. Goubin, L., Courtois, N.: Cryptanalysis of the ttm cryptosystem. In Okamoto, T., ed.: ASIACRYPT. Volume 1976 of Lecture Notes in Computer Science., Springer (2000) 44–57
11. Tsujii, S., Gotaishi, M., Tadaki, K., Fujita, R.: Proposal of a signature scheme based on sts trapdoor. In Sendrier, N., ed.: PQCrypto. Volume 6061 of Lecture Notes in Computer Science., Springer (2010) 201–217
12. Porras, J., Baena, J., Ding, J.: Zhfe, a new multivariate public key encryption scheme. [22] 229–245
13. Tao, C., Diene, A., Tang, S., Ding, J.: Simple matrix scheme for encryption. [23] 231–242
14. Ding, J., Petzoldt, A., Wang, L.: The cubic simple matrix encryption scheme. [22] 76–87
15. Szepieniec, A., Ding, J., Preneel, B.: Extension field cancellation: A new central trapdoor for multivariate quadratic systems. [24] 182–196
16. Perlner, R.A., Smith-Tone, D.: Security analysis and key modification for ZHFE. [24] 197–212
17. Moody, D., Perlner, R.A., Smith-Tone, D.: An asymptotically optimal structural attack on the ABC multivariate encryption scheme. [22] 180–196
18. Moody, D., Perlner, R.A., Smith-Tone, D.: Key recovery attack on the cubic abc simple matrix multivariate encryption scheme. In: Selected Areas in Cryptography – SAC 2016: 23rd International Conference, Revised Selected Papers, LNCS, Springer (2017)
19. Barker, E., Roginsky, A.: Transitions: Recommendation for transitioning the use of cryptographic algorithms and key lengths. NIST Special Publication (2015) http://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-131Ar1.pdf.
20. Diene, A., Tao, C., Ding, J.: Simple matrix scheme for encryption (abc). Presentation: PQCRYPTO 2013 (2013) http://pqcrypto2013.xlim.fr/slides/05-06-2013/Diene.pdf.
21. Developers, T.S.: SageMath, the Sage Mathematics Software System (Version x.y.z). (YYYY) `http://www.sagemath.org`.
22. Mosca, M., ed.: Post-Quantum Cryptography - 6th International Workshop, PQCrypto 2014, Waterloo, ON, Canada, October 1-3, 2014. Proceedings. Volume 8772 of Lecture Notes in Computer Science., Springer (2014)
23. Gaborit, P., ed.: Post-Quantum Cryptography - 5th International Workshop, PQCrypto 2013, Limoges, France, June 4-7, 2013. Proceedings. In Gaborit, P., ed.: PQCrypto. Volume 7932 of Lecture Notes in Computer Science., Springer (2013)
24. Takagi, T., ed.: Post-Quantum Cryptography - 7th International Workshop, PQCrypto 2016, Fukuoka, Japan, February 24-26, 2016, Proceedings. Volume 9606 of Lecture Notes in Computer Science., Springer (2016)