Analysis of the MQQ Public Key Cryptosystem

Rune Ødegård[1] ⋆, Ludovic Perret[2], Jean-Charles Faugère[2], and Danilo Gligoroski[3]

[1] Centre for Quantifiable Quality of Service in Communication Systems at the Norwegian University of Science and Technology in Trondheim, Norway.
rune.odegard@q2s.ntnu.no
[2] Institut National de Recherche en Informatique et en Automatique, Solving ALgebraic Systems and Applications Project
Laboratoire d'Informatique de Paris 6, Universite Pierre et Marie Curie, France
ludovic.perret@lip6.fr
Jean-Charles.Faugere@grobner.org
[3] Department of Telematics at the Norwegian University of Science and Technology in Trondheim, Norway
danilog@item.ntnu.no

Abstract. MQQ is a multivariate cryptosystem based on multivariate quadratic quasigroups and the Dobbertin transformation [18]. The cryptosystem was broken both by Gröbner bases computation and MutantXL [27]. The complexity of Gröbner bases computation is exponential in the degree of regularity, which is the maximum degree of polynomials occurring during the computation. The authors of [27] observed that the degree of regularity for solving the MQQ system is bounded from above by a small constant. In this paper we go one step further in the analysis of MQQ. We explain why the degree of regularity for the MQQ system is bounded. The main result of this paper is how the complexity of solving the MQQ system is the minimum complexity of solving just one quasigroup block and solving the Dobbertin transformation. Furthermore, we show that the degree of regularity for solving the Dobbertin transformation is bounded from above by the same constant as the bound on the MQQ system. We then investigate the strength of a tweaked MQQ system where the input to the Dobbertin transformation is replaced with random linear equations. We find that the degree of regularity for this tweaked system varies both in the size of the quasigroups and the number of variables. We conclude that if a suitable replacement for the Dobbertin transformation is found, MQQ can possibly be made strong enough to resist pure Gröbner attack for correct choices of quasigroups size and number of variables.

Keywords: multivariate cryptography, Gröbner bases , public-key, multivariate quadratic quasigroups, algebraic cryptanalysis

⋆ Rune Steinsmo Ødegård was visiting the SALSA team at LIP6 during the research of this paper.

1 Introduction

Multivariate cryptography comprises all the cryptographic schemes that use multivariate polynomials. At first glance, many aspects of such systems are tempting for cryptographers. Basing schemes on the hard problem of solving a system of multivariate equations is very appealing for multiple reasons. Most importantly, generic algorithms to solve this problem are exponential in the worst case, and solving random system of algebraic equations is also known to be difficult (i.e. exponential) in the average case. Moreover, no quantum algorithm allowing to solve non linear equations exists. Finally, multivariate schemes usually require computations with rather small integers leading to rather efficient smart-card implementations (see for example [7]).

The use of polynomial systems in cryptography dates back to the mid eighties with the design of Matsumoto and Imai [26], later followed by numerous other proposals. Two excellent surveys on the current state of proposals for multivariate asymettric cryptosystems has been made by Wolf and Prenel [34] and Billet and Ding [6]. Basicly the current proposals can be classified into four main categories, some of which combine features from several categories: Matsumoto-Imai like schemes [29,31], Oil and Vinegar like schemes [30,21], Stepwise Triangular Schemes [32,19] and Polly Cracker Schemes [12]. In addition Gligoroski et al. has proposed a fifth class of trapdoor functions based on multivariate quadratic quasigroups [18].

Unfortunately, it appears that most multivariate public-key schemes suffer from obvious to less obvious weaknesses. This is evident in [6] where a nice overview of the cryptanalysis techniques in multivariate asymmetric cryptography is given. Some attacks are specific attacks which focus on one particular variation and breaks it due to specific properties. One example of this is the attack of Kipnis and Shamir against Oil and Vinegar [22]. However most attacks use general purpose algorithms that solve multivariate system of equations. As mentioned, algorithms for solving random system of equations are known to be exponential in the avarage case. However in the case of multivariate public-key schemes the designer has to embed some kind of trapdoor function to enable efficient decryption and signing. To acheive this the public-key equations are constructed from a highly structured system of equations. Although the structure is hidden, it can be exploited for instance via differential or Gröbner based techniques.

Gröbner basis [9] is a well established and general method for solving polynomial systems of equations. The complexity of Gröbner bases computation is exponential in the degree of regularity, which is the maximum degree of polynomials occurring during the computation [4]. The first published attack on multivariate public-key cryptosystems using Gröbner basis is the attack by Paterin on the Matsumoto-Imai scheme [28]. In the paper Patarin explains exactly why Gröbner bases is able solve the system. The key remark is that there exists bilinear equations relating the input and the output of the system [6]. This low degree relation between the input and the output means that only polynomials

of low degree will occur during the computation of Gröbner bases. As a result the complexity of solving the system is bounded in this low degree.

Another multivariate cryptosystem that has fallen short for Gröbner bases cryptanalysis is the MQQ public key block cipher [18]. The cipher was solved both by Gröbner bases and MutantXL independently in [27]. However [27] did not theoretically explain why the algebraic systems of MQQ are easy to solve in practice. In this paper we explain exactly why the MQQ cryptosystem is susceptible to algebraic cryptanalysis. This is of course interesting from a cryptanalysis perspective, but also from a design perspective. If we want to construct strong multivariate cryptographic schemes we must understand why the weak schemes have been broken.

1.1 Organisation of the paper

This paper is organized as follows. In Section 2 we give an introduction to multivariate quadratic quasigroups. After that we describe the MQQ public key cryptosystem. In Section 3 we give a short introduction to the theory of Gröbner basis and recall the theoretical complexity of computing such bases. The complexity of computing Gröbner bases is exponential in the degree of regularity, which is the maximal degree of the polynomials occurring during computation. In Section 4 we show that the degree of regularity of MQQ systems is bounded from above by a small constant. We then explain this behavior thanks to the shape of the inner system. In Section 5 we further elaborate on the weaknesses of the MQQ system, and investigate if some tweaks can make the system stronger. Finally, Section 6 concludes the paper.

2 Description of MQQ public key cryptosystem

In this section we give a description of the multivariate quadratic quasigroup public key cryptosystem [18]. The system is based on previous work by Gligoroski and Markovski who introduced the use of quasigroup string processing in cryptography [24,25].

2.1 Multivariate quadratic quasigroups

We first introduce the key building block namely multivariate quadratic quasigroups. For a detailed introduction to quasigroups in general we refer the interested reader to [33].

Definition 1 *A quasigroup is a set Q together with a binary operation $*$ such that for all $a, b \in Q$ the equations $\ell * a = b$ and $a * r = b$ have unique solutions ℓ and r in Q. A quasigroup is said to be of order n if there are n elements in the set Q.*

Let $(Q, *)$ be a quasigroup of order 2^d, and β be a bijection from the quasigroup to the set of binary strings of length d, i.e

$$\begin{array}{rl} \beta : Q \to & \mathbb{Z}_2^d \\ a \mapsto & (x_1, \ldots, x_d) \end{array} \tag{1}$$

Given such a bijection we can naturally define a vector valued Boolean function

$$\begin{array}{rl} *_{vv} : & \mathbb{Z}_2^d \times \mathbb{Z}_2^d \to \mathbb{Z}_2^d \\ & (\beta(a), \beta(b)) \mapsto \beta(a * b) \end{array} \tag{2}$$

Now let $\beta(a * b) = (x_1, \ldots, x_d) *_{vv} (x_{d+1}, \ldots, x_{2d}) = (z_1, \ldots, z_d)$. Note that each z_i can be regarded as a $2d$-ary Boolean function $z_i = f_i(x_1, \ldots, x_{2d})$, where each $f_i : \mathbb{Z}_2^d \to \mathbb{Z}_2$ is determined by $*$. This gives us the following lemma [18].

Lemma 1 *For every quasigroup $(Q, *)$ of order 2^d and for each bijection $\beta : Q \to \mathbb{Z}_2^d$ there is a unique vector valued Boolean function $*_{vv}$ and d uniquely determined $2d$-ary Boolean functions $f_1, f_2, \ldots, f_d$ such that for each $a, b, c \in Q$:*

$$\begin{array}{c} a * b = c \\ \Updownarrow \\ (x_1, \ldots, x_d) *_{vv} (x_{d+1}, \ldots, x_{2d}) = (f_1(x_1, \ldots, x_{2d}), \ldots, f_d(x_1, \ldots, x_{2d})). \end{array} \tag{3}$$

This leads to the following definition for multivariate quadratic quasigroups.

Definition 2 *([18]) Let $(Q, *)$ be a quasigroup of order 2^d, and let $f_1, \ldots, f_d$ be the uniquely determined Boolean functions under some bijection β. We say that the quasigroup is multivariate quadratic quasigroup (MQQ) of type $Quad_{d-k}Lin_k$ (under β) if exactly $d - k$ of the corresponding polynomials f_i are of degree 2 and k of them are of degree 1, where $0 \leq k \leq d$.*

Gligoroski et al. mention [18] that quadratic terms might cancel each other. By this we mean that some linear transformation of $(f_i)_{1 \leq i \leq n}$ might result in polynomials where the number of linear polynomials is larger than k, while the number of quadratic polynomials is less than $d - k$. Later Chen et al. [10] have shown that this is more common than previously expected. In their paper they generalizes the definition of MQQ above to a family which is invariant by linear transformations in $\mathbb{Z}_2[x_1, \ldots, x_{2d}]$.

Definition 3 *Let $(Q, *)$ be a quasigroup of order 2^d, and let $f_1, \ldots, f_d$ be the unique Boolean functions under some bijection β. We say that the quasigroup is a multivariate quadratic quasigroup (MQQ) of strict type $Quad_{d-k}Lin_k$ (under β), denoted by $Quad^s_{d-k}Lin^s_k$, if there are at most $d-k$ quadratic polynomials in $(f_i)_{1 \leq i \leq d}$ whose linear combination do not result in a linear form.*

Chen et al. also improves Theorem 2 from [18] which gives a sufficient condition for a quasigroup to be MQQ. We restate this theorem below.

Theorem 1 *Let* $\mathbf{A}_1 = [f_{ij}]_{d\times d}$ *and* $\mathbf{A}_2 = [g_{ij}]_{d\times d}$ *be two* $d\times d$ *matrices of linear Boolean expressions with respect to* $x_1, \ldots, x_d$ *and* $x_{d+1}, \ldots, x_{2d}$ *respectively. Let* $\mathbf{c}$ *be a binary column vector of* d *elements. If* $det(\mathbf{A}_1) = det(\mathbf{A}_2) = 1$ *and*

$$\mathbf{A}_1 \cdot (x_{d+1}, \ldots, x_{2d})^T + (x_1, \ldots, x_d)^T = \mathbf{A}_2 \cdot (x_1, \ldots, x_d)^T + (x_{d+1}, \ldots, x_{2s})^T, \quad (4)$$

then the vector valued Boolean operation

$$(x_1, \ldots, x_d) *_{vv} (x_{d+1}, \ldots, x_{2d}) = \mathbf{B}_1\mathbf{A}_1 \cdot (x_{d+1}, \ldots, x_{2d})^T + \mathbf{B}_2 \cdot (x_1, \ldots, x_d)^T + \mathbf{c} \quad (5)$$

defines a quasigroup $(Q, *)$ *of order* 2^d *which is MQQ for any non-singular Boolean matrices* $\mathbf{B}_1$ *and* $\mathbf{B}_2$

In addition Chen et al. proved [10] that no MQQ as in Theorem 1 can be of strict type $\mathrm{Quad}_d^s\mathrm{Lin}_0^s$. This result uncovered a possible weakness in [18] since the proposed scheme is using 6 quasigroups of type $\mathrm{Quad}_5\mathrm{Lin}_0$.

Notice that the vector valued Boolean function defining the MQQ in Theorem 1 have no terms of the form x_ix_j with $i, j \leq d$ or $i, j > d$. This means that if we set the first or the last half of the variables to a constant, we end up with only linear terms in the MQQ. It is still an open question if there exists MQQ that are not as in Theorem 1.

The MQQs used in this paper has been produced using the algorithm provided in Appendix A. The algorithm is based on the paper [10], and produces MQQs that are more suitable for encryption since they are guaranteed to be of strict type $\mathrm{Quad}_{d-k}^s\mathrm{Lin}_k^s$.

2.2 Dobbertin bijection

In addition to MQQs, the public key cryptosystem [18] also uses a bijection introduced by Dobbertin in [13]. Dobbertin proved that the following function, in addition to being multivariate quadratic, is a bijection in $\mathbb{Z}_{2^{2r+1}}$.

$$\begin{aligned} D_r : \mathbb{Z}_{2^{2r+1}} &\to \mathbb{Z}_{2^{2r+1}} \\ x &\mapsto x^{2^{r+1}+1} + x^3 + x \end{aligned} \quad (6)$$

2.3 Public key cryptosystem based on MQQ

We are now ready to describe the public key cryptosystem presented by Gligoroski et al. in [18]. Let $N = nd$ be the desired number of variables $(x_1, \ldots, x_N)$, and let $\{*_{vv}^1, \ldots, *_{vv}^k\}$ be a collection of MQQs of size 2^d represented as $2d$-ary vector valued Boolean functions. The public key is constructed as follows.

Algorithm *MQQ public key construction.*

1. Set $\mathbf{X} = [x_1, \ldots, x_N]^T$. Randomly generate a $N \times N$ non-singular Boolean matrix $\mathbf{S}$, and compute $\mathbf{X} \leftarrow \mathbf{S} \cdot \mathbf{X}$.
2. Construct an n-tuple $I = \{i_1, \ldots, i_n\}$, where $i_j \in \{1, \ldots, k\}$. The tuple I will decide which MQQ, $*_{vv}^{i_j}$, to use at each point of the quasigroup transformation.

3. Represent $\mathbf{X}$ as a collection of vectors of length d, $\mathbf{X} = [X_1, \ldots, X_n]^T$. Compute $\mathbf{Y} = [Y_1, \ldots, Y_n]^T$ where $Y_1 = X_1, Y_2 = X_1 *_{vv}^{i_1} X_2$, and $Y_j = X_j *_{vv}^{ij} X_{j+1}$ for $j = 1, \ldots, n$.
4. Set $\mathbf{Z}$ to be the vector of all the linear terms of $Y_1, \ldots, Y_j$. Here Y_1 will be all linear terms, while each Y_j has between 1 and k linear terms depending on the type $\text{Quad}_{d-k}^s \text{Lin}_k^s$ of MQQ used. Transform $\mathbf{Z}$ with one or more Dobbertin bijections of appropriate size. For example if $\mathbf{Z}$ is of size 27 we can use one Dobbertin bijection of dimension 27, three of dimension 9, or any other combination adding up to 27. $\mathbf{W} \leftarrow \text{Dob}(\mathbf{Z})$.
5. Replace the linear terms of $\mathbf{Y} = [Y_1, \ldots, Y_n]^T$ with the terms in $\mathbf{W}$. Randomly generate a $N \times N$ non-singular Boolean matrix $\mathbf{T}$, and compute $\mathbf{Y} \leftarrow \mathbf{T} \cdot \mathbf{Y}$
6. **return** the public key $\mathbf{Y}$. The private key is $\mathbf{S}, \mathbf{T}, \{*_{vv}^1, \ldots, *_{vv}^k\}$ and I.

3 Gröbner basis cryptanalysis

In recent years Gröbner bases has been used as a tool to mount efficient algebraic cryptanalysis [6] In particular, Gröbner bases has been used to attack MQQ [27]. In this paper we go one step further by explaining the weakness found in [27]. In addition we investigate the possibility of constructing stronger MQQ systems.

3.1 Short introduction to Gröbner bases

This section introduces the concept of Gröbner bases. We refer to [11] for basic definitions, and a more detailed description of the concepts.

Let $\mathbb{K}$ be a field and $\mathbb{K}[x_1, \ldots, x_n]$ the polynomial ring over $\mathbb{K}$ in the variables $x_1, \ldots, x_n$. Recall that a *monomial* in a collection of variables is a product $x^\alpha = x_1^{\alpha_1} \cdots x_n^{\alpha_n}$ where $\alpha_i \geq 0$. Let $>$ be an admissible *monomial order* on $k[x_1, \ldots, x_n]$. The most common example of monomial order is the *lexicographical order* where $x^\alpha > x^\beta$ if in the difference $\alpha - \beta \in \mathbb{Z}^n$ the leftmost nonzero entry is positive. Another frequently encountered order is the *graded reverse lexicographical* order where $x^\alpha > x^\beta$ iff $\sum_i \alpha_i > \sum_i \beta_i$ or $\sum_i \alpha_i = \sum_i \beta_i$ and in the difference $\alpha - \beta \in \mathbb{Z}^n$ the rightmost nonzero entry is negative. For different orders Gröbner bases has specific theoretical property and different practical behaviors. Given a monomial order $>$ the *leading term* of a polynomial $f = \sum_\alpha c_\alpha x^\alpha$, denoted $LT_>(f)$, is the product $c_\alpha x^\alpha$ where x^α is the largest monomial appearing in f in the ordering $>$.

Definition 4 *([11]) Fix a monomial order $>$ on $k[x_1, \ldots, x_n]$, and let $I \subset k[x_1, \ldots, x_n]$ be an ideal. A Gröbner basis for I (with respect to $>$) is a finite collection of polynomials $G = \{g_1, \ldots, g_t\} \subset I$ with the property that for every nonzero $f \in I$, $LT_>(f)$ is divisible by $LT_>(g_i)$ for some i.*

Let

$$f_1(x_1, \ldots, x_n) = \cdots = f_m(x_1, \ldots, x_n) = 0 \tag{7}$$

by a system of m polynomials in n unknowns over the field $\mathbb{K}$.The set of solutions in $\mathbb{K}$, which is the algebraic variety, is defined as

$$V = \{(z_1, \ldots, z_n) \in k | f_i(z_1, \ldots, z_n) = 0 \forall 1 \leq i \leq n\} \tag{8}$$

In our case we are interested in the solutions of the MQQ system, which is defined over $\mathbb{Z}_2$.

Proposition 1 *([16]) Let G be a Gröbner bases of $[f_1, \ldots, f_m, x_1^2 - x_1, \ldots, x_n^2 - x_n]$. It holds that:*

1. *$V = \emptyset$ (no solution) iff $G = [1]$.*
2. *V has exactly one solution iff $G = [x_1 - a_1, \ldots, x_n - a_n]$ where $a_i \in \mathbb{Z}_2$. Then $(a_1, \ldots, a_n)$ is the solution in $\mathbb{Z}_2$ of the algebraic system.*

From the proposition we learn that in order to solve a system over $\mathbb{Z}_2$ we should add the field equations $x_i^2 = x_i$ for $i = 1, \ldots, n$. This means that we have to compute a Gröbner bases of $m + n$ polynomials and n variables. This is quite helpful, since the more equations you have, the more able you are to compute Gröbner bases [16].

3.2 Complexity of computing Gröbner bases

Historically the concept of Gröbner bases, together with an algorithm for computing them, was introduced by Bruno Buchberger in his PhD-thesis [9]. Buchberger's algorithm is implemented in many computer algebra systems. However, in the last decade, more efficient algorithms for computing Gröbner bases have been proposed. Most notable are Jean-Charles Faugère's F_4[14] and F_5 [15] algorithms. In this paper we have used the magma [23] 2.16-1 implementation of the F_4 algorithm on a 4 core Intel Xeon 2.93GHz computer with 128GB of memory.

The complexity of computing a Gröbner basis of an ideal I depends on the maximal degree of the polynomials appearing during the computation. This degree, called *degree of regularity*, is the key parameter for understanding the complexity of Gröbner basis computations [4]. Indeed, the complexity of the computation is polynomial in the degree of regularity D_{reg}, more precisely the complexity is:

$$\mathcal{O}(N^{\omega D_{\text{reg}}}), \tag{9}$$

which basically correspond to the complexity of reducing a matrix of size $N^{D_{\text{reg}}}$. Here $2 < \omega \leq 3$ is the "linear algebra constant", and N the number of variables of the system. Note that D_{reg} is also a function of N, where the relation between D_{reg} and N depends on the specific system of equations. This relation is well understood for regular (and semi-regular) systems of equations [1,4,2,5]. On the contrary, as soon as the system has some kind of structure, this degree is much more difficult to predict. In some particular cases, it is however possible to bound the degree of regularity (see the works done on HFE [16,20]). But it is a hard task in general.

Note that the degree of regularity is related to the ideal $I = \langle f_1, \ldots, f_n \rangle$ and not the equations $f_1, \ldots, f_n$ themselves. This means given any non-singular matrix S and linear transformation $[f'_1, \ldots, f'_n]^T = S \cdot [f_1, \ldots, f_n]^T$, the degree of regularity for solving equations $f'_1, \ldots, f'_n$ with Gröbner bases is the same as for equations $f'_1, \ldots, f'_n$ since $\langle f'_1, \ldots, f'_n \rangle = \langle f_1, \ldots, f_n \rangle$. More generally, we can assume that this degree is invariant for a (invertible) linear change of variables, and (invertible) combination of the polynomials. These are exactly the transformations performed to mask the MQQ structure.

4 Why MQQ is susceptible to algebraic cryptanalysis

In [27] MQQ systems with up to 160 variables was broken using both the MutantXL and the F_4 algorithm independently. The most important remark by [27] is that the degree of regularity is bounded from above by 3. This is much lower than a random system of quadratic equations where the degree of regularity increases linearly in the number of equations N. Indeed, for a random system it holds that D_{reg} is asymptotically equivalent to $\frac{N}{11.114}$ [2]. The authors of [27] observed that the low degree for MQQ is due to the occurrence of many new low degree relations during the computation of the Gröbner basis. Here, we go one step further in the analysis. We explain precisely why low-degree relations appear. This is due to the very structure of the MQQ system as we explain in detail in Section 4.2. First, we show that we observe the same upper bound on the degree of regularity using the improved quasigroups described in Section 2.1.

4.1 Experimental results on MQQ

To test how the complexity of Gröbner bases computation of MQQ public key systems is related to the number of variables, we constructed MQQ systems of size $30, 60, 120, 180$ following the procedure described in Section 2.3. In this construction we used 17 MQQs of strict type $\text{Quad}_8^s\text{Lin}_2^s$ and Dobbertin bijections over different extension fields of dimension 7 and 9 respectively. The results of this test are presented in Table 1. From the table we see that the degree of regularity does not increase with the number of variables, but remains constant at 3.

Table 1. Results for MQQ-(30,60,120,180). Computed with magma 2.16-1's implementation of the F_4 algorithm on a 4 processor Intel Xeon 2.93GHz computer with 128GB of memory.

Variables	D_{reg}	Solving Time (s)	Memory (b)
30	3	0,06	15,50
60	3	1,69	156,47
120	3	379,27	4662,00
180	3	4136,31	28630,00

Once again, this is not the behaviour of a random system of equations for which the degree of regularity is asymptotically linear in the number of variables. We explain the reason of such difference in the next section.

4.2 Shape of the MQQ system

This non-random behavior can be explained by considering the shape of the "unmasked" MQQ system. By unmasked we mean the MQQ system without the linear transformation S and T. The maximal degree of the polynomials occurring in the computation of a Gröbner basis is invariant under the linear transformation S and T as explained in Section 3.2. In Figure 1 we show what variables appear in each equation for an unmasked MQQ system of 60 variables. The staircase shape comes from the cascading use of quasigroups, while the three blocks of equations at the bottom are from Dobbertin bijection of size 7. A

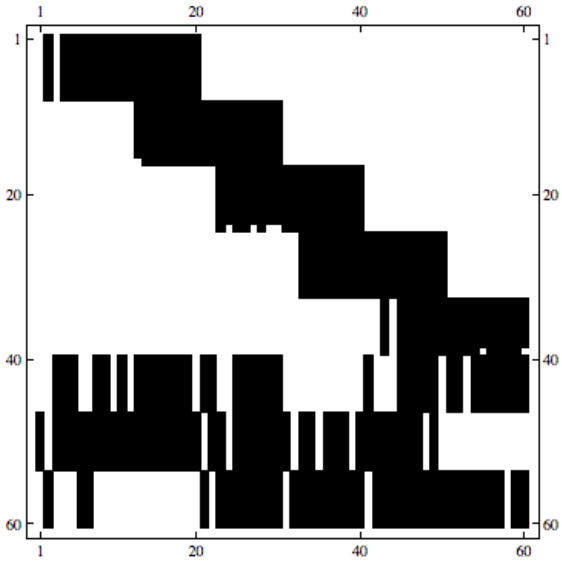

Fig. 1. Shape of 60 variable MQQ public key system without the use of S and T transformation. Black means that the corresponding variables is used in the equation. The system was constructed with 4 MQQs of type $\mathrm{Quad}_8^s\mathrm{Lin}_2^s$, one MQQ of type $\mathrm{Quad}_7^s\mathrm{Lin}_3^s$, and 3 Dobbertin bijections defined over 3 different extension fields of dimension 7.

random multivariate system would use all 60 variables in all equations. For the MQQ system in this example only $\frac{1}{3}$ of the variables are used in each quasigroup and about $\frac{2}{3}$ is used in each block of Dobbertin transformation.

Now assume the Gröbner basis algorithm somewhere during the calculation has found the solution for one of the quasigroup blocks $Y_j = X_j *_{vv}^{ij} X_{j+1}$. Due to the cascading structure of the MQQ system the variables of X_j are used in

the block $Y_{j-1} = X_{j-1} *_{vv}^{i_{j-1}} X_j$ and the variables of X_{j+1} are used in the block. $Y_{j+1} = X_{j+1} *_{vv}^{i_{j+1}} X_{j+2}$. Remember from Section 2.1 that if we set the first or the last half of the variables of an MQQ to constant all equations becomes linear. This means that if we have solved the block Y_j, the equations of the blocks Y_{j-1} and Y_{j+1} becomes linear. The blocks Y_{j-1} and Y_{j+1} can then be solved easily. This gives us solution for the variables X_{j-1} and X_{j+2}, which again makes the equations in the blocks Y_{j-2} and Y_{j+2} linear. Continuing like this we have rapidly solved the whole system.

Similarly, assume the Gröbner basis has solved the Dobbertin blocks at some step. This gives us the solution to all the variables in X_1 which makes the first quasigroup block $Y_1 = X_1 *_{vv}^{i_1} X_2$ linear. Solving this gives us the first half of the equations of the block Y_2 and so on. This means that the solution of the whole MQQ system is reduced to either solving just one block of quasigroup equations, or solving the Dobbertin transformation. The security of solving the MQQ system is therefore the minimum complexity of solving Dobbertin transformation and one MQQ block.

5 Further analysis of MQQ

Using the knowledge from Section 4.2 we investigate if it is possible to strengthen the MQQ system. To do this we have to determine the weakest part of the system; the Dobbertin transformation or the quasigroup transformation.

5.1 The Dobbertin transformation

Recall that the Dobbertin transformation is a bijection over $\mathbb{Z}_2^{2r+1}$ defined by the function $D_r(x) = x^{2^{r+1}+1} + x^3 + x$. For any r we can view this function as $2r+1$ Boolean functions in $2r+1$ variables. In Table 2 our experimental results on the degree of regularity for solving this system of equations is listed for various r. From the table we see that the degree of regularity for the Dobbertin transformation seems to be bounded from above by 3. This means Dobbertin's transformation is not "random" and can be easily solved by Gröbner bases computation. In addition we learn that tweaking the MQQ system by increasing the size of the extension field, over which the transformation is defined, will have no effect on strengthening the system.

Proving mathematically (if true) that the degree of regularity for $D_r(x)$ is constant at 3 for all r is difficult. We can however give show that the degree of regularity is low for all practical r. Let $\mathbb{K} = \mathbb{F}_q$ be a field of q elements, and let $\mathbb{L}$ be an extension of degree n over $\mathbb{K}$. Recall that an HFE polynomial f is a low-degree polynomial over $\mathbb{L}$ with the following shape:

$$f(x) = \sum_{\substack{0 \le i,j \le n \\ q^i + q^j \le d}} a_{i,j} x^{q^i + q^j} + \sum_{\substack{0 \le k \le n \\ q^k \le d}} b_k x^{q^k} + c, \tag{10}$$

Table 2. The observed degree of regularity, D_{reg}, for the Dobbertin bijection over $\mathbb{Z}_2^{2r+1}$ for $r = 2, \ldots, 22$ when computed with magma 2.16-1's implementation of the F_4 algorithm on a 4 processor Intel Xeon 2.93GHz computer with 128GB of memory

r	D_{reg}	r	D_{reg}	r	D_{reg}
2	3	9	3	16	3
3	3	10	3	17	3
4	3	11	3	18	3
5	3	12	3	19	3
6	3	13	3	20	3
7	3	14	3	21	3
8	3	15	3	22	3

where $a_{i,j}, b_k$ and c all lie in $\mathbb{L}$. The maximum degree d of the polynomial has to be chosen such that factorization over $\mathbb{L}$ is efficient [8]. Setting $q = 2$ and $n = 2r + 1$ we notice that the Dobbertin transformation is actually an HFE polynomial, $D_r(x) = x^{2^{r+1}+2^0} + x^{2^1+2^0} + x^{2^0}$. This is very helpfull since a lot of work has been done on the degree of regularity for Gröbner basis compuation of HFE polynomials [16,8]. Faugère and Joux showed that the degree of regularity for Gröbner bases computation of an HFE polynomial of degree d is bounded from above by $\log_2(d)$ [16,17]. For the Dobbertin transformation this means the degree of regularity is bounded from above by $r + 1$.

However, since the coefficients of the Dobbertin transformation all lie in $GF(2)$, we can give an even tighter bound on the degree of regularity. Similar to the weak-key polynomials in [8] the Dobbertin transformation commutes with the Frobenius automorphism and its iterates $F_i(x) : x \mapsto x^{2^i}$ for $0 \leq i \leq n$.

$$D_r \circ F_i(x) = F_i \circ D_r(x) \tag{11}$$

From the equations we see that when $D_r(x) = 0$ we have $F_i \circ D_r(x) = 0$. This means for each i we can add the n equations over $GF(2)$ corresponding to the equation $D_r \circ F_i(x) = 0$ over $GF(2^n)$ to the ideal. However, many of these equations are similar. Actually, we have that F_i and F_j are similar if and only if $gcd(i, n) = gcd(j, n)$ [8]. Worst case scenario is when n is prime. The Frobenius automorphism then gives us $2n$ equations in n variables. From [3] we have the following formula for the degree of regularity for a random system of multivariate equations over $GF(2)$ when the number of equations m is a multiple of the number of variables n. For $m = n(N + o(1))$ with $N > 1/4$ the degree of regularity is

$$\frac{D_{\text{reg}}}{n} = \frac{1}{2} - N + \frac{1}{2}\sqrt{2N^2 - 10N - 1 + 2(n+2)\sqrt{N(N+2)}} + o(1) \tag{12}$$

Setting $N = 2$ we get $D_{\text{reg}} = -\frac{3}{2} + \frac{1}{2}\sqrt{-13 + 16\sqrt{2}} \cdot n \approx 0.051404 \cdot n$. This is the upper bound for a *random* multivariate system with the same number of equations and variables as the Dobbertin transformation. This provides us a good indication that the degree of regularity for Dobbertin (which is not random

at all) should be small, as observed in the experiments, and even smaller than a regular HFE polynomial.

5.2 The quasigroup transformation

To get an idea how strong the quasigroup transformation is, we decided to run some experiments where we replaced the input to the Dobbertin transformation with random linear equations. This means that solving the Dobbertin block will no longer make all the equations in the quasigroup transformation linear. The result of our experiment on this special MQQ system where the linear equations are perfectly masked is listed in Table 3. From the table it appears that both the quasigroup size and the number of variables have an effect on the degree of regularity. This tells us that if we replace Dobbertin transformation with a stronger function, the MQQ system can possibly be made strong enough to resist pure Gröbner attack for correct choices of quasigroups size and number of variables.

Table 3. Effects of quasigroup size and the Dobbertin transformation on the observed degree of regularity for Gröbner bases computations of 60 variable MQQ systems. D_{reg} is the observed degree of regularity of normal MQQ systems, while D^*_{reg} is the observed degree of regularity for the same system where the input to Dobbertin has been replaced with random linear equations.

Variables	Quasigroup size	Quasigroups type	Dobbertin	D_{reg}	D^*_{reg}
30	2^5	4 $\text{Quad}^s_3\text{Lin}^s_2$ and 1 $\text{Quad}^s_2\text{Lin}^s_3$	7,9	3	3
	2^{10}	2 $\text{Quad}^s_8\text{Lin}^s_2$	7,7	3	4
40	2^5	5 $\text{Quad}^s_3\text{Lin}^s_2$ and 2 $\text{Quad}^s_2\text{Lin}^s_3$	7,7,7	3	4
	2^{10}	3 $\text{Quad}^s_8\text{Lin}^s_2$	7,9	3	4
	2^{20}	1 $\text{Quad}^s_{17}\text{Lin}^s_3$	7,7,9	3	4
50	2^5	9 $\text{Quad}^s_3\text{Lin}^s_2$	7,7,9	3	3
	2^{10}	4 $\text{Quad}^s_8\text{Lin}^s_2$	9,9	3	4
60	2^5	11 $\text{Quad}^s_3\text{Lin}^s_2$	9,9,9	3	3
	2^{10}	4 $\text{Quad}^s_8\text{Lin}^s_2$ and 1 $\text{Quad}^s_7\text{Lin}^s_3$	7,7,7	3	5
	2^{20}	1 $\text{Quad}^s_{18}\text{Lin}^s_2$ and 1 $\text{Quad}^s_{17}\text{Lin}^s_3$	7,9,9	3	5

6 Conclusion

We have confirmed the results of [27] showing that the degree of regularity for MQQ systems are bounded from above by a small constant, and therefore MQQ systems in large number of variables can easily be broken with Gröbner bases cryptanalysis. The main result of this paper is our explanation of the underlying reason for this bound on the degree of regularity. We explained this by showing how the complexity of solving MQQ systems with Gröbner bases is equal to the minimum of the complexity of solving the Dobbertin transformation and

the complexity of solving one MQQ block. Furthermore, our experimental data showed that the degree of regularity for solving the Dobbertin transformation is bounded from above by 3, the same as the bound on the MQQ system. It is natural to conclude that the Dobbertin transformation is a serious weakness in the MQQ system.

We also showed that if the Dobbertin transformation is replaced with an ideal function, which perfectly hides the linear parts of the system, the degree of regularity varies in the size of the quasigroups and the number of variables. We conclude that if a suitable replacement for the Dobbertin transformation is found, MQQ can possibly be made strong enough to resist pure Gröbner attack for correct choices of quasigroups size and number of variables.

References

1. Magali Bardet. *Étude des systèmes algébriques surdéterminés. Applications aux codes correcteurs et à la cryptographie*. PhD thesis, Université de Paris VI, 2004.
2. Magali Bardet, Jean-Charles Faugère, and Bruno Salvy. Complexity study of Gröbner basis computation. Technical report, INRIA, 2002. `http://www.inria.fr/rrrt/rr-5049.html`.
3. Magali Bardet, Jean-Charles Faugère, and Bruno Salvy. Complexity of Gröbner basis computation for semi-regular overdetermined sequences over F2 with solutions in F2. Technical report, Institut national de recherche en informatique et en automatique, 2003.
4. Magali Bardet, Jean-Charles Faugère, and Bruno Salvy. On the complexity of Gröbner basis computation of semi-regular overdetermined algebraic equations. In *Proc. International Conference on Polynomial System Solving (ICPSS)*, pages 71–75, 2004.
5. Magali Bardet, Jean-Charles Faugère, Bruno Salvy, and Bo-Yin Yang. Asymptotic behaviour of the degree of regularity of semi-regular polynomial systems. In *Proc. of MEGA 2005, Eighth International Symposium on Effective Methods in Algebraic Geometry*, 2005.
6. Olivier Billet and Jintai Ding. Overview of cryptanalysis techniques in multivariate public key cryptography. In Massimiliano Sala, Teo Mora, Ludovic Perret, Shojiro Sakata, and Carlo Traverso, editors, *Gröbner bases, coding and cryptography*, pages 263–283. Springer Verlag, 2009.
7. Andrey Bogdanov, Thomas Eisenbarth, Andy Rupp, and Christopher Wolf. Time-Area Optimized Public-Key Engines: MQ -Cryptosystems as Replacement for Elliptic Curves? In *Cryptographic Hardware and Embedded Systems (CHES)*, volume 5154, pages 145–61. Lecture Notes in Computer Science, 2008.
8. Charles Bouillaguet, Pierre-Alain Fouque, Antoine Joux, and Joana Treger. A family of weak keys in hfe (and the corresponding practical key-recovery). Cryptology ePrint Archive, Report 2009/619, 2009.
9. Bruno Buchberger. *Ein Algorithmus zum Auffinden der Basiselemente des Restklassenringes nach einem nulldimensionalen Polynomideal*. PhD thesis, Leopold-Franzens University, 1965.
10. Yanling Chen, Svein Johan Knapskog, and Danilo Gligoroski. Multivariate Quadratic Quasigroups (MQQ): Construction, Bounds and Complexity. Submitted to ISIT 2010, 2010.

11. David Cox, John Little, and Donal O'Shea. *Using Algebraix Geometry*. Springer, 2005.
12. Francoise Levy dit Vehel, Maria Grazia Marinari, Ludovic Perret, and Carlo Traverso. A survey on polly cracker system. In Massimiliano Sala, Teo Mora, Ludovic Perret, Shojiro Sakata, and Carlo Traverso, editors, *Gröbner bases, coding and cryptography*, pages 263–283. Springer Verlag, 2009.
13. Hans Dobbertin. One-to-one highly nonlinear power functions on GF(2^n). *Appl. Algebra Eng. Commun. Comput.*, 9(2):139–152, 1998.
14. Jean-Charles Faugère. A new efficient algorithm for computing Gröbner bases (F_4). *Journal of Pure and Applied Algebra*, 139(1-3):61–88, June 1999.
15. Jean-Charles Faugère. A new efficient algorithm for computing Gröbner bases without reduction to zero (F_5). In *Proceedings of the 2002 International Symposium on Symbolic and Algebraic Computation*, New York, 2002. ACM.
16. Jean-Charles Faugère and Antoine Joux. Algebraic cryptanalysis of Hidden Field Equation (HFE) cryptosystems using Gröbner bases. In Dan Boneh, editor, *Advances in Cryptology - CRYPTO 2003*, volume 2729 of *LNCS*, pages 44–60. Springer, 2003.
17. Pierre-Alain Fouque, Gilles Macario-Rat, and Jacques Stern. Key recovery on hidden monomial multivariate schemes. In Nigel P. Smart, editor, *EUROCRYPT*, volume 4965 of *Lecture Notes in Computer Science*, pages 19–30. Springer, 2008.
18. Danilo Gligoroski, Smile Markovski, and Svein Johan Knapskog. Multivariate quadratic trapdoor functions based on multivariate quadratic quasigroups. In *MATH'08: Proceedings of the American Conference on Applied Mathematics*, pages 44–49, Stevens Point, Wisconsin, USA, 2008. World Scientific and Engineering Academy and Society (WSEAS).
19. Louis Goubin, Nicolas T. Courtois, and Schlumbergersema Cp. Cryptanalysis of the ttm cryptosystem. In *Advances of Cryptology, Asiacrypt2000*, pages 44–57. Springer, 2000.
20. Louis Granboulan, Antoine Joux, and Jacques Stern. Inverting HFE is quasipolynomial. In *CRYPTO*, pages 345–356, 2006.
21. Aviad Kipnis, Hamarpe St. Har Hotzvim, Jacques Patarin, and Louis Goubin. Unbalanced oil and vinegar signature schemes. In *In Advances in Cryptology EUROCRYPT 1999*, pages 206–222. Springer, 1999.
22. Aviad Kipnis and Adi Shamir. Cryptanalysis of the oil & vinegar signature scheme. In *CRYPTO '98: Proceedings of the 18th Annual International Cryptology Conference on Advances in Cryptology*, pages 257–266, London, UK, 1998. Springer-Verlag.
23. MAGMA. High performance software for algebra, nuber theory, and geometry — a large commercial software package. http://magma.maths.usyd.edu.au.
24. Smile Markovski. Quasigroup string processing and applications in cryptography. In *Proc. 1-st Inter. Conf. Mathematics and Informatics for industry MII 2003, 1416 April, Thessaloniki, 278290*, page 278290, 2003.
25. Smile Markovski, Danilo Gligoroski, and Verica Bakeva. Quasigroup string processing. In *Part 1, Contributions, Sec. Math. Tech. Sci., MANU, XX*, pages 1–2, 1999.
26. Tsutomu Matsumoto and Hideki Imai. Public quadratic polynomial-tuples for efficient signature-verification and message-encryption. In *Advances in Cryptology – EUROCRYPT 1988*, volume 330 of *LNCS*, pages 419–453. Springer–Verlag, 1988.
27. Mohamed Saied Mohamed, Jintai Ding, Johannes Buchmann, and Fabian Werner. Algebraic attack on the MQQ public key cryptosystem. In *CANS '09: Proceedings*

of the 8th International Conference on Cryptology and Network Security, pages 392–401, Berlin, Heidelberg, 2009. Springer-Verlag.
28. Jacques Patarin. Cryptanalysis of the matsumoto and imai public key scheme of eurocrypt'88. In *Lecture Notes in Computer Science*, pages 248–261, 1995.
29. Jacques Patarin. Hidden field equations (hfe) and isomorphisms of polynomials (ip): two new families of asymmetric algorithms. In *EUROCRYPT*, pages 33–48. Springer-Verlag, 1996.
30. Jacques Patarin. The oil & vinegar signature scheme, 1997.
31. Jacques Patarin, Louis Goubin, and Nicolas Courtois. C $^*-+$ and hm: Variations around two schemes of t.matsumoto and h.imai. In *Advances in Cryptology - Asiacrypt'98*, volume 1514, pages 35–49. Springer, 1998.
32. Adi Shamir. Efficient signature schemes based on birational permutations. In *CRYPTO '93: Proceedings of the 13th annual international cryptology conference on Advances in cryptology*, pages 1–12, New York, NY, USA, 1994. Springer-Verlag New York, Inc.
33. J. D. H. Smith. *An introduction to quasigroups and their representations*. Chapman & Hall/CRC, 2007.
34. Christopher Wolf and Bart Preneel. Taxonomy of public key schemes based on the problem of multivariate quadratic equations. Cryptology ePrint Archive, Report 2005/077, 2005.

A Algorithm for generating random MQQ

In this section we present the pseudo-code for how the MQQs used in this paper have been generated. The code was implemented in magma.

Algorithm *MQQ algorithm*
1. $n \leftarrow$\{size of quasigroup\}
2. $L \leftarrow$\{number of linear terms\}
3. **if** $L \leq 2$
4. **then** $Q = n$
5. **else** $Q = n - L$
6. CorrectDeg $\leftarrow$True
7. **while** CorrectDeg
8. **do** $A1 \leftarrow$IdentityMatrix(n) ($*$ The identity matrix of size n $*$)
9. $X1 \leftarrow [x_1, \ldots, x_n]^T$
10. $X2 \leftarrow [x_{n+1}, \ldots, x_{2n}]^T$
11. **for** $i \leftarrow 1$ **to** Q
12. **do for** $j \leftarrow i+1$ **to** n
13. **do for** $k \leftarrow i+1$ **to** (n)
14. $r \in_R \{0,1\}$ ($*$ random element from the set {0,1} $*$)
15. $A1_{(i,j)} = A1_{(i,j)} + r * X1_k$
16. $B \leftarrow$RandomNonSingularBooleanMatrix(n) ($*$ Random non singular Boolean matrix of size n $*$)
17. $C \leftarrow$RandomBooleanVector(n) ($*$ Random Boolean vector of size n $*$)

18. $A1 \leftarrow B * A1$
19. $X1 \leftarrow B * X1 + C$
20. $L1 \leftarrow$RandomNonSingularBooleanMatrix(n) ($*$ Random non singular Boolean matrix of size n $*$)
21. $L2 \leftarrow$RandomNonSingularBooleanMatrix(n) ($*$ Random non singular Boolean matrix of size n $*$)
22. $A1 \leftarrow$LinTrans$(A1, L1)$ ($*$ Lineary transform the indeterminates of $A1$ according to $L1$ $*$)
23. $X1 \leftarrow$LinTrans$(X1, L1)$ ($*$ Lineary transform the indeterminates of $X1$ according to $L1$ $*$)
24. $X2 \leftarrow$LinTrans$(X2, L2)$ ($*$ Lineary transform the indeterminates of $X2$ according to $L2$ $*$)
25. MQQ $\leftarrow A1 * X2 + X1$
26. GBMQQ $\leftarrow$Gröbner(MQQ,2) ($*$ The truncated Gröbnerbasis of degree 2 under graded reverse lexicographical ordering. $*$)
27. Deg $\leftarrow$\{number of linear terms in GBMQQ\}
28. **if** Deg$= L$
29. **then** CorrectDeg $\leftarrow$False
30. **return** GBMQQ