Analysis of the MQQ Public Key Cryptosystem

Jean-Charles Faugère, Rune Ødegard, Ludovic Perret, Danilo Gligoroski

Analysis of the MQQ Public Key Cryptosystem

Jean-Charles Faugère[2], Rune Steinsmo Ødegård[1] ⋆, Ludovic Perret[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] SALSA Project - INRIA (Centre Paris-Rocquencourt)
UPMC, Univ Paris 06 - CNRS, UMR 7606, LIP6
104, avenue du Président Kennedy 75016 Paris, France
jean-charles.faugere@inria.fr, ludovic.perret@lip6.fr
[3] Department of Telematics at the Norwegian University of Science and Technology in Trondheim, Norway
danilog@item.ntnu.no

Abstract. MQQ is a multivariate public key cryptosystem (MPKC) based on multivariate quadratic quasigroups and a special transform called "*Dobbertin transformation*" [17]. The security of MQQ, as well as any MPKC, reduces to the difficulty of solving a non-linear system of equations easily derived from the public key. In [26], it has been observed that that the algebraic systems obtained are much easier to solve that random non-linear systems of the same size. In this paper we go one step further in the analysis of MQQ. We explain why systems arising in MQQ are so easy to solve in practice. To do so, we consider the so-called the degree of regularity; which is the exponent in the complexity of a Gröbner basis computation. For MQQ systems, we show that this degree is bounded from above by a small constant. This is due to the fact that the complexity of solving the MQQ system is the minimum complexity of solving just one quasigroup block or solving the Dobbertin transformation. Furthermore, we show that the degree of regularity of the Dobbertin transformation is bounded from above by the same constant as the bound observed on MQQ system. We then investigate the strength of a tweaked MQQ system where the input of the Dobbertin transformation is replaced with random linear equations. It appears that the degree of regularity of this tweaked system varies both with 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 attacks for adequate choices of quasigroup 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

The use of polynomial systems in cryptography dates back to the mid eighties with the design of Matsumoto and Imai [25], later followed by numerous other proposals. Two excellent surveys on the current state of proposals for multivariate asymmetric cryptosystems has been made by Wolf and Preneel [33] as well as Billet and Ding [6]. Basically the current proposals can be classified into four main categories, some of which combine features from several categories: Matsumoto-Imai like schemes [28,30], Oil and Vinegar like schemes [29,20], Stepwise Triangular Schemes [31,18] and Polly Cracker Schemes [11]. In addition Gligoroski et al. has proposed a fifth class of trapdoor functions based on multivariate quadratic quasigroups [17].

As pointed out in [6], it appears that most multivariate public-key cryptosystems (MPKC) suffer from obvious to less obvious weaknesses. Some attacks are specific and focus on one particular variation and breaks it due to specific properties. One example is the attack of Kipnis and Shamir against the Oil and Vinegar scheme [21]. However, most attacks use general purpose algorithms that solve multivariate system of equations. 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. 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 achieve 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 basis based techniques.

Using Gröbner basis [8] is a well established and general method for solving polynomial systems of equations. The complexity of a Gröbner basis 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 Patarin on the Matsumoto-Imai scheme [27]. In this paper Patarin explains exactly why one is able to solve the system by using Gröbner bases. The key aspect is that there exists bilinear equations relating the input and output of the system [6]. This low degree relation between the input and the output means that only polynomials of a low degree will appear during the computation of the Gröbner basis. Consequently, the complexity of solving the system is bounded by this low degree.

Another multivariate cryptosystem which has been broken by Gröbner bases cryptanalysis is the MQQ public key block cipher [17]. The cipher was broken both by Gröbner bases and MutantXL independently in [26]. Given a ciphertext encrypted using the public key, the authors of [26] were able to compute the corresponding plaintext. However, the paper 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 bases and reiterate the generic complexity of computing such bases. 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 charcteristic by looking at the shape of the inner system. In Section 5 we further elaborate on the weaknesses of MQQ, and investigate if some tweaks can make the system stronger. Finally, Section 6 concludes the paper.

2 Description of the MQQ public key cryptosystem

In this section we give a description of the multivariate quadratic quasigroup public key cryptosystem [17]. The system is based on previous work by Gligoroski and Markovski who introduced the use of quasigroup string processing in cryptography [23,24].

2.1 Multivariate quadratic quasigroups

We first introduce the key building block of the MQQ PKC, namely multivariate quadratic quasigroups. For a detailed introduction to quasigroups in general, we refer the interested reader to [32].

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 respectively. 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

$$\begin{array}{rl} \beta : Q \rightarrow & GF(2^d) \\ a \mapsto & (x_1,\ldots,x_d) \end{array} \tag{1}$$

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

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

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 $f_i : GF(2^d) \rightarrow GF(2)$ is determined by $*$. This gives us the following lemma [17].

Lemma 1 *For every quasigroup $(Q,*)$ of order 2^d and for each bijection $\beta : Q \rightarrow GF(2^d)$ 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 *([17]) 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 a 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. [17] mention 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. [9] 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, namely:

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 improved Theorem 2 from [17] which gives a sufficient condition for a quasigroup to be MQQ. We restate this result 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 two non-singular Boolean matrices $\mathbf{B}_1$ and $\mathbf{B}_2$

In addition Chen et al. [9] proved that no MQQ as in Theorem 1 can be of strict type $\mathrm{Quad}^s_d\mathrm{Lin}^s_0$. This result uncovered a possible weakness in [17] as the proposed scheme used 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 have been produced using the algorithm provided in Appendix A. The algorithm is based on the paper [9], and produces MQQs that are more suitable for encryption since they are guaranteed to be of strict type $\mathrm{Quad}^s_{d-k}\mathrm{Lin}^s_k$ for $0<k\leq d$.

2.2 The Dobbertin bijection

In addition to MQQs, [17] also uses a bijection introduced by Dobbertin in [12]. Dobbertin proved that the following function, in addition to being multivariate quadratic over $GF(2)$, is a bijection in $GF(2^{2r+1})$:

$$\begin{array}{rcl} D_r : GF(2^{2r+1}) & \rightarrow & GF(2^{2r+1}) \\ x & \mapsto & x^{2^{r+1}+1} + x^3 + x \end{array} \tag{6}$$

2.3 A Public Key Cryptosystem Based on MQQ

We are now ready to describe the public key cryptosystem presented by Gligoroski et al. in [17]. 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 an $N \times N$ non-singular Boolean matrix $\mathbf{S}$, and compute $\mathbf{X} \leftarrow \mathbf{S} \cdot \mathbf{X}$.
2. Randomly choose a 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 $Y_{j+1} = X_j *_{vv}^{i_j} X_{j+1}$ for $j = 1, \ldots, n-1$.
4. Set $\mathbf{Z}$ to be the vector of all the linear terms of $Y_1, \ldots, Y_n$. Here Y_1 will be all linear terms, while each Y_j has between 1 and k linear terms depending on the type $\mathrm{Quad}_{d-k}^s \mathrm{Lin}_k^s$ 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 summing up to 27. Finally, set $\mathbf{W} \leftarrow \mathrm{Dob}(\mathbf{Z})$.
5. Replace the linear terms of $\mathbf{Y} = [Y_1, \ldots, Y_n]^T$ with the terms in $\mathbf{W}$. Randomly generate an $N \times N$ 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 bases

This section introduces the concept of Gröbner bases as well as a complexity bound to compute such bases. We refer to (for instance) [10] for basic definitions, and a more detailed description of the concepts.

Let $\mathbb{K}$ be a field and $\mathbb{K}[x_1, \ldots, x_N]$ be 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 $\mathbb{K}[x_1, \ldots, x_n]$. The most common example of such ordering 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 monomial orderings Gröbner bases hold specific theoretical properties and show 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 *([10]) Fix a monomial order $>$ on $\mathbb{K}[x_1,\ldots,x_N]$, and let $I \subset \mathbb{K}[x_1,\ldots,x_N]$ 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

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

In our case we are interested in the solutions of the MQQ system, which are defined over $GF(2)$.

Proposition 1 *([15]) Let G be a Gröbner basis of $[f_1,\ldots,f_m,x_1^2-x_1,\ldots,x_N^2-x_N]$. Then the following holds:*

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 $a_i \in GF(2)$. Then $(a_1,\ldots,a_N)$ is the solution in $GF(2)$ of the algebraic system.*

It is clear that as we are solving systems over $GF(2)$ we have to add the field equations $x_i^2 = x_i$ for $i = 1,\ldots,N$. This means that we have to compute 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 [15].

3.1 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 [8]. 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[13] and F_5 [14] algorithms. In this paper we have used the magma [22] 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 maximum degree of the polynomials appearing during the computation. This degree, called *degree of regularity*, is the key parameter for understanding the complexity of a Gröbner basis computation [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 $\approx N^{D_{\text{reg}}}$. 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]. However, as soon as the system has some kind of structure, this degree is much more difficult to predict. In some particular cases, it is actually possible to bound the degree of regularity (see the works done on HFE [15,19]). But this is a hard task in general.

As already pointed out, the degree of regularity is abnormally small for algebraic systems occuring in MQQ. This fact explains the weakness observed in [26]. In this paper, we go one step further in the security analysis by explaining why the degree of regularity is so small for MQQ.

Note that the degree of regularity is related to the ideal $I = \langle f_1, \ldots, f_m \rangle$ and not the equations $f_1, \ldots, f_m$ themselves. In particular, for any non-singular matrix T, the degree of regularity of $[f'_1, \ldots, f'_m]^t = T \cdot [f_1, \ldots, f_m]^t$ is similar to the degree of regularity of $[f_1, \ldots, f_m]$. More generally, we can assume that this degree is generically (i.e. with high probability) invariant for a random (invertible) linear change of variables, and an (invertible) combination of the polynomials. These are exactly the transformations performed to mask the MQQ structure. Note that such a hypothesis has already been used for instance in [19].

4 Why MQQ is Susceptible to Algebraic Cryptanalysis

In [26], MQQ systems with up to 160 variables was broken usin MutantXL (the same result has aslo been obtained independently with F_4). The most important point made by [26] 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 with the number of variables N. Indeed, for a random system it holds that D_{reg} is asymptotically equivalent to $\frac{N}{11.114}$ [2]. The authors of [26] observed that this low degree is due to the occurrence of many new low degree relations during the computation of a Gröbner basis. In Section 4.2, we will explain in detail how the very structure of the MQQ system results in the apparance of the low degree relations. First, however, we will show that same upper bound on the degree of regularity is obtained 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 systems is related to the number of variables, we constructed MQQ systems in $30, 60, 120$ and 180 variables 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. We then tried to compute the plaintext given a ciphertext encrypted with the public key. 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. This means breaking the MQQ system is only polynomial in

Table 1. Results for MQQ-(30,60,120,180). Computed with magma 2.16-1's implementation of the F_4 algorithm on a Intel Xeon 2.93GHz quad core 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

the number of variables. Once again, this is not the behaviour of a random system of equations, for which the degree of regularity increases linearly with the number of variables, and the solving time therefore increases exponentially. We explain the reason of such difference in the next section.

4.2 Shape of the MQQ system

The non-random behavior described above can be explained by considering the shape of the "unmasked" MQQ system. By unmasked we mean the MQQ system without the linear transformations S and T. As already explained in Section 3.1, the maximum degree of the polynomials occurring in the computation of a Gröbner basis is invariant under the linear transformation S and T.

In Figure 1 we show which 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 the Dobbertin bijection of size 7. Obviously, a random multivariate system would use all 60 variables in all equations. For this instance of MQQ, only $\frac{1}{3}$ of the variables are used in each quasigroup and about $\frac{2}{3}$ is used in each block of the Dobbertin transformation.

Now assume that the Gröbner basis algorithm somewhere during the calculation has found the solution for one of the quasigroup blocks $Y_j = X_j *_{vv}^{i_j} X_{j+1}$. 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 $Y_{j+1} = X_{j+1} *_{vv}^{i_{j+1}} X_{j+2}$. In Section 2.1 we showed that if we set the first or the last half of the variables of an MQQ to constant, all equations become 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 a 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. As a conclusion, solving a MQQ system is reduced to either solving just one block of quasigroup equations, or solving the Dobbertin transformation. The security of solving an MQQ system is therefore the minimum complexity between solving the Dobbertin transformation or one MQQ block.

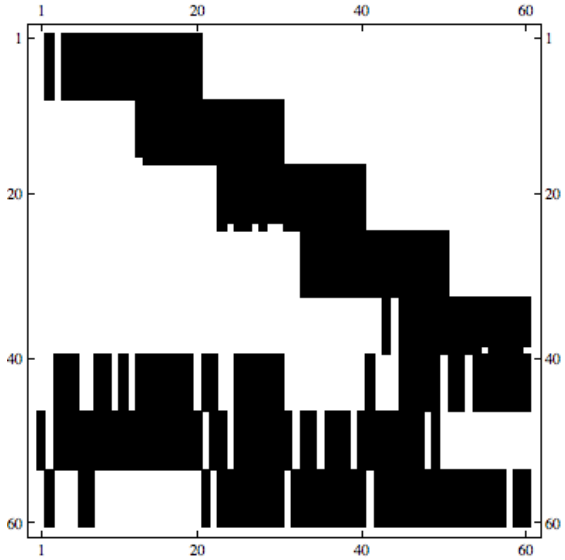

Fig. 1. Shape of 60 variable MQQ public key system without the use of S and T transformations. The black color means that the corresponding variables is used in the equation. The system was constructed using 4 MQQs of type $\text{Quad}_8^s\text{Lin}_2^s$, one MQQ of type $\text{Quad}_7^s\text{Lin}_3^s$,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

5 Weaknesses of MQQ

The goal of this part is to determine the weakest part of the system; the Dobbertin transformation or the quasigroup transformation. We first look closer at the Dobbertin block of equations. Since these equations constitutes a square system of equations, we expect them to be easier to solve then the quasigroup block of equations, which is an undetermined system of equations.

5.1 The Dobbertin transformation

Recall that the Dobbertin transformation is a bijection over $GF(2^{2r+1})$ 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 equations in $2r+1$ variables. Using magma 2.16-1's implementation of the F_4 algorithm[4], we experimentally computed the degree of regularity for solving this system of equations for $r = 2, \ldots, 22$. We observed that the degree of regularity was 3 for all computed instances. Therefore the Dobbertin transformation can be easily solved by a Gröbner basis 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.

[4] The computer used was 4 processor Intel Xeon 2.93GHz computer with 128GB of memory

Proving mathematically (if true) that the degree of regularity of $D_r(x)$ is constant at 3 for all r is difficult. We can, however, explain why 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:

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

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 [7]. 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 helpful since a lot of work has been done on the degree of regularity for Gröbner basis compuation of HFE polynomials [15,7]. Indeed, it has been proved that the degree of regularity for HFE polynomial of degree d is bounded from above by $\log_2(d)$ [15,16]. For Dobbertin's transformation this means the degree of regularity is bounded from above by $r+1$ at least.

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. Similarly to the weak-key polynomials in [7], the Dobbertin transformation commutes with the Frobenius automorphism and its iterates $F_i(x) : x \mapsto x^{2^i}$ for $0 \le i \le n$, namely

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

Thus $D_r(x) = 0$ implies that $F_i \circ D_r(x) = 0$. This means for each i we can add the $2r+1$ equations over $GF(2)$ corresponding to the equation $D_r \circ F_i(x) = 0$ over $GF(2^{2r+1})$ 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 $gcd(i, 2r+1) = gcd(j, 2r+1)$ [7]. Worst case scenario is when $2r+1$ is prime. The Frobenius automorphism then gives us (only) $2(2r+1)$ equations in $2r+1$ 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 $m = N(k+o(1))$ with $k > 1/4$ the degree of regularity is

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

Setting $k = 2$ we get $D_{\text{reg}} = -\frac{3}{2} + \frac{1}{2}\sqrt{-13 + 16\sqrt{2}} \cdot (2r+1) \approx 0.051404 \cdot (2r+1) = 0.102808 \cdot r + 0.051404$. Note that the degree of regularity can not be smaller then 3. This means we have $max(3, 0.102808 \cdot r + 0.051404)$ as an upper bound for a *random* multivariate system with the same number of equations and variables as the Dobbertin transformation. This provides 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 performed some experiments where we replaced the input of the Dobbertin transformation by random linear

equations. This means that solving a Dobbertin transformation block will no longer make all the equations in the first quasigroup transformation linear. The result of our experiment on this special MQQ system where the linear equations are perfectly masked is listed in Table 2. Note that the degree of regularity of 5 is still too small to prevent Gröbner bases attacks. What is important is how the degree of regularity increases when we increase different parameters. 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 the Dobbertin transformation with a stronger function, the MQQ system can possibly be made strong enough to resist pure Gröbner attacsk for adequate choices of quasigroup size and number of variables.

Table 2. Effects of quasigroup size and the Dobbertin transformation on the observed degree of regularity for different MQQ. 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 further explained the results of [26] by showing that the degree of regularity for MQQ systems are bounded from above by a small constant. Therefore even MQQ systems with large number of variables can easily be broken with Gröbner bases cryptanalysis. The main result of this paper is an explanation of the underlying reason for this abnormal degree of regularity. We demonstrated 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. These experimental results were also explained mathematically. A natural interpretation of the results of our investigation is that the Dobbertin transformation employed is a serious weakness in the MQQ system.

From a design point of view, we also showed that if Dobbertin's transformation is replaced with an ideal function – which perfectly hides the linear parts of the system

– the degree of regularity varies with the size of the quasigroups and the number of variables. We conclude that if a suitable replacement for Dobbertin's transformation is found, MQQ can possibly be made strong enough to resist pure Gröbner attacsk for adequate choices of quasigroup size and number of variables. This remains an interesting open problem.

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. 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.
8. Bruno Buchberger. *Ein Algorithmus zum Auffinden der Basiselemente des Restklassenringes nach einem nulldimensionalen Polynomideal*. PhD thesis, Leopold-Franzens University, 1965.
9. Yanling Chen, Svein Johan Knapskog, and Danilo Gligoroski. Multivariate Quadratic Quasigroups (MQQ): Construction, Bounds and Complexity. Submitted to ISIT 2010, 2010.
10. David Cox, John Little, and Donal O'Shea. *Using Algebraix Geometry*. Springer, 2005.
11. 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.
12. Hans Dobbertin. One-to-one highly nonlinear power functions on GF(2^n). *Appl. Algebra Eng. Commun. Comput.*, 9(2):139–152, 1998.
13. 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.
14. 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.
15. 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.

16. 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.
17. 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).
18. Louis Goubin, Nicolas T. Courtois, and Schlumbergersema Cp. Cryptanalysis of the ttm cryptosystem. In *Advances of Cryptology, Asiacrypt2000*, pages 44–57. Springer, 2000.
19. Louis Granboulan, Antoine Joux, and Jacques Stern. Inverting HFE is quasipolynomial. In *CRYPTO*, pages 345–356, 2006.
20. 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.
21. 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.
22. MAGMA. High performance software for algebra, number theory, and geometry — a large commercial software package. http://magma.maths.usyd.edu.au.
23. 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.
24. Smile Markovski, Danilo Gligoroski, and Verica Bakeva. Quasigroup string processing. In *Part 1, Contributions, Sec. Math. Tech. Sci., MANU, XX*, pages 1–2, 1999.
25. 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.
26. 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.
27. Jacques Patarin. Cryptanalysis of the Matsumoto and Imai public key scheme of Eurocrypt'88. In *Lecture Notes in Computer Science*, pages 248–261, 1995.
28. 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.
29. Jacques Patarin. The oil & vinegar signature scheme, 1997.
30. 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.
31. 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.
32. J. D. H. Smith. *An introduction to quasigroups and their representations*. Chapman & Hall/CRC, 2007.
33. 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