# MULTIVARIATE PUBLIC KEY CRYPTOSYSTEMS FROM DIOPHANTINE EQUATIONS

SHUHONG GAO AND RAYMOND HEINDL

Abstract. At CT-RSA 2006, Wang et al. [WYHL06] introduced the MFE cryptosystem, which was subsequently broken by Ding et al. [DHNW07]. Inspired by their work, we present a more general framework for multivariate public key cryptosystems, which combines ideas from both triangular and oil-vinegar schemes. We also propose a new public key cryptosystem, based on Diophantine equations, which implements the framework.

## 1. Introduction

1.1. **Multivariate Public Key Cryptography.** Public key cryptography plays an integral role in secure digital communication. Cryptosystems such as RSA and ElGamal have gained much popularity; however, if large enough quantum computers can be built, number theoretic systems such as these will be rendered useless due to Shor's algorithm [Sho97]. Also, these systems suffer from slow speeds, so it would be desirable to develop systems which operate more efficiently.

Multivariate public key cryptosystems (MPKC) are one possible alternative to the current public key schemes. The public key of an MPKC is a system of multivariate polynomials, usually quadratic, over a finite field. This idea is based on the fact that solving a multivariate polynomial system over a finite field is an NP-complete problem. In recent years, much inquiry has been made into the subject of multivariate public key cryptography, and several schemes have been proposed. In general, MPKCs have the following structure. Let $k$ be a finite field with $q$ elements. Although the public key, $\bar{F}: k^n \to k^m$ will appear to be a random system of multivariate polynomials, we build it by composing three maps:

$$\bar{F} = L_1 \circ F \circ L_2$$

where $L_1 : k^m \to k^m$ and $L_2 : k^n \to k^n$ are two random invertible affine transformations, and the central map $F : k^n \to k^m$ is a nonlinear multivariate polynomial map which has the property that we can find preimages. This is the trapdoor that will facilitate decryption. (Note that in some systems, $m > n$, but in others, as we shall see in the next section, $m = n$.) The private key consists of $L_1$ and $L_2$, and sometimes $F$. Creating such an $F$ requires adding structure, and though many ideas have been suggested, in most cases, the added structure has led to the discovery of some weakness.

*Key words and phrases.* multivariate public key cryptosystem, Gröbner basis, polynomial identity.

*2000 Mathematics Subject Classification.* Primary 12Y05, 68W30; Secondary 11Y16, 12D05, 13P05.

The authors are grateful to the Institute of Software and the Key Laboratory of Mathematics Mechanization, Academia Sinica, for their hospitality and support while part of this work was being done. The authors were partially supported by National Science Foundation under Grant DMS-0302549 and a US-China collaborate grant from Chinese National Science Foundation.

In this paper, we begin by discussing two existing types of MPKCs: triangular and oil-vinegar systems. Then in Section 2, we introduce a new framework for multivariate systems that combines these two types of systems. We next show that the MFE cryptosystem [WYHL06] can be viewed as an example of our proposed framework. In Sections 3 and 4, we give an implementation of the framework, and Section 5 presents the cryptanalysis of the system. We conclude in Section 6 by posing some open questions and making some final remarks.

1.2. **Triangular Encryption Schemes.** Triangular maps make up one family of easily inverted multivariate maps. A triangular map $F : k^n \to k^n$ has the form:

$$F(x_1,\ldots,x_n) = \begin{pmatrix} x_1 \\ x_2 + g_2(x_1) \\ \vdots \\ x_{n-1} + g_{n-1}(x_1, x_2, \ldots, x_{n-2}) \\ x_n + g_n(x_1, x_2, \ldots, x_{n-2}, x_{n-1}) \end{pmatrix}^T,$$

where each $g_i \in k[x_1,\ldots,x_n]$ is quadratic. Given $(y_1,\ldots,y_n) \in k^n$, it is easy to find $(x_1,\ldots,x_n) \in k^n$ such that $F(x_1,\ldots,x_n) = (y_1,\ldots,y_n)$ by iteratively solving for each component.

Because the transformations $L_1$ and $L_2$ are linear, they cannot hide the linearity of the first equation in $F$, and we cannot build a secure system which simply has a triangular map as the central map. At least two possible ideas have been proposed to circumvent this issue. One is to simply discard several of the initial polynomials and use the remaining system to create a signature scheme [YC05]. Another is to compose more than one triangular system. The inherent difficulty in the latter is that composition in general makes the degree grow very quickly, which is problematic since we desire our central maps to be quadratic. Moh [Moh99] found a way of doing this by composing two triangular maps, one having degree eight, and by using injections (basically adding new variables which are set to zero). However, Moh's original system is susceptible to the minrank attack [GC00], and later modified systems are vulnerable to linearization equation attacks [NJHD07].

Wang et al. [WC04] proposed a generalization of triangular maps that they called tractable rational maps. They define a tractable rational map $F : k^n \to k^n$ as having the form:

$$F(x_1,\ldots,x_n) = \begin{pmatrix} r_1(x_1) \\ r_2(x_2) \cdot \frac{p_2(x_1)}{q_2(x_1)} + \frac{f_2(x_1)}{g_2(x_1)} \\ \vdots \\ r_{n-1}(x_{n-1}) \cdot \frac{p_{n-1}(x_1,x_2,\ldots,x_{n-2})}{q_{n-1}(x_1,x_2,\ldots,x_{n-2})} + \frac{f_{n-1}(x_1,x_2,\ldots,x_{n-2})}{g_{n-1}(x_1,x_2,\ldots,x_{n-2})} \\ r_n(x_n) \cdot \frac{p_n(x_1,x_2,\ldots,x_{n-2},x_{n-1})}{q_n(x_1,x_2,\ldots,x_{n-2},x_{n-1})} + \frac{f_n(x_1,x_2,\ldots,x_{n-2},x_{n-1})}{g_n(x_1,x_2,\ldots,x_{n-2},x_{n-1})} \end{pmatrix}^T,$$

where $p_i, q_i, f_i$, and $g_i$ are polynomials, and $r_i$ is a permutation polynomial over $k$. As in the triangular case, we can find preimages by iteratively solving for each component. However, notice that the rational functions limit the invertibility of $F$ to the set

$$\{(x_1,\ldots,x_n) \in k^n : (p_i q_i g_i)(x_1,\ldots,x_n) \neq 0 \text{ for } i = 2,\ldots,n\}.$$

Rather than composing two maps as Moh did, they introduce the idea of using basic injections and projections to effectively discard the weak top part of the triangle while still being able to compute unique preimages by exploiting other structure. Using this structure, Wang et al. [WYHL06] proposed the MFE cryptosystem, which we will discuss in Section 2.2.

1.3. **Oil-Vinegar Systems.** A second type of MPKC that is interesting for our purposes is called an oil-vinegar signature scheme. Patarin's oil-vinegar polynomial scheme [Pat97] finds its roots in his linearization equation attack [Pat95] on the Matsumoto-Imai cryptosystem. An oil-vinegar polynomial $f \in k[\check{x}_1, \ldots, \check{x}_v, x_1, \ldots, x_o]$ has the form:

$$f = \sum_{i=1}^{o}\sum_{j=1}^{v} a_{ij}x_i\check{x}_j + \sum_{i=1}^{v}\sum_{j=1}^{v} b_{ij}\check{x}_i\check{x}_j + \sum_{i=1}^{o} c_i x_i + \sum_{j=1}^{v} d_j\check{x}_j + e,$$

where $a_{ij}, b_{ij}, c_i, d_j, e \in k$. The variables $x_1, \ldots, x_o$ are called oil variables and the variables $\check{x}_1, \ldots, \check{x}_v$ are called vinegar variables. The important property of these polynomials is that they have no $x_ix_j$ terms (i.e. there are no terms quadratic in the oil variables). So, if we substitute $v$ field values for the vinegar variables, $f$ becomes linear in the oil variables. Basic oil-vinegar systems may be used for signatures as follows: let the private key be given by $F = (f_1, \ldots, f_o)$, where each $f_i$ is a random oil-vinegar polynomial, along with an invertible affine transformation $L : k^{o+v} \to k^{o+v}$. The public key is the map $\bar{F} = F \circ L$. Let $(y_1, \ldots, y_o) \in k^o$ be a document that a user wants to sign. The user chooses $(\check{x}'_1, \ldots, \check{x}'_v) \in k^v$ at random and attempts to compute $(x_1, \ldots, x_o)$ that satisfies the linear system

$$F(\check{x}'_1, \ldots, \check{x}'_v, x_1, \ldots, x_o) = (y_1, \ldots, y_o).$$

A solution will exist as long as the system is nonsingular. If the resulting matrix for the linear system is singular, simply choose a different $(\check{x}'_1, \ldots, \check{x}'_v) \in k^v$ and try again. With high probability, one should be able to compute a solution $(x'_1, \ldots, x'_o) \in k^o$ in very few attempts. Finally, the signature is $(z_1, \ldots, z_{o+v}) = L^{-1}(\check{x}'_1, \ldots, \check{x}'_v, x'_1, \ldots, x'_o)$. Signature verification is done by simply checking that $\bar{F}(z_1, \ldots, z_{o+v}) = (y_1, \ldots, y_o)$.

Kipnis and Shamir [KS98] first broke the system in the case where $o = v$ using the observation that the matrices corresponding to the quadratic forms of the private key have a special form (i.e. a large block of zeros). This allows attackers to separate the oil and vinegar variables and generate an equivalent system that can be used to create forgeries. Later Kipnis et al. [KPG99b] proposed an unbalanced ($o < v$) scheme, extended the original attack to this case, and gave parameters they believed would be good for a secure system. Later, Ding and Schmidt proposed a more efficient "multi-layer" unbalanced oil-vinegar scheme, called Rainbow [DS05].

However, these are signature schemes, and our goal is to build a secure cryptosystem.

## 2. Combining Triangular and Oil-Vinegar Schemes

Recall that the difficulty in creating a secure triangular system is that it is hard to hide the triangular structure, especially the top equations. Though attempts have been made to use high degree "lock polynomials" through composition with another triangular map ([Moh99], [MCY04], [Moh07]), these have been shown to be insecure ([GC00], [DS03], [NJHD07]). However, this method is not the only way to achieve the necessary hiding of the triangular structure. We propose a new way of introducing lock polynomials to completely hide the triangular system by combining the triangular system with a series of oil-vinegar systems.

Let $k$ be a finite field with $q$ elements, and let $\mathbb{F}$ be a degree $d$ extension of $k$. Notice that although we are working in an extension field, our polynomials will be multivariate, as opposed to the univariate polynomials used to build "big-field" systems such as Matsumoto-Imai and HFE. Our approach might be called an "intermediate" (or as Wang et al. [WYHL06] say, "medium") field construction.

In particular, fix a basis $\{\alpha_1, \ldots, \alpha_d\}$ of $\mathbb{F}$ over $k$. We identify $\mathbb{F}$ with $k^d$, via the natural map $\pi : \mathbb{F} \to k^d$ given by

$$\pi(a_1\alpha_1 + \cdots + a_d\alpha_d) = (a_1, \ldots, a_d).$$

Similarly we can view a polynomial $f \in \mathbb{F}[X_1, \ldots, X_n]$ component-wise over $k$ by writing $X_i = x_{i1}\alpha_1 + \cdots + x_{id}\alpha_d$, and then $f = f_1\alpha_1 + \cdots + f_d\alpha_d$ with $f_i \in k[x_{11}, \ldots, x_{nd}]$. Finally, we can extend $\pi$ to the polynomial rings via

$$f \in \mathbb{F}[X_1, \ldots, X_n] \mapsto (f_1, \ldots, f_d) \in k[x_{11}, \ldots, x_{nd}]^d.$$

2.1. **A general framework.** As mentioned above, the public key will be given by $\bar{F} = L_1 \circ F \circ L_2$, where $L_1$ and $L_2$ are invertible affine transformations. Suppose $(Y_1, \ldots, Y_n) = \phi(X_1, \ldots, X_n)$ is a triangular system when viewed component-wise over the base field $k$:

$$\begin{array}{rcl} Y_1 & = & X_1 + \phi_1(X_1) \\ Y_2 & = & X_2 + \phi_2(X_1, X_2) \\ & \vdots & \\ Y_n & = & X_n + \phi_n(X_1, \ldots, X_n). \end{array} \tag{1}$$

More specifically, viewing each polynomial as having $d$ components:

$$Y_i = \begin{pmatrix} Y_{i1} \\ Y_{i2} \\ \vdots \\ Y_{id} \end{pmatrix}^T = \begin{pmatrix} x_{i1} + \phi_{i1}(x_{11}, \ldots, x_{i-1,d}) \\ x_{i2} + \phi_{i2}(x_{11}, \ldots, x_{i-1,d}, x_{i1}) \\ \vdots \\ x_{id} + \phi_{id}(x_{11}, \ldots, x_{i-1,d}, x_{i1}, \ldots, x_{i,d-1}) \end{pmatrix}^T.$$

where each $\phi_{ij}$ is quadratic. To invert, we solve iteratively for $x_{11}, \ldots, x_{1d}, \ldots, x_{n1}, \ldots, x_{nd}$.

Similar to Rainbow [DS05], we will define several, say $\ell$, layers of oil-vinegar systems. However, in our framework, we make the following relaxation: rather than requiring an oil-vinegar system with $o$ oil variables and $v$ vinegar variables to have $o$ oil-vinegar polynomials, we allow more general systems, with $t$ ($\geq o$) polynomials, as long as at least $o$ of them are true oil-vinegar polynomials.

We now build the central map $F : \mathbb{F}^{n+\ell o} \to \mathbb{F}^{n+\ell t}$. Let $\{X_1, \ldots, X_n\}$ be the initial set of vinegar variables, and define the first oil-vinegar system:

$$Y_{n+i} = f_{n+i}(X_1, \ldots, X_n, X_{n+1}, \ldots, X_{n+o}), \quad 1 \leq i \leq t,$$

where $X_{n+1}, \ldots, X_{n+o}$ are the oil variables. In the next layer,

$$Y_{n+i} = f_{n+i}(X_1, \ldots, X_{n+o}, X_{n+o+1}, \ldots, X_{n+2o}), \quad t+1 \leq i \leq 2t,$$

the set of vinegar variables is $\{X_1, \ldots, X_{n+o}\}$, and the set of oil variables is $\{X_{n+o+1}, \ldots, X_{n+2o}\}$. Similarly, we create the other layers, ending with the $\ell$-th layer,

$$Y_{n+i} = f_{n+i}(X_1, \ldots, X_{n+(\ell-1)o}, X_{n+(\ell-1)o+1}, \ldots, X_{n+\ell o}), \quad (\ell-1)t+1 \leq i \leq \ell t,$$

where $\{X_1, \ldots, X_{n+(\ell-1)o}\}$ is the set of vinegar variables, and $\{X_{n+(\ell-1)o+1}, \ldots, X_{n+\ell o}\}$ is the set of oil variables. (Here, we are assuming that each oil-vinegar system has $t$ polynomials. We could be more general by letting the $i$-th system have $t_i$ polynomials.)

We will use these oil-vinegar systems to completely mask the triangular system (1). Decryption will involve first unmasking the triangular system, solving it for the initial set of vinegar variables, then sequentially solving the oil-vinegar systems for the oil variables.

Suppose we can define the $f_i$ in each oil-vinegar system in such a way that there exists nonlinear polynomials

$$g_i \in \mathbb{F}[Y_{n+(i-1)t+1}, \ldots, Y_{n+it}], \quad 1 \le i \le \ell,$$

such that each $g_i(f_{n+(i-1)t+1}, \ldots, f_{n+it})$, $1 \le i \le \ell$, factors as a product of quadratic factors in $\mathbb{F}[X_1, \ldots, X_{n+\ell o}]$. If we have $n$ such quadratic factors, say $\psi_1, \ldots, \psi_n$, then we can use them as lock polynomials by adding one factor to each $Y_i$ in the triangular system (1). That is, let

$$Y_i = f_i(X_1, \ldots, X_{n+\ell o}) = X_i + \phi_i(X_1, \ldots, X_i) + \psi_i(X_1, \ldots, X_{n+\ell o}) \quad 1 \le i \le n. \tag{2}$$

Appending the oil-vinegar systems to the updated triangular system gives our central map:

$$F(X_1, \ldots, X_{n+\ell o}) = (f_1, \ldots, f_{n+\ell t}).$$

Notice that as long as each $\psi_i$ has terms involving at least one of the variables $X_{i+1}, \ldots, X_{n+\ell o}$, the triangular structure of the first $n$ equations is destroyed. Also, we make the observation that we can shrink the size of the triangular system, and hence the number of necessary quadratic factors, to $n-1$, if one of the $\psi_i$ can be viewed as an oil-vinegar polynomial in $X_1, \ldots, X_n$ with a single oil variable $X_n$.

Now, in order to unmask and decrypt the triangular part, we must be able to compute the values of the $\psi_i$. Say there exist functions $h_i$ in the rational function field over $\mathbb{F}$ in $\ell$ variables such that

$$h_i(g_1, \ldots, g_\ell) = \psi_i, \quad 1 \le i \le n.$$

Then during decryption, we simply use $L_1^{-1}$ to compute $Y_{n+1}, \ldots, Y_{n+\ell t}$ from the ciphertext, substitute the values into $g_1, \ldots, g_\ell$, then evaluate each $h_i$, and substitute for each $\psi_i$ in (2), restoring the original triangular structure. There is actually much freedom in the $h_i$ since we can view them as functions of the transformed ciphertext values $Y_{n+1}, \ldots, Y_{n+\ell t} \in \mathbb{F}$, so we are not limited to polynomials, but may also compute inverses and roots (depending on the characteristic of the field). However, we must note that computing inverses will require that the involved $Y_i$'s are nonzero.

So, our proposed framework, which is simply a masked triangular system combined with a series of oil-vinegar systems, requires the existence of two crucial sets of functions:

- Polynomials $f_{n+(i-1)t+j} \in \mathbb{F}[X_1, \ldots, X_{n+io}]$ and $g_i \in \mathbb{F}[Y_{n+(i-1)t+1}, \ldots, Y_{n+it}]$ such that each $g_i(f_{n+(i-1)t+1}, \ldots, f_{n+it})$ factors into quadratics (the $\psi_i$'s) over $\mathbb{F}[X_1, \ldots, X_{n+io}]$.
- Functions $h_i$ which, upon evaluation at the transformed ciphertext values $Y_{n+1}, \ldots, Y_{n+\ell t}$, yield the value of $\psi_i$. We require that there must not exist linear relationships involving the $\psi_i$ and $Y_j$.

Notice that masking the triangular system and adding oil-vinegar polynomials has introduced two possibilities for decryption failure:

- We may not be able to compute inverses needed when evaluating the $h_i$.
- For any of the oil-vinegar systems, after we have computed the values of the vinegar variables, the remaining linear system in the oil variables may not be solvable.

Obviously, any practical cryptosystem must keep decryption failures to a minimum, so for any implementation, the probability of either of the above two problems occurring must be small. One possible solution is to make use of the embedding ($\nearrow$) modifier for MPKCs, first introduced in [DWY07].

2.2. **Example: MFE cryptosystem.** The MFE cryptosystem, although built using tractable rational maps, can be viewed (with slight modification) as an instance of our proposed framework. In fact, it was this system that inspired us to try to develop a more general system that avoids the known flaws of MFE.

We now present the central map of the MFE system in the context of our proposed framework, working only with polynomials over the extension field for ease of exposition. Let $\mathbb{F}$ have characteristic two. MFE's central map will be $F : \mathbb{F}^{12} \to \mathbb{F}^{15}$, where there are three oil-vinegar systems, given by $(Y_4, \dots, Y_7), (Y_8, \dots, Y_{11})$, and $(Y_{12}, \dots, Y_{15})$.

To motivate the definition of the functions $g_i$ and $\psi_i$, define the following matrices:

$$M_1 = \begin{pmatrix} X_1 & X_2 \\ X_3 & X_4 \end{pmatrix}, \quad M_2 = \begin{pmatrix} X_5 & X_6 \\ X_7 & X_8 \end{pmatrix}, \quad M_3 = \begin{pmatrix} X_9 & X_{10} \\ X_{11} & X_{12} \end{pmatrix},$$

and

$$Z_3 = M_1 M_2 = \begin{pmatrix} Y_4 & Y_5 \\ Y_6 & Y_7 \end{pmatrix}, \quad Z_2 = M_1 M_3 = \begin{pmatrix} Y_8 & Y_9 \\ Y_{10} & Y_{11} \end{pmatrix},$$

$$Z_1 = M_2^T M_3 = \begin{pmatrix} Y_{12} & Y_{13} \\ Y_{14} & Y_{15} \end{pmatrix}.$$

The $g_i$ and $\psi_i$ come from relationships between determinants. First notice that $\det(Z_3) = \det(M_1)\det(M_2)$, so letting $g_1 = \det(Z_3)$, $\psi_3 = \det(M_1)$, and $\psi_1 = \det(M_2)$, we have

$$g_1 = Y_4Y_7 + Y_5Y_6 = (X_1X_4 + X_2X_3)(X_5X_8 + X_6X_7) = \psi_3\psi_1.$$

Similarly $\det(Z_2) = \det(M_1)\det(M_3)$ and $\det(Z_1) = \det(M_2)\det(M_3)$ give

$$\begin{array}{rcccccl} g_2 & = & Y_8Y_{11} + Y_9Y_{10} & = & (X_1X_4 + X_2X_3)(X_9X_{12} + X_{10}X_{11}) & = & \psi_3\psi_2, \\ g_3 & = & Y_{12}Y_{15} + Y_{13}Y_{14} & = & (X_5X_8 + X_6X_7)(X_9X_{12} + X_{10}X_{11}) & = & \psi_1\psi_2. \end{array}$$

Also,

$$\begin{array}{rcllll} h_1 & = & (g_1g_3g_2^{-1})^{1/2} & = & ((\psi_3\psi_1)(\psi_1\psi_2)(\psi_3\psi_2)^{-1})^{1/2} & = \psi_1, \\ h_2 & = & g_3h_1^{-1} & & & = \psi_2, \\ h_3 & = & g_1h_1^{-1} & & & = \psi_3. \end{array}$$

Finally, the central map $F : \mathbb{F}^{12} \to \mathbb{F}^{15}$ is given by

$$\begin{array}{rcllrcl} Y_1 & = & X_1 + \phi_1(X_1) + \psi_1 & & & & \\ Y_2 & = & X_2 + \phi_2(X_1, X_2) + \psi_2 & & & & \\ Y_3 & = & X_3 + \phi_3(X_1, X_2, X_3) + \psi_3 & & & & \\ Y_4 & = & X_1X_5 + X_2X_7 & & Y_5 & = & X_1X_6 + X_2X_8 \\ Y_6 & = & X_3X_5 + X_4X_7 & & Y_7 & = & X_3X_6 + X_4X_8 \\ Y_8 & = & X_1X_9 + X_2X_{11} & & Y_9 & = & X_1X_{10} + X_2X_{12} \\ Y_{10} & = & X_3X_9 + X_4X_{11} & & Y_{11} & = & X_3X_{10} + X_4X_{12} \\ Y_{12} & = & X_5X_9 + X_7X_{11} & & Y_{13} & = & X_5X_{10} + X_7X_{12} \\ Y_{14} & = & X_6X_9 + X_8X_{11} & & Y_{15} & = & X_6X_{10} + X_8X_{12} \end{array}$$

The public key is given by $\bar{F} = L_1 \circ F \circ L_2$ where $L_1 : \mathbb{F}^{15} \to \mathbb{F}^{15}$ and $L_2 : \mathbb{F}^{12} \to \mathbb{F}^{12}$ are random invertible affine transformations.

Notice that the framework definition suggests that MFE's central map should be $F : \mathbb{F}^{16} \to \mathbb{F}^{16}$ (since $n = 4, o = t = 4$, and $\ell = 3$), however, it is given as $F : \mathbb{F}^{12} \to \mathbb{F}^{15}$. This is because the third oil-vinegar system does not utilize any new input variables, therefore shrinking the number of input variables by four. Also, $\psi_3$ is actually an oil-vinegar polynomial in $X_1, \dots, X_4$ with single oil variable $X_4$, so the triangular system only needs three polynomials.

To decrypt a ciphertext $(Y_1', \dots, Y_{15}')$, first calculate $(Y_1, \dots, Y_{15}) = L_1^{-1}(Y_1', \dots, Y_{15}')$, then use $h_1, h_2$, and $h_3$ to calculate $\psi_1, \psi_2$, and $\psi_3$. Adding these to $Y_1, Y_2$, and $Y_3$ respectively, restores the triangular structure of the first three polynomials and enables us to recover $X_1, X_2$, and $X_3$. We then use $\psi_3 = X_1X_4 + X_2X_3$ to compute $X_4$. Finally, using the values of the initial oil variables $X_1, \dots, X_4$, we solve in sequence the first two oil-vinegar systems to recover the values of the remaining variables, $X_5, \dots, X_{12}$. Note that the last system is not used in decryption, but is necessary for the $g_i$ and $\psi_i$ polynomials.

**Weakness of MFE.** The creators of the MFE specifically defined $Z_1 = M_2^T M_3$ instead of $Z_1 = M_2M_3$. Otherwise linearization equations (equations linear in both $X$ and $Y$) exist. For instance, the relationship

$$Z_3M_3 = M_1Z_1 \ (= M_1M_2M_3)$$

yields four linearization equations.

However, Ding et al. [DHNW07] showed that other types of linearization equations still exist, called high order linearization equations, where the degree in $Y$ is higher than one. Their second order linearization equations are derived by examining $M_3M_3^*M_1^*M_1M_2$, where $M_i^*$ is the adjoint of $M_i$. In particular,

$$M_3M_3^*M_1^*M_1M_2 = M_3(M_1M_3)^*(M_1M_2) = M_3Z_2^*Z_3$$

and

$$M_3M_3^*M_1^*M_1M_2 = \det(M_3)\det(M_1)M_2 = \det(Z_2)M_2,$$

therefore,

$$M_3Z_2^*Z_3 = \det(Z_2)M_2.$$

This equation gives four equations that are linear in $X$ and quadratic in $Y$. They show that enough of these second order linearization equations exist to break MFE.

We observe that in both cases, the linearization equation attacks result from the fact that the $Z$ matrices are defined as a product of $2 \times 2$ matrices. So, while the determinant relationships are crucial in giving the nice expressions for the $g_i$ and $\psi_i$, the underlying matrix relationships are the critical weakness of the system.

## 3. Polynomial Identities

Although the original form of MFE has been broken, our general framework may be used to create other systems. For instance, notice that each of three $g_i$ in MFE can be viewed as the right hand side of the Diophantine equation (over a polynomial ring):

$$AB = CD + EF, \tag{3}$$

where $C, D, E, F$ are oil-vinegar polynomials in 8 variables. In particular, for $\psi_3\psi_1 = g_1$ in MFE, we have

$$(X_1X_4 + X_2X_3)(X_5X_8 + X_6X_7) = Y_4Y_7 + Y_5Y_6,$$

where $Y_i \in \mathbb{F}[X_1, \ldots, X_8]$. So, solutions to equations like (3) will possibly yield families of cryptosystems in our proposed framework. This is in fact the case, and we now show how to construct a cryptosystem based on a Diophantine equation of the form

$$AB = CD + EF + GH + IJ + KL, \tag{4}$$

where $C, D, \ldots, J$ are oil-vinegar polynomials in 8 oil and 8 vinegar variables, and there are no restrictions on $K$ or $L$. In the context of our framework, we rewrite (4) as

$$\psi_1\psi_2 = f_1 f_2 + \cdots + f_9 f_{10}, \tag{5}$$

where each polynomial has degree two, and

(1) $\psi_1 \in \mathbb{F}[X_1, \ldots, X_n]$, $\psi_2 \in \mathbb{F}[Y_1, \ldots, Y_n]$,
(2) $f_i \in \mathbb{F}[X_1, \ldots, X_n, Y_1, \ldots, Y_n]$, $1 \le i \le 8$, are oil-vinegar polynomials, and
(3) $f_i \in \mathbb{F}[X_1, \ldots, X_n, Y_1, \ldots, Y_n]$, $i = 9, 10$.

Assume $\mathbb{F}$ has characteristic two. We begin our work in the polynomial ring

$$R = \mathbb{F}[x_1, x_2, x_3, x_4, y_1, y_2, y_3, y_4, z_1, z_2, z_3, z_4, w_1, w_2, w_3, w_4],$$

and introduce the following notation:

$$\begin{aligned} p_{xy}^{ij} &= x_i y_j + x_j y_i, & 1 \le i < j \le 4, \\ p^{ij}(x, y, z, w) &= p_{xz}^{ij} + p_{yz}^{ij} + p_{yw}^{ij}, & 1 \le i < j \le 4. \end{aligned}$$

From an algebraic geometry perspective, the $p_{xy}^{ij}$ are simply Plücker coordinates, which are known to satisfy certain quadratic relations. The following identity is easily verified:

$$\begin{aligned} 0 = \; & (p_{xy}^{12} + p_{zw}^{12})p^{34}(x, y, z, w) + (p_{xy}^{13} + p_{zw}^{13})p^{24}(x, y, z, w) + \\ & (p_{xy}^{14} + p_{zw}^{14})p^{23}(x, y, z, w) + (p_{xy}^{23} + p_{zw}^{23})p^{14}(x, y, z, w) + \\ & (p_{xy}^{24} + p_{zw}^{24})p^{13}(x, y, z, w) + (p_{xy}^{34} + p_{zw}^{34})p^{12}(x, y, z, w). \end{aligned} \tag{6}$$

To put (6) in the required oil-vinegar form, define $\rho : R \to \mathbb{F}[X_1, \ldots, X_8, Y_1, \ldots, Y_8]$ as a ring isomorphism induced by

$$\begin{aligned} &(x_1, x_2, x_3, x_4, y_1, y_2, y_3, y_4, z_1, z_2, z_3, z_4, w_1, w_2, w_3, w_4) \mapsto \\ &\quad (X_1, X_3, Y_1 + Y_5, Y_3 + Y_7, X_4, X_2, Y_5, Y_7, X_5, X_7, Y_4 + Y_8, Y_2 + Y_6, X_8, X_6, Y_8, Y_6), \end{aligned}$$

i.e., $\rho(x_1) = X_1, \rho(x_2) = X_3, \rho(x_3) = Y_1 + Y_5$, and so on. Then set

$$\begin{aligned} \psi_1 &= \rho(p_{xy}^{12} + p_{zw}^{12}) &&= X_1X_2 + X_3X_4 + X_5X_6 + X_7X_8 \\ \psi_2 &= \rho(p^{34}(x, y, z, w)) &&= Y_1Y_2 + Y_3Y_4 + Y_5Y_6 + Y_7Y_8 \\ f_1 &= \rho(p_{xy}^{13} + p_{zw}^{13}) &&= X_4Y_1 + X_8Y_4 + (X_1 + X_4)Y_5 + X_5Y_8 \\ f_2 &= \rho(p^{24}(x, y, z, w)) &&= (X_2 + X_3)Y_2 + X_7Y_3 + X_2Y_6 + X_6Y_7 \\ f_3 &= \rho(p_{xy}^{14} + p_{zw}^{14}) &&= X_8Y_2 + X_4Y_3 + X_5Y_6 + (X_1 + X_4)Y_7 \\ f_4 &= \rho(p^{23}(x, y, z, w)) &&= X_7Y_1 + (X_2 + X_3)Y_4 + X_6Y_5 + X_2Y_8 \\ f_5 &= \rho(p_{xy}^{23} + p_{zw}^{23}) &&= X_2Y_1 + X_6Y_4 + (X_2 + X_3)Y_5 + X_7Y_8 \\ f_6 &= \rho(p^{14}(x, y, z, w)) &&= (X_1 + X_4)Y_2 + X_5Y_3 + X_4Y_6 + X_8Y_7 \\ f_7 &= \rho(p_{xy}^{24} + p_{zw}^{24}) &&= X_6Y_2 + X_2Y_3 + X_7Y_6 + (X_2 + X_3)Y_7 \\ f_8 &= \rho(p^{13}(x, y, z, w)) &&= X_5Y_1 + (X_1 + X_4)Y_4 + X_8Y_5 + X_4Y_8 \\ f_9 &= \rho(p_{xy}^{34} + p_{zw}^{34}) &&= Y_1Y_7 + Y_2Y_8 + Y_3Y_5 + Y_4Y_6 \\ f_{10} &= \rho(p^{12}(x, y, z, w)) &&= X_1X_7 + X_2(X_5 + X_8) + X_3X_5 + X_4(X_6 + X_7) \end{aligned} \tag{7}$$

thus satisfying (5).

We now examine the oil-vinegar part of this system, i.e. $f_1, \ldots, f_8$. First consider the case where $X_1, \ldots, X_8$ are the vinegar variables. This yields the following linear system:

$$\begin{bmatrix} X_4 & 0 & 0 & X_8 & X_1+X_4 & 0 & 0 & X_5 \\ 0 & X_2+X_3 & X_7 & 0 & 0 & X_2 & X_6 & 0 \\ 0 & X_8 & X_4 & 0 & 0 & X_5 & X_1+X_4 & 0 \\ X_7 & 0 & 0 & X_2+X_3 & X_6 & 0 & 0 & X_2 \\ X_2 & 0 & 0 & X_6 & X_2+X_3 & 0 & 0 & X_7 \\ 0 & X_1+X_4 & X_5 & 0 & 0 & X_4 & X_8 & 0 \\ 0 & X_6 & X_2 & 0 & 0 & X_7 & X_2+X_3 & 0 \\ X_5 & 0 & 0 & X_1+X_4 & X_8 & 0 & 0 & X_4 \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \\ Y_6 \\ Y_7 \\ Y_8 \end{bmatrix} = \begin{bmatrix} f_1 \\ f_2 \\ f_3 \\ f_4 \\ f_5 \\ f_6 \\ f_7 \\ f_8 \end{bmatrix}$$

When $X_i$ and $f_i$, $1 \leq i \leq 8$, take on values in $\mathbb{F}$, we hope to be able to solve uniquely the above system for the oil variables $Y_1, \ldots, Y_8$. This will be possible whenever the coefficient matrix has nonzero determinant. The determinant is

$$\big((X_1+X_5+X_8)(X_2+X_7)+(X_3+X_6+X_7)(X_4+X_5)\big)^4.$$

On the other hand, if we view $Y_1, \ldots, Y_8$ as the oil variables, the determinant of the resulting linear system in $X_1, \ldots, X_8$ becomes

$$\big((Y_1+Y_8)(Y_2+Y_7)+(Y_3+Y_6)(Y_4+Y_5)\big)^4.$$

In both cases, the probability that the determinant is zero is $\frac{1}{|\mathbb{F}|} + \frac{1}{|\mathbb{F}|^2} - \frac{1}{|F|^3}$.

## 4. Building a cryptosystem

Unfortunately, a cryptosystem based directly on (7) will be susceptible to a separation of oil and vinegar variables attack, so we must introduce three additional, slightly different subsystems. Note that each permutation of $x, y, z$ and $w$ in (6) yields a new identity. In particular, when exchanging $x$ with $y$, or $z$ with $w$, the first factor of each term of (6) remains unchanged. We shall take advantage of this. Renaming each $\psi_j$ and $f_j$ in (7) as $\psi_{1,j}$ and $f_{1,j}$ respectively, we define

$$\psi_{i,1} = \psi_{1,1} \quad \text{and} \quad f_{i,j} = f_{1,j}, \quad i = 2, \ldots, 4, \quad j = 1, 3, \ldots, 9.$$

Then, interchanging $z$ with $w$ in (6), we define

$$\begin{aligned} \psi_{2,2} &= \rho(p^{34}(x,y,w,z)) \\ f_{2,2} &= \rho(p^{24}(x,y,w,z)) \\ f_{2,4} &= \rho(p^{23}(x,y,w,z)) \\ f_{2,6} &= \rho(p^{14}(x,y,w,z)) \\ f_{2,8} &= \rho(p^{13}(x,y,w,z)) \\ f_{2,10} &= \rho(p^{12}(x,y,w,z)). \end{aligned}$$

Similarly, by interchanging $x$ with $y$ in (6), we define $\psi_{3,2}$ and $f_{3,j}$, $j = 2, 4, \ldots, 10$. Finally, by interchanging $x$ with $y$, and $z$ with $w$ in (6), we define $\psi_{4,2}$ and $f_{4,j}$, $j = 2, 4, \ldots, 10$. Then we have four identities:

$$\psi_{i,1}\psi_{i,2} = f_{i,1}f_{i,2} + \cdots + f_{i,9}f_{i,10}, \quad 1 \leq i \leq 4.$$

Using these four subsystems, we define the central map,

$$(Z_1,\dots,Z_{74}) = F(X_1,\dots,X_{24},Y_1,\dots,Y_{32}),$$

by

$$\begin{array}{lcll}
Z_1 & = & X_1+\phi_1(X_1)+\psi_{1,1}(X_1,\dots,X_8) & \\
Z_2 & = & X_2+\phi_2(X_1,X_2)+\psi_{1,2}(Y_1,\dots,Y_8) & \\
Z_3 & = & X_3+\phi_3(X_1,\dots,X_3)+\psi_{2,2}(Y_9,\dots,Y_{16}) & \\
Z_4 & = & X_4+\phi_4(X_1,\dots,X_4)+\psi_{3,2}(Y_{17},\dots,Y_{24}) & \\
Z_5 & = & X_5+\phi_5(X_1,\dots,X_5)+\psi_{2,1}(X_9,\dots,X_{16}) & \\
Z_6 & = & X_6+\phi_6(X_1,\dots,X_6)+\psi_{3,1}(X_{17},\dots,X_{24}) & \\
Z_7 & = & X_7+\phi_7(X_1,\dots,X_7)+\psi_{4,2}(Y_{25},\dots,Y_{32}) & \\
Z_{7+i} & = & f_{1,i}(X_1,\dots,X_8,Y_1,\dots,Y_8) & 1\le i\le 10\\
Z_{17+i} & = & f_{2,i}(X_1,\dots,X_8,Y_9,\dots,Y_{16}) & 1\le i\le 10\\
Z_{27+i} & = & f_{2,i}(Y_1,\dots,Y_8,Y_9,\dots,Y_{16}) & 1\le i\le 8\\
Z_{36} & = & f_{2,10}(Y_1,\dots,Y_8,Y_9,\dots,Y_{16}) & \\
Z_{36+i} & = & f_{3,i}(X_1,\dots,X_8,Y_{17},\dots,Y_{24}) & 1\le i\le 10\\
Z_{46+i} & = & f_{2,i}(X_9,\dots,X_{16},Y_9,\dots,Y_{16}) & 1\le i\le 8\\
Z_{55} & = & f_{2,10}(X_9,\dots,X_{16},Y_9,\dots,Y_{16}) & \\
Z_{55+i} & = & f_{3,i}(X_{17},\dots,X_{24},Y_{17},\dots,Y_{24}) & 1\le i\le 8\\
Z_{64} & = & f_{3,10}(X_{17},\dots,X_{24},Y_{17},\dots,Y_{24}) & \\
Z_{64+i} & = & f_{4,i}(X_9,\dots,X_{16},Y_{25},\dots,Y_{32}) & 1\le i\le 10
\end{array}$$

Notice $f_{2,9}(Y_1,\dots,Y_8,Y_9,\dots,Y_{16})$ has been omitted from the central map to avoid redundancy as

$$f_{2,9}(Y_1,\dots,Y_8,Y_9,\dots,Y_{16}) = f_{2,9}(X_1,\dots,X_8,Y_9,\dots,Y_{16}) = Z_{26}. \tag{8}$$

Similarly, $f_{2,9}(X_9,\dots,X_{16},Y_9,\dots,Y_{16})$ and $f_{3,9}(X_{17},\dots,X_{24},Y_{17},\dots,Y_{24})$ are also omitted. Since the central map is from $\mathbb{F}^{56}$ to $\mathbb{F}^{74}$, the information rate of this cryptosystem is $\frac{56}{74}\approx .76$.

4.1. **Inverting the central map.** Recall that decryption proceeds by unmasking the triangular system $(Z_1,\dots,Z_7)$, and then solving the oil-vinegar systems. We start by focusing on the first three equations of the triangular system. Using the notation we introduced for the general framework, let

$$\begin{array}{lcll}
g_1 & = & Z_8Z_9+Z_{10}Z_{11}+Z_{12}Z_{13}+Z_{14}Z_{15}+Z_{16}Z_{17} & =\psi_{1,1}(X_1,\dots,X_8)\psi_{1,2}(Y_1,\dots,Y_8)\\
g_2 & = & Z_{18}Z_{19}+Z_{20}Z_{21}+Z_{22}Z_{23}+Z_{24}Z_{25}+Z_{26}Z_{27} & =\psi_{2,1}(X_1,\dots,X_8)\psi_{2,2}(Y_9,\dots,Y_{16})\\
g_3 & = & Z_{28}Z_{29}+Z_{30}Z_{31}+Z_{32}Z_{33}+Z_{34}Z_{35}+Z_{26}Z_{36} & =\psi_{2,1}(Y_1,\dots,Y_8)\psi_{2,2}(Y_9,\dots,Y_{16}).
\end{array}$$

Note that $Z_{26}$ appears in both $g_2$ and $g_3$ because of (8). Then, since

$$\psi_{2,1}(X_1,\dots,X_8)=\psi_{1,1}(X_1,\dots,X_8)\quad\text{and}\quad\psi_{2,1}(Y_1,\dots,Y_8)=\psi_{1,2}(Y_1,\dots,Y_8),$$

we have

$$\begin{array}{lcll}
h_1 & = & (g_1g_2g_3^{-1})^{1/2} & =\psi_{1,1}(X_1,\dots,X_8)\\
h_2 & = & g_1h_1^{-1} & =\psi_{1,2}(Y_1,\dots,Y_8)\\
h_3 & = & g_2h_1^{-1} & =\psi_{2,2}(Y_9,\dots,Y_{16}).
\end{array}$$

We can then substitute the transformed ciphertext values into $h_1,h_2,h_3$ and subsequently restore the triangular structure of $Z_1,Z_2,Z_3$, respectively. The next step is to unmask the final four equations in the triangular portion. To do this, we define

$$\begin{array}{lcll} g_4 & = & Z_{37}Z_{38} + Z_{39}Z_{40} + Z_{41}Z_{42} + Z_{43}Z_{44} + Z_{45}Z_{46} & = \psi_{3,1}(X_1,\dots,X_8)\psi_{3,2}(Y_{17},\dots,Y_{24}) \\ g_5 & = & Z_{47}Z_{48} + Z_{49}Z_{50} + Z_{51}Z_{52} + Z_{53}Z_{54} + Z_{26}Z_{55} & = \psi_{2,1}(X_9,\dots,X_{16})\psi_{2,2}(Y_9,\dots,Y_{16}) \\ g_6 & = & Z_{56}Z_{57} + Z_{58}Z_{59} + Z_{60}Z_{61} + Z_{62}Z_{63} + Z_{45}Z_{64} & = \psi_{3,1}(X_{17},\dots,X_{24})\psi_{3,2}(Y_{17},\dots,Y_{24}) \\ g_7 & = & Z_{65}Z_{66} + Z_{67}Z_{68} + Z_{69}Z_{70} + Z_{71}Z_{72} + Z_{73}Z_{74} & = \psi_{4,1}(X_9,\dots,X_{16})\psi_{4,2}(Y_{25},\dots,Y_{32}). \end{array}$$

Then, since $\psi_{3,1}(X_1,\dots,X_8) = \psi_{1,1}(X_1,\dots,X_8)$ and $\psi_{4,1}(X_9,\dots,X_{16}) = \psi_{2,1}(X_9,\dots,X_{16})$, we have

$$\begin{array}{lcll} h_4 & = & g_4h_1^{-1} & = \psi_{3,2}(Y_{17},\dots,Y_{24}) \\ h_5 & = & g_5h_3^{-1} & = \psi_{2,1}(X_9,\dots,X_{16}) \\ h_6 & = & g_6h_4^{-1} & = \psi_{3,1}(X_{17},\dots,X_{24}) \\ h_7 & = & g_7h_5^{-1} & = \psi_{4,2}(Y_{25},\dots,Y_{32}). \end{array}$$

Using $h_4,\dots,h_7$, we can restore the triangular structure of $Y_4,\dots,Y_7$, and easily recover $X_1,\dots,X_7$. To recover $X_8$, we use the value of $h_1 = \psi_{1,1}(X_1,\dots,X_8)$, as long as $X_7$ is nonzero.

We finish the inversion process by solving for the remaining variables $X_9,\dots,X_{24}$ and $Y_1,\dots,Y_{32}$, using 6 of the 7 oil-vinegar systems:

| subsystem | oil-vinegar polynomials | oil variables | vinegar variables |
|---|---|---|---|
| 1 | $Z_8,\dots,Z_{15}$ | $Y_1,\dots,Y_8$ | $X_1,\dots,X_8$ |
| 2 | $Z_{18},\dots,Z_{25}$ | $Y_9,\dots,Y_{16}$ | $X_1,\dots,X_8$ |
| 4 | $Z_{37},\dots,Z_{44}$ | $Y_{17},\dots,Y_{24}$ | $X_1,\dots,X_8$ |
| 5 | $Z_{47},\dots,Z_{54}$ | $X_9,\dots,X_{16}$ | $Y_9,\dots,Y_{16}$ |
| 6 | $Z_{56},\dots,Z_{63}$ | $X_{17},\dots,X_{24}$ | $Y_{17},\dots,Y_{24}$ |
| 7 | $Z_{65},\dots,Z_{72}$ | $Y_{25},\dots,Y_{32}$ | $X_9,\dots,X_{16}$ |

(9)

4.2. **Decryption failures.** In this section, we let $N = |\mathbb{F}|$. From Section 2, we know that decryption may fail 1) if we are unable to perform a necessary inversion in $\mathbb{F}$ while computing the $h_i$'s, or 2) if we are unable to solve an oil-vinegar subsystem. To compute $h_1, h_2, h_3$, notice that we must have $\psi_{1,1}(X_1,\dots,X_8)$, $\psi_{1,2}(Y_1,\dots,Y_8)$, and $\psi_{2,2}(Y_9,\dots,Y_{16})$ all nonzero. It can be shown that when $N$ is large, each of these is zero with probability approximately $\frac{1}{N}$. Notice that to compute $h_4,\dots,h_7$, the only additional requirements are $\psi_{3,2}(Y_{17},\dots,Y_{24}) \neq 0$ and $\psi_{2,1}(X_9,\dots,X_{16}) \neq 0$. Again, each of these are zero with probabilty approximately $\frac{1}{N}$, so the total probability that we will not be able to unmask the triangular system is approximately $\frac{5}{N}$.

Recall that we can use $h_1$ to recover $X_8$ as long as $X_7$ is nonzero. However, if $X_7 = 0$, we can instead use $Z_{17}$ as long as $X_2 \neq 0$. Hence we fail to recover $X_8$ with probability $\frac{1}{N^2}$.

We have already addressed in Section 3 the solvability of the oil-vinegar subsystems. Notice that in (9), we have 6 oil-vinegar systems, but only 4 sets of vinegar variables. Hence the total probability of failing to invert the oil-vinegar systems is approximately $\frac{4}{N}$.

So, we conclude that decryption failure occurs with total probability approximately $\frac{9}{N}$. Practical implementations may avoid this problem by choosing $N$ large enough to ensure that decryption failure is negligible, or by using the embedding ($\nearrow$) modifier.

## 5. Security

Since the theory of provable security for MPKCs has not yet been sufficiently developed, and, not able to make a contribution in that area ourselves, we instead show that our system is safe from known attacks on MPKCs.

Throughout this section, $q$ is the size of the base field, $k$, and $d$ is the degree of $\mathbb{F}$ over $k$.

### 5.1. Attacks based on linear algebra.

We now examine the specific linear algebra-based attacks, examining the minrank and dual rank, separation of oil and vinegar variables, and finally, linearization equations attacks. These attacks have been perhaps the most devastating to attempts at building MPKCs.

**Minrank attack.** We first note that the quadratic part of each polynomial (in the central map, or in the public key) can be viewed as a quadratic form, with which we associate a symmetric matrix. In particular, when the field has characteristic two, given a quadratic form $f = \sum_{i \leq j} f_{ij} X_i X_j$ with $f_{ij} \in \mathbb{F}$, we form the matrix $\bar{A}$ where $\bar{A}_{ij} = f_{ij}$. Then we associate with $f$ the symmetric matrix $A = \bar{A} + \bar{A}^T$, and define the rank of $f$ to be the rank of $A$.

If a variable $X_i$ does not appear in the quadratic part of a polynomial, then the associated matrix will not have full rank, since the $i$-th row and column are zero. So, loosely speaking, if an equation has too few variables, the associated matrix will have small rank. This is the foundation for the minrank attack. Since the public polynomials are just combinations of the central map polynomials (via $L_1$) after a change of variables (via $L_2$), if the matrix for a central map polynomial has low rank, $r$, then some combination of the public key polynomials also must have rank $r$. After such a combination is found, the system may be broken.

From [GC00], the complexity of the attack is $q^{\lceil \frac{m}{n} \rceil r}$, where $m$ is the number of central map polynomials, $n$ is the number of variables, and $r$ is the smallest rank. Considering the ranks of the central map polynomials, viewed component-wise over $k$, the smallest rank is $8d$, hence the complexity of attack is $q^{\lceil \frac{74}{56} \rceil 8d} = q^{16d}$.

**Dual rank attack.** While minrank succeeds when an equation has too few variables, the dual rank attack is effective when a variable appears in too few equations. In this case, the matrix corresponding to the quadratic part of a polynomial in which the variable does not appear will have less than full rank. In particular, if a variable only appears in the quadratic part of $u$ central map equations, then some combination of $(u+1)$ of the public polynomials must have less than full rank. Finding such a combination will again enable us to break the system. The complexity of this attack is at least $n^3 q^u$ [YC04].

In our case, viewing the central map polynomials component-wise over $k$, each of the $56d$ variables appears in at least $6d$ equations, so the complexity of the attack is $(56d)^3 q^{6d}$.

**Separation of oil and vinegar variables attack.** As mentioned in Section 1, the goal of this attack is to find the space of the transformed oil variables (i.e. after $L_2$ has been applied). Kipnis et al. [KPG99b] give a complexity of $o^4 q^{v-o-1}$ where $o$ and $v$ are the number of oil and vinegar variables, respectively. Determining the transformed oil space may possibly lead to breaking the system, so we show that the complexity for our system is sufficiently high.

Recall that in an oil-vinegar system, no terms in the system may be quadratic in the oil variables. This means that given a system of polynomials, adding terms may result in a decrease of the size of the oil set, but never an increase, hence the vinegar set cannot possibly shrink by adding terms. So, disregarding the $\phi_i$ of the triangular system and viewing our central map $F$ as a system of oil-vinegar polynomials with coefficients in $\mathbb{F}$, we can determine the size of the minimal vinegar set by computing the maximum size of an oil set. This can be done by finding the clique number of the graph with vertices given by the 56 variables and edges occuring whenever the product of two variables does *not* appear in any polynomials of $F$. Using Magma, we found that the clique number is 20, so the smallest vinegar set has 36 variables. This gives a complexity of $20^4 q^{15d}$.

**Linearization equations attack.** We have verified using Magma that there are no first order linearization equations. Regarding second order linearization equations, we point out an important contrast between our system and MFE. In MFE, each $\psi_i$ has rank 4, and can therefore be expressed as the determinant of a $2 \times 2$ matrix. The $f_i$ are defined as elements of the product of two of these matrices, and the identity (3) holds by the multiplicative property of the determinant. Since the $\psi_i$ in our system have rank 8, no simple matrix decomposition exists, as each $\psi_i$ has an expression as the sum of two $2 \times 2$ determinants. Further, the $f_i$ are not defined as elements of a matrix product, so the multiplicative property of the determinant is of no use.

However, the above argument obviously cannot completely rule out the possibility of second order linearization equations. A search for second order equations would involve solving a linear system in approximately $\binom{56d+1}{1}\binom{74d+2}{2}$ coefficients. So, for $d = 1$, naive Gaussian elimination would require $> 2^{51}$ field operations, and for $d = 2$, it would require $> 2^{60}$ field operations.

5.2. **Algebraic attacks.** At the heart of these attacks are the $F_4$ and $F_5$ algorithms of Faugére [Fau99] and [Fau02], as well as the XL algorithm of Courtois et al. [CKPS00]. There have been some recent contributions to complexity estimates for these algorithms, assuming general systems.

Barget et al. [BFSY05] give the total number of operations in $k$ for $F_5$ (and hence XL) as

$$O\left(\binom{n + d_{reg}}{n}^{\omega}\right),$$

where $\omega$ is the exponent in Gaussian reduction and $d_{reg}$ is the degree of regularity of the ideal formed by the polynomials in the system, given by the degree of the first term with negative coefficient in the expansion of

$$\frac{\prod_{i=1}^{m}(1 - z^{d_i})}{(1 - z)^n}, \tag{10}$$

where $d_i$ is the total degree of the $i$-th polynomial. But since each of our polynomials have total degree two, (10) simplifies to $(1 - z)^{m-n}(1 + z)^m$. For us, if we take the degree of $\mathbb{F}$ over $k$ to be 1 (so $m = 74$ and $n = 56$), we have $d_{reg} = 15$. Hence the attack requires $2^{\omega \log \binom{71}{56}} > 2^{49\omega}$ operations in $k$. If we instead take the degree of $\mathbb{F}$ over $k$ to be 2, $d_{reg}$ becomes 26, and the attack requires $> 2^{92\omega}$ operations.

One other attack of note in this category is the attack of Joux et al. [JJMR05] against an earlier tractable rational map cryptosystem of Wang and Chen [WC04]. The authors exploit the fact that within the central map, there is a smaller subsystem of 11 equations in 7 variables. However, because of the design of our system, there appears to be no such overdetermined subsystem.

5.3. **Parameters.** Based on the preceding discussion, Table 1 presents security levels for different choices of $q = |k|$ and $d = [\mathbb{F} : k]$. To compute the complexity of $F_5$, we have used $\omega = 2.3$.

TABLE 1. Security

| Claimed Security | Input [bits] | Output [bits] | Parameters $q$ | $d$ | Complexity $F_5$ | Rank/UOV | Key Size [kBytes] Public | Private |
|---|---|---|---|---|---|---|---|---|
| $2^{113}$ | 896 | 1184 | $2^{16}$ | 1 | $2^{114}$ | $2^{113}$ | 245 | 18 |
| $2^{212}$ | 1792 | 2368 | $2^{16}$ | 2 | $2^{213}$ | $2^{212}$ | 1907 | 70 |
| $2^{114}$ | 1792 | 2368 | $2^{32}$ | 1 | $2^{114}$ | $2^{209}$ | 490 | 36 |

5.4. **Efficiency.** Consider our proposed system with $q = 2^{16}$ and $d = 1$. We compare it to MFE-1, which was shown to have a significant advantage over RSA-1024 in decryption speed [WYHL06]. Since MFE-1 uses a degree 4 extension of $\mathbb{F}_{2^{16}}$, multiplications in the extension field require $4^2$ operations over the base field. A rough count of multiplications over $\mathbb{F}_{2^{16}}$ yields about 2400 for MFE-1 and 3200 for our system. We implemented both systems in a straightforward way using Magma on a 1600MHz UltraSPARC IIIi. The results are recorded in Table 2.

TABLE 2. Implementation results

| System | Input [bits] | Output [bits] | Encryption Time | Decryption Time Central Map | Total |
|---|---|---|---|---|---|
| MFE-1 | 768 | 960 | 52ms | 2ms | 2.7ms |
| Our System | 896 | 1184 | 94ms | 1.4ms | 2.3ms |

As expected, encryption in our system is slower since it uses 74 equations in 56 variables over $\mathbb{F}_{2^{16}}$, whereas MFE-1 uses 60 equations in 48 variables. However, even though the multiplication count for our system is larger, decryption is actually faster. This is because the division and square root operations are slower in the large extension field of MFE-1; furthermore, decrypting MFE-1 requires converting back and forth between the base field and the extension field.

## 6. CONCLUSION

6.1. **Open questions.** We pose the following open questions:

- How can we find all quadratic solutions to the general Diophantine equations (3) and (4)?
- One solution to (4) gives several possible cryptosystems. How can we effectively choose the best one?
- What strategies can be formulated to help minimize the decryption failure rate?
- What other polynomial identities may be used to construct cryptosystems in the proposed framework?
- Rather than having $g_i$ factor into two distinct quadratics, can we find $g_i = \sum \alpha_{jk} Y_j Y_k = \psi_i$, or $g_i = \psi_i^2$? This would make $h_i$ much simpler: $h_i = g_i$ and $h_i = g_i^{1/2}$, respectively, and the first type of decryption failure would become irrelevant.
- A further generalization of the framework would be to omit the $g_i$'s and simply require the existence of rational functions $h_i \in \mathbb{F}(Y_{n+1}, \ldots, Y_{n+\ell t})$ such that each $h_i(f_{n+1}, \ldots, f_{n+\ell t})$ is quadratic in $X_1, \ldots, X_{n+\ell o}$, and could be used as a lock polynomial. Can other systems be successfully created using this generalization?

6.2. **Concluding remarks.** We have presented a new framework for multivariate public key cryptosystems that combines the ideas of triangular and oil-vinegar systems. Also, we have proposed a cryptosystem, based on a Diophantine equation, which implements the framework, and have shown the system to be secure against known MPKC attacks. Our framework has much freedom, and should provide a fertile ground for new research in the area of multivariate public key cryptography.

6.3. **Acknowledgment.** The authors would like to thank Jintai Ding, Zhuojun Liu, Zuowen Tan and Mingsheng Wang for their help and encouragement throughout this work, and Jerome Hoffman for a particularly inspiring lecture given at the Chinese Academy of Sciences in Beijing.

## References

[BFSY05] Magali Barget, Jean-Charles Faugére, Bruno Salvy, and Bo-Yin Yang, "Asymptotic Expansion of the Degree of Regularity for Semi-Regular Systems of Equations", Proceedings of MEGA'05: 8th International Symposium on Effective Methods in Algebraic Geometry (2005)

[Cou01] Nicolas Courtois, "The Security of Hidden Field Equations (HFE)," Topics in Cryptology - CT-RSA 2001: The Cryptographers' Track at RSA Conference 2001. LNCS 2020, 266-281. Springer (2001)

[CKPS00] Nicolas Courtois, Alexander Kilmov, Jacques Patarin, and Adi Shamir, "Efficient Algorithms for Solving Overdefined Systems of Multivariate Polynomial Equations," Advances in Cryptology - EUROCRYPT 2000: International Conference on the Theory and Application of Cryptographic Techniques. LNCS 1807, 392-407 (2000)

[CSV93] Don Coppersmith, Jacques Stern and Serge Vaudenay, "Attacks on the Birational Permutation Signature Schemes," Advances in Cryptology - CRYPTO'93: 13th Annual International Cryptology Conference. LNCS 773, 435-443. Springer (1994)

[CSV97] Don Coppersmith, Jacques Stern and Serge Vaudenay, "The Security of the Birational Permutation Signature Schemes," *Journal of Cryptology.* 10 no. 3, 207-221 (1997)

[DGS06] Jintai Ding, Jason Gower, and Deiter Schmidt, *Multivariate Public Key Cryptosystems.* Springer (2006)

[DHNW07] Jintai Ding, Lei Hu, Xuyun Nie, Jianyu Li and John Wagner, "High Order Linearization Equation (HOLE) Attack on Multivariate Public Key Cryptosystems," Public Key Cryptography - PKC 2007: 10th International Conference on Practice and Theory in Public-Key Cryptography. LNCS 4450, 233-248. Springer (2007)

[DS03] Jintai Ding and Dieter Schmidt, "The new TTM implementation is not secure," *Proceedings of International Workshop on Coding, Cryptography, and Combinatorics (CCC 2003).* 106-121 (2003)

[DS05] Jintai Ding and Dieter Schmidt, "Rainbow, a New Multivariate Polynomial Signature Scheme," Applied Cryptography and Network Security: Third International Conference (ANCS 2005). LNCS 3531, 164-175. Springer (2005)

[DWY07] Jintai Ding, Christopher Wolf, and Bo-Yin Yang, "$\ell$-Invertible Cycles for Multivariate Quadratic Public Key Cryptography," Public Key Cryptography - PKC 2007: 10th International Conference on Practice and Theory in Public-Key Cryptography, LNCS 4450, 266-281, Springer (2007)

[Fau99] Jean-Charles Faugére, "A new efficient algorithm for computing Gröbner bases ($F_4$)," *Journal of Pure and Applied Algebra.* 139, 61-68 (1999)

[Fau02] Jean-Charles Faugére, "A new efficient algorithm for computing Gröbner bases without reduction to zero ($F_5$)," ISSAC '02: Proceedings of the 2002 International Symposium on Symbolic and Algebraic Computation. 75-83. ACM (2002)

[GC00] Louis Goubin and Nicolas Courtois, "Cryptanalysis of the TTM Cryptosystem," Advances in Cryptology - ASIACRYPT 2000: 6th International Conference on the Theory and Application of Cryptology and Information Security. LNCS 1976, 44-57. Springer (2000)

[JJMR05] Antoine Joux, Sébastien Kunz-Jacques, Frédéric Muller, and Pierre-Michel Ricordel, "Cryptanalysis of the Tractable Rational Map Cryptosystem," Public Key Cryptography - PKC 2005: 8th International Workshop on Theory and Practice in Public Key Cryptography. LNCS 3386, 258-274. Springer (2005)

[KS98] Aviad Kipnis and Adi Shamir, "Cryptanalysis of the oil and vinegar signature scheme," Advances in Cryptology - CRYPTO '98: 18th Annual International Cryptology Conference. LNCS 1462, 257-266. Springer (1998)

[KPG99a] Aviad Kipnis, Jacques Patarin, and Louis Goubin, "Unbalanced oil and vinegar signature schemes," Advances in Cryptology - EUROCRYPT '99: International Conference on the Theory and Application of Cryptographic Techniques. LNCS 1592, 206-222. Springer (1999)

[KPG99b] Aviad Kipnis, Jacques Patarin, and Louis Goubin, "Unbalanced oil and vinegar signature schemes," Extended version from citeseer/kipnis99unbalanced.html.

[MAGMA] Wieb Bosma, John Cannon, and Catherine Playoust. The Magma algebra system. I. The user language. *J. Symbolic Comput.* 24 (3-4), 235-265 (1997)

[Moh99] Tzuong Tsieng Moh, "A public key system with signature and master key functions," *Communications in Algebra.* 27 no. 5, pp.2207-2222 (1999)

[Moh07] Tzuong Tsieng Moh, "Two New Examples of TTM," Cryptology ePrint Archive, Report 2007/144. http://eprint.iacr.org (2007)

[MCY04] Tzuong Tsieng Moh, Jiun-Ming Chen, and Bo-Yin Yang, "Building Instances of TTM Immune to the Goubin-Courtois Attack and the Ding-Schmidt Attack," Cryptology ePrint Archive, Report 2004/168. http://eprint.iacr.org (2004)

[NJHD07] Xuyun Nie, Xin Jiang, Lei Hu, and Jintai Ding, "Cryptanalysis of Two New Instances of TTM Cryptosystem," Cryptology ePrint Archive, Report 2007/381. http://eprint.iacr.org (2007)

[Pat95] Jacques Patarin, "Cryptanalysis of the Matsumoto and Imai Public Key Scheme of Eurocrypt '88," Advances in Cryptology - CRYPTO '95: 15th Annual International Cryptology Conference. LNCS 963, 248-261. Springer (1995)

[Pat97] Jacques Patarin, "The oil and vinegar signature scheme," Presented at the Dagstuhl Workshop on Cryptography (1997)

[Sha94] Adi Shamir, "Efficient Signature Schemes Based on Birational Permutations," Advances in Cryptology - CRYPTO'93: 13th Annual International Cryptology Conference. LNCS 773, 1-12. Springer (1994)

[Sho97] Peter Shor, "Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer," *SIAM Journal on Computing.* 26 no. 5, 1484-1509 (1997)

[WC04] Lih-Chung Wang and Fei-Hwang Chang, "Revision of Tractable Rational Map Cryptosystem," Cryptology ePrint Archive, Report 2004/046. http://eprint.iacr.org (2004)

[WP05] Christopher Wolf and Bart Preneel, "Large Superfluous Keys in Multivariate Quadratic Asymmetric Systems," Public Key Cryptography - PKC 2005: 8th International Workshop on Theory and Practice in Public Key Cryptography. LNCS 3386, 275-287. Springer (2005)

[WYHL06] Lih-Chung Wang, Bo-Yin Yang, Yuh-Hua Hu and Feipei Lai, "A Medium-Field Multivariate Public-Key Encryption Scheme," Topics in Cryptology - CT-RSA 2006: The Cryptographers' Track at the RSA Conference 2006. LNCS 3860, 132-149. Springer (2006)

[YC04] Bo-Yin Yang and Jiun-Ming Chen, "TTS: Rank Attacks in Tame-Like Multivariate PKCs," Cryptology ePrint Archive, Report 2004/061. http://eprint.iacr.org (2004)

[YC05] Bo-Yin Yang and Jiun-Ming Chen, "Building Secure Tame-like Multivariate Public-Key Cryptosystems: The New TTS," Information Security and Privacy: 10th Australasian Conference (ACISP 2005). LNCS 3574, 518-531. Springer (2005)

Department of Mathematical Sciences, Clemson University, Clemson, SC 29634-0975 USA *E-mail address*: {sgao, rheindl}@clemson.edu