Hash-based Multivariate Public Key Cryptosystems

WANG Hou-Zhen and ZHANG Huan-Guo*

The Key Laboratory of Aerospace Information Security and Trusted Computing
Ministry of Education, School of Computer, Wuhan University, Wuhan 430072, China
wanghouzhen@126.com
liss@whu.edu.cn
http://liss.whu.edu.cn

Abstract. Many efficient attacks have appeared in recent years, which have led to serious blow for the traditional multivariate public key cryptosystems. For example, the signature scheme SFLASH was broken by Dubois *et al.* at CRYPTO'07, and the Square signature (or encryption) scheme by Billet *et al.* at ASIACRYPTO'09. Most multivariate schemes known so far are insecure, except maybe the sigature schemes UOV and HFEv-. Following these new developments, it seems that the general design principle of multivariate schemes has been seriously questioned, and there is a rather pressing desire to find new trapdoor construction or mathematical tools and ideal. In this paper, we introduce the hash authentication techniques and combine with the traditional MQ-trapdoors to propose a novel hash-based multivariate public key cryptosystems. The resulting scheme, called EMC *(Extended Multivariate Cryptosystem)*, can also be seen as a novel hash-based cryptosystems like Merkle tree signature. And it offers the double security protection for signing or encrypting. By the our analysis, we can construct the secure and efficient not only signature scheme but also encryption scheme by using the EMC scheme combined some modification methods summarized by Wolf. And thus we present two new schems: EMC signature scheme (with the Minus method "-") and EMC encryption scheme (with the Plus method "+"). In addition, we also propose a reduced scheme of the EMC signature scheme (a light-weight signature scheme). Precise complexity estimates for these schemes are provided, but their security proofs in the random oracle model are still an open problem.

Keywords: multivariate public key cryptosystems, hash function, digital signatures, tame transformation

1 Introduction

Multivariate Polynomials Public Key Cryptography (hereafter MPKC for short) is an area of research which attempts to build asymmetric key schemes, based

* This work was supported by the National High-Tech Research and Development Plan of China (Grant No.2007AA01Z411), the National Natural Science Foundation of China (Grant No.60673071, 6097115, 60970116).

on solving randomly chosen systems of multivariate quadratic polynomials (MQ problem) over a finite field, the MQ problem which in general is an NP-hard problem. Multivariate schemes have recently received much attention, for several reasons. First, the hard problems of reference are not known to be polynomial in the quantum model, unlike integer factorization and the discrete logarithm problems [33]. More importantly, MPKC schemes are in general much more computationally efficient than number theoretic-based schemes such as RSA. And some of these schemes seem to be very suitable for constrained environments such as smart card, active RFID tags, wireless sensor networks and other embedded devices. Indeed, the SFLASH signature scheme [1], one of the best-known multivariate cryptosystems, had stood for a decade and was even accepted as a security standard for use in low-cost smart cards by the New European Schemes for Signatures, Integrity and Encryption. However, this scheme and some related schemes had recently been broken [28,18,13,12,3].

Questions. As we will see in Section 2, most multivariate schemes known so far were insecure because many efficient attacks have been found in succession. Despite some countermeasures proposed for resisting these new attacks, unfortunately, most of the countermeasures make the related schemes too slow. Hence there is a rather pressing need to find new MQ-trapdoors or new modifiers for constructing secure and efficient multivariate schemes. In addition, multivariate quadratic equations can be used to easily design cryptographic primitives for signing applications. By now, how to design the secure and efficient multivariate encryption schemes is an open question.

Our Results. In this paper, we construct a *hash-tame (HT) transformation* based on the hash function and the tame transformation appeared in [17]. And then we propose a hash-based multivariate public key cryptosystems based on a HT transformation and the traditional MQ-trapdoors. As the special construction, we call our new construction the EMC (*Extended Multivariate Cryptosystem*) scheme. and it offers the double security protection for signing or encrypting by combining the traditional multivariate public key checking and a HT transformation as hash authentication checking. By the our analysis, we can construct the secure and efficient not only signature scheme but also encryption scheme by using some modification methods appeared in [35], combined with our EMC scheme. And thus we present two new schems: EMC signature scheme (with the Minus method "-") and EMC encryption scheme (with the Plus method "+"). In addition, we also propose a reduced scheme of the EMC signature scheme (a light-weight signature scheme). This scheme combines the ideals of McEliece cryptosystems and the Minus method "-", and uses an affine bijective map instead of the nonlinear bijective map of EMC to hide the process of finding the inverse of the HT transformation. Precise complexity estimates for these schemes are provided, but their security proof in the random oracle model is still an open problem.

Organization of the Paper. In Section 2, we recall the traditional MQ-trapdoors and related attack methods. Next, in Section 3, we describe a novel hash-based multivariate cryptosystem (called EMC), and propose the EMC signature scheme and the EMC signature scheme respectively. We give a general cryptanalysis in Section 4. We propose two practical instances, and analyze their implementation efficiency in Section 5. We propose a reduced scheme of the EMC signature scheme in Section 6. We conclude the paper in Section 7.

2 Constructions for MQ-trapdoors

We denote by $\mathbb{F}_q^n$ the n-dimensional vector space over the finite field $\mathbb{F}_q$ with q elements. A function F from $\mathbb{F}_q^n$ to $\mathbb{F}_q^m$ is defined by m quadratic polynomials in n variables and coefficients are in $\mathbb{F}_q$, called the *central map* and its components *central polynomials.* Extant multivariate schemes almost always hide the cental map $F = (f_1, \cdots, f_m)$ via composition with two affine maps U, T, and obtain the public map $P : \mathbb{F}_q^n \to \mathbb{F}_q^m$,

$$P = (p_1, \cdots, p_m) = T \circ F \circ U \tag{1}$$

We usually write, for $1 \le j \le k \le n, 1 \le i \le m$,

$$p_i(x_1, \cdots, x_n) = \sum_{1 \le j \le k \le n} c_{ijk} x_j x_k + \sum_{j=1}^{n} b_{ij} x_j + a_i$$

where a_i is usually normalized to zero and coefficients $c_{ijk}, b_{ij} \in \mathbb{F}_q$.

In any given scheme, the central map F belongs to a certain class of quadratic maps whose inverse can be computed relatively easily. The maps U, T are affine (sometimes linear) and full-rank. The key of a MPKC is the design of the central map. The public key consists of the polynomials in P, In practice, this is always the collection of the coefficients of the p_i's. The secret key consists of the information in U, T and F, that is U^{-1}, T^{-1} and F^{-1} (sometimes F can be discarded).

To encrypt a block or verify a signature x, one simply computes $y = P(x)$, that is

$$y = P(x) \Longleftarrow T\Big(F\big(U(y)\big)\Big)$$

To decrypt or sign a block, one can compute

$$x = U^{-1}\Big(F^{-1}\big(T^{-1}(y)\big)\Big)$$

So far, there are the following four previously known basic trapdoors for the central map of MPKCs:

(I) *Matsumoto-Imai Scheme A (MIA).* The scheme MIA was proposed by Matsumoto and Imai [24], also called the C^* scheme. It is the first scheme which uses two different finite fields. Its central map F is defined from a monomial over the degree n extension field of $\mathbb{F}_q$, denoted $\mathbb{F}_{q^n}$, of the form $F(x) = x^{1+q^\theta}$, where θ

satisfies $gcd(q^\theta+1, q^n-1) = 1$. The C^* scheme was broken by Patarin in 1995 [28]. Subsequently, Patarin *et al.* proposed in 2001 [29,30] to remove from the public key the last r quadratic polynomials (out of the initial n) using Shamir's minus method [31], and called the resulting scheme C^{*-}. Furthermore, if the value of r is chosen such that $q^r \geq 2^{80}$, then the variant scheme is termed C^{*--}. SFLASH [1] belongs to the C^{*--} family and has been chosen as a candidate for the NESSIE selection, and finally accepted. SFLASH has been entirely broken: the r missing equations can be recovered in most cases as explained in [12] and the private key of the C^* family can be recovered following the cryptanalysis described in [19]. Recently, two new proposals were based on internal transformations that are not only quadratic on the base field, but also on the extension field: a signature scheme called square-vinegar was proposed in [2] and an encryption scheme called square appeared in [4]. However, they were broken by Billet *et al* at ASIACRYPT 2009 [3]. In addition, PMI+ [8] and ℓ-invertible cycle [11] can be best considered improved versions or extensions of the C^* scheme.

(II) *Hidden Field Equations (HFE).* After breaking MIA in 1995, Patarin generalized the underlying trapdoor to "Hidden Field Equations" [27]. This generalization aims at the central equations and uses a univariate polynomial rather than a univariate monomial here, that is, the central map of C^* was changed into

$$F(X) = \sum_{\substack{0\leq i,j\leq d,\\ q^i+q^j\leq d}} c_{ij}X^{q^i+q^j} + \sum_{\substack{0\leq k\leq d,\\ q^k\leq d}} b_k X^{q^k} + a$$

From a cryptanalytic point of view, the basic HFE scheme is broken: an efficient key recovery attack, using the MinRank-problem, has been demonstrated in [22]. In 2002, Faugère reported have broken the HFE-Challenge I. Since the fact that the central map of HFE is not injective, his attacks have improved and in 2003, Faugère and Joux published their results on the security of HFE [16]. But other variants such as HFEv- are still secure so far.

(III) *The Oil-Vinegar (OV) scheme.* The Oil and Vinegar and later derived unbalance Oil and Vinegar (UOV) schemes [21] are suitable for signatures. The original OV scheme was broken by Shamir and Kipnis [32]. This construction is inspired by the Patarin's of linearization attacks [28]. This scheme uses two sets of unknowns $(x_1, \cdots, x_o)$ and $(\hat{x}_1, \cdots, \hat{x}_v)$ respectively called the oil and the vinegar variables. The central map then consists of an o-tuple of polynomials $F = (f_1, \cdots, f_o)$ of the special form:

$$f_i(x,\hat{x}) = \sum_{j=1}^{o}\sum_{k=1}^{v} e_{ijk}x_j\hat{x}_k + \sum_{j=1}^{v}\sum_{k=1}^{v} d_{ijk}\hat{x}_j\hat{x}_k + \sum_{j=1}^{o} c_{ij}x_j + \sum_{j=1}^{v} b_{ij}\hat{x}_j + a_i$$

where all corresponding coefficients are randomly chosen from the base field $\mathbb{F}_q$.

The message size over signature size for the UOV signature scheme is not optimal since the number of vinegar unknowns must be at least twice big as the number of oil unknowns for it to be secure [32,21,3].

(IV) *The Stepwise Triangular System (STS).* This scheme first appeared for lectures in Japanese [34] and in English [31]. Generalized later to its present form

[36]. A *tame map* $G : \mathbb{F}_q^n \to \mathbb{F}_q^n$ used in [17] are a special case of the triangular maps from algebraic geometry, which are more generally defined by:

$$G(x_1, \cdots, x_n) = \begin{pmatrix} x_1 \\ x_2 + f_1(x_1) \\ \quad\vdots \\ x_{n-1} + f_{n-2}(x_1, \cdots, x_{n-2}) \\ x_n + f_{n-1}(x_1, \cdots, x_{n-1}) \end{pmatrix}^T \tag{2}$$

where the f_i are arbitrary polynomial functions. Obviously, G can be easily inverted assuming that the f_i are known.

The Tame Transformation Method (TTM) cryptosystem was first proposed by T. T. Moh [25]. But most of these schemes are shown to be not secure such as [20,9]. The original TTM schemes were intended for the purpose of public key encryption. Attempts were made to apply a similar but simpler idea for signatures. It was called the TTS (Tamed Transformation Signature) scheme, As with the TTM schemes, most of the original TTS versions [37,38]are broken by Ding *et al.* in [20,10].

By now, all basic trapdoors (MIA, HFE, OV and STS) are insecure, and must be modified using some effective measures for enhancing their security. A summary of modifications is given by Wolf and Preneel [35], including minus method "-", plus method "+", subfield method "/", branching "⊥", fixing "f", sparse polynomials "s", vinegar variables "v", internal perturbation "i", homogenising "h" and masking "m". Therefore, to construct secure MPKCs, we can make use of these modifications such as QUARTZ [26] (HFE, with vinegar variables and minus method).

3 Hash-based Multivariate Public Key Cryptosystems

3.1 Hash-based Multivariate Cryptosystems

We first construct a hash-based tame transformation and also called *a HT transformation.* Let $\mathbb{F}_q$ be a finite field with $q = 2^k$ elements and $\mathbb{F}_q^n$ the n-dimensional vector space over $\mathbb{F}_q$. A HT transformation is defined by $L : \mathbb{F}_q^n \to \mathbb{F}_q^n$,

$$\begin{cases} \begin{pmatrix} y_1 \\ \vdots \\ y_{n-\delta} \end{pmatrix} = A \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_{n-\delta} \end{pmatrix} + \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_{n-\delta} \end{pmatrix} \\ y_{n-\delta+1} = \gamma_1 x_{n-\delta+1} + \sum\limits_{j=1}^{n-\delta} a_{ij} x_j + \gamma_{\delta+1} \boldsymbol{x}_{n+1} + \beta_{n-\delta+1} \\ \quad\vdots \\ y_n = \gamma_\delta x_n + \sum\limits_{j=1}^{n-1} a_{nj} x_j + \gamma_{2\delta} \boldsymbol{x}_{n+\delta} + \beta_n \end{cases} \tag{3}$$

where δ is the number of the *extended variables* $x_{n+i} \in \mathbb{F}_q$ $(1 \leq i \leq \delta)$, $\gamma_i \neq 0$ $(1 \leq i \leq 2\delta)$, $(n-\delta) \times (n-\delta)$ matrix must be full-rank, other corresponding coefficients are randomly chosen in $\mathbb{F}_q$; in particular, the *extended variables* $x_{n+i}(1 \leq i \leq \delta)$ are defined by

$$x_{n+i} = H_k(x_1||x_2||\cdots||x_{n-\delta+i-1}) \tag{4}$$

Here, the notations $||$, $H(\cdot)$ denote the "concatenation" operation and the standard hash function such as SHA-1, respectively; and $H_k(\cdot)$ represents to extract the first k bits of $H(\cdot)$ and maps the bitstring into an element in $\mathbb{F}_q$.

Obviously, L is also bijective (but not linear) as same as the map G like Eq. (2). We can compute the preimage $x = L^{-1}(y)$ as easily as $y = L(x)$, but it is difficult to write x explicitly as a function of y variable because of the nonlinear property of hash function. In addition, if we look upon the δ extended variables $x_{n+1}, \cdots, x_{n+\delta}$ as the new variables identified with the first n variables $x_1, \cdots, x_n$, then L can be seen as a compression map from $\mathbb{F}_q^{n+\delta}$ to $\mathbb{F}_q^n$, that is

$$(y_1, \cdots, y_n) = L(x_1, \cdots, x_n, x_{n+1}, \cdots, x_{n+\delta})$$

The construction of our new schemes is the same as the traditional MQ-trapdoors. The only difference is that the transformation U in Eq. (2) is replaced by U' $(= U \circ L)$ the combination of an affine bijective map and a HT transformation, obviously, U' is also a bijective map. More precisely, the generic construction of our schemes has the canonical decomposition of most MPKC's as follows. The public map P is the form of multivariate quadratic polynomials from $\mathbb{F}_q^{n+\delta}$ to $\mathbb{F}_q^n$, and defined by

$$P = (p_1, \cdots, p_n) = T \circ F \circ U' \xlongequal{def\ U'=U\circ L} T \circ F \circ (U \circ L) \tag{5}$$

The public key consists of the polynomials in P, In practice, this is always the collection of the coefficients of the p_i's. The secret key consists of L^{-1}, U^{-1}, T^{-1} and F^{-1}. In addition, from the theoretical point of view, we can be free to choose one of the four basic MQ-trapdoors appeared in Section 2.

As the special construction of the new scheme, we call our proposed algorithm the **EMC** (Extended Multivariate Cryptosystems) scheme.

3.2 The EMC Signature Scheme

The EMC signature scheme combines the HT transformation defined in this paper and the Shamir's Minus method [31] as follows.

The Secret Parameters. We randomly choose two affine bijective maps U, T from $\mathbb{F}_q^n$ to $\mathbb{F}_q^n$ and a HT transformation L from $\mathbb{F}_q^{n+\delta}$ to $\mathbb{F}_q^n$, and find their inverse map U^{-1}, T^{-1} and L^{-1} respectively. In addition, we need to choose an appropriate central map $F : \mathbb{F}_q^n \to \mathbb{F}_q^n$ from the four basic MQ-trapdoors (see Section 2). Consequently, the private key consists of the information U^{-1}, T^{-1}, L^{-1} and F^{-1}, where the size of F^{-1} depends on the MQ-trapdoor used.

The Public Parameters. By Eq. (5), the public map P can be obtained by the composition of the above four maps. Once we apply the Shamir's minus method to P, example by deleting the last r components $(0 \leq r < n)$, we will have a new map of the EMC signature scheme $P^{-} : \mathbb{F}_q^{n+\delta} \to \mathbb{F}_q^{n-r}$ defined by

$$P^{-}(x_1, \cdots, x_{n+\delta}) = (p_1, \cdots, p_{n-r}) \tag{6}$$

Apart from this, the public key includes the field structure of $\mathbb{F}_q$ and the standard hash function used.

The Signing Process. The message (or its hash value) is $Y^{-} = (y_1, \cdots, y_{n-r})$ in $\mathbb{F}_q^{n-r}$. The signer first chooses r random elements $y_{n-r+1}, \cdots, y_n \in \mathbb{F}_q$ which are appended to Y^{-} to obtain $Y = (y_1, \cdots, y_n)$ in $\mathbb{F}_q^n$. Then the signature X is obtained by

$$X = (x_1, \cdots, x_{n+\delta}) = L^{-1}\Bigg(U^{-1}\Big(F^{-1}\big(T^{-1}(Y)\big)\Big)\Bigg)$$

The verifying Process. Anyone who receives the message Y^{-} and its signature $X = (x_1, \cdots, x_{n+\delta})$ first applies the public hash function to check if indeed

$$x_{n+i} = H_k(x_1||x_2|| \cdots ||x_{n-\delta+i-1}), 1 \leq i \leq \delta \tag{$V1$}$$

If equality holds, then continues to check

$$\big(p_1(X), \cdots, p_{n-r}(X)\big) = Y^{-} \tag{$V2$}$$

We can conclude that a signature X of Y^{-} is valid if and only if the two conditions $(V1)$ and $(V2)$ are simultaneously satisfied. As the use of hash function, the $(V1)$ is called the *hash authentication checking.* the checking method in $(V2)$ is the same as the traditional MPKC's. Of course, the steps $(V1)$ and $(V2)$ can be permuted, it is not required that they are executed in the order shown above.

We will discuss that in Section 4, both $(V1)$ and $(V2)$ offer the double-protection for the security of the EMC signature scheme.

3.3 The EMC Encryption Scheme

Contrary to the EMC signature scheme, the EMC encryption scheme combines the Plus method [27,35]. the Plus method amounts to adding r randomly chosen polynomial components as follows.

The Secret Parameters. We randomly choose three maps: two affine bijective maps $U : \mathbb{F}_q^n \to \mathbb{F}_q^n$ and $T : \mathbb{F}_q^{n+r} \to \mathbb{F}_q^{n+r}$, and a HT transformation $L : \mathbb{F}_q^{n+\delta} \to \mathbb{F}_q^n$. and find their inverse map U^{-1}, T^{-1} and L^{-1} respectively. In addition, we need to choose an appropriate central map $F : \mathbb{F}_q^n \to \mathbb{F}_q^n$ from the four basic MQ-trapdoors (see Section 2). Consequently, the private key consists of the information U^{-1}, T^{-1}, L^{-1} and F^{-1}, where the size of F^{-1} depends on the MQ-trapdoor used.

The Public Parameters. By Eq. (5), the public map P can be obtained by the composition of the above four private maps. Once we apply the Plus method to the internal map $F' = F \circ U \circ L = (f'_1, \cdots, f'_n)$, example by adding the r random chosen quadratic polynomials $(0 \leq r < n)$, we have

$$F'^{+} = (f'_1, \cdots, f'_n, f'_{n+1}, \cdots, f'_{n+r})$$

Therefore, the public key of the EMC encryption scheme $P^{+} : \mathbb{F}_q^{n+\delta} \rightarrow \mathbb{F}_q^{n+r}$ can be obtained by

$$P^{+}(x_1, \cdots, x_{n+\delta}) = T \circ F'^{+} = (p_1, \cdots, p_{n+r}) \tag{7}$$

In addition, the public key also includes the field structure of $\mathbb{F}_q$ and the standard hash function used.

The Encyption Process. Given a plaintext $(x_1, \cdots, x_n) \in \mathbb{F}_q^n$. one first calculates the extended variables

$$x_{n+i} = H_k(x_1||x_2|| \cdots ||x_{n-\delta+i-1}), 1 \leq i \leq \delta$$

and applies the public polynomials P^{+} to encrypt the plaintext. Then the corresponding ciphertext is easily obtained by

$$(y_1, \cdots, y_{n+r}) = P^{+}(x_1, \cdots, x_{n+\delta})$$

The Decyption Process. To decrypt the ciphertext $(y_1, \cdots, y_{n+r})$, we successively execute the follwing steps:

1. Compute $(a_1, \cdots, a_{n+r}) = T^{-1}(y_1, \cdots, y_{n+r})$, and then discard the r redundant values $a_{n+1}, \cdots, a_{n+r}$ to produce a n-dimensional vector $(a_1, \cdots, a_n)$.
2. Compute $(b_1, \cdots, b_n) = F^{-1}(a_1, \cdots, a_n)$.
3. Compute $(c_1, \cdots, c_n) = U^{-1}(b_1, \cdots, b_n)$.
4. Finally, the plaintext can be obtained by $(x_1, \cdots, x_n) = L^{-1}(c_1, \cdots, c_n)$.

Obviously, comparison with the decryption process of the traditional multivariate public key cryptosystems, SEMC only increases a linear multiplication with regard to L^{-1}.

4 Security Analysis

4.1 Linearization Equations Attack

Linearization equations attack used to cryptanalyse the C^* scheme was proposed in 1995 by Patarin [28]. This attack relies on using the public key P to generate a large set of equations in the plaintext indeterminates $x_1, x_2, \cdots, x_n$ and the ciphertext indeterminates $y_1, y_2, \cdots, y_n$. The equations to be generated in this attack all have the "bilinear" form

$$\sum_{i=1}^{n}\sum_{j=1}^{n} \gamma_{ij} x_i y_j + \sum_{i=1}^{n} \delta_i x_i + \sum_{i=1}^{n} \epsilon_i + \eta = 0 \tag{8}$$

It is shown in [28] that the linear equations in $\gamma_{ij}, \delta_i, \epsilon_i$ and η provided by a sufficient number of P input-output pairs allow to recover these unknown coefficients, and that once this has been done, the obtained vector space of solutions can be used to compute the inverse by P of any $\mathbb{F}_q^n$ element Y at the expense of sovling a small linear system. The complexity of the attack is approximately $\mathcal{O}(n^6 log^2 q)$.

By Section 3, we know that the HT transformation L is a nonlinear bijective function with regard to input variables $x_1, \cdots, x_n$ (but its degree is unknown by the properties of hash function). In other words, L further increases the nonlinear degree of MPKCs. We checked experimentally, for a large number of system parameters chosen, there only exists trivial solution for the linear equations with regard to the all coefficients of Eq. (8) constructed by a sufficient number of P input-output pairs.

4.2 Direct Attacks

The most natural ideal of attack on any public-key cryptosystems is to find a plaintext x for a given ciphertext y without using any information beyond the public key itself. In the case of MPKCs, the intention is to equivalent to solving an instance of the MQ problem over a finite field. However this problem is known to be NP-hard, even when restricted to quadratic eauations over $GF(2)$ and over any finite field [27]. Several major methods based on Gröbner bases have been developed to solve the MQ problem, such as XL [7] and improved variants of Buchbergers's Gröbner bases computation algorithm such as Faugère's F4 and F5 algorithms [14,15], but these algorithms are exponential in time and memory. It is worth mentioning that in 2002, Faugère reported to have broken the HFE-Challenge I, and subsequently improved his attacks in [16]. Generally speaking, these attacks can be easily avoided by choosing appropriate scale parameters.

For the EMC cryptosystems, we transform the original public key $P : \mathbb{F}_q^n \rightarrow \mathbb{F}_q^n$ into the new public key $\tilde{P} : \mathbb{F}_q^{n+\delta} \rightarrow \mathbb{F}_q^n$ using a HT map L, and obtained by $\tilde{P} = P \circ L$, where $P = T \circ F \circ U$ (see Section 3). According to [27], to find P and L from $\tilde{P}$ belongs to the IP problem and is NP-hard. Emerging a HT transformation into the traditional MPKCs is equivalent to increasing the number of input variables, which is difficult to be eliminated by the attacker because of the hash-value relationship like (4) among these input variables. Obviously, to find a preimage x from the new public key $\tilde{P}$ is more difficult than from the original public key P.

4.3 Structure-based Attacks and Security Estimates

The structure-based attack type relies solely on the specific structures of the corresponding multivariate public key schemes. The preliminary security analysis suggests that we can choose any one of the four basic MQ-trapdoors (see Section 2) to be used as the central map F of our proposed EMC schemes. As space is limited, we do not intend to discuss each scheme based on a basic MQ-trapdoor in detail. However, we have shown that the original public key P was disguised as our new public key $\tilde{P}$ by a HT map L, and to find P and L from $\tilde{P}$ is a computationally intractable problem. Hence the EMC schemes, including signature and encryption schemes, can effectively resist all known attacks for the original public key P.

In addition, it offers the double security protection for the EMC schemes using the "Chemical Synthesis" of the HT map and traditional MQ-trapdoors. For the EMC signature, if $\mathcal{O}(T_s)$ denotes the complexity of forging a signature $X' = (x'_1, \cdots, x'_{n+\delta})$ of the message Y such that $\tilde{P}(X') = Y$ (i.e., passing (V2) in the verifying process), then the success probability of X' satisfying Eq. (4) is approximately $1/q^{\delta}$. Thus the complexity of X' simultaneously satisfying (V1) and (V2) in the verifying process is approximately $\mathcal{O}(T_s \cdot q^{\delta})$, which is replaced by conservative security level of the EMC signature scheme is

$$C_{\text{sign}} = \mathcal{O}\big(q^{min(r,\delta)}\big)$$

where r, δ can be considered as the dimension of solution (or signature) space of $\tilde{P}$ and the number of hash-value authentication respectively.

As for the EMC encryption scheme, from the security analysis above, the new public key $\tilde{P}$ can resist all known cryptanalysis techniques for the original public key P. Thus the complexity of the best known attack is equal to of the exhaustive search, that is $\mathcal{O}(q^n)$. We also propose that an even more conservative security level of the EMC encryption scheme is

$$C_{\text{encrypt}} = \mathcal{O}\big(q^{\delta - r}\big)$$

which also is the solution (or preimage) space of the new public key $\tilde{P}$.

5 Practical-sized Instances of the EMC Schemes

In order to facilitate the discussion, we can denote the our proposed EMC scheme by EMC(q, n, δ, r), where the parameters q, n, δ and r have been defined in Section 3. For the parameter r, if $r < 0$, then the scheme can be only used as signing. Conversely, if $r > 0$, then it can be only used as encrypting. Of course, when $r = 0$, it can be use to both sign and encrypt a message (but generally not recommended).

Based on the security analysis above, we propose a signature scheme of at least 80-bit security level called EMC$(2^8, 37, 10, -10)$, where the central map F in Eq. (5) also uses the MIA construction, the hash function uses SHA-256. And we summarize operating details as follows.

- Message size: 27 bytes
- Signature size: 47 bytes
- Public key size: 31K bytes
- Private key size: 3.8K bytes

We also propose EMC$(2^8, 37, 20, 10)$ for an encryption scheme of at least 80-bit security level, and summarize operating details as follows.

- Plaintext size: 27 bytes
- Ciphertext size: 37 bytes
- Public key size: 42.5K bytes
- Private key size: 2.5K bytes

The concrete computing process of the above schemes is the same as the SFLASH^{v3} scheme [6], which are more efficient than SFLASH^{v3}. Comparison to the traditional MQ-schemes, the only difference is an increase of 20 times hash-value operations.

6 SEMC: a Light-weight Signature Scheme

From the security analysis above, we know that the hash authentication combined the traditional MQ-trapdoors to offer the double security protection for the EMC signature scheme. In order to better adapt to the very constrained devices such as passive RFID tags. We propose a simplified EMC signature scheme (SEMC for short) as follows.

The Secret Parameters. We randomly choose an affine bijective maps T from $\mathbb{F}_q^n$ to $\mathbb{F}_q^n$ and a HT transformation L from $\mathbb{F}_q^{n+\delta}$ to $\mathbb{F}_q^n$ respectively. Consequently, The private key consists of the corresponding inverse maps T^{-1}, L^{-1}.

The Public Parameters. The public map P can be obtained by the composition of the above two maps T and L, that is $P = T \circ L$. And we also apply the Shamir's Minus method to P, example by deleting the last r components $(0 \leq r < n)$, we will have a new map of the SEMC signature scheme $P^- : \mathbb{F}_q^{n+\delta} \to \mathbb{F}_q^{n-r}$ defined by

$$P^-(x_1, \cdots, x_{n+\delta}) = (p_1, \cdots, p_{n-r}) \tag{9}$$

Of course, the public key also includes the field structure of $\mathbb{F}_q$ and the standard hash function used.

The Signing Process. The message (or its hash value) is $Y^- = (y_1, \cdots, y_{n-r})$ in $\mathbb{F}_q^{n-r}$. The signer first chooses r random elements $y_{n-r+1}, \cdots, y_n \in \mathbb{F}_q$ which are appended to Y^- to obtain $Y = (y_1, \cdots, y_n)$ in $\mathbb{F}_q^n$. Then the signature X is obtained by

$$X = (x_1, \cdots, x_{n+\delta}) = L^{-1}\big(T^{-1}(Y)\big)$$

The verifying Process. Anyone who receives the message Y^- and its signature $X = (x_1, \cdots, x_{n+\delta})$ first applies the public hash function to check if indeed

$$x_{n+i} = H_k(x_1||x_2||\cdots||x_{n-\delta+i-1}), 1 \leq i \leq \delta \tag{$V'1$}$$

If equality holds, then continues to check

$$\big(p_1(X), \cdots, p_{n-r}(X)\big) = Y^- \tag{$V'2$}$$

A signature X of Y^- is valid if and only if the two conditions $(V'1)$ and $(V'2)$ are simultaneously satisfied. And the $(V'1)$ is also called the *hash authentication checking.*

The construction method of SEMC is inspired from McEliece-type cryptosystems [5]. We use a random chosen linear bijection T to hide a HT map L defined in this paper. In addition, according to Eq. (3), there exists a certain linear relationship between part of input and output variables of L. Hence we use the Minus method to avoid the linear attacks like Patarin's in [28]. The above analysis indicates that the complexity of the best-known attack for SEMC is $\mathcal{O}(q^{min(r,\delta)})$.

7 Conclusion

In this paper, we introduce the hash authentication techniques and combine with the traditional MQ-trapdoors to propose a novel hash-based multivariate public key cryptosystems. The new scheme EMC can also be seen as a novel hash-based cryptosystems like Merkle tree signature [23]. And it offers the double security protection for signing or encrypting. And we present two new schems: EMC signature scheme and EMC encryption scheme. In addition, we also propose a light-weight signature scheme, which is a reduced scheme of the EMC signature scheme. By the our security analysis, these schemes above can avoid the all known attacks, but their security proof in the random oracle model is still an open problem. We believe that our hash-based MQ-trapdoor construction can easily produce excellent multivariate schemes for practical applications.

References

1. Akkar, M., Courtois, N.: A fast and secure implementation of SFLASH. In: PKC 2003. in: LNCS, vol. 2567, pp. 267-278. Springer, Heidlberg(2003)
2. Baena, J., Clough, C., Ding, J.: Square-vinegar signature scheme. In: Buchmann, J., Ding, J. (eds.) PQCrypto 2008. LNCS, vol. 5299, pp. 17-30. Springer, Heidelberg (2008)
3. Billet, O. Macario-Rat, G.: Cryptanalysis of the Square Cryptosystems. ASIACRYPT 2009, LNCS, vol. 5912, pp. 451-468. Springer, Heidlberg(2009).
4. Clough, C., Baena, J., Ding, J., Yang, B. Y. and Chen, M. S.: Square, a New Multivariate Encryption Scheme. In: Fischlin, M. (ed.) CT-RSA 2009, LNCS, vol. 5473, pp. 252-264. Springer, Heidlberg(2009)
5. Courtois, N., Finiasz, M., and N.Sendrier: How to achieve a McEliece-based digital signature scheme. In ASIACRYPT 2001. LNCS, vol. 2248, pp. 157-174. Springer, Heidlberg(2001).
6. Courtois, N., Goubin, L. and Patarin, J.: SFLASHv3, a fast asymmetric signature scheme. Cryptology ePrint Archive: Report 2003/211, 2003, Revised Specication of SFLASH, version 3.0., October 17th, 2003. Available at `http://eprint.iacr.org/2003/211/`
7. Diem, C.: The XL-algorithm and a conjecture from commutative algebra. In: ASIACRYPT 2004. LNCS, vol. 3329, pp. 323-337. Springer, Heidlberg(2004).
8. Ding, J., Gower, J.: Inoculating multivariate schemes against differential attacks. In: Yung, M., Dodis, Y., Kiayias, A., Malkin, T.G. (eds.) PKC 2006. LNCS, vol. 3958, pp. 290C301. Springer, Heidelberg (2006)

9. Ding, J., Hu, L., Nie, X., Li, J., Wagner, J.: High order linearization Equation (HOLE) attack on multivariate public key cryptosystems. In: Okamoto, T., Wang, X. (eds.) PKC 2007. LNCS, vol. 4450, pp. 233C248. Springer, Heidelberg (2007)
10. Ding, J. Schmidt, D. and Yin, Z.: Cryptanalysis of the new TTS scheme in CHES 2004, Int. J. Inf. Sec. 2006, 5(4), pp. 231-240. Springer, Heidlberg(2006)
11. Ding, J., Wolf, C. and Yang, B. Y.: ℓ-invertible cycles for multivariate quadratic public key cryptography. In: PKC 2007. LNCS, vol. 4450, ppa. 266-281. Springer, Heidlberg(2007).
12. Dubois, V., Fouque, P.A., Shamir, A. and Stern, J.: Practical Cryptanalysis of SFLASH, In: Crypto 2007. LNCS, vol. 4622, pp.1-12. Springer, Heidlberg(2007).
13. Dubois, V., Fouque, P.A. and Stern, J.: Cryptanalysis of SFLASH with Slightly Modified Parameters. In: Eurocrypt2007, LNCS, vol. 4145, pp. 264-275. Springer, Heidlberg(2007)
14. Faugère, J.C.: A new efficient algorithm for computing Gröbner bases (F4), Journal of Pure and Applied Algebra 139(61), 88 (1999).
15. Faugère, J.C. : A new efficient algorithm for computing Gröbner bases without reductions to zero (F5), in: ISSAC 2002, 75-83. ACM Press, New York (2002).
16. Faugère, J. C. and Joux, A.: Algebraic cryptanalysis of Hidden Field Equations (HFE) using gröbner bases. In: Boneh, D. (ed.) CRYPTO 2003. LNCS, vol. 2729 , pp. 44-60. Springer, Heidlberg(2003).
17. Fell, H. and Diffie, W.: Analysis of public key approach based on polynomial substitution. In: Hugh,Williams (ed.) CRYPTO 1985. LNCS, vol. 218, pp. 340-349. Springer, Heidlberg(1985).
18. Fouque, P.A., Granboulan, L. and Stern, J.: Differential Cryptanalysis for Multivariate Schemes. In: Eurocrypt 2005. LNCS, vol. 3494, pp. 341-353.Springer, Heidlberg(2005)
19. Fouque, P.A., Macario-Rat, G., Stern, J.: Key Recovery on Hidden Monomial Multivariate Schemes. In: Smart, N.P. (ed.) EUROCRYPT 2008. LNCS, vol. 4965, pp. 19-30. Springer, Heidelberg (2008)
20. Goubin, L. and Courtois, N.: Cryptanalysis of the TTM cryptosystem. In: ASIACRYPT 2000. LNCS, vol. 1976, pp. 44-57. Springer, Heidlberg(2000).
21. Kipnis, A., Patarin, J. and Goubin, L.: Unbalanced Oil and Vinegar signature schemes. In: EUROCRYPT 1999. LNCS, vol. 1592, pp. 206-222. Springer, Heidlberg(1999)
22. Kipnis, A. and Shamir, A.: Cryptanalysis of the HFE public key cryptosystem. In: Wiener, M. (edi.) CRYPTO 1999. LNCS, vol. 1666, pp. 19-30. Springer, Heidlberg(1999)
23. Merkle, R.C. A certified digital signature. In: CRYPTO1989. LNCS, vol. 435, pp. 218-238. Springer, Heidlberg(1989)
24. Matsumoto, Tsutomu, Imai nd Hideki: Public quadratic polynomial-tuples for efficient signature verification and message encryption. In: EUROCRYPT1988, LNCS, vol. 330, pp. 419-453. Springer, Heidlberg(1988).
25. Moh, T.T.: A fast public key system with signature and master key functions. Comm. in Algebra, 27: 2207-2222. `http://www.usdsi.com/ttm.html`.
26. Patarin, J., Courtois, N. and Goubin, L.: QUARTZ, 128-bit long digital signatures. In: CT-RSA 2001, LNCS, vol. 2020, pp. 298-307. Springer, Heidlberg(2001).
27. Patarin, J.: Hidden Field Equations (HFE) and Isomorphism of Polynomials (IP): Two new families of asymmetric algorithms. In: Eurocrypt 1996. LNCS, vol 1070, pp. 33-48. Springer, Heidlberg(1996)
28. Patarin, J.: Cryptanalysis of the Matsumoto and Imai public key scheme of Eurocrypt1988. In: Crypto1995. LNCS, vol. 963, pp. 248-261. Springer, Heidlberg(1998)

29. Patarin, J., Courtois, N. and Goubin,L.: FLASH, a Fast Multivariate Signature Algorithm. In: CT-RSA 2001. LNCS, vol. 2020, pp. 297-307. Springer, Heidlberg(2001)
30. Patarin, J., Goubin, L. and Courtois, N.: C^{*}_{-+} and HM: Variations Around Two Schemes of T. Matsumoto and H. Imai. In: Asiacrypt1998. LNCS, vol. 1514, pp. 35-49. Springer, Heidlberg(1998)
31. Shamir, A.: Efficient Signature Scheme Based on Birational Permutations. In: Crypto 1993. LNCS, vol. 773, pp. 1-12. Springer, Heidlberg(1993)
32. Shamir, A., Kipnis, A.: Cryptanalysis of the Oil & Vinegar Signature Scheme. In: Krawczyk, H. (ed.) CRYPTO 1998. LNCS, vol. 1462, pp. 257C266. Springer, Heidelberg (1998)
33. P.: Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. SIAM Journal on Computing 26(5), 1484-1509 (1997)
34. Tsujii, S., Kurosawa, K.,Itoh, T., Fujioka, A. and Matsumoto, T.: A public key cryptosystem based on the difficulty of solving a system of nonlinear equations. ICICE Transactions (D) J69-D, 12:1963-1970(1986).
35. Wolf, C. and Preneel, B.: Taxonomy of Public Key Schemes based on the problem of Multivariate Quadratic equations. `http://eprint.iacr.org/2005/077`
36. Wolf, C., Braeken, A., and Preneel, B.: Efficient cryptanalysis of RSE(2)PKC and RSSE(2)PKC. In Conference on Security in Communication Networks-SCN 2004, LNCS, vol. 3352, pp. 294-309. Springer, Heidlberg(2004). Extended version: `http://eprint.iacr.org/2004/237`.
37. Yang, B.Y., Chen, J.M. and Chen, Y. H.: TTS: High-speed signatures on a low-cost smart card. In: Marc Joye Jean-Jacques Quisquater (ed.) CHES 2004. LNCS, vol 3156, pp. 371-385, Springer, Heidlberg(2004).
38. Yang, B.Y. and Chen, J.M: Building secure tame-like multivariate public-key cryptosystems- the new TTS. In: Boyd, Gonzalez, (ed.) ACISP 2005, LNCS, vol. 3574, pp. 518-531. Springer, Heidlberg(2005).