Multivariate Cryptography

Louis Goubin, Versailles St-Quentin-en-Yvelines University, France,
Jacques Patarin, Versailles St-Quentin-en-Yvelines University, France, and
Bo-Yin Yang, Institute of Information Science, Academia Sinica, Taipei, Taiwan

Synonyms

- MPKC
- $\mathcal{MQ}$
- Multivariate Quadratic Public-Key Cryptosystem (MQPKC)

Related Concepts

- Differential cryptanalysis
- Digital signature schemes
- Gröbner basis
- Post-quantum cryptography
- Public-key cryptography

Key words

- MQ problem (solving a system multivariate quadratic equations)
- IP problem (Isomorphism of polynomials)
- MP problem (Morphism of Polynomials)
- MinRank
- Polar form of quadratic equations over finite fields
- Functional decomposition

Definition

A *Multivariate Public-Key Cryptosystem* (MPKC) is a public-key cryptosystem where the public map $\mathcal{P}$, or trapdoor one-way function, is given as a set of m polynomial of a small degree d equations over n variables in a finite field F. Usually $d = 2$, hence the alternate name "Multivariate Quadratic" (MQ).

To decrypt, authenticate or sign digitally, a user must, for a given m-tuple $\mathbf{z} = (z_1, \ldots, z_m)$, find a solution for $\mathbf{w} = (w_1, \ldots, w_n)$ the system

$$(\mathcal{P}) \begin{cases} p_1(w_1, \ldots, w_n) = z_1 \\ \cdots \\ p_m(w_1, \ldots, w_n) = z_m \end{cases}.$$

For a *digital signature*, a challenge-response *authentication* scheme, and an *encryption* scheme, $\mathbf{z} = (z_1, \ldots, z_m)$ is respectively the hash of the message to be signed, the challenge, and the ciphertext, while $\mathbf{w} = (w_1, \ldots, w_n)$ is respectively the signature of this message, the response, and the cleartext.

Anyone can easily verify whether a given pair $(\mathbf{w}, \mathbf{z})$ satisfies the system $(\mathcal{P})$, or compute $\mathbf{z} = \mathcal{P}(\mathbf{w})$ from any given $\mathbf{w}$. However, finding at least one solution $\mathbf{w} = (w_1, \ldots, w_n)$ of the system $\mathcal{P}$ should be difficult for most $\mathbf{z} = (z_1, \ldots, z_m)$ without the knowledge of a secret key, but easy with the knowledge of a secret key. In other words, without the knowledge of a secret key, it should be infeasible to decrypt a message, forge a signature successfully, or pass an authentication.

Background

In general, solving a set of quadratic equations over a finite field is NP-hard (see computational complexity) for any finite field. This problem, known as $\mathcal{MQ}$, is even conjectured to be probabilistically hard, *i.e.* as $m, n \to \infty$, $\forall \varepsilon > 0$, and for any probabilistic Turing Machine $\mathcal{A}$, $\Pr(\mathcal{P}(x) = y$ be solved by $\mathcal{A}$ in $\text{poly}(m, n)) < \varepsilon$.

However, it is difficult to obtain a proof of security for multivariate schemes: since the system $(\mathcal{Q})$ has to be easy to solve, it is never random, and neither is $(\mathcal{P})$ which is obtained by linear changes of variables from $(\mathcal{Q})$.

One can always try to solve the system of equations directly using system solvers based on Gröbner bases (XL, $\mathbf{F_4}$, $\mathbf{F_5}$). For more or less "generic" or "random" equations, most current methods take exponential time when m and n are of the same size. Some systems appear very non-random under Gröbner basis methods. Recently differential analysis techniques were applied to MPKCs resulting in the break of the SFLASH system[**?**,**?**,**?**].

The first proposals of cryptosystems based on $\mathcal{MQ}$ date back to the early 1980's, see Imai et al. [**?**], Matsumoto and Imai [**?**], Tsujii et al. [**?**], Matsumoto and Imai [**?**].

Shor's algorithm cannot be used to speed up solving $\mathcal{MQ}$, so unlike the RSA cryptosystem, elliptic curve cryptography and the digital signature standard which are known to be breakable by quantum computers, MPKCs are a candidate for post-quantum cryptography.

There are variants in which more general systems of polynomials are used with the $\mathbf{w}$ and $\mathbf{z}$ variables mixed ("implicit MPKC"), or where not all the equations of the system $(\mathcal{P})$ have to be valid, but only a given fixed percentage ("probabilistic MPKC"). Some authors also include in the category of multivariate cryptography other schemes like the zero-knowledge schemes IP [**?**] or MinRank [**?**]. These variants are not considered in this entry.

Theory

Typically, a multivariate scheme is built from an easy-to-solve system $\mathcal{Q}(\mathbf{x}) = \mathbf{y}$, the *central map*, which is then "hidden" by two secret random linear (or affine) transformations $S : \mathbf{w} \mapsto \mathbf{x}$ and $T : \mathbf{y} \mapsto \mathbf{z}$ to obtain a new system $(\mathcal{P})$, by composing them (on the left and on the right) with $\mathcal{Q}$. The system $(\mathcal{P})$ is the

public key, the original system ($\mathcal{Q}$) and the transformations S and T form the secret key.

More precisely, if ($\mathcal{Q}$) is given by m polynomials $(q_1, \ldots, q_m)$ in $(x_1, \ldots, x_n)$, the new system ($\mathcal{P}$) is given by m polynomials $(p_1, \ldots, p_m)$ in $(w_1, \ldots, w_n)$ as in:

$$(p_1, \ldots, p_m)(w_1, \ldots, w_n) = T(q_1(S(w_1, \ldots, w_n)), \ldots, q_m(S(w_1, \ldots, w_n)))$$

where S and T are secret random linear (or affine) bijective changes of variables.

The new system $\mathcal{P} := T \circ \mathcal{Q} \circ S$ is also quadratic and finding a solution of

$$(\mathcal{P}) \begin{cases} p_1(w_1, \ldots, w_n) = z_1 \\ \cdots \\ p_m(w_1, \ldots, w_n) = z_m \end{cases}$$

for a given m-tuple $(z_1, \ldots, z_m)$ is expected to be difficult without the knowledge of S and T. In the following, the system ($\mathcal{P}$) will also be denoted by $\mathcal{P}(\mathbf{w}) = \mathbf{z}$.

There exist several families of multivariate schemes, corresponding to several choices for the system $\mathcal{Q}$ (some examples are given below). Some schemes have been broken; for others, only generic Gröbner-basis attacks apply and the parameters are chosen large enough to make those infeasible.

Variants (Perturbations)

In order to increase the security of an MPKC or to repair a broken scheme, some variants or "perturbations" of the polynomials are often introduced, either on the system ($\mathcal{P}$) or on the system ($\mathcal{Q}$). The most classical variants are denoted $+$, $\oplus$, $-$, V and F.

- $+$: The public key of a "plus" variant is not given by the system of polynomials $(p_1(\mathbf{w}), \ldots, p_m(\mathbf{w}))$ but by $(p_1(\mathbf{w}) + L_1(R_1(\mathbf{w}), \ldots, R_\alpha(\mathbf{w})), \ldots, p_m(\mathbf{w}) + L_m(R_1(\mathbf{w}), \ldots, R_\alpha(\mathbf{w})), R_1(\mathbf{w}), \ldots, R_\alpha(\mathbf{w}))$, where $L_1, \ldots, L_m$ are secret linear forms and $R_1, \ldots, R_\alpha$ are α truly random secret quadratic polynomials in $(w_1, \ldots, w_n)$, with $\alpha \ll m$.
- $\oplus$: The public key is not given by the system of polynomials $(p_1(\mathbf{w}), \ldots, p_m(\mathbf{w}))$ but by $(p_1(\mathbf{w}) + R_1(L_1(\mathbf{w}), \ldots, L_\alpha(\mathbf{w})), \ldots, p_m(\mathbf{w}) + R_m(L_1(\mathbf{w}), \ldots, L_\alpha(\mathbf{w})))$, where $R_1, \ldots, R_m$ are m truly random secret quadratic polynomials, and $L_1, \ldots, L_\alpha$ are secret linear forms, with $\alpha \ll m$.
- $-$: The "minus" variant consists in keeping secret some of the m polynomials $(p_1, \ldots, p_m)$. Only $m - \beta$ polynomials are made public and will be used to check the validity of the equations.
- V : In the "vinegar" variant—also called "extra variables perturbation"—α new variables are introduced in the system ($\mathcal{Q}$), and secretly combined with the usual variables $(w_1, \ldots, w_n)$. This mixing is not always linear, but is such that the new obtained system is still quadratic.
- F : In the "fix" variant, the public key polynomials $(p'_1, \ldots, p'_m)$ are formed by replacing α of the variables in $(p_1(\mathbf{w}), \ldots, p_m(\mathbf{w}))$ by arbitrary linear combinations in $\mathbf{w} = (w_1, \ldots, w_n)$.

Applications

Among the most famous multivariate public-key schemes are C^* by Matsumoto and Imai [**?**,**?**]; HFE (Hidden Field Equations) by Patarin [**?**]; UOV (Unbalanced Oil and Vinegar) by Kipnis, Patarin, and Goubin [**?**]; Rainbow/TTS by Ding and Schmidt [**?**] and modified by Ding *et al*[**?**]; and IFS (Intermediate Field System, nicknamed "chocolate bar") by Billet, Patarin, Seurin [**?**].

Many other schemes have been presented by various authors from around the world, but most of them have been broken. For a recent overview article see Ding and Yang [**?**].

C^* and SFLASH

C^* [**?**] was originally designed by Matsumoto and Imai in 1985. The idea is to create a quadratic system ($\mathcal{Q}$) from a monomial transformation $\mathbf{x} \mapsto \mathbf{y} = \mathbf{x}^{1+q^\theta}$, for some θ, over an extension $F = \mathrm{GF}(q^n)$ of the finite field $K = \mathrm{GF}(q)$, where the elements $\mathbf{x}$ and $\mathbf{y}$ of F correspond to the usual vectors over K by using an explicit basis (see vector space).

It was broken by Patarin in 1995 [**?**], using the following idea: if $\mathbf{y} = \mathbf{x}^{1+q^\theta}$ then $\mathbf{x}\mathbf{y}^{q^\theta} = \mathbf{x}^{q^{2\theta}}\mathbf{y}$. As $\mathbf{v} \mapsto \mathbf{v}^q$ is K-linear, regardless of S and T, the variables $\mathbf{v}$ and $\mathbf{w}$ satisfy a system of equations of the form $\sum \alpha_{ij} w_i z_j + \sum \beta_i w_i + \sum \gamma_j z_j + \delta = 0$. The coefficients can be determined by evaluating the original system at many random $\mathbf{w}$s and inserting the resulting pairs $(\mathbf{w}, \mathbf{z})$. Finally, for a target $\mathbf{z}$ this linear system can be solved for $\mathbf{w}$ by Gaussian elimination.

Many ways to repair the original scheme have been suggested. Most famous among these is the SFLASH scheme [**?**], which is a $(C^*)^-$ instance (*i.e.* a "minus" variant of the C^* scheme). It has been broken in 2007 by Dubois, Fouque, Shamir and Stern [**?**]. Some variants of C^* are still under investigation, for example in characteristic 3 instead of characteristic 2.

HFE and QUARTZ

Instead of using a monomial over an extension $F = \mathrm{GF}(q^n)$ of the finite field $\mathrm{GF}(q)$, HFE [**?**] uses a *polynomial* transformation

$$F \to F; \mathbf{x} \mapsto \mathbf{y} = \sum_{0 \le i \le j < r} a_{ij} \mathbf{x}^{q^i + q^j} + \sum_{0 \le i < r} b_i \mathbf{x}^{q^i} + c.$$

and builds ($\mathcal{Q}$) by using an explicit basis of F over $\mathrm{GF}(q)$. Operations involving the secret key work in F while the public key is given over $\mathrm{GF}(q)$. By construction ($\mathcal{Q}$) is quadratic in $\mathbf{x}$. Finding $\mathbf{w}$ given $\mathbf{z}$ corresponds to finding roots of the polynomial *e.g.* using Berlekamp's Algorithm (see Berlekamp Q matrix).

When the maximum degree D is fixed, polynomial time attacks exist, as mentioned in the original paper. When D increases with n, no polynomial attack is known. Nevertheless, subexponential and superpolynomial-time attacks have been found by Faugère and Joux: the basic HFE central map has (in some sense) a rank which is decided by D, which in turn bounds the operating degree in a direct algebraic attack if D is too small. Therefore, it is recommended to use HFE with some perturbations. For example, only exponential attacks are known

against HFE$^-$. QUARTZ [**?**] is an HFE^{V-} scheme with $q = 2$, $d = 129$, $n = 103$, $r = 8$, $v = 3$ and the central polynomial is of the form below with 4 polynomials removed from the system:

$$\sum_{\substack{0\le i,j<r\\ q^i+q^j\le d}} a_{ij}\mathbf{x}^{q^i+q^j} + \sum_{\substack{0\le i,j<r\\ q^i+q^j\le d}} b_{ij}\mathbf{x}^{q^i}\bar{\mathbf{x}}^{q^j} + \sum_{\substack{0\le i,j<r\\ q^i+q^j\le d}} \alpha_{ij}\bar{\mathbf{x}}^{q^i+q^j} + \sum_{i=0}^{r-1} b_i\mathbf{x}^{q^i} + \sum_{i=0}^{r-1} \beta'_i\bar{\mathbf{x}}^{q^i} + c,$$

where $\bar{\mathbf{x}}$ denotes the vinegar variables.

UOV

In UOV [**?**], the system $(\mathcal{Q})$ is given by m quadratic polynomials in $n = m+v$ variables $(w_1, \ldots, w_m, w_{m+1}, \ldots, w_{m+v})$. The first m variables $w_1, \ldots, w_m$ are called "oil variables" and the v other variables $w_{m+1}, \ldots, w_{m+v}$ are called "vinegar" variables. Each polynomial may contain quadratic terms of the form "oil×vinegar" (*i.e.* w_iw_j with $1 \le i \le m$ and $m+1 \le i \le m+v$) or "vinegar×vinegar" (*i.e.* w_iw_j with $m+1 \le i, j \le m+v$), but must not contain quadratic terms of the form "oil×oil" (*i.e.* w_iw_j with $1 \le i, j \le m$). As a consequence, the system $(\mathcal{Q})$ is easy to solve, by fixing arbitrary values for the vinegar variables, and solving the obtained system (in the m oil variables) by Gaussian elimination, however $(\mathcal{P})$ should hide the distinction between oil and vinegar variables. An early version of this scheme ("Oil and Vinegar", corresponding to the particular choice $v = m$ of the parameters) was broken by Kipnis and Shamir in 1998 [**?**]. However the "unbalanced" version (UOV) remains unbroken as long as $v > 2m$ (*i.e.* $n > 3m$).

Rainbow/TTS

The idea of the Rainbow scheme [**?**,**?**] is to use u UOV instances, in an iterative way. The first UOV instance contains $v_2 - v_1$ polynomials in $v_2 - v_1$ oil variables and v_1 vinegar variables. The second UOV instance contains $v_3 - v_2$ polynomials in $v_3 - v_2$ oil variables and v_2 vinegar variables. The last UOV instance contains $v_{u+1} - v_u$ polynomials in $v_{u+1} - v_u$ oil variables and v_u vinegar variables. Denoting v_{u+1} by n, the obtained system $(\mathcal{Q})$ thus contains $m = n - v_1$ polynomials in n variables, and is easily solved by recursively applying the UOV principle: fix arbitrarily the v_1 vinegar variables of the first UOV instance, solve this instance in the $v_2 - v_1$ oil variables, then consider all these $v_1 + (v_2 - v_1) = v_2$ variables as the vinegar variables of the second UOV instance, and so on. This means that the legitimate signer has to solve more but smaller systems compared to UOV. For carefully chosen parameters, the Rainbow scheme remains unbroken. TTS [**?**], a predecessor of Rainbow, can be considered an aggressive variant of the latter with sparse coefficients. This makes it faster in exchange of opening more possible pitfalls.

"Medium-Sized" Extension Fields and IFS

Attempts were made with use several variables in a extension field such as [**?**,**?**]. An unbroken one is the "Intermediate Field System" (IFS, [**?**]) where $(\mathcal{Q})$

is constructed using a system ($\mathcal{Q}'$) of k polynomial equations in k variables over an algebraic extension field of F. This system ($\mathcal{Q}'$) is chosen to be solvable using tailored Gröbner bases algorithms, and such that when rewritten in terms of variables $\mathbf{x}$ using a basis over the "small" field F, it is quadratic. For carefully chosen parameters, the IFS scheme remains unbroken.

Implementations

Multivariate digital signature schemes can have very short signatures as in QUARTZ [**?**], which allows approximately 100 bit long signatures.

Multivariate cryptography is in general quite fast both for the public and private maps. For instance, Rainbow/TTS are among the fastest digital signature schemes and can be implemented cheaply on ASICs.

Another advantage of multivariate schemes is their flexibility in the design of various schemes, with ad-hoc properties.

Open problems

The main drawbacks of MPKCs are large keys and the uncertainty about their security. It is difficult to obtain (relative) security proofs as in schemes based on factorization of discrete logarithm problem. The confidence in a given scheme relies on its resistance to all known attacks that have been developed against multivariate systems. Most needed are some provable security results; also useful are management of large keys and continuing optimizations on current hardware.

Recommended Readings

[1] O. Billet, J. Patarin, Y. Seurin, *Analysis of Intermediate Field Systems*. In Proceedings of SCC 2008. 2008.

[2] Chen, A. I.-T., Chen, M.-S., Chen, T.-R., Cheng, C.-M., Ding, J., Kuo, E. L.-H., Lee, F. Y.-S., Yang, B.-Y., *SSE Implementation of Multivariate PKCs on Modern x86 CPUs*, Proc. CHES 2009, Lecture Notes in Computer Science 5747, pp. 33-48.

[3] Courtois, N.T. *Efficient Zero-knowledge authentication based on a linear algebra problem MinRank*. In Proceedings of ASIACRYPT'2001, LNCS 2248, pp. 402-421, Springer-Verlag, 2001.

[4] Courtois, N.T. Goubin, L., Patarin, J. *QUARTZ, 128-bit long digital signatures*. In Proceedings of CT-RSA'2001, LNCS 2020, pp. 282-297, Springer-Verlag, 2001.

[5] Ding, J., Dubois, V., Yang, B.-Y., Chen, C.-H., Cheng, C.-M., *Can SFLASH be Repaired*, Proc. ICALP 2008, Lecture Notes in Computer Science 5126, pp. 691-701.

[6] Ding, J., Schmidt, D. *Rainbow, a New Multivariable Polynomial Signature Scheme*. In Proceedings of ACNS 2005, LNCS 3531, pp. 164-175, Springer-Verlag, 2005.

[7] Ding, J., Werner, F., Yang, B.-Y., Chen, C.-H., Chen, M.-S., *Odd-Char Multivariate Hidden Field Equations*, Cryptology ePrint Archive Report 2008/543 version 20081229:161921.

[8] Ding, J., Wolf, C., Yang, B.-Y., *ℓ-Invertible Cycles for Multivariate Quadratic Public Key Cryptography*, Proc. PKC 2007, Lecture Notes in Computer Science 4450, pp. 266-281.

[9] Ding, J., Yang, B.-Y., Chen, C.-H., Chen, M.-S., Cheng, C.-M., *New Differential-Algebraic Attacks and Reparametrization of Rainbow*, Proc. ACNS 2008, Lecture Notes in Computer Science 5037, pp. 242–257.

[10] Ding, J. and Yang, B.-Y., *Multivariate Public-Key Cryptography*, in Bernstein, D. J., Buchmann, J., and Dahmen, E., eds, *Post-Quantum Cryptography*, Springer-Verlag, 2009. ISBN:978-3-540-88701-0, e-ISBN:978-3-540-88702-7.

[11] Dubois, V., Fouque, P.-A., Shamir, A., and Stern, J., *Practical Cryptanalysis of SFLASH*, Proc. Crypto 2007, Lecture Notes in Computer Science 4622, pp. 1–12.

[12] Faugère, J.-C., and Joux, A., *Algebraic Cryptanalysis of Hidden Field Equation (HFE) Cryptosystems Using Gröbner Bases.* Proc. Crypto 2003, Lecture Notes in Computer Science 2729, pp. 44-60.

[13] Faugère, J.-C., and Perret, L., *Polynomial Equivalence Problems: Algorithmic and Theoretical Aspects.* Proc. Eurocrypt 2006, Lecture Notes in Computer Science 4004, pp. 30-47.

[14] Garey, M. R., and Johnson D. S., *Computers and Intractability — A Guide to the Theory of NP-Completeness*, W.H. Freeman and Company, 1979.

[15] Kipnis, A., Patarin, J., and Goubin, L., *Unbalanced Oil and Vinegar signature schemes*, Proc. Eurocrypt 1999, Lecture Notes in Computer Science 1592, pp. 206–222.

[16] Kipnis, A., Shamir, A. *Cryptanalysis of the oil and vinegar signature scheme.* in: Krawczyk, H. (editor). Advances in cryptology — CRYPTO '98, 18th annual international cryptology conference, Santa Barbara, California, USA, August 23, proceedings. Lecture Notes in Computer Science 1462. pp. 257–266, 1998.

[17] Matsumoto, T., Imai, H., Harashima, H., and Miyakawa, H. *A class of asymmetric cryptosystems using obscure representations of enciphering functions*. In Proceedings of the 1983 National Convention Record on Information Systems, IECE, 1983.

[18] Matsumoto, M. and Imai, H., *Algebraic Methods for Constructing Asymmetric Cryptosystems*, Algebraic Algorithms and Error-Correcting Codes: 3rd International Conference, AAECC-3, Grenoble, France, July 15-19, 1985, proceedings, Lecture Notes in Computer Science 229, pp. 108–119, 1986.

[19] Matsumoto, M. and Imai, H., *Public quadratic polynomial-tuples for efficient signature verification and message-encryption*, Proc. Eurocrypt 1988, Lecture Notes in Computer Science 330, pp. 419–545.

[20] Patarin, J. *Cryptanalysis of the Matsumoto and Imai public Key Scheme of Eurocrypt'88*. Proceedings of CRYPTO'95, Lecture Notes in Computer Science 963, pp. 248-261, Springer-Verlag, 1995.

[21] Patarin, J., *Hidden Fields Equations (HFE) and Isomorphisms of Polynomials (IP): two new families of Asymmetric Algorithms.* Proc. Eurocrypt 1996, Lecture Notes in Computer Science 1070, pp. 33-48.

[22] Patarin, J., Courtois, N., Goubin, L., *FLASH, a Fast Multivariate Signature Algorithm*, Proceedings of the 2001 Conference on Topics in Cryptology: The Cryptographer's Track at RSA, Lecture Notes in Computer Science 2020, pp.298–307, 2001.

[23] Tsujii, S., Itoh, T., Fujioka, A., Kurosawa, K., and Matsumoto, T. *A public-key cryptosystem based on the difficulty of solving a system of nonlinear equations*. Systems and Computers in Japan 19, 10–18, 1988.

[24] Yang, B.-Y., Chen, J.-M., Chen, Y.-H. *TTS: high-speed signatures on a low-cost smart card*. in: Joye, M., Quisquater, J.-J. (editors). Cryptographic hardware and embedded systems—CHES 2004, 6th international workshop, Cambridge, MA, USA, August 11–13, 2004, proceedings. Lecture Notes in Computer Science 3156. pp. 371–385, 2004.