On the security of UOV

Jean-Charles Faugère, Ludovic Perret

Abstract. In this short note, we investigate the security of the Unbalanced Oil and Vinegar Scheme [15]. To do so, we use a hybrid approach for solving the algebraic systems naturally arising when mounting a signature-forgery attack. The basic idea is to compute Gröbner bases of several modified systems rather than a Gröbner basis of the initial system. It turns out that our approach is efficient in practice. We have obtained a complexity bounded from above by $2^{40.3}$ (or 9 hours of computation) to forge a signature on a set of parameters proposed by the designers of UOV.

1. Introduction

Multivariate Cryptography is the set of all the cryptographic primitives using multivariate polynomials. The use of algebraic systems in cryptography dates back to the mid eighties, and was initially motivated by the need for alternatives to number theoretic-based schemes. Indeed, although quite a few problems have been proposed to construct public-key primitives, those effectively used are essentially factorization (e.g. in RSA [17]) and discrete logarithm (e.g. in Diffie-Hellman key-exchange [11]). It has to be noted that multivariate systems enjoy low computational requirements; moreover, such schemes are not concerned with the quantum computer threat, whereas it is well known that number theoretic-based schemes like RSA, DH, or ECDH are [18].

Multivariate cryptography has become a dynamic research area, as reflected by the ever growing number of papers in the most famous cryptographic conferences. This is mainly due to the fact that an European project (NESSIE[1]) has advised in 2003 to use such a scheme (namely, SFLASH [8]) in the smart-card context. Unfortunately, Dubois, Fouque, Shamir and Stern [10] discovered a sever flaw in the design of SFLASH, leading to an efficient cryptanalysis of this scheme. In this paper, we investigate the security of another multivariate signature scheme, the so-called Unbalanced Oil and Vinegar UOV scheme [15]. To this end, we have used

[1] https://www.cosic.esat.kuleuven.be/nessie/

Gröbner bases [6, 7] and efficient algorithms for computing such bases; namely F_5 [14].

1.1. Organization of the Paper

After this introduction, the paper is organized as follows. In Section 2, we introduce the main concern of this paper, namely the Unbalanced Oil and Vinegar UOV scheme [15]. To our knowledge, no successful attack has been reported on this scheme. In section 3, we will present an efficient algebraic attack against UOV. Our attack can be viewed as a natural extension of [5, 4].

2. Unbalanced Oil and Vinegar Scheme

The most interesting type of one-way function used in multivariate cryptography is based on the evaluation of a set of algebraic polynomials $\mathbf{p} = \big(p_1(x_1, \ldots, x_n), \ldots, p_m(x_1, \ldots, x_n)\big) \in \mathbb{K}[x_1, \ldots, x_n]^m$, namely :

$$\mathbf{m} = (m_1, \ldots, m_n) \in \mathbb{K}^n \longmapsto \mathbf{p}(\mathbf{m}) = \big(p_1(\mathbf{m}), \ldots, p_m(\mathbf{m})\big) \in \mathbb{K}^m.$$

The mathematical hard problem underlying this one-way function is :
Polynomial System Solving (PoSSo**)**
INSTANCE : polynomials $p_1(x_1, \ldots, x_n), \ldots, p_m(x_1, \ldots, x_n)$ of $\mathbb{K}[x_1, \ldots, x_n]$.
QUESTION : Does there exists $(z_1, \ldots, z_n) \in \mathbb{K}^n$ s. t. :

$$p_1(z_1, \ldots, z_n) = 0, \ldots, p_m(z_1, \ldots, z_n) = 0.$$

To introduce a trapdoor, we start from a carefully chosen algebraic system :

$$\mathbf{f}(\mathbf{x}) = \big(f_1(x_1, \ldots, x_n), \ldots, f_m(x_1, \ldots, x_n)\big) \in \mathbb{K}[x_1, \ldots, x_n]^m,$$

which is *easy* to solve. That is, for all $\mathbf{c} = (c_1, \ldots, c_m) \in \mathbb{K}^m$, we have an efficient method for describing/computing the zeroes of :

$$f_1(x_1, \ldots, x_n) = c_1, \ldots, f_m(x_1, \ldots, x_n) = c_m.$$

In order to hide the specific structure of $\mathbf{f}$, we usually choose two linear transformations – given by invertible matrices – $(S, U) \in GL_n(\mathbb{K}) \times GL_m(\mathbb{K})$ and set

$$\big(p_1(\mathbf{x}), \ldots, p_m(\mathbf{x})\big) = \big(f_1(\mathbf{x} \cdot S), \ldots, f_m(\mathbf{x} \cdot S)\big) \cdot U,$$

abbreviated by $\mathbf{p}(\mathbf{x}) = \mathbf{f}(\mathbf{x} \cdot S) \cdot U \in \mathbb{K}^m$ to shorten the notation.

The public-key of such systems will be the polynomials of $\mathbf{p}$ and the secret-key is the two matrices $(S, U) \in GL_n(\mathbb{K}) \times GL_m(\mathbb{K})$ and the polynomials of $\mathbf{f}$.

To generate a signature $\mathbf{s} \in \mathbb{K}^n$ of a digest $\mathbf{m} \in \mathbb{K}^m$, we compute $\mathbf{s}' \in \mathbb{K}^n$ such that $\mathbf{f}(\mathbf{s}') = \mathbf{m} \cdot U^{-1}$. This can be done efficiently due to the particular choice of $\mathbf{f}$. Finally, the signature is $\mathbf{s} = \mathbf{s}' \cdot S^{-1}$ since :

$$\mathbf{p}(\mathbf{s}) = \mathbf{f}(\mathbf{s}' \cdot S^{-1} \cdot S) \cdot U = \mathbf{m} \cdot U^{-1} \cdot U = \mathbf{m}.$$

To verify the signature $\mathbf{s} \in \mathbb{K}^n$ of the digest $\mathbf{m} \in \mathbb{K}^m$, we check whether the equality : "$\mathbf{p}(\mathbf{s}) = \mathbf{m}$" holds. We would like to emphasize that most of the multivariate

signature schemes proposed so far (e.g. [8, 19]), including UOV [15], follow this general principle. For UOV, the matrix U is simply equal to the identity matrix.

The main specificity of UOV lies in the way of constructing the inner polynomials $\mathbf{f}(\mathbf{x}) = \big(f_1(x_1,\ldots,x_n),\ldots,f_m(x_1,\ldots,x_n)\big) \in \mathbb{K}[x_1,\ldots,x_n]^m$. Kipnis, Patarin, and Goubin [15] proposed the following construction. The n variables $x_1,\ldots,x_n$ are partitioned in two sets $\{x_i\}_{i\in V}$ and $\{x_i\}_{i\in O}$ where $V = \{1,\ldots,n-m\}$ is the set of *vinegar* indices and $O = \{n-m+1,\ldots,n\}$ the set of *oil* indices. Each polynomial f_k of the secret mapping has a quadratic part $f_k^{(2)}$ of the special form :

$$f_k^{(2)}(x_1,\ldots,x_n) = \sum_{(i,j)\in V\times V | i\leq j} \alpha_{i,j}^{(k)} x_i x_j + \sum_{(i,j)\in V\times O} \beta_{i,j}^{(k)} x_i x_j \ . \qquad (1)$$

As a consequence of this very special structure, the equations induced by equations (1) when fixing the vinegar variables to constant values are linear in the remaining (oil) variables. This provides an efficient method for inverting system $\mathbf{f}(\mathbf{x})$ with a high probability — the linear system in the oil variables we obtain is invertible with high probability. For more details, we refer the reader to the initial paper [15].

2.1. Recommended Values for UOV

The authors of UOV have recommended in [15, 16] to choose parameter values such that $n > 3m$. In particular, they proposed [16] the following set of parameters : $\mathbb{K} = \mathbb{F}_{2^4}, m = 16, n = 32$ (or 48). We will show that this set of parameters does not guaranty a sufficient level of security.

3. Description of the Attack

We now describe our attack against UOV [15]. Our goal is to forge a valid signature $\mathbf{s}' \in \mathbb{K}^n$ for a given digest $\mathbf{m} = (m_1,\ldots,m_m) \in \mathbb{K}^m$. In other words, we want to find an element of the variety :

$$V_{\mathbb{K}}(p_1 - m_1,\ldots,p_m - m_m) \subseteq \mathbb{K}^n,$$

with $p_1,\ldots,p_m \in \mathbb{K}[x_1,\ldots,x_n]$ the polynomials of UOV public-key. We recall that the parameters are $\mathbb{K} = \mathbb{F}_{2^4}, m = 16$ and $n = 32$ (or 48) .

The main limitation for computing directly this variety is due to the fact that the number of equations (m) is smaller that the number of variables (n). As a consequence, there is at least $(\#\mathbb{K})^{n-m}$ valid solutions to the signature-forgery system. Hence, even if you suppose that you have been able to compute a Gröbner basis for the degree reverse lexicographical ordering (DRL), you will probably not be able to recover efficiently the Lex-Gröbner basis using FGLM [12]. The reason is that the complexity of FGLM is polynomial in the size of the variety.

A natural way to overcome these practical limitations is to randomly specialize (i.e. fix) $n-m$ variables. We will have to solve a system having the same number of variables and equations (m). For each specification of the $n-m$ variables, we

can always find a solution of the new system yielding to a valid signature. We also mention that the specialized system will have very few solutions in practice. Thus, the cost of computing the variety will be now essentially the cost of computing a Gröbner basis. It is exactly at this point that the authors of [5], which also tried to attack UOV using Gröbner basis, stopped their analysis.

The important observation here is that – after having specified $n-m$ variables – the new system will behave like a semi-regular system [1, 3, 2]. We will present latter in this section experimental results supporting this claim. Note that such a behavior has been also observed, in a different context, in [20]. The degree of regularity of a semi-regular system of m variables and equations is equal to $m+1$. In our context ($m=16$), this remains out of the scope of the F_5 algorithm. This is exactly the reason explaining why Braeken, Wolf, Preneel concluded that their attack can not be efficient.

To avoid this difficulty, we can use the fact that $\#\mathbb{K}$ is relatively small and try to decrease the degree of regularity by specializing $r \geq 0$ more variables (in addition of the $n-m$ variables already fixed). Thus, we will have to solve a systems of m equations with $m-r$ variables, which behave like semi-regular systems. This allows to decrease the degree of regularity, and thus the complexity of F_5. For instance, the degree of regularity of a semi-regular system of $m-1$ variables and m equations is approximately equal to $\left\lceil \frac{(m+1)}{2} \right\rceil$. More generally, the degree of regularity is given by the index of the first non-positive coefficient of the series[1, 3, 2] :

$$\frac{\prod_{i=1}^{m}(1-z^2)}{(1-z)^{m-r}}.$$

In the following table, we have quoted the degree of regularity observed in our experiments. Namely, the maximum degree reached during F_5 on systems obtained by fixing $n-m+r$ variables ($r \geq 0$) on signature-forgery systems. We have also quoted the theoretical degree of regularity of a semi-regular system of m equations in $m-r$ variables. These experiments suggest that the systems obtained when mounting a specify+solve signature forgery attack against UOV behave like semi-regular systems.

m	$m-r$	r	d_{reg} (theoretical)	d_{reg} (observed)
16	15	1	9	9
16	14	2	7	7
16	13	3	6	6

By fixing variables, we can obtain a significant gain on the complexity the F_5. On the other hand, as soon as $r > 0$, each specification of the r variables will not necessarily lead to an algebraic system whose set of solutions is not empty . But, we know that there exists a least one guess of the r variables (in practice exactly one) leading to a system whose zeroes allow to construct a valid signature. Thus, we have to perform an exhaustive search on the r new variables. In other words, instead of computing one Gröbner basis of a system of m equations and variables, we compute $(\#\mathbb{K})^r$ Gröbner bases of "easier" systems (m equations with $m-r$

variables). We have then to find an optimal tradeoff between the cost of F_5 and the number of Gröbner basis that we have to compute.

In the table above, we have quoted practical results that we have obtained with F_5 when solving systems obtained by fixing $n-m+r$ variables ($r \geq 0$) on signature-forgery systems. In this table, T_{F_5} is the time of computing one Gröbner basis with F_5. We also included the corresponding number of operations (field multiplications) Nop_{F_5} performed by F_5, and the total number N of operations of our attack (i.e. the cost of computing $2^{4 \cdot r}$ Gröbner bases). Finally, we have quoted the maximum memory, denoted Mem, used during the Gröbner basis computation. The experimental results have been obtained using a bi-pro Xeon 2.4 Ghz with 6 Gb. of Ram.

m	$m-r$	r	T_{F_5}	Mem	Nop_{F_5}	N
16	15	1	≈ 1 h.	3532 Mb.	$2^{36.9}$	$2^{40.9}$
16	14	2	126 s.	270 Mb.	$2^{32.3}$	$2^{40.5}$
16	13	3	9.41 s.	38 Mb.	$2^{28.7}$	$2^{40.7}$

Once can see that the most interesting tradeoff is obtained with $r = 2$. In this case, we obtain a complexity of $2^{40.3}$, which is approximatively equivalent to 9 hours of computations. It is interesting to remark that the complexity of our attack not relies on the number of variables n, but only on the number of polynomials. So, the result presented here are valid for $n = 32$, or $n = 48$ (and for any $n = k \cdot m$, with $k \geq 1$).

As the consequence, the parameters of UOV [16] that we have studied should be no longer recommended. We believe that one should use the other set of parameters, with $\mathbb{K} = \mathbb{F}_2$, proposed in [16].

References

[1] M. Bardet. *Etude des systèmes algébriques surdéterminés. Applications aux codes correcteurs et à la cryptographie.* Thèse de doctorat, Université de Paris VI, 2004.

[2] M. Bardet, J-C. Faugère, B. Salvy *On the complexity of Grbner basis computation of semi-regular overdetermined algebraic equations.* In Proc. International Conference on Polynomial System Solving (ICPSS), pp. 71–75, 2004. Available at *http://www-calfor.lip6.fr/ICPSS/papers/43BF/43BF.htm.*

[3] M. Bardet, J-C. Faugère, B. Salvy and B-Y. Yang. *Asymptotic Behaviour of the Degree of Regularity of Semi-Regular Polynomial Systems.* In Proc. of MEGA 2005, Eighth International Symposium on Effective Methods in Algebraic Geometry, 2005.

[4] L. Bettale, J.-C. Faugère, L. Perret. *Cryptanalysis of the TRMS System of PKC'05.* Africacrypt'2008, Lecture Notes in Computer Science,vol. 5023. *to appear.*

[5] A. Braeken, C. Wolf, B. Preneel. *A Study of the Security of Unbalanced Oil and Vinegar Signature Schemes.* The Cryptographer's Track at RSA Conference

2005 (CT'RSA 2005), Lecture Notes in Computer Science, vol. 3376 , Springer-Verlag, pp. 29–43 ,2005.

[6] B. Buchberger, G.-E. Collins, and R. Loos. *Computer Algebra Symbolic and Algebraic Computation.* Springer-Verlag, second edition, 1982.

[7] B. Buchberger. *Gröbner Bases : an Algorithmic Method in Polynomial Ideal Theory.* Recent trends in multidimensional systems theory. Reider ed. Bose, 1985.

[8] N. Courtois, L. Goubin, and J. Patarin. *SFLASH, a Fast Symmetric Signature Scheme for low-cost Smartcards – Primitive Specification and Supporting documentation.* Available at `www.minrank.org/sflash-b-v2.pdf`.

[9] V. Dubois, P.-A. Fouque, and J. Stern. *Cryptanalysis of SFLASH with Slightly Modified Parameters.* Advances in Cryptology – EUROCRYPT 2007, to appear.

[10] V. Dubois, P.-A. Fouque, A. Shamir, and J. Stern. *Practical Cryptanalysis of SFLASH.* Advances in Cryptology – CRYPTO 2007.

[11] W. Diffie, and M.E. Hellman. *New Directions in Cryptography.* IEEE Transactions on Information Theory, IT–22(6), pp. 644–654, 1976.

[12] J. C. Faugère, P. Gianni, D. Lazard, and T. Mora. *Efficient Computation of Zero-Dimensional Gröbner Bases by Change of Ordering.* Journal of Symbolic Computation, 16(4), pp. 329–344, 1993.

[13] J.-C. Faugère. *A New Efficient Algorithm for Computing Gröbner Basis:* F_4. Journal of Pure and Applied Algebra, vol. 139, pp. 61–68, 1999.

[14] J.-C. Faugère. *A New Efficient Algorithm for Computing Gröbner Basis without Reduction to Zero:* F_5. Proceedings of ISSAC, pp. 75–83. ACM press, July 2002.

[15] A. Kipnis, J. Patarin, and L. Goubin. *Unbalanced Oil and Vinegar Signature Schemes.* Advances in Cryptology – EUROCRYPT 1999, Lecture Notes in Computer Science, vol. 1592 , Springer-Verlag, pp. 206–222,1999.

[16] A. Kipnis, J. Patarin, and L. Goubin. *Unbalanced Oil and Vinegar Signature Schemes.* Extended version available at http://citeseer.ist.psu.edu/231623.html.

[17] R. Rivest, A. Shamir and L. Adleman. *A Method for Obtaining Digital Signatures and Public-Key Cryptosystems.* Communications of the ACM, 21(2), pp. 120–126, 1978.

[18] P. W. Shor. *Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer.* SIAM J. Computing 26, pp. 1484-1509 (1997).

[19] C. Wolf. *Multivariate Quadratic Polynomials in Public Key Cryptography.* Ph.D. thesis, Katholieke Universiteit Leuven, B. Preneel (supervisor), 156+xxiv pages, November 2005.

[20] B.-Y. Yang, J.-M. Chen, and N. T. Courtois. *On Asymptotic Security Estimates in XL and Gröbner Bases-Related Algebraic Cryptanalysis.* In proc. of ICICS 2004, Lecture Notes in Computer Science, vol. 3269, Springer–Verlag, pp. 401413, 2004.

Jean-Charles Faugère, Ludovic Perret
INRIA, Centre Paris-Rocquencourt, SALSA Project
UPMC, Univ Paris 06, LIP6
CNRS, UMR 7606, LIP6
Université Pierre et Marie Curie Paris 6
UFR Ingénierie 919
LIP6 Passy Kennedy, bureau 733
Boite Courrier 169
4, Place Jussieu 75252 Paris cedex 05

e-mail: Jean-Charles.Faugere@inria.fr
e-mail: ludovic.perret@lip6.fr