A PUBLIC KEY, SIGNATURE AND KEM SYSTEM BASED ON DOUBLE EXPONENTIATION WITH MATRIX EXPONENTS

I. LUENGO[1]

The system that we persent to the NIST call it is a multivariate public key cryptosystem based on a new construction of the central maps, that allow the polynomials of the public key to be of arbitrary degree. In order to get a reasonable size of of the public key one has to use a small number of variables and special non dense linear maps at both ends of the composition.

We will present the algorithms and construction of the system in general, but for the implementation we will choose parameter that give polinomials with 6 to 12 variables. We will build the central map using a vectorial exponentiation with matrix exponents as follows:

Let us take a finite field $\mathbb{F}_q$, $q = p^e$, and a matrix $A = (a_{ij}) \in \mathcal{M}_{n\times n}(\mathbb{Z}_{q-1})$ one can define a kind of exponentiation of vectors by using a monomial map G_A asociated to the matrix A as follows

$$G_A : {\mathbb{F}_q}^n \to {\mathbb{F}_q}^n : \quad G_A(x_1, \ldots, x_n) = (x_1^{a_{11}} \ldots x_n^{a_{1n}}, \ldots, x_1^{a_{n1}} \ldots x_n^{a_{nn}}). \tag{1}$$

The following two facts are easy to verify:

a) If $A, B. \in \mathcal{M}_{n\times n}(\mathbb{Z}_{q-1})$ and $C = BA$ then the composition $G_C = G_B \circ G_A$.
b) if $\det(A) = \pm 1$ and the inverse matrix $A^{-1} \in \mathcal{M}_{n\times n}(\mathbb{Z})$ then G_A is invertible on $(\mathbb{F}_q \setminus \{0\})^n$ and the inverse is given by $G_{A^{-1}}$.

Notice that if in a column k has r entries different from zero the product of r copies of $\mathbb{F}_q$ is maped to $O = (0, \ldots, 0)$ so G_A as a map in ${\mathbb{F}_q}^n$ is a univariate polynomial of degree at least q^r.

This kind of maps are extensibe used in Algebraic Geometry, they produce the birrational maps. In Projective Geometry that are called Cremona transformations. In [1] this Cremona transformations are use to produce multivariate public key cryptosystem

If $\det(A) \neq \pm 1$ the monomial mad is not birrational, in fact one has,

Proposition 1. *Let $G_A : {\mathbb{F}_q}^n \to {\mathbb{F}_q}^n$ be a monomial map as (1) and K an algebraically closed field of any characteristic then the monomial map G_A has geometric degree $d := |\det(A)|$ on $(K \setminus \{0\})^n$, that is, for $x \in (K \setminus \{0\})^n$, $G_A^{-1}(x))$ has generically d preimages.*

Now if we take $A \in \mathcal{M}_{n\times n}(\mathbb{Z}_{q-1})$ we have:

Theorem 0.1. *Let $A \in \mathcal{M}_{n\times n}(\mathbb{Z}_{q-1})$ and $G_A : {\mathbb{F}_q}^n \to {\mathbb{F}_q}^n$ be the corresponding monomial map. If $\gcd(\det(A), q-1) = 1$ and $b := \det(A)^{-1} \in \mathbb{Z}_{q-1}$, $B := bAd(A)$ then $A^{-1} = B \in \mathcal{M}_{n\times n}(\mathbb{Z}_{q-1})$.*

This is easy to verify because $b\det(A) = 1 + \lambda(q-1)$ and if I_n is the identity matrix then $AB = AbI_n (\cong I_n q - 1)$.

We can use this fact to build a multivariate PKC in the standard way by putting in the entries of the matrix A powers of q. If each row has 2 entries $q^{a_{ij}}$ then after composing with two linear maps at both ends one get a cuadratic public key (see [1]). In our case we made extensive computer test and we arrive to the conclusion that those systems are not safe agains Grobner bases attack for reazonable key size, what it happens with most multivariate PKC.

In order to make an stronger system against algebraic cryptoanalisys we will produce a system with the following design options:

- We allow the entries of the matrix A to be of the for p^a instead of $q^a (q = p^a)$, this will make the final polynomials with arbitrary degree up to q,
- the determinat $d = \det(A)$ has an expansion in base d with many non zero digits.

These two conditions make the resulting system safe against Grobner basis attack but in order to make it safe against structural attacks we propose as central maps to use to exponentials in two differnt intermediate fields, ${\mathbb{F}_q}^n$ and ${\mathbb{F}_q}^m$ and the resultant public key will be polinomials in $n-m$ variables with degree up to q in each variable. In the system we implemented we use the parameters $m = 3, n = 2$ and the the public key F has 6 polinomials with 64 monomials each.

For convenience we denote the coordinates in $(\mathbb{F}_q)^{nm}$ as

$$\underline{x} = (x_{11}, \ldots, x_{1n}, \ldots, x_{n1}, \ldots, x_{nm}).$$

We will use a padding $H : \mathbf{F}_p^N \to \mathbf{F}_q^{nm}$ by adding $S \geq m$ random elements of $\mathbf{F}_p$ in such a way that the coordinates of $H(u) = (x_1, \ldots, x_{nm}, x_{1n}, x_{2n}, \ldots, x_{nm}$ are different from zero. The padding can be chosen in several different ways, for instance one can add only one bit in each x_{in} and the encryption is deterministic or we can add randon bits to each component x_{ij} in oder to adrees the IND-CPA security.

The public key is $K_P = (h, \pi_0, F)$, where $F : \mathbb{F}_q^{nm} \to \mathbb{F}_q^{nm}$ is a map obtained as composition of five maps, $F = L_3 \circ G_2 \circ L_2 \circ G_1 \circ L_1$, according to the diagram:

The maps L_1, L_2 and L_3 are $\mathbb{F}_q$-linear isomorphims and L_1 satisfies that for every $x \in H(\mathbb{F}_p^N)$, $L_1(x) \in (\mathbb{F}_q^n \setminus \{0\})^m$. The map L_2 is designed to verify the

condition:

$$\forall y \in (\mathbb{F}_q^n \setminus \{0\})^m, \; L_2(y) \in (\mathbb{F}_q^m \setminus \{0\})^n.$$

$$\underbrace{\mathbb{F}_q^{nm} \xrightarrow{L_1} (\mathbb{F}_{q^n})^m \xrightarrow{G_1} (\mathbb{F}_{q^n})^m \xrightarrow{L_2} (\mathbb{F}_{q^m})^n \xrightarrow{G_2} (\mathbb{F}_{q^m})^n \xrightarrow{L_3} \mathbb{F}_q^{mn}}_{F}$$

The maps G_1 and G_2 are monomial maps with the invertible determinant and entries powers of p. With all the above conditions it is clear that F is injective in $H(\mathbb{F}_p^N)$ and the components of F and F^{-1} are given by polynomials in $\mathbb{F}_q[x_1, \ldots, x_{mn}]$. The maps G_1 and G_2 are chosen in such a way that the polynomial F have few monomials and the polynomial F^{-1} has a huge number of monomials. For instance, for the parameters that we choose for the implementation, $(M = 3, n = 2, s = 2, t = 2)$, we get that each component of F has 64 monomials and that each component of F^{-1} has at least 2^{100} monomials.

Let's describe the five maps in detail.

The map $L_1 = \tilde{\pi_1} \circ \tilde{L}_1 \circ \tilde{l}$ is obtained as composition of three linear $\mathbb{F}_q$-isomorphims according to the diagram (2).

$$\underbrace{\mathbb{F}_q^{nm} \xrightarrow[\sim]{} (\mathbb{F}_q^n)^m \xrightarrow{\tilde{L}_1} (\mathbb{F}_q^n)^m \xrightarrow[\sim]{\tilde{\pi}_1} (\mathbb{F}_{q^n})^m}_{L_1}$$

The map $\tilde{\pi_1} = (\pi_1, \ldots, \pi_m)$ is defined by using an $\mathbb{F}_q$ linear isomorphims $\pi_1 : \mathbb{F}_q^n \to \mathbb{F}_{q^n}, \pi_1(v_1, \ldots, v_n) = \alpha_1 v_1 + \ldots + \alpha_n v_m$, where $\{\alpha_1, \ldots, \alpha_n\}$ is a fixed $\mathbb{F}_q$ basis of $\mathbb{F}_{q^n}$.

The isomorfism $\tilde{L}_1 = (L_{11}, \ldots, L_{1m})$ is defined by its components $L_{1i} : \mathbb{F}_q^n \to .\mathbb{F}_q^n$ given by $L_{1i}(\underline{x}_i) = \underline{x}_i A_{1i}$, where $A_{1i} \in GL_n(\mathbb{F}_q)$.

The isomorfism $\tilde{l}$ is obtained by grooping the components of x in m vectors according to its index $h(x_1, \ldots, x_{nm}) = (\underline{x}_1, \ldots, \underline{x}_m))$, where $\underline{x}_i = (x_{i1}, \ldots, x_{in})$.

The $\mathbb{F}_q$-linear isomorphism $L_2 = \tilde{\pi_1}^{-1} \circ M \circ \tilde{L}_2 \circ \tilde{\pi}_2$ is a composition according to the diagram (3).

$$\underbrace{(\mathbb{F}_{q^n})^m \xrightarrow[\sim]{\tilde{\pi}_1^{-1}} (\mathbb{F}_q^n)^m \xrightarrow{M} (\mathbb{F}_q^m)^n \xrightarrow{\tilde{L}_2} (\mathbb{F}_q^m)^n \xrightarrow[\sim]{\tilde{\pi}_2} (\mathbb{F}_{q^m})^n}_{L_2}$$

The "mixing" isomorphism M transforms the m vectors of ${\mathbb{F}_q}^n$ in n vectors of ${\mathbb{F}_q}^m$ in such a way that the components of $\underline{x}_1, \ldots, \underline{x}_n$ are placed in the first n components of $\underline{x}'_1, \ldots, \underline{x}'_n$ and the components of $\underline{x}_{m-n+1}, \ldots, \underline{x}_m$ are placed in the last $m - n$ components of $\underline{x}'_1, \ldots, \underline{x}'_n$. For instance a way to produce such mixing is diagrama 4.

$$(\mathbb{F}_{q^m})^n \xrightarrow[\sim]{\tilde{\pi}_2} (\mathbb{F}_q^m)^n \xrightarrow{\tilde{L}_3} (\mathbb{F}_q^m)^n \xrightarrow[\sim]{} \mathbb{F}_q^{mn}$$
$$\overset{\curvearrowright}{L_3}$$

That is, if we write the first matrix in two blocks $\binom{M_1}{M_2}$, where M_1 is given by the first n rows and M_2 is given by the mixim M the last rows the mixing map send $\binom{M_1}{M_2}$ to $(M_1, M_2^t) = (M_1, M_2')$ Any bijective map that send the $(m-n)\times n$ entries of M_2 in the $n \times (m-n)$ entries of M_2'will be also valid, but the final number of monomial depends on the mixing.

For instance if we take $m = 4$ and $n = 2$ we can have the next two mixing of $\binom{x_{11}x_{12}}{x_{41}x_{42}}$

$$\begin{pmatrix} x_{11}x_{12}x_{31}x_{41} \\ x_{21}x_{22}x_{32}x_{42} \end{pmatrix} \tag{2}$$

or

$$\begin{pmatrix} x_{11}x_{12}x_{31}x_{32} \\ x_{21}x_{22}x_{41}x_{42} \end{pmatrix} \tag{3}$$

When we explain below how to calculate the monomials of F_i that (2) produces more mixing and 144 monomials and (3) produce less mixing but 64 monomials in each component.

This construction of the mixing map M guarantees that if $x \in (\mathbb{F}_{q^n} \setminus \{0\})^m$ then $x' \in (\mathbb{F}_{q^m} \setminus \{0\})^n$ but there is not implication in the other sense, that is $x' \in (\mathbb{F}_{q^m} \setminus \{0\})^n$ do not implies $x \in (\mathbb{F}_{q^n} \setminus \{0\})^m$. This fact means that one can always encript and decrypt a message but there are messages that can not be signed.

The $\mathbb{F}_q$-linear isomorphism $L_3 = e \circ \tilde{L}_3 \circ \pi_2^{-1}$ is defined as

$$(\mathbb{F}_{q^n})^m \xrightarrow[\sim]{\pi_2^{-1}} (\mathbb{F}_q^n)^m \xrightarrow[\tilde{L}_3]{} (\mathbb{F}_q^n)^m \xrightarrow[\ell^{-1}]{} \mathbb{F}_q^{nm}$$
$$\overset{\curvearrowright}{L_3}$$

The morphism $\tilde{L}_3$ is defined as $\tilde{L}_3 = (L_{31}, \ldots, L_{3n})$ where $L_{3j}(x_j') = x_j' A_{3j}$, $A_{3j} \in \mathscr{M}_{n\times n}(\mathbb{F}_q)$ and $\det(A_{3j}) \neq 0$.

The main part of the design of the system are the two exponential maps G_1 and G_2 build with monomial maps as follows:

$G_1(x_1, \ldots, x_m) = (x_1^{a_{11}} \cdot \ldots \cdot x_m^{a_{1m}}, \ldots, x_1^{a_{m1}} \cdot \ldots \cdot x_m^{a_{mm}})$, $\qquad G_1 : (\mathbb{F}_{q^n})^m \to (\mathbb{F}_{q^n})^m$

where $A_1 = (a_{ij}) \in \mathscr{M}_{m\times m}(\mathbb{Z}_{q^n-1})$ such that $d_1' = \det(A_1)$ is prime with $q^n - 1$;

$G_2(x_1', \ldots, x_n') = ({x_1'}^{b_{11}} \cdot \ldots \cdot {x_n'}^{b_{1n}}, \ldots, {x_1'}^{b_{n1}} \cdot \ldots \cdot {x_n'}^{b_{nn}})$, $\qquad G_1 : (\mathbb{F}_{q^m})^n \to (\mathbb{F}_{q^m})^n$

where $B_2 = (b_{ij}) \in \mathscr{M}_{n\times n}(\mathbb{Z}_{q^m-1})$ such that $d_2' = \det(B_2)$ is prime with $q^m - 1$

If $\underline{x} = (x_{11}, \ldots, x_{nm}) \in {\mathbb{F}_q}^{nm}$ are the inicial coordinates, then the composition of the five maps L_1, G_1, L_2, G_2 and G_3 allow us to compute the components of $F(\underline{x})$ as polynomials $F_i \in \mathbb{F}_q[x_{11}, \ldots, x_{nm}]$. In order to keep small the number of monomials, we choose the matrices A_1 and B_2 with the following properties:

(1) The entries of A_1 and B_2 are of the form p^a.
(2) We fix two integers s and t such that the rows of A_1 have at most s non zero entries and the rows of B_2 have at most t non zero entries. One can compute the monomials in the F_i with the algorithm described below, resulting that the total number of monomials is $MON = (b \cdot n^s)^t$ where b depends on the mixing map M.
(3) The inverse maps G_1^{-1} and G_2^{-1} can be computed in the same way from the inverse matrix of A_1 and B_2 respectively and F_1^{-1} is also polynomial.

If the number of monomials in F^{-1} is not very big, one can get the coefficient of the polynomial by computing enough number of pairs $(x, F(x))$. To avoid this attack we tak A_1 such that $d_1 = \frac{1}{\det(A_1)} \mod q^n - 1$ has a expansion in base p with $d_1 = [K_0, \ldots, K_e]$ with at least s_1 non vanishing digist and the same with B_2 and $d_2 = \frac{1}{\det(B_2)}$ (with at least t_1 non vanishing digits). The details of values of t_1, s_1 will be given when discussing the security of the system.

The public key of the system is $K_P = (h, \pi_0, F)$ and the private key is given by h, π_0 and the five maps $L_1, \ldots, L_3$ and their inverses that can be used to encrypt and decrypt. Given an encrypted message $z = F(\underline{x}) = DM(\overline{x})$, one compute $\underline{x} = F^{-1}(z)$ and discard the random entries with the use of h.

It is possible to get the monomials of the F_i without computing the composition of the five maps as follows: we start with m lists that contain the coordinates of the $\underline{x}_i$, $M_{11} = [x_{11}, \ldots, x_{1n}]$, $\ldots$, $M_{mn} = [x_{m1}, \ldots, x_{mn}]$, and we define the operations on lists: multiplication and exponentiation. If $S = [s_1, \ldots, s_m]$, $T = [t_1, \ldots, t_m]$ then $S \cdot T = [s_i \cdot t_i]$ and $S^a = [s_i^a]$.

With these notations, one can see that the exponential G_1 produce, on each component, polynomials whose list of monomials is $N_{0k} = M_{01}^{a_{k1}} \cdot \ldots \cdot M_{0n}^{a_{kn}}$.

The mixing map M determines that in the list of monomials of each x'_k appears the list N_{0k}, joint with the list N_{0j} of the vectors that are placed at the $m - n$ last entreis of x'_k. If b_k is the number of vectors adjoined to x'_k then, if we denote by P_{0k} $(k = 1, \ldots, n)$ such list, then the final list of monomials of each component after G_2 to each monomial ${x'_1}^{b_{k1}} \cdot \ldots \cdot {x'_n}^{b_{kn}}$ gives $Q_{0k} = P_{01}^{b_{k1}} \cdot \ldots \cdot P_{0n}^{b_{kn}}$.

Notice that when we apply the final $\mathbb{F}_q$-linear bijection $\tilde{L}_3$, each component still have the same monomial, than means that there are n groups of m polynomials $F_{k1}, \ldots, F_{km}$ such that they have the same monomials, namely the list Q_{0k}.

It is clear that the number of monomials of Q_{0k} is at most $((1+b_k)\cdot n^s)^t$. So if we denote by $b_{\max} = \max_k(1+b_k)$, we get on each component at most $(b_{\max}\cdot n^s)^t$ monomials.

Once we get the list of monomials of the F_i one gets the coefficient of each group of polynomials by evaluating the polynomials $F_{k1}, ...F_{km}$ set of pairs $(\underline{c}, F_{ki}(\underline{c}))$ big enough to guarantee that the corresponding linear equation are independendents. That is if $Q_k = [q_1...q_d]$ and $F_{kj} = \sum_{i=1}^{d} f_{ji}q_i(x)$ we take vector $\underline{c}_1, \ldots, \underline{c}_R$ such that the linear equations (on the f_{ij}) $F_k(c_e) = \sum f_{ji}q_i(c_e)$ are independent and can be resolved to get coefficient of the polynomials $F_{k1}, \ldots, F_{km}$. This algorithm is implemented in the system to get the public key from the private key.

It is also possible to use this algorithm to get a fast evaluation of the $F_{ij}(\underline{c})$ to encript a message. If we start with the list of the coordinates of $\underline{c}$ instead of the list of variables in the algorithm we get at the end a list of the evaluated monomials $[q_j(\underline{c})]$. In order to evaluate the polynomials $F_{kj}(\underline{c}) = \sum_{i=1}^{d} f_{ji}q_i(\underline{c})$ one needs only to write their cofficients f_{ij} in a matrix $MF_k = (f_{ji})$ and compute a matrix multiplication $b_i(x)\cdot MF_k$.

Summary of the system DME

Fix parameters (m, n, s, t, N, S), a field $\mathbb{F}_q$ with $q = p^e$ and an $\mathbb{F}_q$-isomorphism $\pi_0 : \mathbb{F}_p^e \to \mathbb{F}_{p^e}$. The public key is $K_P = (h, \pi_0, F)$ or $K_P = (h, \pi_0, F, A_1, B_2)$ if we allow to use the fast evaluation algorithm. The private key are the maps L_1, G_1, L_2, G_2, L_3 defined by the matrices A_{1i}, A_{2j}, A_{3j}, the exponent matrices A_1 and B_2 and the mixing map M. The $\mathbb{F}_q$-linear isomorphisms $\pi_1 : \mathbb{F}_q^n \to \mathbb{F}_{q^n}$ and $\pi_2 : \mathbb{F}_q^m \to \mathbb{F}_{q^m}$ are not needed for encryption and can be chosen once for all users of the system or individually for its user and form part of the private key.

The exponent matrices A_1 and B_2 can be deduced from the exponents of the monomials in F_i so there is no need to hide them and can be made public in order to use them for the fast method to evaluate the polinomials of the public key.

Digital signature with the system

The system can be used to sign a message in $(\mathbb{F}_q^m \setminus \{0\})^n$ by computing $F^{-1}(z)$. As F is not surjective onto $(\mathbb{F}_q^m \setminus \{0\})^n$ there are messages that can not be signed. One need to add some randoness to the message. Given $z \in (\mathbb{F}_q^m \setminus \{0\})^n$ there exists $x \in F^{-1}(z)$ if $(L_3 \circ F \circ L_2)^{-1}(z) \in (\mathbb{F}_q^n \setminus \{0\})^m$, so the probability for $z \notin \mathrm{Im}(F)$ is of order $\frac{1}{q^n}$.

One can sign a message v in $(F_p)^{N_1}$, $N_1 < en$, by padding it in a similar way that we do for encript a message. We need to coose a map $h_1 : \{1, \ldots, N_1\} \to \{1, \ldots, e\cdot n\cdot m\}$ and fill the entries not in $\mathrm{Im}(h_1)$. There is a difference with the

encrpytation it is the fact that N_1 need not to be fixed a priori. The signature of a message $z_0 \in (\mathbb{F}_p)^{N_1}$ is $\mathrm{sig}(z_0) = (x, z_0, h_1)$ such that there exist $x = F^(-1)(z)$. If it does not exist we padd again z_0 to get a different z. For the verification of the signature one computes $F(x) = z$ and trows away the random digits to get z_0.

If given two parties A and B, A want to send an encrypted message x to B, A encrypt x with the public key of B obtainig $z \in (\mathbb{F}_q)^{nm}$ that can not be padded because $N_1 = e \cdot n \cdot m$. If is not possible to get the signature $y = (F_A)^{-1}(z)$ one can encrypt x again (because the system is not deterministic) up to get a message that can be signed.

The system can be used for KEM in a standard way but for KEM there is no need to use the padding. If two parties want to share a key for a symmetric system like AES they pick up a hash fonction and one of them A choose a random $x \in (\mathbb{F}_q)^{nm}$ with $x_i \neq 0$ and send $z = F_B(x)$ to B who decrypt z and both parties compute the common hash.

The setting of the system DME that is implemented in the proposal:

We take $m = 3$, $n = 2$, $s = t = 2$ and $q = 2^e$. The number of monomials of each component is $(2 \cdot n^s)^t = 64$. The polynomial map of the public key is $F = (F_1, \ldots, F_6) : (\mathbb{F}_2^e)^6 \to (\mathbb{F}_2^e)^6$ where F_1, F_2, F_3 share 64 monomials and F_4, F_5, F_6 share other 64 monomials. For 128 bit security we propose $q = 2^{24}$ that is the message space is $(\mathbb{F}_2)^{144}$. We will justify this choice when we discuss the security in the corresponding paragraph. For the padding we can add from 3 to 16 bits. For instance if we add only 3 bits one $'1'$ in each coordinate x_{12}, x_{22} and x_{32}; one gets a deterministic public key system. We choose to add 12 random bits, 4 bits in each coordinate so the encryption map are $DM : (\mathbb{F}_2)^{132} \to (\mathbb{F}_2)^{144}$

$$\mathbb{F}_2^{132} \xrightarrow{H} \mathbb{F}_{2^{24}}^6 \xrightarrow{F} \mathbb{F}_{2^{24}}^6$$

for 128 bit.

For 256 bit security we propose $q = 2^{48}$ that is the message space is $(\mathbb{F}_2)^{288}$ with 24 random bits that is the encryption map is $DM : (\mathbb{F}_2)^{132} \to (\mathbb{F}_2)^{144}$.

$$\mathbb{F}_2^{264} \xrightarrow{H} \mathbb{F}_{2^{48}}^6 \xrightarrow{F} \mathbb{F}_{2^{48}}^6$$

References

[1] I. Luengo .

Departamento de Álgebra, Universidad Complutense, Plaza de las Ciencias s/n, Ciudad Universitaria, 28040 Madrid, SPAIN

E-mail address: iluengo@ucm.es