Multivariate Cryptography
Part 1: Basics

Albrecht Petzoldt

PQCrypto Summer School 2017
Eindhoven, Netherlands
Tuesday, 20.06.2017

Multivariate Cryptography [DS06]

MPKC: Multivariate Public Key Cryptosystem
Public Key: System of nonlinear multivariate polynomials

$$
\begin{aligned}
p^{(1)}(x_1,\ldots,x_n) &= \sum_{i=1}^{n}\sum_{j=i}^{n} p_{ij}^{(1)}\cdot x_i x_j \quad + \quad \sum_{i=1}^{n} p_i^{(1)}\cdot x_i + p_0^{(1)} \\
p^{(2)}(x_1,\ldots,x_n) &= \sum_{i=1}^{n}\sum_{j=i}^{n} p_{ij}^{(2)}\cdot x_i x_j \quad + \quad \sum_{i=1}^{n} p_i^{(2)}\cdot x_i + p_0^{(2)} \\
&\vdots \\
p^{(m)}(x_1,\ldots,x_n) &= \sum_{i=1}^{n}\sum_{j=i}^{n} p_{ij}^{(m)}\cdot x_i x_j \quad + \quad \sum_{i=1}^{n} p_i^{(m)}\cdot x_i + p_0^{(m)}
\end{aligned}
$$

$d :=$ degree of the polynomials in the system
$m :=$ # equations
$n :=$ # variables

Public Key Size

$$\text{size}_{\text{ public key}} = m \cdot T \text{ field elements}$$

with $T = \#$ monomials of degree $\leq d$.

$$\#\text{ monomials of degree } d = \binom{n+d-1}{d}$$

$$\#\text{ monomials of degree} \leq d = \binom{n+d}{d}$$

$$\Rightarrow \text{size}_{\text{ public key}} = m \cdot \binom{n+d}{d} \overset{m \approx n}{\sim} O(n^{d+1})$$

$\Rightarrow$ For $d \geq 2$ the public key size gets very big

$\Rightarrow$ Most MPKCs use for efficiency reasons $d = 2$.

Security

The security of multivariate schemes is based on the

Problem MQ: Given m multivariate quadratic polynomials $p^{(1)}(\mathbf{x}), \ldots, p^{(m)}(\mathbf{x})$, find a vector $\bar{\mathbf{x}} = (\bar{x}_1, \ldots, \bar{x}_n)$ such that $p^{(1)}(\bar{\mathbf{x}}) = \ldots = p^{(m)}(\bar{\mathbf{x}}) = 0$.

- proven to be NP hard [GJ78]
- believed to be hard on average (both for classical and quantum conputers) [BB08]
- also known as the PoSSo Problem (especially for $d > 2$)

However: no direct reduction

Construction

- Easily invertible quadratic map $\mathcal{F} : \mathbb{F}^n \to \mathbb{F}^m$
- Two invertible linear maps $\mathcal{S} : \mathbb{F}^m \to \mathbb{F}^m$ and $\mathcal{T} : \mathbb{F}^n \to \mathbb{F}^n$
- *Public key*: $\mathcal{P} = \mathcal{S} \circ \mathcal{F} \circ \mathcal{T}$ supposed to look like a random system
- *Private key*: $\mathcal{S}$, $\mathcal{F}$, $\mathcal{T}$ allows to invert the public key

Isomorphism of Polynomials

Definition

Two polynomial systems $\mathcal{G} : \mathbb{F}^n \to \mathbb{F}^m$ and $\mathcal{H} : \mathbb{F}^n \to \mathbb{F}^m$ are called isomorphic

$$\Leftrightarrow \exists \text{linear (affine) maps } \mathcal{L}_1 \text{ and } \mathcal{L}_2 \text{ s.t. } \mathcal{H} = \mathcal{L}_1 \circ \mathcal{G} \circ \mathcal{L}_2.$$

$\Rightarrow$ The central map $\mathcal{F}$ and the public key $\mathcal{P}$ of an MPKC are isomorphic.

Isomorphism of Polynomials (2)

Due to their construction, the security of MPKCs is also based on the

Problem EIP (Extended Isomorphism of Polynomials): Given the public key $\mathcal{P}$ of a multivariate public key cryptosystem, find affine maps $\bar{\mathcal{S}}$ and $\bar{\mathcal{T}}$ as well as an easily invertible quadratic map $\bar{\mathcal{F}}$ such that $\mathcal{P} = \bar{\mathcal{S}} \circ \bar{\mathcal{F}} \circ \bar{\mathcal{T}}$.

$\Rightarrow$ Hardness of the problem depends heavily on the structure of the central map
$\Rightarrow$ In general, not much is known about the complexity
$\Rightarrow$ Security analysis of multivariate schemes is a hard task

Encryption Schemes ($m \geq n$)

Encryption: Given message $\mathbf{z} \in \mathbb{F}^n$, compute the ciphertext $\mathbf{w} \in \mathbb{F}^m$ by $\mathbf{w} = \mathcal{P}(\mathbf{z})$.

Decryption: Given ciphertext $\mathbf{w} \in \mathbb{F}^m$, compute recursively $\mathbf{x} = \mathcal{S}^{-1}(\mathbf{w}) \in \mathbb{F}^m$, $\mathbf{y} = \mathcal{F}^{-1}(\mathbf{x}) \in \mathbb{F}^n$ and $\mathbf{z} = \mathcal{T}^{-1}(\mathbf{y})$.
The condition $(m \geq n)$ guarantees that $\mathcal{F}$ is more or less injective, i.e. we do not get too many possible plaintexts.

Important Schemes

- PMI+, IPHFE+
- ZHFE ($\rightarrow$ this conference)
- Simple Matrix ($\rightarrow$ this conference)

Signature Schemes ($m \leq n$)

Signature Generation: Given message d, use a hash function $\mathcal{H} : \{0,1\}^\star \to \mathbb{F}^m$ to compute $\mathbf{w} = \mathcal{H}(d) \in \mathbb{F}^m$. Compute recursively $\mathbf{x} = \mathcal{S}^{-1}(\mathbf{w}) \in \mathbb{F}^m$, $\mathbf{y} = \mathcal{F}^{-1}(\mathbf{x}) \in \mathbb{F}^n$ and $\mathbf{z} = \mathcal{T}^{-1}(\mathbf{y})$. The signature of the message d is $\mathbf{z} \in \mathbb{F}^n$.
The condition $(m \leq n)$ is needed for the surjectivity of the map $\mathcal{F}$, i.e. every message has a signature.

Signature Verification: To check the authenticity of a signature $\mathbf{z} \in \mathbb{F}^n$ for a message d, compute $\mathbf{w} \in \mathcal{H}(d) \in \mathbb{F}^m$ and $\mathbf{w}' = \mathcal{P}(\mathbf{z}) \in \mathbb{F}^m$. If $\mathbf{w}' = \mathbf{w}$ holds, the signature is accepted, otherwise rejected.

Important Schemes

- UOV, Rainbow
- HFEv-, Gui
- MQDSS
- pFLASH ($\to$ this conference), TTS

Signature Schemes ($m \leq n$)

Signature Generation: Given message d, use a hash function $\mathcal{H} : \{0,1\}^\star \to \mathbb{F}^m$ to compute $\mathbf{w} = \mathcal{H}(d) \in \mathbb{F}^m$. Compute recursively $\mathbf{x} = \mathcal{S}^{-1}(\mathbf{w}) \in \mathbb{F}^m$, $\mathbf{y} = \mathcal{F}^{-1}(\mathbf{x}) \in \mathbb{F}^n$ and $\mathbf{z} = \mathcal{T}^{-1}(\mathbf{y})$. The signature of the message d is $\mathbf{z} \in \mathbb{F}^n$.
The condition $(m \leq n)$ is needed for the surjectivity of the map $\mathcal{F}$, i.e. every message has a signature.

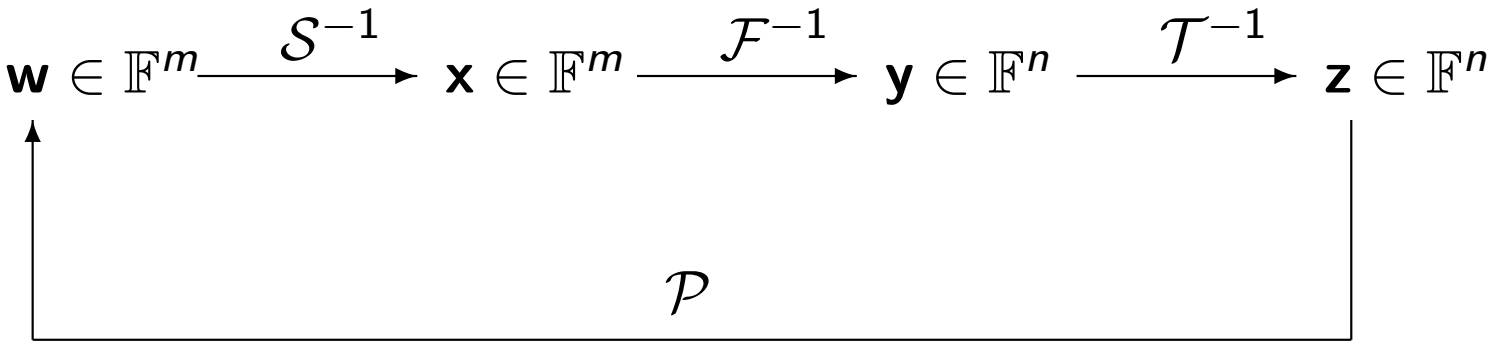
Signature Verification: To check the authenticity of a signature $\mathbf{z} \in \mathbb{F}^n$ for a message d, compute $\mathbf{w} \in \mathcal{H}(d) \in \mathbb{F}^m$ and $\mathbf{w}' = \mathcal{P}(\mathbf{z}) \in \mathbb{F}^m$. If $\mathbf{w}' = \mathbf{w}$ holds, the signature is accepted, otherwise rejected.

Important Schemes

- UOV, Rainbow
- HFEv-, Gui
- MQDSS
- pFLASH ($\to$ this conference), TTS

Workflow

Decryption / Signature Generation



Encryption / Signature Verification

Attacks

Direct Attacks: Try to solve the public equation $\mathcal{P}(\mathbf{z}) = \mathbf{w}$ as an instance of the MQ-Problem

all algorithms have exponential running time (for $m \approx n$)

XL -Algorithm

Given: nonlinear polynomials $f_1, \ldots, f_m$

1. **eXtend** multiply each polynomial $f_1, \ldots, f_m$ by every monomial of degree $\leq D$
2. **Linear Algebra Step**: Apply Gaussian Elimination on the extended system to generate a univariate polynomial p
3. **Solve**: Use Berlekamps algorithm to solve the polynomial p.
4. **Repeat**: Substitute the solution of p into the system and continue with the simplified system.

many variations, e.g. FXL, MutantXL

$$\text{Complexity} = 3 \cdot \binom{n + d_{\text{reg}}}{d_{\text{reg}}}^2 \cdot \binom{n}{2}$$

Gröbner Bases Algorithms

- find a "nice" basis of the ideal $\langle f_1, \ldots, f_m \rangle$
- first studied by B. Buchberger
- later improved by Faugère et al. (F_4, F_5) [Fa99]
- currently fastest algorithms to solve random systems (Hybrid F_5 [BFP09])

$$\text{Complexity}(q, m, n) = \min_k \; q^k \cdot O\left(m \cdot \binom{n - k + d_{\text{reg}} - 1}{d_{\text{reg}}}^{\omega} \right)$$

with $2 < \omega \leq 3$.

Complexity of Direct Attacks

How many equations are needed to meet given levels of security?

security level (bit)	number of equations		
	GF(16)	GF(31)	GF(256)
80	30	28	26
100	39	36	33
128	51	48	43
192	80	75	68
256	110	103	93

Remark

Every cryptosystem can be represented as a set of nonlinear multivariate equations

$\Rightarrow$ Direct attacks are used in the cryptanalysis of many cryptographic schemes (in particular block and stream ciphers)

$\Rightarrow$ The MQ (or PoSSo) Problem can be seen as one of the central problems in cryptography

Structural Attacks

Try to decompose the public key $\mathcal{P}$ into $\mathcal{P} = \mathcal{S} \circ \mathcal{F} \circ \mathcal{T}$ by using the known structure of the central map $\mathcal{F}$

MinRank attack [CSV94]: For many multivariate schemes (certain) central equations have low rank

$\Rightarrow$ look for a linear combination of the public key polynomials of low rank

$\Rightarrow$ this linear combination corresponds to a central equation

$\Rightarrow$ this linear combination yields (parts) of an equivalent affine map $\mathcal{S}$

$\Rightarrow$ further analysis: recover equivalent maps $\mathcal{S}$, $\mathcal{F}$ and $\mathcal{T}$

MinRank Attack

Problem MinRank: Given m $n \times n$ matrices $G_1, \ldots, G_m$, find a linear combination

$$H = \sum_{i=1}^{m} \lambda_i G_i$$

such that $\mathrm{Rank}(H) \leq r$.

$$\mathrm{Complexity}(\mathrm{MinorsModelling}) = O\binom{n+r}{r}^{\omega}$$

with $2 < \omega \leq 3$.

Other Attacks

- **HighRank Attack**: Try to recover the linear transformation of the variables appearing the lowest time in the central equations. This yields information about the affine transformation $\mathcal{T}$ and therefore the private key.
- **Differential Attacks**: Look for invariants or symmetries of the differential

$$\mathcal{G}(x, y) = \mathcal{P}(x + y) - \mathcal{P}(x) - \mathcal{P}(y) + \mathcal{P}(0)$$

These symmetries yield information about the private key.

Advantages

- resistant against attacks with quantum computers
- very fast (much faster than RSA)
- only simple arithmetic operations required
  $\Rightarrow$ can be implemented on low cost devices
  $\Rightarrow$ suitable for security solutions for the IoT
- many practical signature schemes (UOV, Rainbow, HFEv-, ...)
- very short signatures (e.g. 120 bit signatures for 80 bit security)

Disadvantages

- large key sizes (public key size $\sim 10 - 100$ kB)
- no security proofs
  But: Practical Security (attack complexities) follows closely theoretical estimations
- mainly restricted to digital signatures (and public key encryption)

Lessons Learned

Multivariate Cryptography

- deals with systems of nonlinear (usually quadratic) multivariate polynomials
- one of the main candidates for post-quantum cryptosystems
- very efficient signature schemes (e.g. Rainbow, HFEv-) with short signatures
- not so good for encryption schemes
- large public key sizes, no security proofs
- But: Theoretical Security estimates match very well with experimental data

References

BB08 D.J. Bernstein, J. Buchmann, E. Dahmen (eds.): Post Quantum Cryptography. Springer, 2009.

DG06 J. Ding, J. E. Gower, D. S. Schmidt: Multivariate Public Key Cryptosystems. Springer, 2006.

GJ79 M. R. Garey and D. S. Johnson: Computers and Intractability: A Guide to the Theory of NP-Completeness.

References (2)

BF09 L. Bettale, L.C Faugère, L. Perret: Hybrid approach for solving multivariate systems over finite fields. Journal of Mathematical Cryptology 3, pp. 177-197 (2009).

Fa99 J.C. Faugère: A new efficient algorithm for computing Gröbner bases (F4). Journal of Pure and Applied Algebra 139, pp. 61-88 (1999).

CS94 D. Coppersmith, J. Stern, S. Vaudenay: Attacks on the Birational Signature Scheme. CRYPTO 1994, LNCS vol. 773, pp. 435 - 443. Springer, 1994.