Multivariate Quadratic Cryptography

Ward Beullens

August 8, 2022

Excercise session 1

Exercise 1. Are multivariate quadratic maps collision resitant? I.e., given a random quadratic map $\mathcal{P} : \mathbb{F}_q^n \to \mathbb{F}_q^n$, is it hard to find $\mathbf{x}, \mathbf{x}'$ such that $\mathbf{x} \neq \mathbf{x}'$ and $\mathcal{P}(\mathbf{x}) = \mathcal{P}(\mathbf{x}')$?

Hint: Suppose there is a collision x, x' and you are given $\Delta = x - x'$, can you find x, x' more easily?

Definition 1 (Macaulay matrix)**.** Let $p_1, \ldots, p_m \in \mathbb{F}_q[x_1, \ldots, x_n]$ be a sequence of multivariate quadratic polynomials. We say the Macaulay matrix of $p_1, \ldots, p_m$ at degree D is the matrix whose $\binom{n+D}{D}$ collumns correspond to monomials of degree at most D in the variables $x_1, \ldots, x_n$, and whose $m\binom{n+D-2}{D-2}$ rows correspond to the polynomials of the form Mp_i, where M is a monomial of degree at most $D-2$ and $i \in \{1, \ldots, m\}$.

Exercise 2. Suppose $p_1(x) = \cdots = p_m(x) = 0$ is a system of quadratic polynomials with a solution $x' \in \mathbb{F}_q^n$. Prove that the Macaulay matrix of $p_1, \ldots, p_m$ has a vector in its right kernel.

Exercise 3 (Rank of Macaulay matrices of random quadratic polynomials)**.** Let $p_1, \ldots, p_m \in \mathbb{F}_q[x_1, \ldots, x_n]$ be a sequence of non-zero multivariate quadratic polynomials. Let $[p_1, \ldots, p_k]_{\leq d}$ be the vectorspace spanned by all the polynomials of the form $x^\alpha p_i$, where x^α is a monomial of degree at most $d-2$, and where $1 \leq i \leq k$. That is, $[p_1, \ldots, p_k]_{\leq d}$ corresponds to the span of the rows of the Macaulay matrix of $p_1, \ldots, p_k$ at degree D.

Clearly, we have $[p_1, \ldots, p_k]_{\leq d-2} \cdot p_{k+1} \subset [p_1, \ldots, p_k]_{\leq d} \cap [p_{k+1}]_{\leq d}$. Suppose that this is an equality for all $k \in \{0, \ldots, m-1\}$ and all d, such that $[p_1, \ldots, p_m]_{\leq d} \neq \mathbb{F}_q[x_1, \ldots, x_n]_{\leq d}$. (Random systems satisfy this property with high probability.)

- Prove that $\dim(\mathbb{F}_q[x_1, \ldots, x_n]_{\leq d})$ is equal to the coefficient of t^d in the power series expansion of
$$\frac{1}{(1-t)^{n+1}}.$$
- Prove that $\dim([p_1, \ldots, p_m]_{\leq d})$ is equal to the coefficient of t^d in the power series expansion of
$$\frac{1-(1-t^2)^m}{(1-t)^{n+1}},$$
for all d such that $[p_1, \ldots, p_m]_{\leq d} \neq \mathbb{F}_q[x_1, \ldots, x_n]_{\leq d}$
- Conclude that the Macaulay matrix of $p_1, \ldots, p_m$ at degree D has full rank if there exists $d \leq D$ such that the coefficient of t^d in the power series expansion of
$$\frac{(1-t^2)^m}{(1-t)^{n+1}}$$
has a non-positive coefficient.

XL algorithm. If $p_1(x) = \cdots = p_m(x) = 0$ is a random system with a solution, then heuristically, the ranks of Macaulay matrices of this system are the same as those in Exercise 3, except that when the Macaulay matrix from Exercise 3 has full rank, the Macaulay matrix of a system with a solution has corank 1 instead. The XL algorithm works by constructing the Macaulay matrix at a degree D that is high enough such that the Macaulay matrix has a kernel of rank 1. Then the algorithm does linear algebra to find the vector from Exercise 2, from which the solution x can be recovered easily.

A naive implementation of Gaussian Elimination would require $O(\binom{n+D}{D}^3)$ multiplications. But the Macaulay matrix is very sparse (each row has at most $\binom{n+2}{2}$ non-zero entries), so with sparse linear algebra methods the kernel vector can be found with roughly
$$3\binom{n+2}{2}\binom{n+D}{D}^2 \tag{1}$$

multiplications instead.

It is often beneficial to guess the values of a few variables before applying the XL algorithm. This reduces the number of variables, which often allows the algorithm to run at a lower degree D, which makes it much more efficient. The drawback is that if you make k guesses, the algorithm needs to be repeated roughly q^k times, so guessing k variables is beneficial if the cost of the XL algorithm is reduced by more than a factor q^k. This variant of the XL algorithm is often called HybridXL, because it is a hybrid between XL ($k = 0$) and exhaustive search ($k = n$).

Exercise 4 (Estimate the cost of solving the MQ problem)**.** We estimate the cost of solving some multivariate quadratic systems, to illustrate the fact that finding a solution becomes much easier if more equations are given. Use Exercise 3 to find D, and use formula (1) for the cost of the linear algebra.

- Let $\mathcal{P} : \mathbb{F}_q^n \to \mathbb{F}_q^m$, be a random quadratic map with $n = 40$ and $m = 80$, and $q = 256$. Give an estimate of the cost (number of field multiplications) of the XL algorithm to find $\mathbf{x}$, given $\mathcal{P}(\mathbf{x})$.
- Let $\mathcal{P} : \mathbb{F}_q^n \to \mathbb{F}_q^m$, be a random quadratic map with $n = 40$ and $m = 40$, and $q = 256$. Find the optimal number of guesses for the HybridXL algorithm, and estimate the cost of running the algorithm.

You might want to use a computer algebra system for your calculations.

Answer: Solving the first system takes 2^{68} multiplications, the operating degree is $D = 8$. Solving the second algorithm takes 2^{129} multiplications for $k = 3$ guesses and $D = 18$.

Excercise session 2: Breaking a simplified version of the Matsumoto-Imai scheme.

Let $K = \mathbb{F}_q$ be a finite field of order q, and let L be a field extension of degree n. Let θ be an integer such that $\gcd(1 + q^\theta, q^n - 1) = 1$.

Exercise 5. Consider the exponentiation map $E_\theta : L \to L : x \mapsto x^{q^\theta+1}$. Prove that E_θ is a bijection. Give a polynomial-time algorithm that given θ and $y \in L$, outputs $E_\theta^{-1}(y) \in L$.

Exercise 6. Let $T : L \to K^n$ and $S : K^n \to L$ be invertible K-linear maps (L is a K-vector space of dimension n). Prove that $F = T \circ E_\phi \circ S$ is a multivariate quadratic map.

In 1988, Matsumoto and Imai [8] proposed a variant of the following public-key cryptosystem: Fix public parameters q, n, θ. The private key consists of two randomly chosen invertible linear maps $T : L \to K^n$ and $S : K^n \to L$, the public key is the multivariate map $P : K^n \to K^n = T \circ E_\theta \circ S$. To encrypt a message $m \in K^n$, a user just evaluates $P(m)$, which he can send over the wire. Given, T and S, one can efficienly decrypt the ciphertext $P(m) = T \circ E_\theta \circ S(m)$ by first undoing T, then undoing E_θ, and finally undoing S.

Exercise 7. Show that the Matsumoto-Imai scheme is not secure with the parameters $q = 256, n = 41, \theta = 1$. That is, give an efficient algorithm that given a public key $P : K^n \to K^n$, and a ciphertext $c = P(m) \in K^n$ outputs the message $m \in K^n$.

Hint 1: We saw that the relation $y = x^{q^\theta+1}$ (over L) becomes quadratic when viewed over K, wouldn't it be nice if this implied some other equation that becomes bi-linear in the coefficients of x and y instead?

Hint 2: Raise both sides of the equation to the power $q^\theta - 1$ and multiply both sides by xy.

Hint 3: If you know that input-output pairs of the cryptosystem satisfy some polynomial equations with (not too many) unknown coefficients, you can just evaluate P on a lot of inputs, and solve for the coefficients.

Exercise 8. Implement your attack in SAGE. Download a public key and ciphertext and a SAGE file to get you started, and recover the message.

Some solutions

Exercise 1. Random multivariate quadratic maps $\mathcal{P} : \mathbb{F}^n \to \mathbb{F}^n$ are not collision resistant! We define the differential $\mathcal{P}'(\mathbf{x}, \Delta) := \mathcal{P}(\mathbf{x} + \Delta) - \mathcal{P}(\mathbf{x}) - \mathcal{P}(\Delta) + \mathcal{P}(0)$. Observe that this is bi-linear in $\mathbf{x}$ and Δ. If you fix a random

$\Delta \in \mathbb{F}^n$, you can solve a linear system to find x such that $\mathcal{P}(\mathbf{x}) = \mathcal{P}(\mathbf{x}+\Delta)$, because

$$\mathcal{P}(\mathbf{x}) - \mathcal{P}(\mathbf{x}+\Delta) = \cancel{\mathcal{P}(\mathbf{x})} - \mathcal{P}'(\mathbf{x},\Delta) - \cancel{\mathcal{P}(\mathbf{x})} - \mathcal{P}(\Delta) + \mathcal{P}(0) = 0\,,$$

is linear in $\mathbf{x}$. For each choice of Δ we get a random system of n linear equations in n variables, so it has a solution with large probability. If the system doesn't have a solution, try again with a different choice of Δ.

Exercise 3.

- The dimension of $\mathbb{F}_q[x_1,\ldots,x_n]_{\le d}$ is the number of monomials of degree at most d, because these monomials form a basis. The number of monomials is $\binom{n+d}{d}$, which has generating function $(1-t)^{-n-1}$. (See https://en.wikipedia.org/wiki/Stars_and_bars.)

- Proof by induction on m.

  Base case $m = 1$: The power series evaluates to $\frac{t^2}{(1-t)^{n+1}}$, and indeed the vectorspace $[p_1]_{\le d}$ is generated by all the polynomials $M \cdot p_1$. There are $\binom{n+d-2}{d-2}$ of these polynomials, and they are all linearly independent. The generating function of $\binom{n+d-2}{d-2}$ is $\frac{t^2}{(1-t)^{n+1}}$.

  Induction case: Suppose the statement is true for all m' less than $m+1$. For general subspaces A, B we have $\dim(A+B) = \dim(A) + \dim(B) - \dim(A \cap B)$. We apply this to $A = [p_1,\ldots,p_{m-1}]_{\le d}$ and $B = [p_m]_{\le d}$. We get

  $$\dim([p_1,\ldots,p_m]_{\le d}) = \dim([p_1,\ldots,p_{m-1}]_{\le d}) + \dim([p_m]_{\le d}) - \dim([p_1,\ldots,p_{m-1}]_{\le d} \cap [p_m]_{\le d})\,.$$

  Multiplication by p_m is injective, so $\dim([p_1,\ldots,p_{m-1}]_{\le d-2} \cdot p_m) = \dim([p_1,\ldots,p_{m-1}]_{\le d-2})$. Using our assumption on the intersection we get

  $$\dim([p_1,\ldots,p_m]_{\le d}) = \dim([p_1,\ldots,p_{m-1}]_{\le d}) + \dim([p_m]_{\le d}) - \dim([p_1,\ldots,p_{m-1}]_{\le d-2})\,.$$

  Using the induction hypothesis for $m' = 1$ and $m' = m-1$, this is equal to the coefficient of t^d in the power series expansion of

  $$\frac{1}{(1-t)^{n+1}}\left[1 - (1-t^2)^{m-1} + 1 - (1-t^2) - t^2(1-(1-t^2)^{m-1})\right] = \frac{1-(1-t^2)^m}{(1-t)^{n+1}}$$

- The rows of the Macaulay matrices at degree D correspond to the generators of $[p_1, \ldots, p_m]_{\leq D}$, so rank of the Macaulay matrix is $\dim([p_1, \ldots, p_m]_{\leq D})$ equals the number of collumns of the Macaulay matrix $\binom{n+D}{D}$. The power series is valid as long as $[p_1, \ldots, p_m]_{\leq d} \neq \mathbb{F}_q[x_1, \ldots, x_n]_{\leq d}$, i.e., as long as the dimension of $[p_1, \ldots, p_m]_{\leq D}$ is less than $\binom{n+D}{D}$, which is as long as the coefficient of t^D in

$$\frac{1}{(1-t)^{n+1}} - \frac{1-(1-t^2)^m}{(1-t)^{n+1}} = \frac{(1-t^2)^m}{(1-t)^{n+1}}$$

is positive. If the coefficient of some t^d is non-positive we must have $[p_1, \ldots, p_m]_{\leq d} = \mathbb{F}_q[x_1, \ldots, x_n]_{\leq d}$, so the Macaulay matrix is full rank at degree d, and all degrees higher than d.

Solution of session 2. We have the equation $xy^{q^\theta} - yx^{q^{2\theta}} = 0$ which is bi-linear (over K) in the n coefficients of $x = \sum x_i t^i \in K[t]/f(t)$ and $y = \sum_i y_i t^i$. Moreover, the coefficients of x are linear in the message m, and the coefficients y is linear in $\mathcal{P}(m)$. So there are some bi-linear equations in m and $\mathcal{P}(m)$. These equations are of the form

$$\sum_{i,j} \alpha_{i,j} m_i \mathcal{P}(m)_j \,.$$

The coefficients $\alpha_{i,j}$ depend on S and T, so they are not known to us as attackers. But we can evaluate $\mathcal{P}$ at many inputs to get many $(m, \mathcal{P}(m))$ pairs. We can plug those in the above equation and solve for the $\alpha_{i,j}$. It turns out there is a $n-1$ dimensional solutions space of $\alpha_{i,j}$, so we obtain $n-1$ linearly independent bilinear equations. To decrypt a ciphertext $c = \mathcal{P}(m')$, we just plug c into the bilinear equations, and solve for m'. There is a one dimensional space of solutions, consisting of m' and all the multiples of m'. Since q is small we can check which of the q multiples is the correct message by brute force.

You can download a SAGE implementation of this attack here.

Further reading:

Algorithms for solving systems of multivariate equations:

- Polynomial-time algorithm for solving a system witn $n \geq m(m+1)$ variables in m equations in fields of characteristic 2. [7] (section 7.)
- Algorithm that reduces solving a multivariate quadratic system with $n = \omega m$ variables in m equations to solving a system of $m + 1 - \lfloor \omega \rfloor$ equations and variables. [9]
- Fast exhaustive search. ($O(\log(n)q^n)$ instead of naive exhaustive search which has complexity $O(mn^2q^n)$) [3]
- Paper discussing an optimized implementation of XL with sparse linear algebra methods. [5]
- Algorithm for solving systems over $\mathbb{F}_2$ based on the polynomial method. [6]

Multivariate quadratic signature schemes:

- The Oil and Vinegar algorithm. [7]
- The Rainbow signature scheme and how to break it. [2]
- MQDSS (an MQ signature without trapdoors). [4]
- MAYO: a relatively new MQ signature with very small keys. (Try to break it!) [1]

References

[1] Ward Beullens. MAYO: Practical post-quantum signatures from oil-and-vinegar maps. Cryptology ePrint Archive, Report 2021/1144, 2021. https://eprint.iacr.org/2021/1144.

[2] Ward Beullens. Breaking rainbow takes a weekend on a laptop. Cryptology ePrint Archive, Paper 2022/214, 2022. https://eprint.iacr.org/2022/214.

[3] Charles Bouillaguet, Hsieh-Chung Chen, Chen-Mou Cheng, Tung Chou, Ruben Niederhagen, Adi Shamir, and Bo-Yin Yang. Fast exhaustive

search for polynomial systems in $\mathbb{F}_2$. In Stefan Mangard and François-Xavier Standaert, editors, *CHES 2010*, volume 6225 of *LNCS*, pages 203–218, Santa Barbara, CA, USA, August 17–20, 2010. Springer, Heidelberg, Germany.

[4] Ming-Shing Chen, Andreas Hülsing, Joost Rijneveld, Simona Samardjiska, and Peter Schwabe. From 5-pass MQ-based identification to MQ-based signatures. In Jung Hee Cheon and Tsuyoshi Takagi, editors, *ASIACRYPT 2016, Part II*, volume 10032 of *LNCS*, pages 135–165, Hanoi, Vietnam, December 4–8, 2016. Springer, Heidelberg, Germany.

[5] Chen-Mou Cheng, Tung Chou, Ruben Niederhagen, and Bo-Yin Yang. Solving quadratic equations with XL on parallel architectures. In Emmanuel Prouff and Patrick Schaumont, editors, *CHES 2012*, volume 7428 of *LNCS*, pages 356–373, Leuven, Belgium, September 9–12, 2012. Springer, Heidelberg, Germany.

[6] Itai Dinur. Cryptanalytic applications of the polynomial method for solving multivariate equation systems over GF(2). In Anne Canteaut and François-Xavier Standaert, editors, *EUROCRYPT 2021, Part I*, volume 12696 of *LNCS*, pages 374–403, Zagreb, Croatia, October 17–21, 2021. Springer, Heidelberg, Germany.

[7] Aviad Kipnis, Jacques Patarin, and Louis Goubin. Unbalanced oil and vinegar signature schemes. In Jacques Stern, editor, *EUROCRYPT'99*, volume 1592 of *LNCS*, pages 206–222, Prague, Czech Republic, May 2–6, 1999. Springer, Heidelberg, Germany.

[8] Tsutomu Matsumoto and Hideki Imai. Public quadratic polynominal-tuples for efficient signature-verification and message-encryption. In C. G. Günther, editor, *EUROCRYPT'88*, volume 330 of *LNCS*, pages 419–453, Davos, Switzerland, May 25–27, 1988. Springer, Heidelberg, Germany.

[9] Enrico Thomae and Christopher Wolf. Solving underdetermined systems of multivariate quadratic equations revisited. In Marc Fischlin, Johannes Buchmann, and Mark Manulis, editors, *PKC 2012*, volume 7293 of *LNCS*, pages 156–171, Darmstadt, Germany, May 21–23, 2012. Springer, Heidelberg, Germany.