Cryptanalysis of the Huang–Liu–Yang Cryptosystem from PKC 2012

Yosuke Todo and Keita Xagawa

NTT Secure Platform Laboratories
{todo.yosuke, xagawa.keita}@lab.ntt.co.jp

Abstract. This short note describes a key-recovery attack against a multi-variate quadratic cryptosystem proposed by Huang, Liu, and Yang (PKC 2012). Our attack is running lattice-basis reduction algorithms on a lattice constructed from the keys in the cryptosystem. The attack takes less than 20 minutes for the proposed parameter sets which are expected to be 80-bit and 128-bit security.

1 Introduction

Provably-secure MQ-based cryptography: Post-quantum cryptosystems have been studied extensively in this decade. Cryptosystem based on multivariate quadratic-polynomials (MQ, in short) is a candidate for quantumly-secure cryptosystem.

Recently there are a few MQ-based scheme with *provable security*: In 2006, Berbain, Gilbert, and Patarin proposed a stream cipher, whose name is QUAD, based on the hardness of the random MQ problem [4]. In 2011, Sakumoto, Shirai, and Hiwatari proposed an identification scheme based on the hardness of the random MQ problem [9].

Albrecht, Farshim, Faugère, and Perret [1] revisited the "Polly Cracker" scheme and showed its security from reasonable assumptions. Herold [6] pointed out some assumptions are insecure and the others are at most as secure as Regev's scheme [8], which is based on the LWE assumption. See [2] for the details of the assumptions themselves and the relations between them.

The HLY12 cryptosystem: Along this line, Huang, Liu, and Yang [7] proposed public-key encryption schemes based on the hardness of the MQ problem with a certain distribution. We call a basic public-key encryption scheme as the HLY12 cryptosystem in short. They put forward a new assumption on the MQ problem. Informally speaking, the assumption states that it is infeasible to solve the MQ system for a specified distribution instead of a uniform distribution. They gave two parameter sets (Case 1 and Case 2) which are expected to be 80-bit and 128-bit security. See Table 1 for the details of the parameter sets.

Our Contribution. We here examine the security of the MQ-based cryptosystem proposed by Huang, Liu, and Yang [7] and show that the security of the scheme is far from that expected. Although Huang et al. treated their problem as a variant of the LWE problem, their security estimation ignored attacks employing lattice algorithms. Therefore, we examine the security of the HLY12 cryptosystem from the view of lattice problems.

We convert a key-recovery problem into the well-known shortest-vector problem. Exploiting the extreme shortness of the secret key, our attack successfully find the secret key in less than 5 minutes for Case 1 and 30 minutes for Case 2.

Independent Work: Very recently, Albrecht, Faugère, Fitzpatrick, and Perret [3] cryptanalyzed of the HLY12 cryptosystems. They also mount their key-recovery attack in 1 day for Case 1 and 3 days for Case 2. We note that our attack is faster than theirs because we exploit the extreme shortness of the secret key.

2 Preliminaries

Let us recall the definition of multivariate quadratic polynomials.

Definition 2.1. *Let $q \in \mathbb{N}$ be a power of prime and let $\mathbb{F}_q$ be the finite field of order q. Let n be a positive integer. A quadratic n-variate polynomial can be denoted by the following form: $f(x_1, \ldots, x_n) = \sum_{1 \le j,k \le n} r_{j,k} x_j x_k + \sum_{1 \le j \le n} l_j x_j + c$, where $r_{j,k}, l_j, c \in \mathbb{F}_q$.*

Following Huang et al., we use $S[\cdot]$, $R[\cdot]$, and $L[\cdot]$ to denote systems of polynomials and $S(\cdot)$, $R(\cdot)$, and $L(\cdot)$ to denote the corresponding functions.

Let us consider a system with n variables and m polynomials, $S = (f_1, \ldots, f_m)$ which defined by $r_{i,j,k}, l_{j,i}, c_i \in \mathbb{Z}_q$ for $i \in [m]$ and $j, k \in [n]$. Notice that we can write

$$f_i(x_1, \ldots, x_n) = \boldsymbol{x} \cdot \boldsymbol{R}_i \cdot \boldsymbol{x}^\top + \boldsymbol{x} \cdot \boldsymbol{l}_i^\top + c_i,$$

where $\boldsymbol{x} = (x_1, \ldots, x_n)$, $\boldsymbol{R}_i = (r_{i,j,k})_{1 \le j,k \le n} \in \mathbb{Z}_q^{n \times n}$, and $\boldsymbol{l}_i = (l_{1,i}, \ldots, l_{n,i}) \in \mathbb{Z}_q^n$. For compactness, we employ $R[\boldsymbol{x}]$ and $L[\boldsymbol{x}]$ to denote a collection of the quadratic and linear part of the polynomials in S, respectively. We also employ $R(\cdot)$ and $L(\cdot)$ to denote corresponding functions.

By using this notation, we can write

$$S(\boldsymbol{x}) = R(\boldsymbol{x}) + \boldsymbol{x}\boldsymbol{L} + \boldsymbol{c},$$

where $\boldsymbol{L} = [\boldsymbol{l}_1^\top, \ldots, \boldsymbol{l}_m^\top] \in \mathbb{Z}_q^{n \times m}$.

Distributions: In the following, we let n, m, β, q be natural numbers and let α be a positive real. For any positive integer β, we define $H_\beta = \{-\beta, -\beta - 1, \ldots, \beta - 1, \beta\}$.

We first review the folded Gaussian distribution and its discretized version in [8].

Definition 2.2 ($\Psi_{q,\alpha}$ **and** $\bar{\Psi}_{q,\alpha}$**:).** *The distribution $\Psi_{q,\alpha}$ is defined as following procedure: 1) Sample $z \leftarrow N(0, \alpha^2)$, which is the normal distribution with mean 0 and standard deviation α. 2) Output $z \bmod q$.*

The distribution $\bar{\Psi}_{q,\alpha}$ is defined as following procedure: 1) Sample $z \leftarrow N(0, \alpha^2)$, which is the normal distribution with mean 0 and standard deviation α. 2) Output $\lfloor z \rceil \bmod q$.

3 The HLY12 Cryptosystem

Preliminaries: Huang, Liu, and Yang introduced the following multivariate problem.

Definition 3.1 (MQ(n, m, q, χ, H) **[7]).** *Let n be a positive integer, m a positive integer of order $\Theta(n)$, q a prime, χ a distribution over $\mathbb{Z}_q$, and $H \subseteq \mathbb{Z}_q$.*

We define a distribution of n-variate polynomials, denoted by $M_{n,q,\chi}$: 1) For $1 \le j, k \le n$, choose $r_{j,k}$ according to the distribution χ, 2) for $1 \le j$, choose l_j from $\mathbb{Z}_q$ uniformly at random, and 3) output $f(x_1, \ldots, x_n) = \sum_{1 \le j,k \le n} r_{j,k} x_j x_k + \sum_{1 \le j \le n} l_j x_j$.

The problem MQ(n, m, q, χ, H) *is, given a system of m quadratic n-variate polynomials $S = (f_1, \ldots, f_m)$ and $\boldsymbol{y} = S(\boldsymbol{x})$, finding $\boldsymbol{x} \in H^n$, where $f_i \leftarrow M_{n,q,\chi}$ and $\boldsymbol{x} \leftarrow H^n$.*

We next review the MQ hardness assumption introduced in [7].

Definition 3.2 (MQ$(n, m, q, \bar{\Psi}_{q,\alpha}, H_\beta)$ **assumption).** *Let k be the security parameter. For every constant $c > 1 \in \mathbb{N}$, every efficiently computable and polynomially bounded $n, m, q : \mathbb{N} \to N$, $\alpha : \mathbb{N} \to [-q/2, q/2]$ and every $0 < \beta \le [q/2]$ such that (1) $m = cn$, (2) q is prime, and (3) $\alpha = O(1)$. We say that the* MQ$(n, m, q, \bar{\Psi}_{q,\alpha}, H_\beta)$ *assumption holds if the problem* MQ$(n, m, q, \bar{\Psi}_{q,\alpha}, H_\beta)$ *is hard to solve for any polynomial-time adversary.*

The HLY12 Cryptosystem Huang et al. proposed the following public-key encryption scheme for bits in [7, Section 3.2].

Parameters: Let k be the security parameter. Let $n, m, q, \beta \in \mathbb{N}$ and let α and λ be positive reals.

Key Generation: Sample an MQ system $S = (L, R)$ with n variables and $m = \Theta(n)$ polynomials according to a distribution $M^m_{n,q,\bar{\Psi}_{q,\alpha}}$ and sample $\boldsymbol{x} \leftarrow H^n_\beta$. The public key is $(S, \boldsymbol{y} = S(\boldsymbol{x}))$ and the secret key is $\boldsymbol{x}$.

Encryption: On input a bit b, sample $\boldsymbol{r} \leftarrow H^m_{n^\lambda}$, compute $(\boldsymbol{c}_1, c_2) = (\boldsymbol{L} \cdot \boldsymbol{r}^\top, \boldsymbol{y} \cdot \boldsymbol{r}^\top + b \cdot \lfloor q/2 \rfloor)$, and output $(\boldsymbol{c}_1, c_2)$.

Decryption: On input secret key $\boldsymbol{x}$ and ciphertext $(\boldsymbol{c}_1, c_2)$, output $\left[\left\lfloor \frac{2}{q}[c_2 - \boldsymbol{x} \cdot \boldsymbol{c}_1]_q \right\rceil\right]_2$, where $\lfloor a \rceil$ denotes the nearest integer of a ($\lfloor a \rceil = \lfloor a - 1/2 \rfloor$) and $[a]_p$ denotes the residue of a modulo integer p.

Let us review their correctness. From the description, we have $\boldsymbol{y} = S(\boldsymbol{x}) = \boldsymbol{x}\boldsymbol{L} + R(\boldsymbol{x}) = \boldsymbol{x}\boldsymbol{L} + \boldsymbol{e}$, where we set an error vector $\boldsymbol{e} = R(\boldsymbol{x}) \in \mathbb{Z}^m_q$ by using the quadratic part R of S. Notice that

$$c_2 - \boldsymbol{x} \cdot \boldsymbol{c}_1 = b \lfloor q/2 \rfloor + \boldsymbol{y} \cdot \boldsymbol{r}^\top - \boldsymbol{x} \cdot \boldsymbol{L} \cdot \boldsymbol{r}^\top = b \lfloor q/2 \rfloor + \boldsymbol{e} \cdot \boldsymbol{r}^\top.$$

The proposers carefully set parameters to make $\boldsymbol{e} = R(\boldsymbol{x})$ short, $\frac{2}{q}(b \lfloor q/2 \rfloor + \boldsymbol{e} \cdot \boldsymbol{r}^\top) \approx b$. Precisely speaking, they showed that

Theorem 3.1. *The HLY12 cryptosystem is (statistically) correct if* $k\alpha n^{(2+\lambda)} m\beta^2 \leq q/4$.

They proved the security of the above public-key encryption scheme.

Theorem 3.2 ([7, Theorem 2]). *Assume that the* $\mathrm{MQ}(n, m, q, \bar{\Psi}_{q,\alpha}, H_\beta)$ *assumption holds and the parameters satisfy the relation* $m \cdot \log(2n^\lambda + 1) \geq (n+1)\log q + 2k$. *Then, the HLY12 cryptosystem is semantically secure.*

Finally, we review the parameter sets proposed by Huang et al., which are summarized in Table 1.

Case	k	n	m	α	β	q	Hardness (T, ν)
1	12	200	400	10	2	$1803131754697263278851 9 \approx 2^{74}$	$2^{156}, 2^{-100}$
2	12	256	512	10	2	$5232440279576267872 4873 \approx 2^{76}$	$2^{205}, 2^{-104}$

Table 1. Concrete Parameters

4 Our Attack

The public key $\boldsymbol{y}$ can be written as

$$\boldsymbol{y} = \boldsymbol{x}\boldsymbol{L} + \boldsymbol{e} \bmod q,$$

where $\boldsymbol{e} = R(\boldsymbol{x}) \bmod q$. Since $\boldsymbol{e}$ is relatively short, the problem can be considered as the LWE problem [8]. Therefore, it is natural to consider the key-recovery problem as a lattice problem.

In this section, we first take a direct approach to recover $\boldsymbol{e}$ as a warm up. We then introduce a more practical approach to recover $\boldsymbol{x}$ which exploits the extreme shortness of $\boldsymbol{x}$.

4.1 Direct Approach

Let us consider q-ary lattice $\Lambda_q(\boldsymbol{L})$ spanned by $\boldsymbol{L}$:

$$\Lambda_q(\boldsymbol{L}) := \{\boldsymbol{z} \in \mathbb{Z}^m : \exists \boldsymbol{s} \in \mathbb{Z}^n_q \text{ such that } \boldsymbol{z} \equiv \boldsymbol{s}\boldsymbol{L} \bmod q\}.$$

It is easy to verify that $\boldsymbol{y}$ is a vector close to $\Lambda_q(\boldsymbol{L})$ if $\boldsymbol{e}$ is short.

Therefore, the problem is reduced to the CVP problem on instance $(\Lambda_q(\boldsymbol{L}), \boldsymbol{y})$. To solve this CVP problem, we first compute a basis $\boldsymbol{B} \in \mathbb{Z}^{m \times m}$ of $\Lambda_q(\boldsymbol{L})$ by reducing $\binom{q\boldsymbol{I}_m}{\boldsymbol{L}}$. Following the Kannan embedding technique, we next apply the basis reduction algorithms such as LLL and BKZ to the matrix

$$\begin{pmatrix} \boldsymbol{B} & \boldsymbol{0} \\ \boldsymbol{y} & \gamma \end{pmatrix},$$

where γ is an estimated value of $\boldsymbol{e}$.

We note that the independent work by Albrecht et al. took this approach.

4.2 More Practical Approach:

We notice that $\boldsymbol{x}$ is *extremely short*, since its coefficients lie in $H = [-\beta, \beta]$ and Huang et al. set $\beta = 2$. To exploit this shortness of $\boldsymbol{x}$, we consider the following basis $\boldsymbol{B}$ and spanned lattice $\Lambda(\boldsymbol{B})$:

$$\boldsymbol{B} = \begin{pmatrix} q\boldsymbol{I}_m & \boldsymbol{O} & \boldsymbol{0} \\ \boldsymbol{L} & \boldsymbol{I}_n & \boldsymbol{0} \\ -\boldsymbol{y} & \boldsymbol{0} & 1 \end{pmatrix}.$$

Since $\boldsymbol{y} \equiv \boldsymbol{x}\boldsymbol{L} + \boldsymbol{e} \bmod q$, there exists $\boldsymbol{k} \in \mathbb{Z}^m$ satisfying $\boldsymbol{y} = \boldsymbol{x}\boldsymbol{L} + \boldsymbol{e} + q\boldsymbol{k} \in \mathbb{Z}^m$. Notice that, using this $\boldsymbol{k}$, the lattice $L(\boldsymbol{B})$ contains a short vector $\boldsymbol{w} = [-\boldsymbol{e} \mid \boldsymbol{x} \mid 1] \in \mathbb{Z}^{m+n+1}$;

$$[\boldsymbol{k} \mid \boldsymbol{x} \mid 1] \cdot \begin{pmatrix} q\boldsymbol{I}_m & \boldsymbol{O} & \boldsymbol{0} \\ \boldsymbol{L} & \boldsymbol{I}_n & \boldsymbol{0} \\ -\boldsymbol{y} & \boldsymbol{0} & 1 \end{pmatrix} = [-\boldsymbol{e} \mid \boldsymbol{x} \mid 1].$$

Since the short vector $\boldsymbol{w} = [-\boldsymbol{e} \mid \boldsymbol{x} \mid 1] \in \mathbb{Z}^{m+n+1}$ contains whole $\boldsymbol{e}$, $\boldsymbol{w}$ is longer than $\boldsymbol{e}$ and, thus, it is less efficient than the direct approach.

Hence, we consider *a truncated lattice* defined by a $(m' + n + 1)$-dimension right-bottom submatrix $\boldsymbol{B}'$ of $\boldsymbol{B}$. By this truncation, we have the following relations:

$$[\boldsymbol{k}' \mid \boldsymbol{x} \mid 1] \cdot \begin{pmatrix} q\boldsymbol{I}_{m'} & \boldsymbol{O} & \boldsymbol{0} \\ \boldsymbol{L}' & \boldsymbol{I}_n & \boldsymbol{0} \\ -\boldsymbol{y}' & \boldsymbol{0} & 1 \end{pmatrix} = [-\boldsymbol{e}' \mid \boldsymbol{x} \mid 1] \in \mathbb{Z}^{m'+n+1}.$$

We note that $\boldsymbol{w}' = [-\boldsymbol{e}' \mid \boldsymbol{x} \mid 1]$ is relatively shorter than the previous $\boldsymbol{w}$. Hence, we can expect that the basis reduction algorithm retrieves this short vector $\boldsymbol{w}'$.

5 Experiment

We examine our attack against the proposed parameter sets.

Settings: We will take $m' \approx n/3$ for clarity and construct $n + m' + 1$-dimensional lattices for attack. For each case, we choose $S = (L, R)$ only once and generate ten public keys by choosing ten random vector $\boldsymbol{x} \leftarrow H^n_\beta$. We mount the attack on a Core i7 PC using the NTL librarywith GMP. In each case, we run the BKZ algorithm (G_BKZ_FP with $\delta = 0.99$, block size = 30, and prune = 10) on lattices constructed from the public keys.

Case 1 ($n = 200$): We set $m' = 66 \approx 200/3$. Our attack can recover the secret key $\boldsymbol{x}$ from every public key $\boldsymbol{y}$. The running times vary from 268.69 to 295.34 seconds and the average of them is 278.16 seconds.

Case 2($n = 256$): Recall that Huang et al. set $n = 256$, $m = 512$, and as Case 2. We set $m' = 90$ as a slightly larger integer than a third of n. Our attack successfully recover the secret keys from all public keys. The running times vary from 898.14 to 1119.53 seconds and the average of them is 964.83 seconds ($\approx$ 16 minutes).

Comparison to Albrecht et al.'s attack: Albrecht et al. reported their running time as "roughly one day for the first challenge (i.e. Case 1)" and "roughly three days for the second challenge (i.e. Case 2)." Precisely speaking, their computation took $\sim$ 26 hours for the Case 1 and $\sim$ 98 hours for the Case 2 on a single core [3, Section 4], while our computation took 0.07–0.08 hours for the Case 1 and 0.25–0.31 hours for the Case 2. Our attack is faster than Albrecht et al.'s attack in approximately 100x–300x factor.

References

1. Albrecht, M. R., Farshim, P., Faugère, J.-C., and Perret, L. Polly cracker, revisited. In *ASIACRYPT 2011* (2011), D. H. Lee and X. Wang, Eds., vol. 7073 of *LNCS*, Springer, Heidelberg, pp. 179–196. 1
2. Albrecht, M. R., Faugère, J.-C., Farshim, P., Herold, G., and Perret, L. Polly cracker, revisited. Cryptology ePrint Archive, Report 2011/289, 2011. 1
3. Albrecht, M. R., Faugère, J.-C., Fitzpatrick, R., and Perret, L. Practical cryptanalysis of a public-key encryption scheme based on new multivariate quadratic assumptions. Cryptology ePrint Archive, Report 2013/470, 2013. 1, 5
4. Berbain, C., Gilbert, H., and Patarin, J. QUAD: A multivariate stream cipher with provable security. *Journal of Symbolic Computation 44*, 12 (2009), 1703–1723. A preliminary version appeared in *EUROCRYPT 2006*, 2006. 1
5. Fischlin, M., Buchmann, J., and Manulis, M., Eds. *Public Key Cryptography - PKC 2012 - 15th International Conference on Practice and Theory in Public Key Cryptography, Darmstadt, Germany, May 21-23, 2012, Proceedings* (2012), vol. 7293 of *LNCS*, Springer, Heidelberg. 5
6. Herold, G. Polly cracker, revisited, revisited. In Fischlin et al. [5], pp. 17–33. 1
7. Huang, Y.-J., Liu, F.-H., and Yang, B.-Y. Public-key cryptography from new multivariate quadratic assumptions. In Fischlin et al. [5], pp. 190–205. 1, 2, 3
8. Regev, O. On lattices, learning with errors, random linear codes, and cryptography. *Journal of the ACM 56*, 6 (2009), Article 34. A preliminary version appeared in *STOC 2005*, 2005. 1, 2, 3
9. Sakumoto, K., Shirai, T., and Hiwatari, H. Public-key identification schemes based on multivariate quadratic polynomials. In *CRYPTO 2011* (2011), P. Rogaway, Ed., vol. 6841 of *LNCS*, Springer, Heidelberg, pp. 706–723. 1