Lifting approach against the SNOVA scheme

Shuhei Nakamura[1], Yusuke Tani[1], and Hiroki Furue[2]

[1] Department of Computer and Information Sciences, Ibaraki University
[2] NTT Social Informatics Laboratories, Tokyo

Abstract. In 2022, Wang et al. proposed the multivariate signature scheme SNOVA as a UOV variant over the non-commutative ring of $\ell \times \ell$ matrices over $\mathbb{F}_q$. This scheme has small public key and signature size and is a first round candidate of NIST PQC additional digital signature project. Recently, Ikematsu and Akiyama, and Li and Ding show that the core matrices of SNOVA with v vinegar-variables and o oil-variables are regarded as the representation matrices of UOV with ℓv vinegar-variables and ℓo oil-variables over $\mathbb{F}_q$, and thus we can apply existing key recovery attacks as a plain UOV. In this paper, we propose a method that reduces SNOVA to smaller UOV with v vinegar-variables and o oil-variables over $\mathbb{F}_{q^\ell}$. As a result, we show that the previous first round parameter sets at $\ell = 2$ do not meet the NIST PQC security levels. We also confirm that the present parameter sets are secure from existing key recovery attacks with our approach.

Keywords: Post-Quantum Cryptography, Multivariate Cryptography, UOV, SNOVA, Key Recovery Attack

1 Introduction

Standard RSA and EC cryptosystems are designed based on difficult mathematical problems such as prime factorization and discrete logarithm problems. However, these mathematical problems are known to be solved in polynomial time using a large-scale quantum computer. It is therefore necessary to construct cryptography based on new mathematical problems resistant to quantum computers. Such cryptography is referred to as post-quantum cryptography. In 2016, the National Institute of Standards and Technology (NIST) started public recruitment of such cryptography candidates [9].

Multivariate public key cryptography is based on an NP-hard problem of solving a system of quadratic equations, called the MQ problem [4], and is one of the main categories of the NIST PQC standardization project. UOV is a multivariate signature scheme proposed by Kipnis, Patarin, and Goubin [7], and has essentially not been broken over 20 years. It and its variants provide a faster verification and short signature. For example, the Rainbow signature scheme proposed by Ding and Schmidt [2], a multilayer UOV variant, was selected as a third round finalist in the NIST PQC standardization project. However, UOV and Rainbow have a drawback to be a large public key compared to other PQC such as lattice-based cryptosystems.

In 2022, NIST announced that the three signature schemes, Dilithium, Falcon, and SPHINCS+, will be standardized, but also announced to start the additional digital signature schemes competition. In this new process, 40 signature schemes were accepted to the first round and published in July 2023.

SNOVA is a multivariate signature scheme proposed by Wang et al.[11] and is accepted to the first round of the additional digital signature project [12]. It is regarded as a UOV scheme over a non-commutative ring and its parameter sets have short signature and a smaller public key. For example, the key and signature size of the SL1-SNOVA with $\ell = 4$ are close to the already standardized lattice cryptography. Note that, since SNOVA takes the matrix ring as a non-commutative ring, its core matrix is actually regarded as a large UOV matrix over a finite field with a parameter set $(q, \ell v, \ell o)$. Ikematsu and Akiyama [5] and Li and Ding [8] pointed out this fact and showed the parameter set with $\ell = 2$ do not meet the NIST PQC security levels, and the designers change these parameter sets.

1.1 Our contribution

In SNOVA, the core matrix is regarded as the matrix description of a standard UOV polynomial with a parameter set $(q, \ell v, \ell o)$. Then the secret non-singular matrix T is a block matrix and each block component is chosen from the algebra $\mathbb{F}_q[S]$ generated by the symmetric matrix S of size ℓ over $\mathbb{F}_q$. Note that the matrix S presented in [12] is diagonalizable over the splitting field for its characteristic polynomial. By using this property, our proposed method transforms T to a block diagonal matrix $\widehat{T}$ whose diagonal block components has the form of the secret non-singular matrices for smaller UOVs with a parameter set (q^ℓ, v, o). In particular, by concentrating on one diagonal block of $\widehat{T}$, we can perform the key recovery attack on UOV with a parameter set (q^l, v, o) if $mo^2 \geq vo$.

For the proposed parameter sets of SNOVA, since it holds $mo^2 \geq vo$, our method obtains an equivalent key by iterating a key recovery attack against UOV with the parameter set (q^ℓ, v, o). As a result, when $v < 2o$, we can provide a key recovery attack that are as efficient as [5] and [8] and show that the first SNOVA parameter sets at $\ell = 2$ does not meet the NIST PQC security level. On the other hand, for SNOVA parameter sets with $v \geq 2o$, including the new parameter set at $\ell = 2$, we found that SNOVA still maintains the NIST PQC security levels against existing attacks on our method because the iterations are less efficient than [5] and [8] due to the larger field order.

In the specification [12], the characteristic polynomial of S is irreducible over the finite field $\mathbb{F}_q$ defined there, and the matrix S cannot be diagonalized over the finite field. However, if S is diagonalized over the finite field $\mathbb{F}_q$, we are able to use our method without a field extension, and key recovery attacks with this variant show that all the proposed parameter sets will not meet the NIST PQC security level. Therefore, this work gives a reason why the characteristic polynomial of S should be irreducible over $\mathbb{F}_q$.

1.2 Organization

In Section 2, we recall UOV, SNOVA and its security analysis from known key recovery attacks. In Section 3, we describe our lifting method for SNOVA and mention the complexity estimation of existing key recovery attacks on the method. In Section 4, we provide the security analysis for the SNOVA parameter sets proposed in the NIST PQC additional digital signature project.

2 Preliminaries

In this section, we firstly explain the UOV scheme in Subsection 2.1 and key recovery attacks against it in Subsection 2.2. Then, we describe the SNOVA scheme and its security analysis according to [5].

2.1 UOV scheme

Let v, o be positive integers and $\mathbb{F}_q$ be the finite field of order q. Set $n = v + o$ and $m = o$. The UOV scheme is a signature scheme proposed by Kipnis, Patarin, and Goubin [7] and consists of the following algorithms:

Key generation Let $\mathcal{F}_1, \ldots, \mathcal{F}_m$ be a quadratic polynomials in $\mathbb{F}_q[x_1, \ldots, x_n]$ of the form

$$\mathcal{F}_k(x_1, \ldots, x_n) = \sum_{1 \le i \le v, 1 \le j \le n} a_{ij}^{(k)} x_i x_j, \tag{1}$$

where $1 \le k \le m$ and $a_{ij}^{(k)} \in \mathbb{F}_q$. Then these polynomials define a function $\mathcal{F}_i : \mathbb{F}_q^n \to \mathbb{F}_q$ and obtain a polynomial map $\mathcal{F} = (\mathcal{F}_1, \ldots, \mathcal{F}_m) : \mathbb{F}_q^n \to \mathbb{F}_q^m$. Moreover, we randomly choose a linear transformation $\mathcal{T} : \mathbb{F}_q^n \to \mathbb{F}_q^n$ and obtain a quadratic map $\mathcal{P} = \mathcal{F} \circ \mathcal{T}$. We set $\{\mathcal{F}_i\}_{i=1}^m \cup \{\mathcal{T}\}$ as the secret key and $\{\mathcal{P}_i\}_{i=1}^m$ as the public key.

Signature generation and verification For $\boldsymbol{b} \in \mathbb{F}_q^m$, we compute an element of its preimage in $\mathcal{P}$ as follows. Randomly choose $(a'_1, \ldots, a'_v) \in \mathbb{F}_q^v$ and then $\{\mathcal{F}_1(a'_1, \ldots, a'_v, x_{v+1}, \ldots, x_n), \ldots, \mathcal{F}_m(a'_1, \ldots, a'_v, x_{v+1}, \ldots, x_n)\}$ is a system of m linear polynomials in o variables $x_{v+1}, \ldots, x_n$. We then solve this linear system and set its solution as $(a'_{v+1}, \ldots, a'_n)$ if any. Otherwise, retake the value $(a'_1, \ldots, a'_v) \in \mathbb{F}_q^v$. As a result, we obtain a preimage element $\boldsymbol{a}$ under $\mathcal{P}$ by computing $\boldsymbol{a} = \mathcal{T}^{-1}(\boldsymbol{a}')$ where $\boldsymbol{a}' = (a'_1, \ldots, a'_n)$. The verification process confirms whether $\mathcal{P}(\boldsymbol{a}) = \boldsymbol{b}$ holds.

Matrix description Note that, for any quadratic polynomial $f(x_1, \ldots, x_n)$, there exists a matrix M such that $f(x_1, \ldots, x_n) = {}^t\boldsymbol{x} M \boldsymbol{x}$ where $\boldsymbol{x} =$

${}^t(x_1,\ldots,x_n)$. For example, we can take the following matrix F_k for the quadratic polynomial $\mathcal{F}_k(x_1,\ldots,x_n)$ of the form (1):

$$F_k = \begin{pmatrix} a'_{11} & \cdots & a'_{1v} & a'_{1\,v+1} & \cdots & a'_{1n} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ a'_{v1} & \cdots & a'_{vv} & a'_{v\,v+1} & \cdots & a'_{vn} \\ a'_{v+1\,1} & \cdots & a'_{v+1\,v} & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ a'_{n\,1} & \cdots & a'_{n\,v} & 0 & \cdots & 0 \end{pmatrix}, \tag{2}$$

where $a'_{ij} \in \mathbb{F}_q$. Namely, $\mathcal{F}_k(x_1,\ldots,x_n) = {}^t\boldsymbol{x}F_k\boldsymbol{x}$. In this paper, we introduce the following definition.

Definition 1. *Denote by* $\mathrm{Mat}_{UOV(v,o)}(\mathbb{F}_q)$ *the set of a square matrix of size* $v+o$ *over* $\mathbb{F}_q$ *whose lower right* $o\times o$ *submatrix is zero, i.e. of the form* (2). *We call an element of* $\mathrm{Mat}_{UOV(v,o)}(\mathbb{F}_q)$ *a* UOV matrix *with a parameter set* (q,v,o).

Meanwhile, it is well-known that the linear transformation $\mathcal{T}:\mathbb{F}_q^n \to \mathbb{F}_q^n$ also has a matrix description and there exists a non-singular matrix $T \in GL(n,\mathbb{F}_q)$ such that $\mathcal{T}(\boldsymbol{x}) = T\boldsymbol{x}$. In the UOV scheme, it is sufficient to take T as the following form:

$$T = \begin{pmatrix} 1 & & & t_{1\,v+1} & \cdots & t_{1n} \\ & \ddots & & \vdots & \ddots & \vdots \\ & & 1 & t_{v\,v+1} & \cdots & t_{vn} \\ & & & 1 & & \\ & O & & & \ddots & \\ & & & & & 1 \end{pmatrix}. \tag{3}$$

Then we can obtain the matrix description $P_1,\ldots,P_m$ of the public key $\mathcal{P}_1,\ldots,\mathcal{P}_m$ by

$$P_k = {}^tTF_kT \quad (1 \le k \le m).$$

2.2 Key recovery attacks against UOV

In this subsection, we describe known key recovery attacks against UOV and give these complexity estimations.

Kipnis-Shamir attack The Kipnis-Shamir attack is a key recovery attack proposed by Kipnis and Shamir [6]. It randomly constructs two matrices M_1 and M_2 as a linear combination of public keys where M_1 is non-singular, and finds a vector of the twisted oil space $T^{-1}\mathcal{O}$ from the stable subspace under $M_1^{-1}M_2$ where the oil space $\mathcal{O} = \mathrm{Span}(\{\boldsymbol{e}_i \mid v+1 \le i \le n\})$. For example, $T^{-1}\boldsymbol{e}_i$ coincides with the i-th column of the secret key T^{-1} for $v+1 \le i \le n$. In [12], the complexity of the attack is estimated as

$$O(q^{v-o}).$$

Intersection attack Based on the Kipnis-Shamir attack, Beullens proposed the intersection attack [1]. The attack randomly constructs two non-singular matrices M_1 and M_2 as a linear combination of public keys. If the intersection $M_1T^{-1}\mathcal{O} \cap M_2T^{-1}\mathcal{O}$ has a non-zero vector $\boldsymbol{x}$, then $M_1^{-1}\boldsymbol{x}, M_2^{-1}\boldsymbol{x} \in T^{-1}\mathcal{O}$ and we obtain

$$\begin{cases} {}^t\boldsymbol{x}{}^tM_1^{-1}P_kM_1^{-1}\boldsymbol{x} = 0 \\ {}^t\boldsymbol{x}{}^tM_2^{-1}P_kM_2^{-1}\boldsymbol{x} = 0 \\ {}^t\boldsymbol{x}{}^tM_1^{-1}(P_k + {}^tP_k)M_2^{-1}\boldsymbol{x} = 0 \end{cases}.$$

If $v < 2o$, the dimension of the intersection $M_1T^{-1}\mathcal{O} \cap M_2T^{-1}\mathcal{O}$ is at least $2o-v$ and we can impose $2o-v$ affine constraints. Hence, we solve the system of $3m-2$ quadratic equations in $2v-o$ variables by omitting two redundant polynomials. The complexity is estimated by

$$O\left(3 \cdot \binom{N + D_{M,N}}{D_{M,N}}^2 \binom{N+2}{2}\right),$$

where $M = 3m - 2$ and $N = 2v - o$,

$$D_{M,N} = \min\{i > 0 \mid \mathrm{Coeff}((1-t^2)^M(1-t)^{-N-1}, t^i) \leq 1\}.$$

If $v \geq 2o$, [1] mentions that the probability of the existence of a non-trivial intersection is $1/q^{v-2o+1}$ and estimates the complexity as

$$O\left(3 \cdot q^{v-2o+1} \cdot \binom{N + D_{M,N}}{D_{M,N}}^2 \binom{N+2}{2}\right),$$

where $M = 3m - 2$ and $N = 2v - o$.

Reconciliation attack The reconciliation attack is proposed by Ding et al.[3] Set t_{ij} in (3) as variables and $\boldsymbol{o}_i = T^{-1}\boldsymbol{e}_{v+i}$. Since $\boldsymbol{o}_i \in T^{-1}\mathcal{O}$, for $1 \leq c \leq o$, we obtain quadratic equations

$$\mathcal{S}_c = \{{}^t\boldsymbol{o}_iP_k\boldsymbol{o}_i = 0, {}^t\boldsymbol{o}_i(P_k + {}^tP_k)\boldsymbol{o}_j = 0 \mid 1 \leq i < j \leq c, 1 \leq k \leq m\}.$$

When $m \geq v$, the reconciliation attack solves a overdetermined system $\mathcal{S}_1$ of m quadratic equations in v variables. Then we can solve $\mathcal{S}_2$ more efficiently because it contains m linear equations in v variables. After that, it can be solved easily as well. Hence, the complexity of solving $\mathcal{S}_1$ is dominant and estimated by

$$O\left(3 \cdot \binom{N + D_{M,N}}{D_{M,N}}^2 \binom{N+2}{2}\right), \tag{4}$$

where $M = m$ and $N = v$. When $m\binom{o+1}{2} \geq ov$, the (full) reconciliation attack solves a overdetermined system $\mathcal{S}_o$ and its complexity is estimated by (4) with $M = m\binom{o+1}{2}$ and $N = ov$. In this paper, setting $c = \min\{1 \leq i \leq o \mid m\binom{i+1}{2} \geq vi\}$, we solve the overdetermined subsystem $\mathcal{S}_c$ first and other subsystems subsequently.

2.3 SNOVA and its security analysis

In this subsection, we describe the SNOVA scheme proposed in the first round of NIST PQC additional digital signature scheme project [12]. Let o and v be two positive integers such that $v > o$ and $\mathbb{F}_q$ be the finite field of order q. Denote by $\mathcal{R}$ the ring of all $\ell \times \ell$ matrices over $\mathbb{F}_q$, i.e. $\mathcal{R} = \mathrm{Mat}_{\ell\times\ell}(\mathbb{F}_q)$.

Key generation Let S be a square matrix of size ℓ. Denote by $\mathbb{F}_q[S]$ the algebra generated by S over $\mathbb{F}_q$, i.e.

$$\mathbb{F}_q[S] = \{a_0 I_\ell + a_1 S + \cdots + a_{\ell-1} S^{\ell-1} \mid a_i \in \mathbb{F}_q\} \subseteq \mathcal{R}.$$

In the specification [12], the finite field is fixed as $\mathbb{F}_{16} = \mathbb{F}_2[x]/\langle x^4 + x + 1\rangle$ and the matrix $S \in \mathrm{Mat}_{\ell\times\ell}(\mathbb{F}_{16})$ is given as follows.

$$S = \begin{cases} \begin{pmatrix} 8 & 7 \\ 7 & 6 \end{pmatrix} & \text{if } \ell = 2, \\ \begin{pmatrix} 8 & 7 & 6 \\ 7 & 6 & 5 \\ 6 & 5 & 4 \end{pmatrix} & \text{if } \ell = 3, \\ \begin{pmatrix} 8 & 7 & 6 & 5 \\ 7 & 6 & 5 & 4 \\ 6 & 5 & 4 & 3 \\ 5 & 4 & 3 & 2 \end{pmatrix} & \text{if } \ell = 4, \end{cases}$$

where the elements of $\mathbb{F}_{16}^\times$ are represented by $\{1, \ldots, 15\}$ with the correspondence $i \leftrightarrow \sigma^i$ and σ is a generator of the cyclic group $\mathbb{F}_{16}^\times$. Then the characteristic polynomial of each matrix S is irreducible over $\mathbb{F}_{16}$.

Let $F_1, \ldots, F_m \in \mathrm{Mat}_{n\times n}(\mathcal{R})$ be randomly chosen matrices such that

$$F_i = \begin{pmatrix} F_i^{11} & F_i^{12} \\ F_i^{21} & 0_o \end{pmatrix}, \tag{5}$$

where $F_i^{11} \in \mathrm{Mat}_{v\times v}(\mathcal{R}), F_i^{12} \in \mathrm{Mat}_{v\times o}(\mathcal{R}), F_i^{21} \in \mathrm{Mat}_{o\times v}(\mathcal{R})$, and 0_o is the zero matrix in $\mathrm{Mat}_{o\times o}(\mathcal{R})$. The non-singular matrix T is given by

$$T = \begin{pmatrix} 1_v & T^{12} \\ 0_{o\times v} & 1_o \end{pmatrix},$$

where $T^{12} \in \mathrm{Mat}_{v\times o}(\mathbb{F}_q[S])$ is randomly chosen, 1_i is the identity matrix in $\mathrm{Mat}_{i\times i}(\mathcal{R})$, and $0_{o\times v}$ is the zero matrix in $\mathrm{Mat}_{o\times v}(\mathcal{R})$. Then we define $P_1, \ldots, P_m \in \mathrm{Mat}_{n\times n}(\mathcal{R})$ as

$$P_i = {}^tT F_i T.$$

We further choose $A_\alpha, B_\alpha, Q_{\alpha 1}, Q_{\alpha 2} \in GL(\ell, \mathbb{F}_q)$ where $Q_{\alpha j} \in \mathbb{F}_q[S]$ and then set $\{F_i\}_{1\le i\le m} \cup \{T\}$ as the secret key and $\{P_i\}_{1\le i\le m} \cup \{A_\alpha, B_\alpha, Q_{\alpha 1}, Q_{\alpha 2}\}_{1\le\alpha\le\ell^2}$ as the public key.

Signature generation and verification In the SNOVA scheme, using the public key, we can construct the polynomial function $\mathcal{P}_k : \mathcal{R}^n \to \mathcal{R}$ as follows:

$$\mathcal{P}_k(Y_1, \ldots, Y_n) := \sum_{\alpha=1}^{\ell^2} \sum_{1 \le i,j \le n} A_\alpha Y_i Q_{\alpha 1} P_{k,[i,j]} Q_{\alpha 2} Y_j B_\alpha \ (1 \le k \le m),$$

where $P_{k,[i,j]}$ is the (i,j)-component of P_k in $\mathrm{Mat}_{n\times n}(\mathcal{R})$. Hence, we obtain the polynomial map $\mathcal{P} = (\mathcal{P}_1, \ldots, \mathcal{P}_m) : \mathcal{R}^n \to \mathcal{R}^m$. Note that a term of degree two in the non-commutative polynomial $\mathcal{R}[Y_1, \ldots, Y_n]$ is of form AY_iCY_jB where $A, B, C \in \mathcal{R}$. In signature generation in SNOVA, for a given hash value, we need to find an element of its preimage under $\mathcal{P}$. The polynomial map $\mathcal{P}$ can be regarded as the composition of a "UOV" map and a linear map. Indeed, the non-singular matrix T induces the linear polynomial map

$$\mathcal{T} : \mathcal{R}^n \to \mathcal{R}^n, {}^t(X_1, \ldots, X_n) \mapsto T \cdot {}^t(X_1, \ldots, X_n), \text{ and}$$

the matrix sequence $(F_1, \ldots, F_m)$ defines the polynomial map $\mathcal{F} = (\mathcal{F}_1, \ldots, \mathcal{F}_m) : \mathcal{R}^n \to \mathcal{R}^m$ where

$$\mathcal{F}_k(X_1, \ldots, X_n) = \sum_{\alpha=1}^{\ell^2} \sum_{1 \le i,j \le n} A_\alpha X_i Q_{\alpha 1} F_{k,[i,j]} Q_{\alpha 2} X_j B_\alpha \ (1 \le k \le m).$$

Then we have $\mathcal{P} = \mathcal{F} \circ \mathcal{T}$ by the commutativity of the elements in $\mathbb{F}_q[S]$. By the UOV-like definition of $F_1, \ldots, F_m$, the map $\mathcal{F}$ is linear when we fix values of $X_1, \ldots, X_v$ and therefore we can compute a preimage element under $\mathcal{F}$ by linear algebra. The verification process performs the substitution in $\mathcal{P}$.

Security analysis as $\mathbf{Mat}_{\ell n\times \ell n}(\mathbb{F}_q)$ Since $\mathcal{R}$ is the matrix ring $\mathrm{Mat}_{\ell\times\ell}(\mathbb{F}_q)$, we can regard T, F_i, and P_i as elements of $\mathrm{Mat}_{\ell n\times \ell n}(\mathbb{F}_q)$. In particular, F_i are regarded as UOV matrices and we can apply key recovery attacks against UOV with a parameter set $(q, \ell v, \ell o)$. Basis on this observation, Ikematsu and Akiyama [5] and Li and Ding [8] showed the parameter sets with $\ell = 2$ do not meet the NIST PQC security levels. In addition, they found a method to strengthen algebraic key recovery attacks. They introduce the following matrix

$$S_{diag} = I_n \otimes S = \begin{pmatrix} S & & \\ & \ddots & \\ & & S \end{pmatrix}$$

and show $TS_{diag} = S_{diag}T$. Hence, if a vector $\boldsymbol{x}$ is contained in the twisted oil space $T^{-1}\mathcal{O}$, then $\boldsymbol{x}S^j_{diag}$ is also contained. Indeed,

$$\boldsymbol{x} \in T^{-1}\mathcal{O} \Rightarrow S^j_{diag}\boldsymbol{x} \in S^j_{diag}T^{-1}\mathcal{O} = T^{-1}S^j_{diag}\mathcal{O} = T^{-1}\mathcal{O}.$$

Thus, we can increase the number of equations to be solved in algebraic key recovery attacks such as the reconciliation attack and the intersection attack.

3 Proposed lifting method

In this section, we describe our lifting method for the SNOVA scheme which reduces it to small UOVs with parameter (q^ℓ, v, o). We explain our lifting method in Subsection 3.1 and mention the complexity estimation of known key recovery attacks on the small UOV obtained by our method in Subsection 3.2.

3.1 Transformation to smaller UOVs

As mentioned in [12], the matrices S given in Subsection 2.3 have an irreducible characteristic polynomial. The following lemma shows that such a matrix is diagonalizable.

Lemma 1. *The non-zero square matrix over a finite field with the irreducible characteristic polynomial is diagonalizable over the splitting field of the characteristic polynomial.*

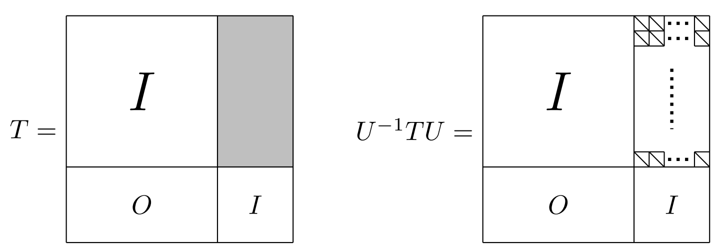
Proof. See Appendix. □

Thus, for the matrix $S \in \mathrm{Mat}_{\ell\times\ell}(\mathbb{F}_{16})$ defined in the specification [12] (see Subsection 2.3), there exists a non-singular matrix B in $GL(\ell, \mathbb{F}_{16^\ell})$ such that $B^{-1}SB$ is a diagonal matrix in $\mathrm{Mat}_{\ell\times\ell}(\mathbb{F}_{16^\ell})$. In particular, all elements of $\mathbb{F}_{16}[S]$ are simultaneously diagonalizable with this matrix B.

Let U be the diagonal block matrix with n copies of the matrix B, namely,

$$U = I_n \otimes B = \begin{pmatrix} B & & \\ & \ddots & \\ & & B \end{pmatrix} \in \mathrm{Mat}_{\ell n\times \ell n}(\mathbb{F}_{16^\ell}).$$

The block components of the matrix T which is an element of $\mathrm{Mat}_{n\times n}(\mathbb{F}_{16}[S])$ are diagonalized with U as follows:



Then the relation $P_i = {}^tTF_iT$ is rewritten as

$$ {}^tUP_iU = ({}^tU\,{}^tT\,{}^tU^{-1})({}^tUF_iU)(U^{-1}TU). \tag{6}$$

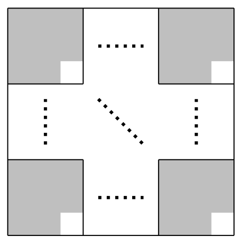
Note that the matrix ${}^tUF_iU$ is still a UOV matrix in $\mathrm{Mat}_{\ell n\times \ell n}(\mathbb{F}_{16^\ell})$.

By the following lemma, we can derive a relation for a smaller UOV matrix from (6):

Lemma 2. *Let* $\mathrm{Diag}_\ell(\mathbb{F})$ *be the set of all diagonal matrices in* $\mathrm{Mat}_{\ell\times\ell}(\mathbb{F})$*, and* $\mathrm{Mat}_{UOV(v,o)}(\mathbb{F})$ *be the set of square matrices of size* $v+o$ *whose* $o \times o$ *submatrix at the lower right is zero. Let* $\overline{F} \in \mathrm{Mat}_{UOV(\ell v,\ell o)}(\mathbb{F})$ *and* $\overline{T} = \begin{pmatrix} I_{\ell v} & \overline{T}^{12} \\ O & I_{\ell o} \end{pmatrix} \in GL(\ell n, \mathbb{F})$ *where* $\overline{T}^{12} \in \mathrm{Mat}_{v\times o}(\mathrm{Diag}_\ell(\mathbb{F}))$*. Then, there exists a permutation matrix* L *such that*

$$ {}^t L \overline{T} L = \begin{pmatrix} \widehat{T}_1 & & O \\ & \ddots & \\ O & & \widehat{T}_\ell \end{pmatrix}, \tag{7} $$

where $\widehat{T}_i = \begin{pmatrix} I_v & \widehat{T}_i^{12} \\ O & I_o \end{pmatrix} \in GL(n,\mathbb{F})$ *and* $\widehat{T}_i^{12} \in \mathrm{Mat}_{v\times o}(\mathbb{F})$*. It deduces* ${}^t L \overline{F} L \in \mathrm{Mat}_{\ell\times\ell}(\mathrm{Mat}_{UOV(v,o)}(\mathbb{F}))$*, i.e.*

${}^t L \overline{F} L =$  .

Note that, when we regard the matrix $\overline{F}$ as the matrix representation of a quadratic form, this lemma means there exists a permutation of the variables such that it separates the variables to smaller variable sets of a UOV polynomial.

By Lemma 2, for $\overline{T} = U^{-1}TU$ and $\overline{F}_i = {}^tUF_iU$, we obtain the permutation matrix $L \in GL(\ell n, \mathbb{F}_{16^\ell})$ satisfying the assertion in Lemma 2. Define

$$ \widehat{T} = {}^t L \overline{T} L, \widehat{F}_i = {}^t L \overline{F}_i L, \text{ and } \widehat{P}_i = {}^t L \overline{P}_i L. $$

Note that $\widehat{T}$ is of the form (7) and $\widehat{F} \in \mathrm{Mat}_{\ell\times\ell}(\mathrm{Mat}_{UOV(v,o)}(\mathbb{F}_{16^\ell}))$. Since ${}^tL = L^{-1}$ for a permutation matrix L, we have a relation

$$ \widehat{P}_k = {}^t\widehat{T}\widehat{F}_k\widehat{T}, $$

but more precisely we have the following relations for block components as a block matrix in $\mathrm{Mat}_{\ell\times\ell}(\mathrm{Mat}_{n\times n}(\mathbb{F}_{16^\ell}))$:

$$ \widehat{P}_k^{[i,j]} = {}^t\widehat{T}_i \cdot \widehat{F}_k^{[i,j]} \cdot \widehat{T}_j, \quad 1 \le i, j \le \ell. \tag{8} $$

3.2 Lifted known attacks on SNOVA

In this subsection, we show it is possible to obtain the key equivalent relation of UOV with the parameter set (q^ℓ, v, o) from the SNOVA scheme proposed in [12].

In the previous section, after our lifted method, we obtain Equation (8) as a relation between the public key and the secret key of SNOVA. Then, as a diagonal component of $\ell \times \ell$ matrix over $\mathrm{Mat}_{n\times n}(\mathbb{F}_{16^\ell})$, we have ℓ relations

$$\widehat{P}_k^{[i,i]} = {}^t\widehat{T}_i \cdot \widehat{F}_k^{[i,i]} \cdot \widehat{T}_i, \quad 1 \le i \le \ell. \tag{9}$$

These relations are regarded as a relation between the public key and the secret key of UOV with the parameter set (q^ℓ, v, o). When $mo^2 \ge vo$, the secret key (3) of UOV with a parameter (v, o) is often unique for a given instance, and we can obtain the secret key $\widehat{T}_1, \ldots, \widehat{T}_\ell$ by applying a key recovery attack to the relations (9) in the diagonal component. Note that, after recovering $\widehat{T}_1$, we obtain linear equations with respect to another $\widehat{T}_j$ with $j \ne 1$ from the $(1, j)$-block relation in (8) and can utilize a more efficient key recovery attack against $\widehat{T}_j$.

Since all SNOVA parameter sets proposed in [12] satisfy the condition $mo^2 \ge vo$, it is sufficient to perform a key recovery attack against ℓ UOV instances with the parameter (q^ℓ, v, o) and its complexity estimation is given as follows (see also Remark 1).

Kipnis-Shamir attack

We can apply the Kipnis-Shamir attack against each UOV public key $\widehat{P}^{[j,j]} = (\widehat{P}_1^{[j,j]}, \ldots, \widehat{P}_m^{[j,j]})$ with the parameter set (q^ℓ, v, o). Hence, by [12], the complexity of the attack is estimated by

$$O((q^\ell)^{v-o}).$$

Note that this coincides with one of the Kipnis-Shamir attack in [5]. However, since our attack is over the extension field, the bit complexity of our attack is larger than their one by (10) in Subsection 4.

Reconciliation attack

Let c be the minimum number such that $c^2 m \ge cv$. Since we consider the case $mo^2 \ge vo$, there exists such a number c. Then the reconciliation attack solves a system of $c^2 m$ quadratic equations in cv variables. Hence, according to [12], the complexity is estimated by

$$O\left(3 \cdot \binom{N + D_{N,M}}{D_{N,M}}^2 \binom{N+2}{2}\right),$$

where $N = vc$ and $M = mc^2$.

Intersection attack

As mentioned in [5], we consider the matrix key equivalent problem (see also Subsection 2.2) and distinguish the two relations ${}^t\boldsymbol{x}^t M_1^{-1} P_k M_2^{-1} \boldsymbol{x} = 0$ and ${}^t\boldsymbol{x}^t M_2^{-1} P_k M_1^{-1} \boldsymbol{x} = 0$. When $v < 2o$, the attack solves the system of $4m - 2$

quadratic equations in $n-(2o-v)$ variables and its complexity is estimated by

$$O\left(3\cdot\binom{N+D_{N,M}}{D_{N,M}}^2\binom{N+2}{2}\right),$$

where $N=n-(2o-v)$ and $M=4m-2$. When $v\geq 2o$, the attack tries to find the non-trivial intersection of two images of the twisted oil-subspace and solves the system of $4m-2$ quadratic equations in n variables. The complexity is estimated by

$$O\left(3\cdot(q^\ell)^{n-3o+1+\max\{n-4m+2,0\}}\cdot\binom{N+D_{N,M}}{D_{N,M}}^2\binom{N+2}{2}\right),$$

where $N=\min\{n,4m-2\}$ and $M=4m-2$.

Note that, by Remark 1 below, we cannot utilize the strengthening of [5] in our lifting method.

Remark 1. We show that the method introduced by [5] (see Subsection 2.3) is trivial after our lifting method and does not contribute to strengthen key recovery attacks. When we apply the discussion in Subsection 3.1 to the matrix $S_{diag}=I_n\otimes S$ considered in [5] (see Subsection 2.3), we obtain $\widehat{S}_{diag}=L^{-1}U^{-1}S_{diag}UL$ and can confirm $\widehat{T}\widehat{S}_{diag}=\widehat{S}_{diag}\widehat{T}$. However, when setting

$$B^{-1}SB=\begin{pmatrix}a_{11} & & \\ & \ddots & \\ & & a_{\ell\ell}\end{pmatrix},$$

we have $\widehat{S}_{diag}^{[j,j]}=a_{jj}I_n$ where $\widehat{S}_{diag}^{[j,j]}$ is the j-th diagonal component of $\widehat{S}_{diag}$ in $\mathrm{Mat}_{\ell\times\ell}(\mathrm{Mat}_{n\times n}(\mathbb{F}_{16^\ell}))$. Thus the commute $\widehat{T}\widehat{S}_{diag}=\widehat{S}_{diag}\widehat{T}$ is trivial in the relation (8).

4 Security analysis

In this section, using the complexity estimation introduced in the previous section, we estimate the complexity of lifted key recovery attacks against for the SNOVA parameter sets in the first round of NIST PQC additional digital signature project.

The parameter sets proposed in the specification [12] are set to satisfy the NIST PQC security levels. The categories of the NIST PQC security level one, three, and five require 143 bit, 206 bit, and 272 bit security, respectively. All proposed parameter sets satisfy the condition $mo^2\geq vo$ and we can apply key recovery attacks in the previous section. Table 1 shows the complexity estimation of key recovery attacks on our lifting method for the proposed parameter sets and the revised parameter sets with $\ell=2$. Here we list the bit complexity for the number of gates which is estimated from the number of multiplications as follows:

$$\sharp gates = \sharp multiplications \times (2(\log_2 q^\ell)^2 + \log_2 q^\ell) \tag{10}$$

In Table 1, "Inter.", "KS", and "Recon." mean the Intersection attack, the Kipnis-Shamir attack, and the reconciliation attack, respectively. Consequently, similar to the results in [5] and [8], the first parameter sets with $\ell = 2$ do not meet the NIST PQC security levels. In particular, the first parameter sets $(v, o, q, \ell) = (28, 17, 16, 2), (43, 25, 16, 2)$, and $(61, 33, 16, 2)$ with $\ell = 2$ are broken within 79.63 bits, 112.74 bits, and 158.99 bits, respectively, by the intersection attack on our lifting method. Those parameter sets were broken within 87 bits, 120 bits, and 167 bits, respectively, in [5] and within 77 bits, 167 bits, and 249 bits, respectively, in [8]. Hence, we are able to provide slightly more efficient attacks against the parameter sets $(v, o, q, \ell) = (43, 25, 16, 2)$ and $(61, 33, 16, 2)$.

Table 1. The bit-complexity estimation for our lifting method against SNOVA in the first round of the NIST PQC additional digital signature project

SL	SNOVA (v, o, q, ℓ)	$\log_2(\sharp gates)$ KS	Recon.	Inter.
1	(28, 17, 16, 2)	**95.08**	**131.38** $(c = 2)$	**79.63**
1	(37, 17, 16, 2)	167.08	196.63 $(c = 3)$	157.66
1	(25, 8, 16, 3)	212.22	195.64 $(c = 4)$	263.77
1	(24, 5, 16, 4)	313.04	319.77 $(c = 5)$	494.56
3	(43, 25, 16, 2)	**151.08**	**192.83** $(c = 2)$	**112.74**
3	(56, 25, 16, 2)	255.08	289.23 $(c = 3)$	232.38
3	(49, 11, 16, 3)	464.22	510.56 $(c = 5)$	693.42
3	(37, 8, 16, 4)	473.04	425.95 $(c = 5)$	704.59
5	(61, 33, 16, 2)	**231.08**	294.15 $(c = 2)$	**158.99**
5	(75, 33, 16, 2)	343.08	376.62 $(c = 3)$	302.49
5	(66, 15, 16, 3)	620.22	655.23 $(c = 5)$	906.03
5	(60, 10, 16, 4)	809.04	785.26 $(c = 7)$	1303.04

5 Conclusion

In this paper, we showed that by using the splitting field $\mathbb{F}_{q^\ell}$ of the characteristic polynomial, it is possible to diagonalize the matrix S, and that the key equivalent problem of SNOVA is reduced to that of UOV with parameters (q^ℓ, v, o). As a result, we were able to provide key recovery attacks that are as efficient as [5] and [8] for the previous SNOVA parameter sets with $v < 2o$. In particular, for the parameter sets $(v, o, q, \ell) = (43, 25, 16, 2)$ and $(61, 33, 16, 2)$, we were able to provide slightly more efficient attacks than [5] and [8]. On the other hand, we showed that the SNOVA parameter sets with $v \geq 2o$, including the new parameter sets with $\ell = 2$, are secure for existing attacks on our lifting method. In general, the disadvantage of our method is that it uses the field of large order q^ℓ and thus requires the large number of iterations in an attack with an exhaustive search. In particular, the condition that $mo^2 \geq vo$ is necessary for the efficient key recovery.

The efficiency of our method depends on the definition of $\mathbb{F}_q$ and S. In the specification [12], the characteristic polynomial of S is irreducible over the finite field $\mathbb{F}_q$ defined there, and the matrix S cannot be diagonalized over the finite field. However, if the roots of the characteristic polynomial of S are different from each other and in $\mathbb{F}_q$, i.e. S is diagonalized over the finite field $\mathbb{F}_q$, we are able to use our method without a field extension. Then the complexity of Kipnis-Shamir attack with this variant is estimated as q^{v-o}, and it shows that all the proposed parameter sets do not meet the NIST PQC security level. Therefore, this work gives a reason why the characteristic polynomial of S should be irreducible over $\mathbb{F}_q$.

Acknowledgement

This work was supported by JST CREST Grant Number JPMJCR2113, and JSPS KAKENHI Grant Numbers 23K16885.

References

1. Beullens, W.: Improved cryptanalysis of UOV and Rainbow. In EUROCRYPT 2021(1), LNCS vol. 12696. Springer, 2021.
2. Ding, J., and Schmidt, D.: Rainbow, a new multivariable polynomial signature scheme. In International conference on applied cryptography and network security, pp. 164–175. Springer, 2005.
3. Ding, J., Yang, B.-Y., Chen, C.-H.O., Chen, M.-S., Cheng, C.-M.: New differential-algebraic attacks and reparametrization of Rainbow, ACNS 2008, in: LNCS, vol. 5037, Springer, 2008, pp. 242–257.
4. Garey, M.-R., Johnson, D.-S.: Computers and intractability: a guide to the theory of NP-completeness. W.H. Freeman (1979).
5. Ikematsu, Y., Akiyama, R.: Revisiting the security analysis of SNOVA, IACR Cryptology ePrint Archive, 2024/096 `https://eprint.iacr.org/2024/096`
6. Kipnis, A., Shamir, A.: Cryptanalysis of the Oil and Vinegar Signature Scheme, CRYPTO 1998, LNCS Vol. 1462, pp. 257–266, Springer (1998)
7. Kipnis, A., Patarin, J., Goubin, L.: Unbalanced oil and vinegar signature schemes. In: Stern, J. (ed.) Advances in Cryptology – EUROCRYPT '99. pp. 206–222. Springer Berlin Heidelberg, Berlin, Heidelberg (1999)
8. Li, P., Ding, J.: Cryptanalysis of the SNOVA signature scheme, IACR Cryptology ePrint Archive, 2024/110 `https://eprint.iacr.org/2024/110`
9. NIST: Post-quantum cryptography CSRC. `https://csrc.nist.gov/projects/post-quantum-cryptography`

10. NIST: Additional Digital Signature Schemes for the Post-Quantum Cryptography Standardization Process. https://csrc.nist.gov/projects/pqc-dig-sig/round-1-additional-signatures
11. Wang, L.C., Tseng, P.E., Kuan, Y.L., Chou, C.Y.: A Simple Noncommutative UOV Scheme, IACR Cryptology ePrint Archive, 2022/1742 (2022)
12. Wang, L.C., Chou, C.Y., Ding, J., Kuan, Y.L., Li, M.S., Tseng, B.S., Tseng, P.E., and Wang, C.C.: SNOVA Proposal for NISTPQC: Digital Signature Schemes project (May 25, 2023) https://csrc.nist.gov/csrc/media/Projects/pqc-dig-sig/documents/round-1/spec-files/SNOVA-spec-web.pdf
13. Yang, B.Y., Chen, J.M.: Theoretical analysis of XL over small fields. In: Wang, H., Pieprzyk, J., Varadharajan, V. (eds.) Information Security and Privacy. pp. 277–288. Springer Berlin Heidelberg, Berlin, Heidelberg (2004)

Appendix (The proof of Lemma 1)

Lemma. *The non-zero square matrix over a finite field with the irreducible characteristic polynomial is diagonalizable over the splitting field of the characteristic polynomial.*

Proof. Let $M \in \mathrm{Mat}_{n\times n}(\mathbb{F}_q)$ and $\Phi(x)$ be its irreducible characteristic polynomial where $q = p^e$. Since $\Phi(x)$ is irreducible, it coincides with the minimal polynomial of M, and its constant term $(-1)^n \det(M)$ is non-zero, namely M is non-singular. Hence M is contained in the finite group $GL(n, \mathbb{F}_q)$ and has a finite order $r := \mathrm{order}(M)$. Since the irreducible characteristic polynomial $\Phi(x)$ is the minimal polynomial of M, it divides $x^r - 1$, i.e. $\Phi(x) \mid (x^r - 1)$. We assume that $p \mid r$. Then $x^r - 1 = (x^c - 1)^p$ where $r = pc$. Since $\Phi(x)$ is irreducible, $\Phi(x) \mid (x^c - 1)^p$ implies $\Phi(x) \mid x^c - 1$ and we have $\mathrm{order}(M) \leq c < r$ which is a contradiction. Therefore, it follows that $p \nmid r$ and the roots of $x^r - 1$ are different each other. Thus, since $\Phi(x) \mid (x^r - 1)$, the characteristic polynomial $\Phi(x)$ does not have a multiple root in its splitting field and the matrix M is diagonalizable. Since the characteristic polynomial is of degree n, the diagonal components as its roots are contained in the splitting field whose degree of the field extension is n. □