Improved Attacks for SNOVA by Exploiting Stability under a Group Action

Daniel Cabarcas¹, Peigen Li², Javier Verbel³, and Ricardo Villanueva-Polanco³

Universidad Nacional de Colombia, Colombia dcabarc@unal.edu.co
Beijing Institute of Mathematical Sciences and Applications, China lpg22@bimsa.cn
Technology Innovation Institute, UAE {javier.verbel,ricardo.polanco}@tii.ae

Abstract. SNOVA is a post-quantum digital signature scheme based on multivariate polynomials. It is a first-round candidate in an ongoing NIST standardization process for post-quantum signatures, where it stands out for its efficiency and compactness. Since its initial submission, there have been several improvements to its security analysis, both on key recovery and forgery attacks. All these works reduce to solving a structured system of quadratic polynomials, which we refer to as SNOVA system.

In this work, we propose a polynomial solving algorithm tailored for SNOVA systems, which exploits the *stability* of the system under the action of a commutative group of matrices. This new algorithm reduces the complexity to solve SNOVA systems, over generic ones. We show how to adapt the *reconciliation* and *direct* attacks in order to profit from the new algorithm. Consequently, we improve the reconciliation attack for all SNOVA parameter sets with speedup factors ranging between 2^3 and 2^{22} . Our algorithm also reduces the complexity of the direct attack for several parameter sets. It is particularly effective for the parameters that give the best performance to SNOVA (l=4), and which were not taken below NIST's security threshold by previous attacks. Our attack brings these parameter sets (l=4) below that threshold with speedup factors between 2^{33} and 2^{52} , over the state-of-the-art.

Keywords: Cryptanalysis, SNOVA, stable ideals, post-quantum, multivariate.

1 Introduction

Digital signatures are essential to ensure the authenticity and integrity of digital communications. The security of widely used digital signature schemes is threatened by quantum computers [14]. Post-quantum cryptography (PQC) is an active area of research aiming at developing cryptographic algorithms that are resilient against quantum attacks. In light of this, NIST has been leading

an effort to evaluate and standardize cryptographic algorithms capable of withstanding quantum adversaries.

Following the success of its first PQC standardization process, in 2023 NIST initiated a second call for submissions, focused on digital signatures. One promising candidate in this process is the SNOVA signature scheme [17], which builds upon the Unbalanced Oil and Vinegar (UOV) signature scheme. SNOVA modifies UOV's structure to reduce the size of the public key and the speed of signing, while keeping UOV's short signatures and fast verification. For example, at security level I, SNOVA can have 1000-byte public keys and 232-byte signatures, while the speed of signing and verification are comparable to those of Dilithium, one of the post-quantum signature schemes standardized by NIST. Due to its efficiency and compactness, SNOVA offers an attractive option for real-world applications.

The main concern about SNOVA is its security, and this is the main focus of this work. There have been several papers since the beginning of 2024 analyzing SNOVA's security from different perspectives. Since our work builds upon those works, it is important to summarize their main findings.

Key-Recovery and reconciliation. In two independent but concurrent works Ikematsu-Akiyama [7] and Li-Ding [8] analysed the security of SNOVA against key-recovery attacks. Both works reached the same conclusion: All known key-recovery attacks for SNOVA with parameters (v, o, l, q) can be seen as attacks to a UOV signature scheme with lo^2 equations and l(v + o) variables over \mathbb{F}_q . In particular, for the *reconciliation* attack, the attacker has to find one specific solution out of many of a quadratic polynomial system of the form

$$\mathbf{x}^t \Lambda_{S^i}^{(n)} P_k \Lambda_{S^j}^{(n)} \mathbf{x} = 0, \quad \forall k = 1, \dots, o, \text{ and } 0 \le i, j < l,$$
 (1)

where n=v+o, \mathbf{x} is a vector of variables, $P_k \in \mathbb{F}_q^{ln \times ln}$ is matrix, $\Lambda_{S^i}^{(n)}$ is a block-diagonal matrix with $S^i \in \mathbb{F}_q^{l \times l}$ along the diagonal, and $S \in \mathbb{F}_q^{l \times l}$ a symmetric matrix with an irreducible characteristic polynomial. Throughout this work, we refer to any system whose quadratic part has the form (1) as a SNOVA system. More recently, Nakamura, Tani, and Furie discuss a similar attack in [10].

Forgery Attacks. On a recent preprint, Beullens proposed a new forgery attack on SNOVA [4]. The main observation is that the SNOVA public key has a similar structure to MAYO's [2]. Specifically, denoting by $\mathcal{B}(\mathbf{u}_i, \mathbf{u}_j)$ the bilinear map associated with the SNOVA system defined by the public key, then the SNOVA public key can be written as:

$$\mathcal{P}(U) = \sum_{i=1}^{l} \sum_{j=1}^{l} E_{i,j} \mathcal{B}(\mathbf{u}_i, \mathbf{u}_j),$$

where $U = [\mathbf{u}_1, \dots, \mathbf{u}_l]$, $\mathbf{u}_i \in \mathbb{F}_q^{nl}$, the $E_{i,j} \in \mathbb{F}_q^{ol^2 \times ol^2}$ are block-diagonal matrices with o copies of a matrix $\tilde{E}_{i,j} \in \mathbb{F}_q^{l^2 \times l^2}$ along the diagonal. Unlike MAYO, there

is a nontrivial linear combination E of the matrices $E_{i,j}$ with a rank defect. Beullens uses this fact to speed up a forgery attack.

1.1 Our Contributions

We propose the first polynomial solving algorithm tailored for SNOVA systems. Our algorithm builds over the observation that the ideal I generated by the quadratic part of Eq. (1) is $\Lambda^{(n)}_{\mathbb{F}_q[S]}$ -stable, meaning, that $f(A\mathbf{x}) \in I$ for any $f \in I$ and $A \in \Lambda^{(n)}_{\mathbb{F}_q[S]}$. In [6], Faugère and Svartz propose a variant of the F5 algorithm to compute a Gröbner basis of a \mathcal{D} -stable ideal, where \mathcal{D} is any commutative group of matrices, and provide an asymptotical analysis of the complexity when the degree of the Macaulay matrix tends to infinity.

We adapt the ideas in [6] for SNOVA systems. Applying an appropriate change of variable to the ideal I, we obtain a stable ideal under the action of a cyclic diagonal group of matrices. The resulting polynomial system has a multi-degree homogeneous structure. Unlike [6], we propose an XL-like algorithm to solve the system and leverage the multi-degree homogeneous structure. Also, we provide a concrete and tight analysis of our algorithm's complexity, which is supported experimentally Our complexity estimates show that solving SNOVA systems is easier than solving a corresponding random quadratic system by a factor of about q^l .

By using our new algorithm to solve SNOVA systems, we improve the reconciliation attack in [7,8], with speedup factors ranging between 2^7 and 2^{22} .

We also show that it is possible to make use of our specialized polynomial solving algorithm to improve the forgery attack in [4]. In order to reduce the forgery of a message msg to the problem of solving a SNOVA system, while maintaining the multi-degree homogeneity of the lifted system, we have to leverage the low-rank of the matrix E in a different way. Instead of finding elements in the left kernel of E to obtain linear equations, we first brute force a value $\mathsf{salt} \in \{0,1\}^*$ such that the target vector $\mathbf{z} = \mathsf{Hash}(\mathsf{msg}\|\mathsf{salt})$ falls in the columns space of E. Then, we find a vector \mathbf{w} such that $\mathbf{z} = E\mathbf{w}$. Finally, we solve for a vector \mathbf{u} over \mathbb{F}_q and $y_1, \ldots, y_p \in \mathbb{F}_q$ such that

$$\mathbf{w} = \mathcal{B}(\mathbf{u}, \mathbf{u}) + \sum_{i=1}^{p} y_i \mathbf{w}_i,$$

where the $\mathbf{w}_1, \dots, \mathbf{w}_p$ are elements in the right-kernel of E.

Our forgery attack is faster than Beullens' [4] for several parameter sets, but it is particularly effective when l=4. We highlight that these are the only parameter sets for which the attack in [4] does not bring the security of the scheme below the threshold defined by NIST for each category. For l=4 parameter sets, the speedup of our attack over Beullens' ranges between 2^{33} and 2^{52} .

In [4], Beullens suggests a modified version of SNOVA called "SNOVA minus whipping". This new version avoids the forgery attacks introduced in the same

paper. The public map of the new scheme is defined as $\mathcal{P}(\mathbf{u}) = \mathcal{B}(\mathbf{u}, \mathbf{u})$, i.e., the public key defines a SNOVA map. The security analysis was left as an open problem. Our new algorithm for SNOVA systems serves as the first security analysis of the SNOVA minus whipping signature scheme.

Our attacks also affect the security of the two new alternative versions of SNOVA proposed in [16] with the goal of counteracting Beullen's attack. As for the first alternative, our forgery attack improves the state-of-the-art for this particular version, and the security of all proposed parameters fall below the threshold for the corresponding security levels. With regards to the second alternative, our forgery attack does not bring the security below the threshold but it improves the original direct attack for all proposed parameters for this particular version, with speedup factors ranging between 2 and 2¹⁶.

This paper is organized as follows. Section 2 describes the SNOVA signature scheme, some of the attacks, and a special XL algorithm for multi-homogeneous systems. In Section 3, we describe some of the algebraic properties of SNOVA systems. Section 4 introduces the proposed polynomial solving algorithm tailored for SNOVA systems and gives a detailed analysis of its complexity. Section 5 presents our adaptation of the reconciliation and direct attacks to profit from the proposed algorithm and shows results for the original version of SNOVA. Finally, Section 6 presents the results of our attacks for the recently proposed alternative versions of SNOVA.

2 Preliminaries

2.1 The SNOVA Signature Scheme

SNOVA is a digital signature scheme based on the ideas of UOV [17], and it is a current candidate to the so-called onramp NIST call for post-quantum digital signature schemes. Here, we describe the necessary objects from the scheme to be able to discuss the attacks, and we refer the reader to [17] for further details.

Let $n, v, o, l \in \mathbb{N}$, with v > o, and n = v + o. Denote by [l] the set $\{1, \ldots, l\}$, by \mathbb{F}_q the finite field with q elements, and by $\mathcal{R} := \mathbb{F}_q^{l \times l}$ the ring of $l \times l$ matrices over \mathbb{F}_q . Furthermore, given a matrix $Q \in \mathbb{F}_q^{l \times l}$, we denote by $\Lambda_Q^{(n)}$, or by $\operatorname{diag}(Q, \ldots, Q)$, the $nl \times nl$ block diagonal matrix with Q-blocks along the diagonal.

Given an $l \times l$ symmetric matrix S, with irreducible characteristic polynomial, $\mathbb{F}_q[S]$ is defined as

$$\mathbb{F}_q[S] := \{a_0 S^0 + a_1 S + \dots + a_{l-1} S^{l-1} : a_0, a_1, \dots, a_{l-1} \in \mathbb{F}_q\}.$$

Note that the elements in $\mathbb{F}_q[S]$ are symmetric and all commute with each other. Additionally, the non-zero elements in $\mathbb{F}_q[S]$ are invertible. In fact, $\mathbb{F}_q[S]$ is a finite field and $\mathbb{F}_q[S] \cong \mathbb{F}_{q^l}$.

The central map of SNOVA is given by $\mathcal{F} = (\mathcal{F}_1, \mathcal{F}_2, \dots, \mathcal{F}_o) : \mathcal{R}^n \to \mathcal{R}^o$, where \mathcal{F}_k is defined as

$$\mathcal{F}_k(X_1, X_2, \dots, X_n) = \sum_{\alpha=1}^{l^2} A_\alpha \cdot \left(\sum_{(i,j) \in \Omega} X_i^t \cdot (Q_{\alpha 1} F_{k,i,j} Q_{\alpha 2}) \cdot X_j \right) \cdot B_\alpha,$$

and $F_{k,i,j}$, A_{α} , $B_{\alpha} \in \mathcal{R}$ and $Q_{\alpha 1}$, $Q_{\alpha 2} \in \mathbb{F}_q[S]$ are invertible matrices, and

$$\varOmega=\{(i,j):\ 1\leq i,j\leq n\}\backslash\{(i,j):\ v+1\leq i,j\leq n\}.$$

The private key is $(\mathcal{F}, \mathcal{T})$. The public map is defined as $\mathcal{P} := \mathcal{F} \circ \mathcal{T} := (\mathcal{P}_1 :=$ $\mathcal{F}_1 \circ \mathcal{T}, \dots, \mathcal{P}_o := \mathcal{F}_o \circ \mathcal{T}$, where $\mathcal{T} : \mathcal{R}^n \to \mathcal{R}^n$ is the invertible linear map, corresponding to a matrix T, made up of blocks from $\mathbb{F}_q[S]$. Then

$$\mathcal{P}_k(U) = \mathcal{F}_k(\mathcal{T}(U)) = \sum_{\alpha=1}^{l^2} \sum_{i=1}^n \sum_{j=1}^n A_\alpha \cdot U_i^t(Q_{\alpha 1} P_{k,i,j} Q_{\alpha 2}) U_j \cdot B_\alpha, \tag{2}$$

where $U = (U_1, \ldots, U_n)^t \in \mathcal{R}^n$, and $P_{k,i,j} = \sum_{(s,t) \in \Omega} T_{i,s} F_{k,s,t} T_{t,j}$. SNOVA signing algorithm is similar to UOV's. To sign a message msg with the private key $(\mathcal{F},\mathcal{T})$, it first samples a salt from $\{0,1\}^{2\lambda}$, and sets Y:= $\mathsf{Hash}(\mathsf{msg}||\mathsf{salt}) \in \mathcal{R}^m$, where Hash is a hash function. Then it chooses random values $V_1, \ldots, V_v \in \mathcal{R}$, solves a linear system to find (V_{v+1}, \ldots, V_n) such that $\mathcal{F}(V_1,\ldots,V_v,V_{v+1},\ldots,V_n)=Y$, and repeats if it has no solution. Finally, the algorithm computes $U = \mathcal{T}^{-1}(V_1, \dots, V_n)$ and outputs $\sigma = (U, \mathsf{salt})$.

Security level	$\begin{array}{c} \text{parameters} \\ (v, o, q, l, \lambda) \end{array}$	Public key (B)	Signature (B)	Private key (B)
I	(37, 17, 16, 2, 128) (25, 8, 16, 3, 128) (24, 5, 16, 4, 128)	9826(+16) 2304(+16) 1000(+16)	$ \begin{array}{c} 108(+16) \\ 148.5(+16) \\ 232(+16) \end{array} $	60008(+48) 37962(+48) 34112(+48)
III	(56, 25, 16, 2, 192) (49, 11, 16, 3, 192) (37, 8, 16, 4, 192)	31250(+16) 5989.5(+16) 4096(+16)	162(+16) 270(+16) 360(+16)	202132(+48) 174798(+48) 128384(+48)
V	(75, 33, 16, 2, 256) (66, 15, 16, 3, 256) (60, 10, 16, 4, 256)	71874(+16) 15187.5(+16) 8000(+16)	216(+16) 364.5(+16) 560(+16)	515360(+48) 432297(+48) 389312(+48)

Table 1. Proposed parameters for SNOVA [17].

Table ${\color{blue}1}$ shows the latest parameters proposed by the SNOVA designers. Note that the parameters for l=2 shown in Table 1 do not match the original parameters in the NIST submission, since the original ones were updated in response to the attacks in [7,8]. The SNOVA scheme is still competitive and it

provides three parameter options for each security level, which yield different public key and signature lengths. This feature gives this scheme flexibility and adaptability to be used in diverse scenarios.

We remark that the SNOVA parameters with l=4 appear to be a good alternative to the standardized scheme FALCON, since they offer similar key sizes and smaller signature sizes than FALCON. Furthermore, for the security levels I, III, and V as specified by the NIST standardization call [12, Section 4B], SNOVA allegedly offers 143, 207, and 272 bits of security respectively.

2.2 Direct attack

Given $Z \in \mathcal{R}^o$, where $Z = (Z_1, \dots, Z_o)^t \leftarrow \mathcal{H}_1(\mathsf{msg}||\mathsf{salt})$, a direct attack consists in finding $U \in \mathcal{R}^n$ such that $\mathcal{P}(U) = Z$, i.e., $\mathcal{P}_i(U) = Z_i$ for all $i \in [o]$. In total, there are $l^2 \cdot o$ quadratic equations $\mathcal{P}_{i,kr}(U) = Z_{i,kr}$ on $l^2 \cdot n$ variables, namely, $U_{j,kr}$ for $j \in [n]$, $k,r \in [l]$ over \mathbb{F}_q . From this observation, the SNOVA authors in [17] claimed that to produce a fake signature, an attacker needs to regard a (v,o,q,l)-SNOVA public map as an (l^2v,l^2o,q) -UOV public map over \mathbb{F}_q and then forge a signature for this UOV instance. However, [4] introduces an improved forgery attack against SNOVA, the best one so far to the best of our knowledge. In particular, the author observes that SNOVA has a similar whipping structure as that of MAYO [3], where the public matrices (E-matrices) defined for SNOVA case may have linear combinations with smaller ranks. The author exploits this observation by applying the ideas described in [3, Section 5].

Let $\mathcal{B}: \mathbb{F}_q^{nl} \times \mathbb{F}_q^{nl} \to \mathbb{F}_q^{ol^2}$ defined by $\mathcal{B}(\mathbf{x}, \mathbf{y}) := (\mathbf{x}^t \Lambda_{S^{i-1}}^{(n)} P_k \Lambda_{S^{j-1}}^{(n)} \mathbf{y})_{i,j \in [l], k \in [o]}^t$ be the bilinear map associated to the SNOVA public polynomials. In [4], it is shown that the SNOVA public key can be written as:

$$\mathcal{P}(U) = \sum_{i=0}^{l-1} \sum_{j=0}^{l-1} E_{i,j} \mathcal{B}(\mathbf{u}_i, \mathbf{u}_j),$$

where $U = [\mathbf{u}_0, \dots, \mathbf{u}_{l-1}]$ with $\mathbf{u}_i \in \mathbb{F}_q^{nl}$ and $E_{i,j} = \Lambda_{\tilde{E}_{i,j}}^{(o)} \in \mathbb{F}_q^{ol^2 \times ol^2}$ for some matrix $\tilde{E}_{i,j} \in \mathbb{F}_q^{l^2 \times l^2}$.

The idea of the attack from [4] is choose random vectors $\mathbf{v}_1, \ldots, \mathbf{v}_{l-1} \in \mathbb{F}_q^{nl}$ and then solve for a solution $U = [\mathbf{u}_0, \ldots, \mathbf{u}_{l-1}]$ satisfying $\mathbf{u}_i = \alpha_i \mathbf{u}_0 + \mathbf{v}_i$ for $i \in \{1, \ldots, l-1\}$ and some scalars $\alpha_1, \ldots, \alpha_{l-1} \in \mathbb{F}_q$. The motivation for doing this change of variables is that the quadratic part of $\mathcal{P}(U)$ becomes $E\mathcal{B}(\mathbf{u}_0, \mathbf{u}_0)$ with $E = \sum_{z=0}^{l-1} \sum_{r=0}^{l-1} \alpha_i \alpha_j E_{i,j}$. Therefore, the attack proceeds as follows

- 1. Find $\alpha_1, \ldots, \alpha_{l-1}$ such that E has small rank or, i.e. $E = \Lambda_{\tilde{E}}^{(o)}$ with $\tilde{E} \in \mathbb{F}_q^{l^2 \times l^2}$ of rank $r \leq l^2$.
- 2. Pick $\mathbf{v}_1, \ldots, \mathbf{v}_{l-1}$ uniformly at random. Use a generic system solving algorithm to solve for \mathbf{u}_0 such that $\mathcal{P}(U) = Z$, where $\mathbf{u}_i = \alpha_i \mathbf{u}_0 + \mathbf{v}_i$ for $i \in \{1, \ldots, l-1\}$.

In the case of SNOVA, the author of [4] sets $\mathbf{u}_i = \Lambda_{R_i}^{(n)} \mathbf{u}_0 + \mathbf{v}_i$, where $R_i = r_{i,0}S^0 + \ldots + r_{i,l-1}S^{l-1} \in \mathbb{F}_q[S]$. The advantage of taking this approach is that it offers extra freedom, i.e. solving with l(l-1) variables instead of l-1, and thus allowing the attacker to find matrices with lower rank at step 1, which means that solving the system at step 2 becomes more efficient.

2.3 Key Recovery Attacks

A key-recovery attack for SNOVA reduces to find a basis to the vector space

$$\mathcal{O} := \{ T^{-1} \cdot (x_1, \dots, x_{ln})^t \in \mathbb{F}_q^{ln} : x_1 = x_2 = \dots = x_{lv} = 0 \},$$

which has dimension lo [17]. In [7,8], it is shown that for any $\mathbf{u}, \mathbf{v} \in \mathcal{O}$, we have

$$\mathbf{v}^t(\Lambda_{S_{i-1}}^{(n)} P_k \Lambda_{S_{j-1}}^{(n)}) \mathbf{u} = 0, \text{ for each } (i, k, j) \in [l] \times [o] \times [l].$$
(3)

The main goal of a key-recovery attack for SNOVA is to find at least one nontrivial element in \mathcal{O} . Once such an element is found, recovering a basis of \mathcal{O} is significantly easier than finding that first element in \mathcal{O} .

In the reconciliation attack, one attempts to find a vector of the form $\mathbf{u}_0 = (u_1, \dots, u_{lv}, 0, \dots, 1)^t \in \mathbb{F}_q^{ln}$ such that

$$\mathbf{u}_{0}^{t}(\Lambda_{S^{i-1}}^{(n)}P_{k}\Lambda_{S^{j-1}}^{(n)})\mathbf{u}_{0} = 0, \text{ for each } (i,k,j) \in [l] \times [o] \times [l].$$
(4)

Since \mathcal{O} is a random vector space of dimension ol, we expect that such a $\mathbf{u}_0 \in \mathcal{O}$ uniquely exists.

The quadratic system in (4) has l^2o equation on lv variables. For all SNOVA parameters it holds that $ol^2 < lv$ [17]. Therefore, the system in (4) has an expected number of $O(q^{lv-ol^2})$ solutions. In [7,8], the authors compute the complexity of this attack by using the hybrid approach [1], assuming that solving an underdetermined system as in (4) is as hard as solving a random one. However, in Section 4, we will introduce an algorithm that solves such systems more efficiently than a generic algorithm.

2.4 An XL-Like Algorithm for Multi-Homogeneous Systems

An important ingredient for our algorithm to solve SNOVA systems is an algorithm that adapts XL for Multi-Homogeneous Systems. Such algorithm has been used for other attacks such as [13,2,9,11]. We describe here its main features for completeness.

Definition 1 (Multi-homogeneous polynomials). Let $P = (X_1, \ldots, X_l)$ be a partition of the set $X = \{x_1, \ldots, x_n\}$. We say a monomial \mathfrak{m} in X has multi-degree $(d_1, \ldots, d_l) \in \mathbb{Z}_{\geq 0}^l$ if the degree of \mathfrak{m} with respect to X_i equals d_i for $i \in [l]$. We say that a polynomial f in X is multi-homogeneous of multi-degree $\mathbf{d} \in \mathbb{Z}_{\geq 0}^l$ if each monomial in the support of f has degree \mathbf{d} . In the particular case, l = 2, we use bi-degree and bi-homogeneous instead of multi-degree and multi-homogeneous.

Let $f_1, \ldots, f_m \in \mathbb{F}_q[x_1, \ldots, x_n]$ be multi-homogeneous quadratic polynomials. Like the XL algorithm, the special XL algorithm aims at finding a common root of the f_i . Unlike the XL, the special XL takes as input a multi-degree (d_1, \ldots, d_l) and is restricted to work with polynomials in $\langle f_1, \ldots, f_m \rangle$ of multi-degree $(e_1, \ldots, e_l) \leq (d_1, \ldots, d_l)$, where \leq means that $e_i \leq d_i$ for each $i \in [l]$.

In the bi-homogeneous case, i.e., l=2, the special XL has been used to estimate the complexity of some attacks in cryptography. See, for instance, the RBS and intersection attacks on Rainbow (see [13, Sec. 5] [2, Sec. 6]).

In the case of bi-homogeneous (l=2) quadratic polynomials, the complexity analysis introduced in [13] suggests that for systems in $n=n_1+n_2$ variables with m_1 equations of bi-degree (2,0), m_{12} of bi-degree (1,1) and m_2 of bi-degree (0,2), the special XL algorithm is expected to effectively work on input bi-degree (a_{sol},b_{sol}) if the coefficient of $t_1^a t_2^b$ in the series

$$\frac{(1-t_1^2)^{m_1}(1-t_1t_2)^{m_{12}}(1-t_2^2)^{m_2}}{(1-t_1)^{n_1+1}(1-t_2)^{n_2+1}},$$

is nonegative for some $a \leq a_{sol}$ and $b \leq b_{sol}$. Note that the series above is also considered in [9] for analysing the complexity of the RSB attack on Rainbow. In this case, the complexity of the special XL algorithm, in number of field multiplications, is upper bounded by

$$3 \cdot \max(n_1, n_2)^2 \cdot \overline{\mathcal{M}}(a_{sol}, b_{sol}),$$

where $\overline{\mathcal{M}}(a_{sol}, b_{sol})$ is the number of monomials with bi-degree $(a, b) \leq (a_{sol}, b_{sol})$. For general multi-homogeneous $(l \geq 2)$ quadratic systems, the complexity is upper bounded by

$$3 \cdot \max(n_1, \dots, n_l)^2 \cdot \left[\overline{\mathcal{M}}(\mathbf{d}_{sol}) \right]^2$$

where $\mathbf{d}_{sol} \in \mathbb{Z}_{\geq 0}^l$, $\overline{\mathcal{M}}(\mathbf{d}_{sol})$ is the number of monomials with multi-degree smaller than \mathbf{d}_{sol} , and there exists $(d_1, \ldots, d_l) \leq \mathbf{d}_{sol}$ such that the coefficient of $t_1^{d_1} t_2^{d_2} \cdots t_l^{d_l}$ in the series

$$\frac{\prod_{k=1}^{m} \left(1 - t_1^{d_1^{(k)}} t_2^{d_2^{(k)}} \cdots t_l^{d_l^{(k)}}\right)}{\prod_{i=1}^{l} (1 - t_i)^{n_i + 1}},$$
(5)

is nonnegative, where $(d_1^{(k)}, d_2^{(k)}, \dots, d_l^{(k)})$ is the multi-degree of f_k and n_i is the size of X_i . Note that the series in Eq. (5) was used in [11] to estimate the complexity of the Kipnis-Shamir attack on the MinRank problem.

3 $\Lambda_{\mathbb{F}_a[S]}^{(n)}$ -Stable Ideals

This section focuses on the system (4) of the key recovery attacks from [7,8]. The ideal generated by its quadratic part is stable under an action by a subgroup of matrices. We will refer to such a system as a $\Lambda_{\mathbb{F},[S]}^{(n)}$ -stable system. Faugère and

Svartz [6] have studied the problem of computing a Gröbner basis for such an ideal. In this section, we describe the particular structure of (4) in the framework introduced in [6]. This will allow us to propose an algorithm to solve it and analyze its complexity in the next section.

As in the description of SNOVA, here $S \in \mathbb{F}_q^{l \times l}$ is a symmetric matrix with irreducible characteristic polynomial, I is the homogeneous ideal generated by the quadratic part of (4), and we define

$$\varLambda^{(n)}_{\mathbb{F}_q[S]} := \left\{ \varLambda^{(n)}_Q \in \mathbb{F}_q^{ln \times ln} : Q \in \mathbb{F}_q[S] \setminus \{0\} \right\}.$$

We use the following definition from [6].

Definition 2 (\mathcal{G} -stable ideals). An ideal $I \subseteq \mathbb{F}_q[x_1, \ldots, x_n]$ is said to be globally stable under a finite matrix group $\mathcal{G} \subseteq GL_n(\mathbb{F}_q)$ if for all $f \in I$ and $G \in \mathcal{G}$, we have $f^G \in I$, where $f^G(\mathbf{x}) := f(G\mathbf{x})$.

It is easy to see that I is $\Lambda^{(n)}_{\mathbb{F}_q[S]}$ -stable.

Proposition 1. Let $F = \{f_{i,j,k} : 0 \leq i, j < l, 1 \leq k \leq o\}$ be the quadratic part of the $m = l^2o$ equations in (4) and $I = \langle F \rangle$ the ideal generated by them. Then, for $f \in F$ and $Q \in \mathbb{F}_q[S]$, it holds that, $f^{A_Q^{(n)}} \in \operatorname{Span}_{\mathbb{F}_q}(F)$. In particular, I is $A_{\mathbb{F}_q[S]}^{(n)}$ -stable.

Proof. For any $0 \le i, j < l, \ 1 \le k \le o$, and $Q = \sum_{h=0}^{l-1} \alpha_h \Lambda_{S^h}^{(n)} \in \mathbb{F}_q[S]$, we have

$$\begin{split} f_{i,j,k}^{\Lambda_Q^{(n)}}(\mathbf{x}) &= \mathbf{x}^t \Lambda_Q^{(n)} \Lambda_{S^i}^{(n)} P_k \Lambda_{S^j}^{(n)} \Lambda_Q^{(n)} \mathbf{x} = \mathbf{x}^t \left(\sum_{h=0}^{l-1} \alpha_h \Lambda_{S^h}^{(n)} \right) \Lambda_{S^i}^{(n)} P_k \Lambda_{S^j}^{(n)} \left(\sum_{h=0}^{l-1} \alpha_h \Lambda_{S^h}^{(n)} \right) \mathbf{x} \\ &= \sum_{a=0}^{l-1} \sum_{h=0}^{l-1} \alpha_{a,b} \mathbf{x}^t \Lambda_{S^a}^{(n)} P_k \Lambda_{S^b}^{(n)} \mathbf{x} = \sum_{a=0}^{l-1} \sum_{h=0}^{l-1} \alpha_{a,b} f_{a,b,k}(\mathbf{x}) \in \operatorname{Span}_{\mathbb{F}_q}(F) \end{split}$$

for some $\alpha_{a,b} \in \mathbb{F}_q$. It immediately follows that I is $\Lambda_{\mathbb{F}_q[S]}^{(n)}$ -stable.

We next expose the structure of $\Lambda_{\mathbb{F}_q[S]}^{(n)}$ that allow us to transform I into an ideal that is stable under the action of a cyclic group. Let $\tau: \mathbb{F}_{q^l} \to \mathbb{F}_{q^l}$ denote the Frobenious map. Abusing notation, we will also denote by τ the function that applies the Frobenious map component-wise to a vector or matrix.

Lemma 1. The matrix S is diagonalizable on an l-extension of \mathbb{F}_q . Specifically, if $\lambda \in \mathbb{F}_{q^l}$ is an eigenvalue of S and $\boldsymbol{\xi} \in \mathbb{F}_{q^l}^l$ its corresponding eigenvector,

$$P := \left[\boldsymbol{\xi} \ \tau(\boldsymbol{\xi}) \cdots \tau^{l-1}(\boldsymbol{\xi}) \right] \in \mathbb{F}_{q^l}^{l \times l}$$
 (6)

is non-singular and $P^{-1}SP$ is diagonal.

Proof. It suffices to prove that the columns of P are eigenvectors corresponding to distinct eigenvalues. Since the entries of S are in \mathbb{F}_q , for $a \in \{0, \ldots, l-1\}$,

$$S\tau^a(\boldsymbol{\xi}) = \tau^a(S\boldsymbol{\xi}) = \tau^a(\lambda \boldsymbol{\xi}) = \lambda^{q^a}\tau^a(\boldsymbol{\xi}),$$

thus $\lambda^{q^a} \in \mathbb{F}_{q^l} \setminus \mathbb{F}_q$ is an eigenvalue of S, corresponding to the eigenvector $\tau^a(\boldsymbol{\xi})$.

Let us show that $\lambda^{q^i} = \lambda^{q^j}$ implies $i \equiv j \pmod{l}$. By [15, Theorem 19.1.], the polynomial $x^{q^l} - x$ is square-free since $x^{q^l} - x$ and its derivative $q^l x^{q^l - 1} - 1 = -1$ are relatively prime. For a contradiction, suppose $\lambda^{q^i} = \lambda^{q^j}$ with i < j and let f be the characteristic polynomial of S. Then λ^{q^j} is a root of f with multiplicity at least 2 and $(x - \lambda^{q^j})^2$ divides f. Since f is a monic irreducible polynomial of degree l, f divides $x^{q^l} - x$ [15, Theorem 19.10.] which would imply $x^{q^l} - x$ is not square-free, contradicting the fact that $x^{q^l} - x$ is square-free. It follows that $\lambda^{q^0}, \ldots, \lambda^{q^{l-1}}$ are l distinct eigenvalues of S.

We can then diagonalize every matrix in $\varLambda^{(n)}_{\mathbb{F}_q[S]}$ to construct a cyclic matrix group.

Proposition 2. The matrix group

$$\mathcal{D} := \{ \Lambda_{P^{-1}}^{(n)} M \Lambda_P^{(n)} : M \in \Lambda_{\mathbb{F}_a[S]}^{(n)} \}. \tag{7}$$

is a cyclic diagonal group generated by $\Lambda_Q^{(n)} \in \mathbb{F}_{q^l}^{ln \times ln}$, where $Q = diag(\beta, \beta^q, \dots, \beta^{q^{l-1}}) \in \mathbb{F}_{q^l}^{l \times l}$ for some $\beta \in \mathbb{F}_{q^l}$.

Proof. Let $\boldsymbol{\xi} \in \mathbb{F}_{q^l}^l$ be an eigenvector of S with corresponding eigenvalue $\lambda \in \mathbb{F}_{q^l}$. For any nonzero $A = a_0 S^0 + a_1 S + \cdots + a_{l-1} S^{l-1} \in \mathbb{F}_q[S]$,

$$AP = P \cdot \operatorname{diag}(\lambda_A, \tau(\lambda_A), \dots, \tau^{l-1}(\lambda_A))$$
 (8)

where $\lambda_A = a_0 + a_1 \lambda + \dots + a_{l-1} \lambda^{l-1}$. Therefore, $P^{-1}AP$ is a diagonal matrix. Let B be a generator of the multiplicative group $\mathbb{F}_q[S]^{\times}$. Clearly, $\Lambda_{P^{-1}}^{(n)} \Lambda_B^{(n)} \Lambda_P^{(n)}$ is a generator of \mathcal{D} and $\boldsymbol{\xi}$ is an eigenvector of B. Let $\beta \in \mathbb{F}_{q^l}$ be the eigenvalue of B associated with $\boldsymbol{\xi}$. Then, $P^{-1}BP = \operatorname{diag}(\beta, \tau(\beta), \dots, \tau^{l-1}(\beta))$. Therefore, $\Lambda_{P^{-1}}^{(n)} \Lambda_B^{(n)} \Lambda_P^{(n)} = \Lambda_Q^{(n)}$, where $Q = \operatorname{diag}(\tau^0(\beta), \dots, \tau^{l-1}(\beta))$.

Applying an appropriate change of variable to the ideal I, yields a stable ideal under the action of a cyclic diagonal group.

Proposition 3. Let $S \in \mathbb{F}_q^{l \times l}$ be a symmetric matrix with irreducible characteristic polynomial. If $I \subset \mathbb{F}_q[x_1, \dots, x_{ln}]$ is $\Lambda_{\mathbb{F}_q[S]}^{(n)}$ -stable, then there exist a matrix $P \in \mathbb{F}_{q^l}^{l \times l}$ such that the ideal

$$I^{\Lambda_P^{(n)}} := \left\{ f^{\Lambda_P^{(n)}} : f \in I \right\} \subseteq \mathbb{F}_{q^l}[x_1, \dots, x_{ln}] \tag{9}$$

is stable under the action of a cyclic diagonal group.

Proof. Let P be a matrix defined as in Lemma 1 and \mathcal{D} a diagonal group of matrices defined as in Proposition 2. Let $f \in I$, $g = f^{\Lambda_P^{(n)}} \in I^{\Lambda_P^{(n)}}$, $M \in \Lambda_{\mathbb{F}_q[S]}^{(n)}$, and $D = \Lambda_{P^{-1}}^{(n)} M \Lambda_P^{(n)} \in \mathcal{D}$. Then,

$$g^D(\mathbf{x}) = g(D \cdot \mathbf{x}) = f(\Lambda_P^{(n)} \cdot D \cdot \mathbf{x}) = f(M \cdot \Lambda_P^{(n)} \cdot \mathbf{x}) = f^M(\Lambda_P^{(n)} \cdot \mathbf{x}).$$

Since I is $\Lambda_{\mathbb{F}_q[S]}^{(n)}$ -stable, then $f^M \in I$, and $g^D = (f^M)^{\Lambda_P^{(n)}} \in I^{\Lambda_P^{(n)}}$. Therefore $I^{\Lambda_P^{(n)}}$ is \mathcal{D} -stable.

The group action of the cyclic diagonal group induces a grading on $\mathbb{F}_{q^l}[x_1,\ldots,x_{nl}]$ that is compatible with the usual degree.

Definition 3 (\mathcal{D} -degree). Let β , $Q = diag(\beta, \beta^q, \dots, \beta^{q^{l-1}}) \in \mathbb{F}_{q^l}^{l \times l}$ and $\mathcal{D} = \langle \Lambda_Q^{(n)} \rangle$ be as in Proposition 2. For a monomial $\mu = x_1^{\alpha_1} \cdots x_{nl}^{\alpha_{nl}}$ in $\mathbb{F}_{q^l}[x_1, \dots, x_{nl}]$, we have

$$\mu^{\Lambda_Q^{(n)}} = (\beta x_1)^{\alpha_1} \cdots (\beta^{q^{l-1}} x_l)^{\alpha_l} (\beta x_{l+1})^{\alpha_{l+1}} \cdots (\beta^{q^{l-1}} x_{ln})^{\alpha_{ln}}$$
$$= \beta^{\sum_{j=1}^l q^{j-1} \cdot \sum_{i=1}^n \cdot \alpha_{(i-1) \cdot l+j} \mu}.$$

Then, we define the \mathcal{D} -degree of μ as

$$\deg_{\mathcal{D}}(\mu) = \sum_{j=1}^{l} q^{j-1} \cdot \sum_{i=1}^{n} \alpha_{(i-1)\cdot l+j} \mod q^l - 1.$$

Example 1. Consider the monomial $\mu = x_1 x_2 \dots x_{ln} \in \mathbb{F}_{q^l}[x_1, \dots, x_{nl}]$. Then

$$\mu^{\Lambda_Q^{(n)}} = \mu(\Lambda_Q^{(n)} \mathbf{x})$$

= $(\beta x_1) \dots (\beta^{q^{l-1}} x_l) (\beta x_{l+1}) \dots (\beta^{q^{l-1}} x_{2l}) \dots (\beta x_{(n-1)l+1}) \dots (\beta^{q^{l-1}} x_{nl}),$

thus, $\deg_{\mathcal{D}}(\mu) = n \sum_{j=1}^l q^{j-1} \mod q^l - 1$. In particular, for any monomial of the form $x_i x_j$, we have $\deg_{\mathcal{D}}(\mu) = q^{(i-1) \mod l} + q^{(j-1) \mod l} \mod q^l - 1$, where $(i-1) \mod l$ refers to the remainder of division by l.

With l=2 and q>2, there are exactly three non-zero \mathcal{D} -degrees for quadratic polynomials in x_1, \ldots, x_{ln} . If q=16 for example, these are, 2, 17 and 32. For any $i, j=1, \ldots, ln$, we have

$$\deg_{\mathcal{D}}(x_i x_j) = \begin{cases} 2 & \text{if } i \equiv j \equiv 1 \mod l \\ 17 & \text{if } i \not\equiv j \mod l \\ 32 & \text{if } i \equiv j \equiv 0 \mod l \end{cases}$$

With n = 2, the variables are x_1, x_2, x_3, x_4 and the quadratic monomials of each \mathcal{D} -degree are

$$-2: x_1^2, x_1x_3, x_3^2,$$

 $\begin{array}{l} -\ 17:\ x_1x_2, x_1x_4, x_2x_3, x_3x_4, \\ -\ 32:\ x_2^2, x_2x_4, x_4^2. \end{array}$

It readily follows that the \mathcal{D} -degree induces a grading on $\mathbb{F}_{q^l}[x_1,\ldots,x_{nl}]$. It is also easy to see that the \mathcal{D} -degree refines the usual degree. We can then refer to the \mathcal{D} -homogeneous components of a polynomial.

Definition 4. A polynomial f in $\mathbb{F}_{q^l}[x_1,\ldots,x_{nl}]$ is said to be \mathcal{D} -homogeneous if all its monomials have the same \mathcal{D} -degree.

Example 2. Let r be a positive integer with $1 \le r < l$. Consider the polynomial

$$f = x_r x_{r+1} + x_{l+r} x_{l+r+1} + x_{2l+r} x_{2l+r+1} + \dots + x_{(n-1)l+r} x_{(n-1)l+r+1}.$$

Note that $\deg_{\mathcal{D}}(x_r x_{r+1}) = q^{r-1 \mod l} + q^{r \mod l} \equiv q^{r-1} + q^r \mod q^l - 1$. Since $r-1 \equiv (i \cdot l + r - 1) \mod l$ and $r \equiv (i \cdot l + r) \mod l$ for $i \in \{1, \ldots, n-1\}$, then

$$\deg_{\mathcal{D}}(x_r x_{r+1}) = \dots = \deg_{\mathcal{D}}(x_{(n-1)l+r} x_{(n-1)l+r+1}),$$

hence, f is \mathcal{D} -homogeneous.

Now, with l=4, consider the polynomial

$$f = x_1 x_2 + x_3 x_4 + x_{l+1} x_{l+2} + x_{l+3} x_{l+4} + \dots + x_{(n-1)l+1} x_{(n-2)l+2} + x_{(n-1)l+3} x_{(n-1)l+4}.$$

Its \mathcal{D} -homogeneous components are

$$h^{(1)} = x_1 x_2 + x_{l+1} x_{l+2} + \dots + x_{(n-1)l+1} x_{(n-2)l+2}$$

and

$$h^{(2)} = x_3 x_4 + x_{l+3} x_{l+4} + \ldots + x_{(n-1)l+3} x_{(n-1)l+4}.$$

Definition 5. An ideal $I \subset \mathbb{F}_{q^l}[x_1, \dots, x_{ln}]$ is said to be \mathcal{D} -homogeneous if for any polynomial $f \in I$, all its \mathcal{D} -homogeneous components are in I.

Theorem 1. Let β , $Q = diag(\beta, \beta^q, \dots, \beta^{q^{l-1}}) \in \mathbb{F}_{q^l}^{l \times l}$ and $\mathcal{D} = \langle \Lambda_Q^{(n)} \rangle$ be as in Proposition 2. An ideal I is \mathcal{D} -stable, if and only if, it is \mathcal{D} -homogeneous.

Proof. It is clear that a \mathcal{D} -homogeneous ideal is \mathcal{D} -stable. For the converse, set $e = q^l - 1$ and fix $g \in I$. For each $i = 0, \ldots, e - 1$, let h_i be the \mathcal{D} -homogeneous component of g of \mathcal{D} -degree i. Note that

$$\begin{bmatrix} g \\ g^{\Lambda_Q^{(n)}} \\ \vdots \\ g^{(\Lambda_Q^{(n)})^{e-1}} \end{bmatrix} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & \beta & \dots & \beta^{e-1} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \beta^{e-1} & \dots & \beta^{(e-1)(e-1)} \end{bmatrix} \begin{bmatrix} h_0 \\ h_1 \\ \vdots \\ h_{e-1} \end{bmatrix},$$
(10)

and since I is stable under \mathcal{D} , then $g^{(\Lambda_Q^{(n)})^d} \in I$, for $d=0,\ldots,e-1$. Note that the matrix of (10) is a Vandermonde matrix $V=[\beta^{(i-1)(d-1)}]_{1\leq i,d\leq e}$. It is known that $\det(V)=\prod_{1\leq i< d\leq e}(\beta^{d-1}-\beta^{i-1})\neq 0$, since β is of order e. Therefore, V is nonsigular and each homogeneous component is in I.

3.1 Distribution of the Basis of Homogeneous Components

One remarkable feature about a SNOVA system, like (4), is that after a suitable change of variables, we can precisely predict the number of homogeneous generators of each \mathcal{D} -degree. Proposition 1 tells us that a SNOVA system is more than $\Lambda^{(n)}_{\mathbb{F}_q[S]}$ -stable. If F is the set of the quadratic parts of a SNOVA system, for $f \in F$ and $A \in \mathbb{F}_q[S]$, it holds that, $f^{\Lambda^{(n)}_A} \in \operatorname{Span}_{\mathbb{F}_q}(F)$. This additional property allows us to easily apply a suitable change of variables to land on a quadratic system that is \mathcal{D} -homogeneous and for which we can precisely predict the number of polynomials of each \mathcal{D} -degree. We first prove that the \mathbb{F}_{q^l} -vector space generated by F equals the \mathbb{F}_{q^l} -vector space generated by their \mathcal{D} -homogeneous components.

Corollary 1. Let P be a matrix of eigenvectors of S, as described in Lemma 1, and β , $Q = diag(\beta, \beta^q, \dots, \beta^{q^{l-1}}) \in \mathbb{F}_{q^l}^{l \times l}$ and $\mathcal{D} = \langle \Lambda_Q^{(n)} \rangle$ as in Proposition 2. Let f_1, \dots, f_m be homogeneous polynomials in $\mathbb{F}_{q^l}[x_1, \dots, x_{ln}]$. Suppose that for $i = 1, \dots, m$ and $A \in \mathbb{F}_q[S]$, it holds that $f_i^{\Lambda_A^{(n)}} \in \operatorname{Span}_{\mathbb{F}_q}(f_1, \dots, f_m)$. Then, denoting by $h_{i,d}$ the \mathcal{D} -degree d homogeneous component of $f_i^{\Lambda_P^{(n)}}$, we have that

$$\mathrm{Span}_{\mathbb{F}_{q^l}}(f_1^{\varLambda_P^{(n)}},\dots,f_m^{\varLambda_P^{(n)}})=\mathrm{Span}_{\mathbb{F}_{q^l}}(h_{1,1},\dots,h_{m,q^l-2}).$$

Proof. Set $\mathcal{V}:=\operatorname{Span}_{\mathbb{F}_{q^l}}(f_1^{\Lambda_P^{(n)}},\ldots,f_m^{\Lambda_P^{(n)}}),\,\mathcal{W}:=\operatorname{Span}_{\mathbb{F}_{q^l}}(h_{1,1},\ldots,h_{m,q^l-2}),$ and $e:=q^l-1.$ It is clear that $\mathcal{V}\subset\mathcal{W},$ so let us prove that $\mathcal{W}\subset\mathcal{V}.$ For $i=1,\ldots,m,$ set $g_i:=f_i^{\Lambda_P^{(n)}}.$ By hypothesis and Proposition 3, $I=\langle g_1,\ldots,g_m\rangle$ is \mathcal{D} -stable. Then, by Theorem 1, I is \mathcal{D} -homogeneous. Moreover, from the proof of Theorem 1, each $h_{i,d}$ is in $\operatorname{Span}_{\mathbb{F}_{q^l}}(g,g_i^{\Lambda_Q^{(n)}},\ldots,g_i^{(\Lambda_Q^{(n)})^{e-1}}).$ Now, note that for any $d,Q^d=P^{-1}AP,$ for some $A\in\Lambda_{\mathbb{F}_q[S]}^{(n)}.$ By hypothesis, there exist $a_1,\ldots,a_m\in\mathbb{F}_q,$ such that, $f_i^{\Lambda_A^{(n)}}=a_1f_1+\cdots+a_mf_m.$ Hence

$$g_{i}^{\Lambda_{Q}^{(n)}} = f_{i}^{\Lambda_{A}^{(n)}\Lambda_{P}^{(n)}} = (a_{1}f_{1} + \dots + a_{m}f_{m})^{\Lambda_{P}^{(n)}} = a_{1}f_{1}^{\Lambda_{P}^{(n)}} + \dots + a_{m}f_{m}^{\Lambda_{P}^{(n)}} \in \mathcal{V},$$
therefore, $\mathcal{W} \subset \mathcal{V}$.

Now, with the notation as in the above corollary, we bound the dimension of the \mathbb{F}_{q^l} -vector space generated by any \mathcal{D} -homogeneous components. For $0 \leq i, j < l, \ 1 \leq k \leq o$, let $f_{i,j,k} := \mathbf{x}^t \Lambda_{S^i}^{(n)} P_k \Lambda_{S^j}^{(n)} \mathbf{x} \in \mathbb{F}_q[x_1, \dots, x_{ln}]$, where $\mathbf{x} = (x_1, \dots, x_{ln})^t$, $P_k \in \mathbb{F}_q^{ln \times ln}$, and $\Lambda_A^{(n)}$ and $S \in \mathbb{F}_q^{l \times l}$ are as in the description of SNOVA. Let $D = \{q^{r_1-1} + q^{r_2-1} \mod q^l - 1 : r_1, r_2 \in \{1, \dots, l\}\}$ be the set of all \mathcal{D} -degrees of quadratic mononomials. We can write $f_{i,j,k}^{\Lambda_p^{(n)}} = \sum_{d \in D} h_{i,j,k,d} \in \mathbb{F}_{q^l}[x_1, \dots, x_{ln}]$, where $h_{i,j,k,d}$ is the \mathcal{D} -degree d homogeneous component of $f_{i,j,k}^{\Lambda_p^{(n)}}$. For $d \in D$, let $\mathcal{H}_d := \{h_{i,j,k,d} \text{ for } i, j \in \{0, \dots, l-1\}, k \in \{1, \dots, o\}\}$,

 $\mathcal{W}_d := \operatorname{\mathsf{Span}}_{\mathbb{F}_{q^l}}(\mathcal{H}_d)$, and $\mathcal{W} := \operatorname{\mathsf{Span}}_{\mathbb{F}_{q^l}}(\bigcup_{d \in D} \mathcal{H}_d)$, so that, $\mathcal{W} = \bigoplus_{d \in D} \mathcal{W}_d$, and $\dim(\mathcal{W}) = \sum_{d \in D} \dim(\mathcal{W}_d)$. By Corollary 1, $\mathcal{W} = \operatorname{\mathsf{Span}}_{\mathbb{F}_{q^l}}(\{f_{i,j,k}^{\Lambda_P^{(n)}} : i, j, k\})$, and hence $\dim(\mathcal{W}) \leq l^2 o$.

Using the notation of Proposition 2, we can express $f_{i,j,k}^{\Lambda_P^{(n)}}$ in terms of a fixed eigenvalue λ of S. Note that

$$f_{i,j,k}^{\Lambda_P^{(n)}}(\mathbf{x}) = \mathbf{x}^t \Lambda_{P^t S^i}^{(n)} P_k \Lambda_{S^j P}^{(n)} \mathbf{x},$$

 $P^tS^i = \operatorname{diag}((\lambda^i)^{q^0}, \dots, (\lambda^i)^{q^{l-1}})P^t$ and $S^jP = \operatorname{diag}((\lambda^j)^{q^0}, \dots, (\lambda^j)^{q^{l-1}})P$, so that with $\Gamma^{(k)} = \Lambda_{P^t}^{(n)} P_k \Lambda_P^{(n)} \in \mathbb{F}_{q^l}^{ln \times ln}$, we can write

$$f_{i,j,k}^{\Lambda_P^{(n)}} = \sum_{a,b \in \{0,\dots,n-1\}} \sum_{r_1,r_2 \in \{1,\dots,l\}} (\lambda^i)^{q^{r_1-1}} (\lambda^j)^{q^{r_2-1}} \Gamma_{al+r_1,bl+r_2}^{(k)} x_{al+r_1} x_{bl+r_2}.$$

As we saw in Example 1, from the perspective of the \mathcal{D} -degree, there are two distinctive types of quadratic monomials. Let $B:=\{d=q^{r_1-1}+q^{r_2-1}\in D: r_1=r_2\in [l]\}$ and $C:=\{d=q^{r_1-1}+q^{r_2-1}\in D: r_1\neq r_2, r_1, r_2\in [l]\}$. Note that $\{B,C\}$ is a partition of D. Moreover, |B|=l and $|C|=\binom{l}{2}$. For $d\in B$,

$$h_{i,j,k,d} = (\lambda^{i+j})^{q^{r_1-1}} \left(\sum_{a=0}^{n-1} \sum_{b=0}^{n-1} \Gamma_{al+r_1,bl+r_1}^{(k)} x_{al+r_1} x_{bl+r_1} \right) = (\lambda^{i+j})^{q^{r_1-1}} h_{0,0,k,d}.$$

hence, for a fixed k and d, all $h_{i,j,k,d}$ are generated by a particular $h_{i_1,j_1,k,d}$, and it follows that $\dim(\mathcal{W}_d) \leq o$.

For $d \in C$,

$$h_{i,j,k,d} = h_{i,j,k,d}^{(1)} + h_{i,j,k,d}^{(2)}$$

where

$$h_{i,j,k,d}^{(1)} = ((\lambda^i)^{q^{r_1-1}}(\lambda^j)^{q^{r_2-1}})(\sum_{a=0}^{n-1}\sum_{b=0}^{n-1}\Gamma_{al+r_1,bl+r_2}^{(k)}x_{al+r_1}x_{bl+r_2})$$

and

$$h_{i,j,k,d}^{(2)} = ((\lambda^i)^{q^{r_2-1}}(\lambda^j)^{q^{r_1-1}})(\sum_{a=0}^{n-1}\sum_{b=0}^{n-1}\Gamma_{al+r_2,bl+r_1}^{(k)}x_{al+r_1}x_{bl+r_2}).$$

Then, for a fixed k and d, all $h_{i,j,k,d}$ are generated by any linearly independent $h_{i_1,j_1,k,d}^{(1)}, h_{i_2,j_2,k,d}^{(2)}$ if such a pair exist. Otherwise, all $h_{i,j,k,d}$ are generated by a particular $h_{i_1,j_1,k,d}$. It follows that $\dim(\mathcal{W}_d) \leq 2o$.

Furthermore, we remark that if $\dim(\mathcal{W}) = l^2 o$, then $\dim(\mathcal{W}_d) = o$ for all $d \in B$ and $\dim(\mathcal{W}_d) = 2o$ for all $d \in C$. Moreover, we observe that after a SNOVA key pair is generated, the event of having $\dim(\mathcal{W}) = l^2 o$ is highly probable.

4 Solving SNOVA Systems

Now we focus on the complexity of solving SNOVA systems, i.e., systems over $\mathbb{F}_q[x_1,\ldots,x_{ln}]$ of the form

$$\mathbf{x}^{t} \left(\Lambda_{S^{i-1}}^{(n)} P_{k} \Lambda_{S^{j-1}}^{(n)} \right) \mathbf{x} = z_{k}^{(i,j)}, \text{ for } (i,k,j) \in [l] \times [o] \times [l],$$
(11)

where $\mathbf{x} = (x_1, \dots, x_{ln})^t$, $P_k \in \mathbb{F}_q^{ln \times ln}$ is a random matrix, and $\Lambda_A^{(n)}$ and $S \in \mathbb{F}_q^{l \times l}$ are as in the description of SNOVA.

In the homogeneous case, i.e., $z_k^{(i,j)} = 0$, if $ln \leq l^2o$, there is a naive way to improve generic solving algorithms. As observed in [8], in this case, if there is a solution **s** to the SNOVA system, then $\Lambda_{S_i}^{(n)}$ **s** is also a solution for each $i \in [l-1]$. Hence, there is either no solution or an l-dimensional vector space of solutions. In this scenario, one can remove l variables and search for a solution of the form $(x_1, \ldots, x_{ln-l}, 1, 0, \ldots, 0)$. Thus, the complexity of solving the SNOVA system would be at least $O(q^l)$ times faster than a random quadratic system with the same dimensions. This approach is less effective when $ln > l^2o$, because we can only specialize $ln - l^2o$ variables for free in the random case.

The algorithm we propose in this section applies to both the underdefined $(ln > l^2o)$ and the overdefined cases $(ln \le l^2o)$. However, we focus on estimating the speed up in the underdefined case, where all the proposed parameters for SNOVA fall into (see Table 1).

4.1 Solving the System of Homogeneous Components

As we will explain in Section 4.2, our strategy to solve underdefined SNOVA systems reduces to solving a multi-homogeneous system. We start by explaining how to solve such a system. More precisely, it is a system of $m = l^2 o$ equations

$$h_1^{(1,1)}(\tilde{\mathbf{x}}) = w_1^{(1,1)}, h_1^{(1,2)}(\tilde{\mathbf{x}}) = w_1^{(1,2)}, \dots, h_o^{(l,l)}(\tilde{\mathbf{x}}) = w_o^{(l,l)},$$
 (12)

over $\mathbb{F}_{q^l}[\tilde{x}_{1,1}, \tilde{x}_{2,1}, \dots, \tilde{x}_{l,n}]$ with the following properties:

- 1. The polynomial set $h_k^{(i,j)}(\tilde{\mathbf{x}})$ is multi-homogeneous with respect to the partition of the variables $\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \dots, \tilde{\mathbf{x}}_l$, where $\tilde{\mathbf{x}}_i := (\tilde{x}_{i,1}, \dots, \tilde{x}_{i,n})$.
- 2. For each $i \in [l]$, the o polynomials $h_1^{(i,i)}, \ldots, h_o^{(i,i)}$ have multi-degree $2\mathbf{e}_i$, where $\mathbf{e}_i \in \mathbb{F}_q^l$ is the i-th canonical vector, that is, they only involve variables from $\mathbf{x}_i := (x_{i,1}, \ldots, x_{i,n})$.
- 3. For each $i \neq j \in [l]$, the 2*o* polynomials $h_1^{(i,j)}, \ldots, h_o^{(i,j)}, h_1^{(j,i)}, \ldots, h_o^{(j,i)}$ have multi-degree $\mathbf{e}_i + \mathbf{e}_j$, that is, they are bilinear in the sets \mathbf{x}_i and \mathbf{x}_j .

In what follows, we abuse of the notation and we write $h_k^{(i,j)}(\tilde{\mathbf{x}}) = h_k^{(i,j)}(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j)$, and $h_k^{(i,i)}(\tilde{\mathbf{x}}) = h_k^{(i,i)}(\tilde{\mathbf{x}}_i)$.

The following algorithm aims at finding a solution to the system in Eq. (12) of the form $\Lambda_{P^{-1}}^{(n)} \mathbf{s} \in \mathbb{F}_{q^l}^{ln}$ for some $\mathbf{s} \in \mathbb{F}_q^{ln}$. It is parameterized by a nonnegative integer k multiple of l.

- 1. Sample $(s_{ol^2+1}, \ldots, s_{ln})^t \in \mathbb{F}_q^{(n-ol)l}$.
- 2. Sample $(s_{ol^2-k+1}, \ldots, s_{ol^2})^t \in \mathbb{F}_q^k$ and compute

$$\tilde{\mathbf{s}}_2 := \Lambda_{P^{-1}}^{(n-ol+\frac{k}{l})} \cdot (s_{ol^2-k+1}, \dots, s_{ln})^t.$$

3. Use the XL algorithm as described in Section 2.4 to find $\tilde{\mathbf{s}}_1 \in \mathbb{F}_{q^l}^{ol^2-k}$ such that

$$h_1(\tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2) = w_1, \dots, h_m(\tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2) = w_m.$$

If no solution is found, go back to Step 2.

4. Set $\mathbf{s} = \Lambda_P^{(n)}(\tilde{\mathbf{s}}_1^t, \tilde{\mathbf{s}}_2^t)^t$. If $\mathbf{s} \in \mathbb{F}_q^{ln}$, output \mathbf{s} . Otherwise, go to Step 2.

We are trying to find a solution $(\tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2)$ to Eq. (12) such that $\Lambda_P^{(n)}(\tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2)$ is a vector in the small field. If $\tilde{\mathbf{s}}_2$ is part of such a solution, since the system $h_1(\mathbf{y}, \tilde{\mathbf{s}}_2) = w_1, \dots, h_m(\mathbf{y}, \tilde{\mathbf{s}}_2) = w_m$ is overdefined, one expects it to have only a few solutions, hence we expect to find $\tilde{\mathbf{s}}_1$ by solving such a partially evaluated system few times.

The number of $\mathbb{F}_{q^l}\text{-multiplications}$ of the algorithm described above is upper bounded by

$$\min_{k \in [ol^2], \ l|k} q^k \cdot 3 \left(ol - \frac{k}{l} \right)^2 \cdot \left[\overline{\mathcal{M}}(\mathbf{d}_{sol}) \right]^2, \tag{13}$$

where $\mathbf{d}_{sol} \in \mathbb{Z}_{\geq 0}^l$ minimizes $\overline{\mathcal{M}}(\mathbf{d}_{sol})$, which is the number of monomials of multi-degree smaller than \mathbf{d}_{sol} , and there exists $(d_1, \ldots, d_l) \leq \mathbf{d}_{sol}$ such that the coefficient of $t_1^{d_1} t_2^{d_2} \cdots t_l^{d_l}$ in the series

$$\frac{\prod_{1 \le i < j \le l} (1 - t_i t_j)^{2o} \cdot \prod_{i=1}^{l} (1 - t_i^2)^o}{\prod_{i=1}^{l} (1 - t_i)^{ol-k/l+1}}$$
(14)

is non-negative.

4.2 Solving Underdefined SNOVA Systems

To simplify the notation, let us write the SNOVA system given in Eq. (11) as

$$f_1(x_1,\ldots,x_{ln})=z_1,\ldots,f_{ol^2}(x_1,\ldots,x_{ln})=z_{ol^2}.$$

We now present an algorithm to solve underdefined SNOVA systems. This algorithm is parameterized by a non-negative k multiple of l, and it performs the following steps:

- 1. For each $i \in [ol^2]$, compute $f_i^{A_P^{(n)}}$, where P is the matrix of eigenvectors of S, as described in Lemma 1.
- 2. Compute a basis $(h_1^{(1,1)}, h_1^{(1,2)}, \dots, h_o^{(l,l)}) \subset \mathcal{H}$ of the vector space spanned by the $f_i^{\Lambda_P^{(n)}}$, where \mathcal{H} is the set of \mathcal{D} -homogeneous components of the $f_i^{\Lambda_P^{(n)}}$. By Corollary 1, such a basis exists.

3. Compute the invertible matrix $B \in \mathbb{F}_{q^l}^{m \times m}$ such that

$$(f_1^{A_P^{(n)}}, \dots, f_n^{A_P^{(n)}})^t = B \cdot (h_1^{(1,1)}, h_1^{(1,2)}, \dots, h_o^{(l,l)})^t,$$

and define $(w_{1,1,1},\ldots,w_{l,l,o})^t = B^{-1}(z_1,\ldots,z_{ol^2})^t$.

4. Use the algorithm described in Section 4.1 to find $\mathbf{s} \in \mathbb{F}_q^{ln}$ such that

$$h_1^{(1,1)}(\Lambda_{P^{-1}}^{(n)}\mathbf{s}) = w_{1,1,1}, \dots, h_o^{(l,l)}(\Lambda_{P^{-1}}^{(n)}\mathbf{s}) = w_{l,l,o}.$$
(15)

5. Output s.

In Section 3.1, we showed that the system Eq. (15) satisfies the first three properties of the system described in Section 4.1 with high probability. Hence, it is correct to apply the algorithm described in Section 4.1 to solve the system at Step 4. Moreover, we expect such an algorithm to successfully output a solution $\mathbf{s} \in \mathbb{F}_q^{ln}$ because after each specialization of the last $ln - ol^2$ variables to any vector $(s_{ol^2+1}, \ldots, s_{ln})$, the specialized SNOVA system

$$f_i(x_1, \dots, x_{ol^2}, s_{ol^2+1}, \dots, s_{ln}) = z_i$$
, for each $i \in [ol^2]$.

has one solution $s_1, \ldots, s_{ol^2} \in \mathbb{F}_q^{ol^2}$ with high probability, since it is a well-defined system. In such a case, the vector $\tilde{\mathbf{s}} = \Lambda_{P^{-1}}^{(n)} \cdot (s_1, \ldots, s_{ln})^t$ is a solution to the system in Eq. (15).

The complexity of the algorithm described above is clearly dominated by the complexity of step 4. Hence, we use the complexity formula given in Eq. (13) to estimate the complexity of solving underdefined SNOVA systems.

Fig. 1 shows bit complexity estimates of the algorithm presented in this section to solve underdefined SNOVA systems with l(o + v) variables and ol^2 equations over \mathbb{F}_{16} , where v is chosen in the same regime of the parameters proposed for SNOVA, see Table 1. For comparison, we include the bit complexity estimates of solving random quadratic systems of the same dimensions.

The estimates for SNOVA systems are computed using Eq. (13). In this case, for every \mathbb{F}_{16^l} -multiplication we assign a cost of $2(\log_2(16^l))^2 + \log_2(16^l)$ bit operations. For a random system, we found that the best strategy for the specific regime of parameters is to fix the $ln-ol^2$ extra variables and then use the hybrid-XL algorithm. These complexity estimates are computed using the MQEstimator [5], which assigns a cost of $2(\log_2 16)^2 + \log_2(16)$ bit operations per \mathbb{F}_{16} -multiplication. From our estimates, we observe that solving an underdefined SNOVA system is easier than its corresponding random version by a factor of $O(q^l)$. For example, for (l,o)=(5,4), the SNOVA systems are 22 bits easier, while for (l,o)=(4,6),(3,10), and (2,18) the differences are 15, 10, and 5, respectively.

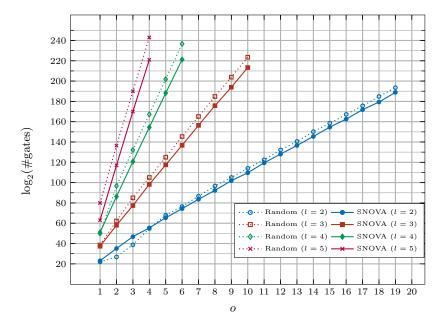


Fig. 1. Comparison of the bit complexity of solving underdefined random systems and SNOVA systems.

5 Attacking the Original Version of SNOVA

5.1 Improved Forgery Attack

As seen in Section 2.2, the SNOVA public key can be written as

$$\mathcal{P}(U) = \sum_{i=0}^{l-1} \sum_{j=0}^{l-1} E_{i,j} \mathcal{B}(\mathbf{u}_i, \mathbf{u}_j),$$

where \mathcal{B} is the bilinear map associated to the SNOVA sequence shown in Eq. (11), and $\mathbf{u}_0, \dots \mathbf{u}_{l-1}$ are ln-tuples of variables. Following a similar approach as Beullens in [4], but using the simpler change of variable $\mathbf{u}_0 = \mathbf{u}$ and $\mathbf{u}_i = \Lambda_{R_i}^{(n)} \mathbf{u}$ for $i = 1, \dots, l-1$, with $R_i = r_{i,0}S^0 + \dots + r_{i,l-1}S^{l-1} \in \mathbb{F}_q[S]$, it follows that

$$\mathcal{P}(\mathbf{u}) = E \cdot \mathcal{F}(\mathbf{u}),$$

where $\mathcal{F}(\mathbf{u}) = \mathcal{B}(\mathbf{u}, \mathbf{u})$, and $E = \Lambda_{\tilde{E}}^{(o)} \in \mathbb{F}_q^{ol^2 \times ol^2}$ is a block-diagonal matrix. Depending on the choice of R_i , \tilde{E} can have rank defect $r \leq l^2$. Besides the advantage of the rank defect of E, this change of variable allows us to exploit the \mathcal{D} -homogeneous structure of the system after we compose with $\Lambda_P^{(n)} \in \mathbb{F}_q^{ln \times ln}$, where P is the matrix of eigenvector described in Lemma 1.

For a target message $\mathsf{msg} \in \{0,1\}^*$, the attacker samples a salt string $\mathsf{salt} \in \{0,1\}^*$ until $\mathbf{z} = \mathsf{Hash}(\mathsf{msg}\|\mathsf{salt}) \in \mathbb{F}_q^{ol^2}$ falls in the column space of E and finds $\mathbf{w} \in \mathbb{F}_q^{ol^2}$ such that

$$z = Ew$$

The cost of finding such salt and \mathbf{z} is $O\left(q^{(l^2-r)\cdot o}\cdot l^6\right)$.

Since E has rank ro, the attacker can efficiently find a set of $p = ol^2 - or$ linearly independent vectors $\mathbf{w}_1, \dots, \mathbf{w}_p \in \mathbb{F}_q^{ol^2}$ in the right-kernel of E. Thus, the attacker obtains a forgery for msg by finding $\mathbf{u} \in \mathbb{F}_q^{ln}$ and $y_1, \dots, y_p \in \mathbb{F}_q$ such that

$$\mathbf{w} = \mathcal{F}(\mathbf{u}) + \sum_{i=1}^{p} y_i \cdot \mathbf{w}_i. \tag{16}$$

If he succeeds, the attacker outputs $U = [\mathbf{u}|\Lambda_{R_1}^{(n)}\mathbf{u}|\cdots|\Lambda_{R_{l-1}}^{(n)}\mathbf{u}]$ along with the salt as a forged signature.

In Section 3.1, we show that, there is a \mathcal{D} -homogeneous sequence \mathcal{H} such that $\mathcal{F}(\Lambda_P^{(n)}\mathbf{v}) = B \cdot \mathcal{H}(\mathbf{v})$, where $B \in \mathbb{F}_{q^l}^{ol^2 \times ol^2}$ is invertible. Then, with the change of variables $\mathbf{u} = \Lambda_P^{(n)}\mathbf{v}$, the system Eq. (16) becomes

$$\mathbf{w} = B \cdot \mathcal{H}(\mathbf{v}) + \sum_{i=1}^{p} y_i \cdot \mathbf{w}_i,$$

$$\underbrace{B^{-1}\mathbf{w}}_{:=\tilde{\mathbf{w}}} = \mathcal{H}(\mathbf{v}) + \sum_{i=1}^{p} y_i \cdot \underbrace{B^{-1}\mathbf{w}_i}_{:=\tilde{\mathbf{w}}_i},$$

a quadratic system whose quadratic part is \mathcal{D} -homogeneous.

Note that the system above has ol^2 equations and ln + p variables, so the attacker can specialize up to $ln + p - ol^2$ variables and still expect a solution to the system. If the attacker specializes $ln - ol^2 + k \ge ln - ol^2 + p$ variables, we expect a solution with probability q^{p-k} , hence he must try q^{k-p} values in order to expect a solution. In order to preserve the \mathcal{D} -homogeneous structure, we choose k to be a multiple of l greater or equal to p and specialize $ln - ol^2 + k$ variables. More precisely, the attacker runs the following algorithm

1. Sample $(s_{ol^2-k+1},\ldots,s_{ln})^t \in \mathbb{F}_q^{(n-ol)l+k}$ and compute

$$\tilde{\mathbf{v}}_2 := \Lambda_{P^{-1}}^{((n-ol)+\frac{k}{l})} \cdot (s_{ol^2-k+1}, \dots, s_{ln})^t;$$

2. Find $(\tilde{\mathbf{v}}_1, (t_1, \dots, t_p)^t) \in \mathbb{F}_{q^l}^{ol^2 - k} \times \mathbb{F}_q^p$ such that

$$\tilde{\mathbf{w}} = \mathcal{H}(\tilde{\mathbf{v}}_1, \tilde{\mathbf{v}}_2) + \sum_{i=1}^p t_i \cdot \tilde{\mathbf{w}}_i;$$

If no solution is found, go back to Step 1.

3. Set $\mathbf{s} = \Lambda_P^{(n)}(\tilde{\mathbf{v}}_1^t, \tilde{\mathbf{v}}_2^t)^t$. If $\mathbf{s} \in \mathbb{F}_q^{ln}$, output \mathbf{s} . Otherwise, go to Step 1.

To solve the quadratic system at step 2, the attacker uses a slight modification of the algorithm described in Section 4.1, in which we restrict the monomials used in the XL part to be monomials in the variables $\mathbf{v}_1 = (v_1, \dots, v_{ol^2-k})^t$.

In this case the working multi-degree would be a $\mathbf{d}^* \in \mathbb{Z}_{>0}^l$ such that

$$0 \ge [\mathbf{t^{d^*}}]H(t_1, \dots, t_l) - p \cdot \sum_{\mathbf{d} < \mathbf{d^*}} [\mathbf{t^d}]H(t_1, \dots, t_l),$$

where

$$H(t_1, \dots, t_l) = \frac{\prod_{1 \le i < j \le l} (1 - t_i t_j)^{2o} \cdot \prod_{i=1}^{l} (1 - t_i^2)^{o}}{\prod_{i=1}^{l} (1 - t_i)^{ol-k/l}}.$$

The number of \mathbb{F}_{q^l} -multiplications of the algorithm described above is upper bounded by

$$\min_{p \le k \le ol^2, \ l|k} q^{k-p} \cdot 3\left(ol - \frac{k}{l}\right)^2 \cdot \left[\widetilde{\mathcal{M}}(\mathbf{d}^*)\right]^2, \tag{17}$$

with $\widetilde{\mathcal{M}}(\mathbf{d}^*) = \mathcal{M}(\mathbf{d}^*) + p \cdot \sum_{\mathbf{d} \leq \mathbf{d}^*} \mathcal{M}(\mathbf{d})$, where $\mathcal{M}(\mathbf{d})$ gives the number of monomials of multi-degree exactly \mathbf{d} .

Table 2 shows the estimated complexity of this approach in comparison to [4]. For the largest values of r, our approach is faster. The cost of finding salt, such that $\mathbf{z} = \mathsf{Hash}(\mathsf{msg}\|\mathsf{salt})$ falls in the column space of E, penalizes our approach for smaller values of r. Our approach tends to be more effective for larger values of l, because we are exploiting the \mathcal{D} -homogeneous structure of the system. This vulnerability puts the parameters sets with l=4 below the security threshold defined by NIST (see Section 2.1). These are precisely the parameters that allow the smallest public keys, public key and signatures sizes, and the ones that are not significantly affected by the attack in [4].

5.2 Improved Reconciliation Attack

The reconciliation attack is a key-recovery attack for SNOVA that involves finding a unique vector in the secret space \mathcal{O} that is solution of an underdefined homogeneous SNOVA system with lv variables and ol^2 equations, see Section 2.3.

Our improved reconciliation attack consists of repitedidly apply the algorithm in Section 4.2 by systematically sampling all possible vector at Steps 1 and 2 in the subroutine described in Section 4.1.

the subroutine described in Section 4.1. We expect to iterate at most q^{lv-ol^2} over Step 1 in the subroutine. Therefore, the complexity of our forgery attack is given by q^{lv-ol^2} multiplied by the complexity of one iteration of the algorithm to solve SNOVA systems. Table 3 shows the bit complexity estimates of our improved reconciliation attack compared with the ones [7,8,10].

6 Attacking the Alternative Versions of SNOVA

While this manuscript was being written the SNOVA team released a preprint [16] with two alternatives to modify their scheme with the goal of counteracting Buellen's attack [4].

		Attack cost				
Security	parameters	r	previous	our	fraction of	cost of finding
level	(v, o, q, l)		best	attack	weak keys	a weak key
		3	137	114	100%	14
I	(37, 17, 16, 2)	2	97	142	$2^{-8.9}$	26
		1	45	210	$2^{-17.1}$	35
		7	150	122	100%	29
I	(25, 8, 16, 3)	6	130	106	$2^{-12.0}$	49
		5	112	138	$2^{-40.0}$	77
		13	167	135	100%	52
I	(24, 5, 16, 4)	12	156	124	2^{-16}	80
		11	145	112	2^{-52}	116
_		3	189	153	100%	14
III	(56, 25, 16, 2)	2	132	206	$2^{-8.9}$	26
		1	68	306	$2^{-17.1}$	35
		7	194	155	100%	29
III	(49, 11, 16, 3)	6	169	142	$2^{-12.0}$	49
		5	143	186	$2^{-40.0}$	77
		13	253	202	100%	52
III	(37, 8, 16, 4)	12	235	179	2^{-16}	80
		11	218	172	2^{-52}	116
		3	240	189	100%	14
V	(75, 33, 16, 2)	2	167	270	$2^{-8.9}$	26
		1	88	402	$2^{-17.1}$	35
		7	253	201	100%	29
V	(66, 15, 16, 3)	6	221	190	$2^{-12.0}$	49
		5	187	250	$2^{-40.0}$	77
		13	307	246	100%	52
V	(60, 10, 16, 4)	12	285	214	2^{-16}	80
		11	264	212	2^{-52}	116

Table 2. Estimated cost of our attack for SNOVA parameters, compared with [4], which indicated as previous best. The estimates in rightmost columns are taken from [4]. The fraction of weak keys indicates the fraction of the keys for which there exist $R_1, \ldots, R_{l-1} \in \mathbb{F}_q[S]$ such that the matrix associated matrix \tilde{E} has a minimal rank r. The cost of finding a weak key indicates the cost to find a public key and the R_i such that the associated matrix \tilde{E} has rank r.

Alternative 1. A first alternative proposed by the SNOVA team is choose random matrices $A_{k,\alpha}, B_{k,\alpha} \in \mathcal{R}$ and $Q_{k,\alpha 1}, Q_{k,\alpha 2} \in \mathbb{F}_q[S]$, for $k \in [o]$ and

Security level	$\begin{array}{c} \text{parameters} \\ (v, o, q, l) \end{array}$	previous best reconciliation attack	our attack
I	(37, 17, 16, 2)	197	194
	(25, 8, 16, 3)	196	185
	(24, 5, 16, 4)	269	249
III	(56, 25, 16, 2)	289	287
	(49, 11, 16, 3)	438	423
	(37, 8, 16, 4)	387	365
V	(75, 33, 16, 2)	377	378
	(66, 15, 16, 3)	574	558
	(60, 10, 16, 4)	695	673

Table 3. Bit complexities of our reconciliation attack compared with the previous best reconciliation attack for SNOVA.

 $\alpha \in [l^2]$, and define the k-th coordinate of the public map $\mathcal{P}(U)$ as

$$\mathcal{P}_{k}(U_{1},\ldots,U_{n}) = \sum_{\alpha=1}^{l^{2}} \sum_{i=1}^{n} \sum_{j=1}^{n} A_{k,\alpha} \cdot U_{i}^{t}(Q_{k,\alpha 1}P_{k,i,j}Q_{k,\alpha 2})U_{j} \cdot B_{k,\alpha}.$$

As pointed out in [17], these changes do not affect the size of the public key, since these extra matrices can be generated from the public seed as previously done. Additionally, they claimed these changes will effectively reduce the impact of Buellen's attack, since the matrix E becomes a block diagonal matrix with different diagonal blocks in general.

Security Level	(v,o,q,l)	lower bound of the rank	r	our attack	previous best
I	(37, 17, 16, 2)	52	52	118	144
	(25, 8, 16, 3)	49	49	102	143
	(24, 5, 16, 4)	52	55	112	146
III	(56, 25, 16, 2)	79	79	166	208
	(49, 11, 16, 3)	83	83	176	210
	(37, 8, 16, 4)	78	91	164	210
V	(75, 33, 16, 2)	106	106	214	272
	(66, 15, 16, 3)	110	110	218	272
	(60, 10, 16, 4)	113	114	196	273

Table 4. Bit complexity of our forgery attack compared with the best attack for SNOVA modified with the alternative 1.

Table 4 shows the complexity estimates of our forgery attack for the alternative 1 version of SNOVA. The lower bound of the rank indicates the minimum rank of the matrix E for the corresponding set of parameters, and these numbers were empirically estimated in [17]. The previous best indicates the complexity of the attack in [4]. Our attacks improves the state-of-the-art for the particular modified version of SNOVA, and the security of all proposed parameters fall below the threshold for the corresponding security level.

Alternative 2. A second alternative proposed in [16] is force the matrix E to always have full rank. The k-th coordinate of the public map $\mathcal{P}(U)$ is defined as

$$\mathcal{P}_{k}(U) = \sum_{\alpha=1}^{l^{4}} \sum_{i=1}^{n} \sum_{j=1}^{n} A_{\alpha} \cdot U_{i}^{t}(Q_{\alpha 1}P_{k,i,j}Q_{\alpha 2})U_{j} \cdot B_{\alpha},$$

where the matrices A_{α} , $B_{\alpha} \in \mathcal{R}$, and $Q_{\alpha 1}$, $Q_{\alpha 2} \in \mathbb{F}_q[S]$, for $\alpha \in [l^4]$, are determined by fixed matrices $\tilde{E}_{i,j}$, for $i, j \in [l]$, so that E has full rank.

Security	(v,o,q,l)	previous best	our
Level		direct attack	attack
I	(37, 17, 16, 2)	165	164
	(25, 8, 16, 3)	171	164
	(24, 5, 16, 4)	184	172
III	(56, 25, 16, 2)	234	232
	(49, 11, 16, 3)	226	221
	(37, 8, 16, 4)	287	271
V	(75, 33, 16, 2)	302	300
	(66, 15, 16, 3)	302	296
	(60, 10, 16, 4)	350	340

Table 5. Bit complexity of our attack compared with the previous best direct attack for SNOVA modified with the alternative 2.

Table 5 shows the complexity estimates of our forgery attack for the alternative 2 version of SNOVA. Our attack improves the previous best direct attack for the particular modified version of SNOVA, with speedup factors ranging between 2 and 2^{16} .

Remark 1. We remark that the proposed alternatives do not affect the complexity of our reconciliation attack. Hence the complexity for both alternative versions are shown in Table 3.

References

- Luk Bettale, Jean-Charles Faugère, and Ludovic Perret. Hybrid approach for solving multivariate systems over finite fields. *Journal of Mathematical Cryptology*, 3(3):177–197, 2009.
- 2. Ward Beullens. Improved cryptanalysis of UOV and Rainbow. In Anne Canteaut and François-Xavier Standaert, editors, *Advances in Cryptology EUROCRYPT 2021*, pages 348–373, Cham, 2021. Springer International Publishing.
- 3. Ward Beullens. MAYO: Practical post-quantum signatures from oil-and-vinegar maps. In Riham AlTawy and Andreas Hülsing, editors, *Selected Areas in Cryptog-raphy*, pages 355–376, Cham, 2022. Springer International Publishing.
- Ward Beullens. Improved cryptanalysis of SNOVA. Cryptology ePrint Archive, Paper 2024/1297, 2024.
- 5. Andre Esser, Javier Verbel, Floyd Zweydinger, and Emanuele Bellini. Sok: Cryptographicestimators a software library for cryptographic hardness estimation. In *Proceedings of the 19th ACM Asia Conference on Computer and Communications Security*, ASIA CCS '24, page 560–574, New York, NY, USA, 2024. Association for Computing Machinery.
- Jean-Charles Faugère and Jules Svartz. Gröbner bases of ideals invariant under a commutative group: the non-modular case. In Proceedings of the 38th International Symposium on Symbolic and Algebraic Computation, ISSAC '13, page 347–354, New York, NY, USA, 2013. Association for Computing Machinery.
- 7. Yasuhiko Ikematsu and Rika Akiyama. Revisiting the security analysis of SNOVA. Proceedings of the 11th ACM Asia Public-Key Cryptography Workshop, 2024.
- 8. Peigen Li and Jintai Ding. Cryptanalysis of the SNOVA signature scheme. In *International Conference on Post-Quantum Cryptography*, pages 79–91. Springer, 2024.
- Shuhei Nakamura, Yasuhiko Ikematsu, Yacheng Wang, Jintai Ding, and Tsuyoshi Takagi. New complexity estimation on the Rainbow-Band-Separation attack. Theoretical Computer Science, 896:1–18, 2021.
- Shuhei Nakamura, Yusuke Tani, and Hiroki Furue. Lifting approach against the SNOVA scheme. Cryptology ePrint Archive, Paper 2024/1374, 2024.
- 11. Shuhei Nakamura, Yacheng Wang, and Yasuhiko Ikematsu. A new analysis of the Kipnis-Shamir method solving the minrank problem. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, E106.A(3):203–211, 2023.
- 12. National Institute of Standards and Technology. Call for additional digital signature schemes for the post-quantum cryptography standardization process. NIST Web Page, 2022. https://csrc.nist.gov/csrc/media/Projects/pqc-dig-sig/documents/call-for-proposals-dig-sig-sept-2022.pdf.
- 13. Ray Perlner and Daniel Smith-Tone. Rainbow band separation is better than we thought. Cryptology ePrint Archive, Paper 2020/702, 2020.
- Peter W Shor. Algorithms for quantum computation: discrete logarithms and factoring. In Proceedings 35th annual symposium on foundations of computer science, pages 124–134. Ieee, 1994.
- 15. Victor Shoup. A Computational Introduction to Number Theory and Algebra. Cambridge University Press, 2005. https://shoup.net/ntb/ntb-v2.pdf.
- Lih-Chung Wang, Chun-Yen Chou, Jintai Ding, Yen-Liang Kuan, Jan Adriaan Leegwater, Ming-Siou Li, Bo-Shu Tseng, Po-En Tseng, and Chia-Chun Wang. A note on the SNOVA security. Cryptology ePrint Archive, Paper 2024/1517, 2024.

17. Lih-Chung Wang, Po-En Tseng, Yen-Liang Kuan, and Chun-Yen Chou. A simple noncommutative UOV scheme. Cryptology ePrint Archive, Paper 2022/1742, 2022.