Designs, Codes and Cryptography (2020) 88:1023–1036
https://doi.org/10.1007/s10623-020-00723-0

All in the C^* family

Ryann Cartor[1] · Daniel Smith-Tone[1,2]

Received: 7 June 2019 / Revised: 9 January 2020 / Accepted: 18 January 2020 / Published online: 26 February 2020

Abstract
The C^* scheme of Eurocrypt 88 is the progenitor of all of the so called "big field" schemes of multivariate cryptography. The introduction, primarily by Patarin, of the numerous modifiers of multivariate schemes have produced several variants that stay faithful to the central structure of the original. From the tumultuous history of C^* derivatives we now see only a very few survivors in the cryptonomy. In this work, we revisit the roots of multivariate cryptography, investigating the viability of C^* schemes, in general, under the entire multidimensional array of the principal modifiers. We reveal that there is a nontrivial space of combinations of modifiers that produce viable schemes resistant to all known attacks. This solution space of seemingly secure C^* variants offers trade-offs in multiple dimensions of performance, revealing a family that can be optimized for disparate applications.

Keywords Multivariate cryptograpy · MinRank · Big field schemes

Mathematics Subject Classification 11T71 · 14G50 · 94A60

1 Introduction

Post-quantum cryptography is now mainstream, with a great deal of attention devoted to the post-quantum standardization "competition" of the National Institute of Standards and Technology (NIST). NIST's project strives to offer protection against prospective quantum adversaries by leveraging expertise in the international community in areas of cryptography not known to be vulnerable to quantum cryptanalysis. As of the writing of this article, NIST is knee deep analyzing the round one candidates of this project; details can be found at [8].

Communicated by O. Ahmadi.

✉ Ryann Cartor
ryann.cartor@louisville.edu

Daniel Smith-Tone
daniel.smith@nist.gov

[1] Department of Mathematics, University of Louisville, Louisville, KY, USA

[2] National Institute of Standards and Technology, Gaithersburg, MD, USA

One family of candidate quantum-safe public key algorithms rely on the difficulty of solving systems of nonlinear multivariate equations. The history of the multivariate family of cryptosystems is dominated by digital signature schemes. This fact is reflected in the multivariate candidate algorithms in the NIST project, essentially all of which are signature schemes.

One subclass of multivariate schemes are the so-called "big field" schemes, descendants of the C^* scheme of [20]. The cryptanalysis of this scheme and the subsequent development of modifiers and their analyses fueled the evolution of multivariate cryptography for a long time. Even a couple of the round one candidates of the NIST project, GeMSS and Gui, see [6] and [13], can be considered of this lineage, though they are truly no longer C^*-like schemes, having scrapped the central monomial map for a more general polynomial.

Some natural questions to ask are: What have we missed? Where are the multivariate encryption schemes? Have we exhausted C^*? In this work, we show that there is a rich array of schemes which are based on C^*, utilize the standard and well studied modifiers, resist known attacks and are waiting to be further analyzed and optimized.

We specify a generalized version of C^* system incorporating the most studied modifiers and derive its security and performance properties based on its multiple parameters. We establish the constraints for these parameters to guarantee security against known attacks and then consider optimizations for various performance criteria in the case of instantiations for signatures as well as encryption. In the case of signatures, the optimizations seem to converge on a single well-known scheme, whereas in the case of encryption there seems to be a multidimensional array of feasible schemes with similar performance properties that can be optimized for different criteria.

2 The C^* scheme and modifiers

At Eurocrypt '88, the first public key cryptosystem that we can call inherently multivariate, in the sense of requiring many variables, was introduced. The C^* scheme, see [20], exploited the property of extensions of finite fields that are simultaneously rings with their own definition of multiplication and vector spaces over the base field. Such a setup became known as the "big field" construction.

Let $\mathbb{F}_q$ be a finite field and let $\mathbb{K}$ be a degree d extension. The C^* scheme uses this construction to obtain a multivariate representation of the univariate map $f : \mathbb{K} \to \mathbb{K}$ defined by $f(x) = x^{q^\theta+1}$, where $\gcd(q^\theta + 1, q^d - 1) = 1$ to guarantee invertibility. Since f is the product of the identity function and a Frobenius automorphism, both of which are $\mathbb{F}_q$-linear, we call such a function $\mathbb{F}_q$-quadratic. The central map f is hidden by a polynomial morphism. In this case, the invertible affine maps T and U both map from $\mathbb{F}_q^d$ to $\mathbb{F}_q^d$. Thus the public key is given by $P = T \circ \phi^{-1} \circ f \circ \phi \circ U$ where $\phi : \mathbb{F}_q^d \to \mathbb{K}$ is an $\mathbb{F}_q$-vector space isomorphism, see Figure 1.

Encryption of a plaintext $x \in \mathbb{F}_q^d$ is accomplished by evaluating the public polynomials P at x. Decryption is accomplished by inverting each of the three component maps individually. The inversion of $v = f(u)$ is performed by solving $h(q^\theta + 1) = 1(\bmod q^d - 1)$, and calculating $u = v^h$. Thus, in a sense, C^* is a multivariate version of RSA, transporting the idea of inverting a power function on the unit group of a finite integer ring $\mathbb{Z}/d\mathbb{Z}$ to the multiplicative group of a field. The original intention for the C^* scheme was encryption, but the primitive could also be utilized in digital signature algorithms.

$$\begin{tikzcd}
 & \mathbb{K} \arrow[r, "f"] & \mathbb{K} \arrow[d, "\phi^{-1}"] & \\
\mathbb{F}_q^d \arrow[r, "U"] \arrow[rrr, dashed, bend right, "P"'] & \mathbb{F}_q^d \arrow[u, "\phi"] & \mathbb{F}_q^d \arrow[r, "T"] & \mathbb{F}_q^d
\end{tikzcd}$$

Fig. 1 The C^* scheme. Given the $\mathbb{F}_q$-quadratic C^* monomial f, the $\mathbb{F}_q$ vector space isomorphism ϕ and affine maps T and U, we compute the vector-valued function P

C^* was broken by Patarin in 1995 using linearization equations in [24]. The goal of the attack is to use the public quadratic formulae to discover bilinear relations between the plaintext and cipher text variables. Modifiers for the scheme were introduced in [25] in the hopes of regaining security.

One such modifier is the minus modifier $(-)$, which eliminates a equations from the public key. The plus modifier (+) adds p random equations to the public key. Another modifier introduced is the projection modifier (p). The idea of projection is to fix the value of $d-n$ input variables. We call the codimension of this projection $t := d-n$.

The vinegar modifier (v) adds extra variables into the public key that can be assigned random values upon inversion. The effect of adding vinegar variables is that new quadratic terms, formed from both products of vinegar variables and C^* variables and products among vinegar variables, increase the Q-rank of the public key, that is, its rank as a quadratic form over $\mathbb{K}$. Vinegar variables are typically applied to Hidden Field Equation (HFE) schemes.

The final modifier we will discuss is internal perturbation (ip). For this modifier, we will consider a public key $P : \mathbb{F}_q^n \to \mathbb{F}_q^m$. We will denote the ith equation of the public key as p_i, and plain-text vector $\overline{x} := (x_1, x_2, \ldots, x_n) \in \mathbb{F}_q^n$. Consider an affine map $S : \mathbb{F}_q^n \to \mathbb{F}_q^s$, and denote $\overline{z} := S(\overline{x})$. Additionally, consider $Q : \mathbb{F}_q^s \to \mathbb{F}_q^m$ to be a set of m quadratic equations, where q_i denotes the ith. equation. We can create the internally perturbed public key $\tilde{P}$ by defining $\tilde{p}_i(\overline{x}) = p_i(\overline{x}) + q_i(\overline{z})$ for each $0 < i \leq m$. The support dimension of (ip) will be denoted as s.

3 Known combinations of modifiers

The unmodified C^* scheme is easy to attack with the help of hindsight and twenty years of serious development in multivariate cryptography. Not only is C^* vulnerable to Patarin's linearization equations attack of [24], the scheme is also weak against differential methods such as [17], can be broken by finding the extension field structure via the techniques of [16] and is easily defeated by MinRank methods such as [2].

Several attempts at encryption and signatures derived from C^* have been proposed over the years using the modifiers of the previous section. Each of the modifiers has a critical weakness. The minus modifier is vulnerable to differential cryptanalysis as the practical attack on SFLASH of [15] shows. The plus modifier does not increase the MinRank of the scheme and so MinRank attacks are effective and reveal the output basis, breaking the scheme. The projected C^* scheme is still vulnerable to a differential attack as shown in [3], as is the internally perturbed C^* scheme, see [17]. The vinegar variant transforms C^* into a particularly bad HFEv scheme unless some hack making it similar to (ip) is applied. Interestingly, none of the modifiers alone are sufficient to secure C^*.

Combinations of these modifiers have seen some greater success. After the attack in [17] of the original PMI scheme of [9], an ipC^* scheme, PMI+, proposing the combination of the (ip) and (+) modifiers was presented in [10]. Similarly, PFLASH, see [7,12] uses both the projection and minus modifiers, as does EFLASH, see [5].

We present in Appendix A a summary of the security properties of C^* schemes with various combinations of modifiers. For brevity we are not exhaustive, and as a general rule it seems that combinations of modifiers each of which are weak against attack A tend to remain weak against attack A.

Two things are important to note about this summary. First is that there is a notable exception to the rule of the previous paragraph. While it has been shown that pC^* schemes and C^{*-} schemes are both vulnerable to differential attacks, it was proven in [28] and generalized in [4] that the combination of projection with the minus modifier renders C^* invincible from differential attacks. Thus the combination (p$-$) is resistant to differential adversaries though (p) and ($-$) are weak. Second, resistance to these attack models is typically not binary but parametrized by the modifier. In particular, though the (ip) modifier provides resistance to MinRank attacks, the original parameters of PMI+ are easily broken by a simple modification of [2] and still larger parameters can be defeated by the new MinRank techniques developed in [14]. (Embarrassingly, we can find no reference to either such attack on PMI+.) We will revisit this analysis in Sect. 5 where we will generalize the analysis of the family of C^* schemes to the full array of possible schemes with all of their parameters.

4 The C^* schema

In this section, we describe a C^* construction that is as general as possible using the modifiers of the previous section. We parametrize this generalized C^* scheme with the values n, the number of variables, d, the degree of the extension, s, the support dimension of the (ip) modifier, a, the number of public equations removed, p, the number of plus polynomials and t, the corank of the projection. We call the resulting scheme a (n, d, s, a, p, t) scheme.

The only modifier whose use we do not consider in the C^* framework is the vinegar modifier, (v). Directly applying the vinegar modifier produces a degenerate HFEv scheme, since inversion must be accomplished via Berlekamp's Algorithm, see [1], or something similar. Such a scheme is not interesting, since it is no better performing than an HFEv scheme while having a smaller key space and worse security properties. Moreover, since the direct application of the vinegar modifiers results in an inversion process that does not use the structure of the C^* map, we do not consider this a C^* scheme, but rather a bad HFE scheme.

It is possible to incorporate the vinegar modifier while honoring the C^* central map by restricting the dimension of the mixing monomials, i.e. the monomials involving both the C^* variables and the vinegar variables, so that there is a low dimensional search space for all values of the vinegar variables. This method, however, requires the user to guess the values of the vinegar variables and the values of the mixing monomials in order to use the C^* central map to verify agreement between the C^* and mixing components of the central map. The performance cost of this technique is quadratic in the size of the vinegar space while the effect on security is similar to the (ip) modifier; thus, we consider the vinegar modifier irrelevant to this investigation of generalized C^* schemes.

Let $\mathbb{K}$ be a degree d extension of $\mathbb{F}_q$, a finite field with q elements and let $\phi : \mathbb{F}_q^d \to \mathbb{K}$ be an $\mathbb{F}_q$-vector space isomorphism. Let $U : \mathbb{F}_q^n \to \mathbb{F}_q^n$ be an invertible $\mathbb{F}_q$-affine map, let

$L_t : \mathbb{F}_q^n \to \mathbb{F}_q^d$ be a corank t embedding, let $\pi : \mathbb{F}_q^n \to \mathbb{F}_q^s$ be the projection onto the first s coordinates and let $T : \mathbb{F}_q^{d+p} \to \mathbb{F}_q^{d+p-a}$ be a full rank $\mathbb{F}_q$-affine map.

Define $f_c : \mathbb{F}_q^n \to \mathbb{F}_q^d$ by

$$f_c = \phi^{-1} \circ f \circ \phi \circ L_t,$$

where $f : \mathbb{K} \to \mathbb{K}$ is given by $f(X) = X^{q^\theta+1}$. Further, let $f_{ip} : \mathbb{F}_q^n \to \mathbb{F}_q^d$ be given by

$$f_{ip} = g \circ \pi,$$

where $g : \mathbb{F}_q^s \to \mathbb{F}_q^d$ is a random quadratic. Finally, let $f_p : \mathbb{F}_q^n \to \mathbb{F}_q^p$ be a random quadratic.

The central map is constructed as

$$F = (f_c + f_{ip}) \| f_p.$$

The public key is then given by $\mathcal{P} = T \circ F \circ U$.

Remark 1 We note that since ϕ is a parameter of the system, that the choice of ϕ amounts to a choice of basis for $\mathbb{K}$ over $\mathbb{F}_q$ and that L_t is random, the choice of defining f_{ip} using only the first s variables still captures the full generality of the (ip) modifier.

The degree of the extension field d is a parameter that can vary widely depending on the application. For most studied signature schemes of this type, $d = n$. For encryption, it is viable to allow d to be much larger than n, as is the case for EFLASH, see [5].

Inversion of the public map given the private key is straightforward. First, at key generation, compute the image of f_{ip} so that its elements can be efficiently enumerated. Given $\mathbf{y} \in \mathbb{F}_q^{d+p-a}$, compute the preimage of $\mathbf{y}$ under T and parse each vector as $\mathbf{w} = \mathbf{w}_1 \| \mathbf{w}_2 = (w_1, \ldots, w_d) \| (w_{d+1}, \ldots, w_{d+p})$. For all elements $\mathbf{w}_s \in \mathrm{Im}(f_{ip})$ and for all such $\mathbf{w}_1$, enumerate all preimages $\mathbf{u}$ of $\mathbf{w}_1 + \mathbf{w}_s$ under f_c. Check that $f_{ip}(\mathbf{u}) = \mathbf{w}_s$ and that $f_p(\mathbf{u}) = \mathbf{w}_2$: if either fails retry with another pair; if the check succeeds then output $U^{-1}(\mathbf{u})$. All of the details are explicitly provided in Algorithm 1, which calls on a subroutine to invert the map f_c, since this process can differ depending on whether the scheme is parameterized as a signature or an encryption scheme.

5 Security analyses of the schema

Here we consider the known attacks on multivariate schemes and discuss their application on the general C^* framework presented in the previous section. In particular, we indicate what combinations of modifiers prevent attacks and highlight what modifiers are vulnerable to attack without a companion modifier.

5.1 Differential analysis

A class of attacks that has proven effective against modified and unmodified C^* schemes is the family of differential attacks. The discrete differential of a function $f : \mathbb{K} \to \mathbb{K}$ is the bivariate function

$$Df(y, x) = f(y + x) - f(y) - f(x) + f(0).$$

Differential attacks have been the basis of several cryptanalyses, see [16–18,21–23,27]. The two basic techniques are the recovery of linear differential symmetry relations and the recovery of differential invariants.

Algorithm 1 (n, d, s, a, p, t) scheme Public Key Inversion

```
Input: $T, \phi, f_c, f_{ip}, f_p, U^{-1}, \mathrm{Im}(f_{ip})$ and $\mathbf{y} \in \mathbb{F}_q^{d+p-a}$.
Output: $\mathbf{x}$ such that $\mathcal{P}(\mathbf{x}) = \mathbf{y}$.
$\mathbf{x} = \perp$
$W_T \leftarrow \{\mathbf{w} \in \mathbb{F}_q^{d+p} : T(\mathbf{w}) = \mathbf{y}\}$
for all $\mathbf{w} \in W_T$ do
    $\mathbf{w}_1 \leftarrow \mathbf{w}[1 : d], \mathbf{w}_2 \leftarrow \mathbf{w}[d + 1 : d + p]$
    for all $\mathbf{w}_s \in \mathrm{Im}(f_{ip})$ do
        $\mathbf{w}_c \leftarrow \mathbf{w}_1 + \mathbf{w}_s$
        $U_c \leftarrow \mathrm{Inv} f_c(\mathbf{w}_c)$
        for all $\mathbf{u} \in U_c$ do
            if $f_{ip}(\mathbf{u}) = \mathbf{w}_s$ and $f_p(\mathbf{u}) = \mathbf{w}_2$ then
                $\mathbf{x} \leftarrow U^{-1}(\mathbf{u})$
                break
            end if
        end for
    end for
end for
return $\mathbf{x}$
```

Differential invariants capture the idea of a subspace being acted on linearly by a large subspace of the public maps. Neither the central C^* map, nor any of the modifiers utilize a linear property of the public map on a subspace in any way, so these techniques are mostly inapplicable. In particular, it is proven in [4,26] that some C^* variants including the (p−) variant are immune from these attacks.

Linear differential symmetry attacks search for linear maps M that can be factored out of the differential of the central map. It can be proven that if any such relation exists in which the coefficients of the unknown linear map M occur linearly, then the following relation must also be satisfied for some linear map Λ_M:

$$Df(My, x) + Df(y, Mx) = \Lambda_M Df(y, x). \tag{1}$$

Thus the recovery of an M satisfying this relation is the most general linear type of differential attack.

If such a map M can be found, it allows one to discover new linear combinations of the central maps that are linearly independent of the public key for schemes with the (−) modifier. This information can be used to effectively “remove” the minus modifier and reduce the security of the modified scheme to that of the original scheme. In some extreme cases, such as the original C^* scheme, the linear maps, M, inducing the symmetry can be used to recover the extension field structure and break the scheme directly, without the need for another step in the attack.

Differential symmetry attacks also broke SFLASH, SQUARE and PMI, see [3,16,17]. The vulnerability of C^* to the differential symmetric attack is provably removed with the combination of the projection and minus modifiers as shown in [7,28] and generalized in [4]. Thus, for any generalized C^* scheme with both the projection and minus modifiers to be vulnerable to such attacks, the additional modifiers must somehow reintroduce this weakness. Since the remaining modifiers introduce random coefficients or equations to the central map, the likelihood of such an occurrence is very remote. In particular, both of the random (ip) and (+) modifiers move the distribution of $(n, d, 0, a, 0, t)$ schemes towards the uniform distribution on quadratic maps in statistical distance. Under the loose heuristic that the distribution of such (n, d, s, a, p, t) schemes is very close to uniform, we can derive an approximation of the probability that the result has a differential symmetry.

To calculate the probability that a random function has a differential symmetry, we note that quantified over all quadratic functions, we obtain all possible differential symmetry relations of the form

$$DP(My, x) + DP(y, Mx) = \Lambda_M DP(y, x). \tag{2}$$

in the unknown coefficients of M and Λ_M. We can see that M is being applied to the plaintext variables a and x, which are vectors of length n. Therefore, M will be an $n \times n$ matrix. Lambda on the other hand, will be applied to the output variables, and will therefore have dimension $d - a + p \times d - a + p$. So the two matrices together will have $(d - a + p)^2 + n^2$ variables. To determine the number of equations, it is easier to think about this relationship coordinate wise. If we denote DP_i as the differential of the ith public key equation, and consider solving $DP_i(My, x) + DP_i(y, Mx) = \Lambda_{M_i} DP(y, x)$, we can see that we will have a set of $d - a + p$ equations. For each coordinate i, we get one equation. But, we can interpret each equation as a matrix equation, $M^\top \mathbf{DP_i} + \mathbf{DP_i} M = \Lambda_{Mi} \mathbf{DP}$. The left hand side of the equation will give us a symmetric matrix as $\mathbf{DP_i}$ is symmetric, so we only need to test the upper triangular section of the matrix, including the diagonal, which gives us $\binom{n+1}{2}$ coordinates. Thus, there are $(d - a + p)$ times $\binom{n+1}{2}$ equations. Therefore the symmetric discrete differential equation has a solution with probability $q^{(d-a+p)^2+n^2-(d-a+p)\binom{n}{2}} = q^{\mathcal{O}(-n^3)}$.

For the sake of caution, we will consider schemes with at most one of the projection and minus modifiers and at most one of the (ip) and (+) modifiers to be insecure. The reason is that schemes with at most one of (p) and (−) exhibit a form of differential symmetry and it is plausible that statistical techniques similar to [17] or projection techniques similar to [14] may be effective against (ip) and (+) modifications of these schemes, respectively.

Thus the space of (n, d, s, a, p, t) schemes resistant to differential attacks are those with $at > 0$ or $sp > 0$.

5.2 MinRank

The idea of the MinRank attack in multivariate cryptography was first presented in [19]. The attack is derived from an invariant of the public key under isomorphisms of polynomials commonly known as Q-rank, but more accurately called min-Q-rank.

Given a big field map $f : \mathbb{K} \to \mathbb{K}$, we may choose a matrix representation $\mathbf{F}$ of f over $\mathbb{K}$ as a quadratic form. Specifically, we may choose

$$f(X) = \left[X \; X^q \; \dots \; X^{q^{d-1}} \right] \mathbf{F} \left[X \; X^q \; \dots \; X^{q^{d-1}} \right]^\top$$

The Q-rank of f is then the rank of $\mathbf{F}$. This quantity is well-defined and invariant under isomorphisms of quadratic forms (change of basis on inputs), but is not invariant under isomorphisms of polynomials, which is to say changes of basis on the output space tend to change the Q-rank. However, the minimum Q-rank in the isomorphism class of f is less subject to change, as discussed below.

Finding a $\mathbb{K}$-linear combination of the public maps that has low Q-rank is a viable attack for a scheme with a low min-Q-rank. This problem, the task of finding a linear combination of matrices $\mathbf{M}_1, \dots, \mathbf{M}_k$ of target rank no greater than r, is known as the MinRank problem. The decisional version of MinRank is known to be NP-complete; however, when the sizes

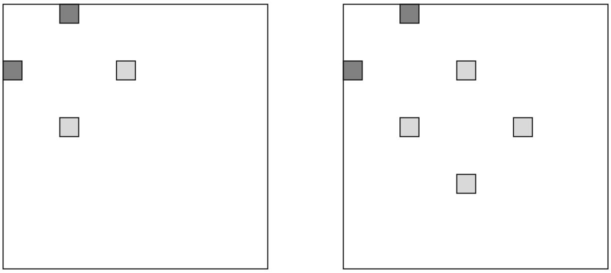

Fig. 2 The shape of the matrix representing the central map of C^{*-}. The darkly shaded regions represent nonzero values of the central map without the minus modifier, the lightly shaded regions represent new nonzero values introduced by the removal of one equation, respectively, two equations. Unshaded areas have coefficients of zero

of the matrices are fixed in relation to the number of matrices, the problem is often tractable for low r.

Many multivariate schemes, including C^*, have low Q-rank, which leaves them vulnerable to MinRank attacks. An unmodified C^* scheme has a Q-rank of 2, as the natural matrix representation $\mathbf{F}$ will have nonzero entries only in the $(\theta, 0)$ and $(0, \theta)$ coordinates.

Since the complexity of MinRank is dependent upon Q-rank, it is clear that if we increase the rank, we decrease the effectiveness of the MinRank attack. As seen in [29], the minus modifier can be seen to increase the rank of the central map. Since the quadratic form associated with the central map is so sparse, in general the removal of one equation affects rank in a different manner than for more general low rank central maps.

Following the analysis in [29], we note that we can pass the minus modifier to the central map by noting that the corank one projection T satisfies $T = S \circ \phi^{-1} \circ \pi_T \circ \phi$, where S is invertible and π_T is a linear polynomial whose roots are exactly $\phi(\ker(T))$. Clearly, if the $\dim(\phi(\ker(T))) = 1$, then the polynomial $\pi_T(x) = x^q + cx$ suffices for some $c \in \mathbb{K}$. One easily checks that $\pi_T \circ f$ has Q-rank 4. However, recall that Q-rank is not invariant under isomorphisms of polynomials. In the case that $\gcd(\theta, d) = 1$, the polynomial $\pi_{T'}(x) = x^{q^\theta} + c^{q^\theta-1}x$ factors

$$\pi_{T'}(x) = x^{q^\theta} + c^{q^\theta-1}x = x(x^{q^\theta-1} + c^{q^\theta-1}).$$

Since for $q = 2^p$ we have that $(q^d - 1, q^\theta - 1) = q - 1$, this polynomial has the same kernel as π_T. Therefore there exists an invertible $\mathbb{F}$-linear map S' such that $T = S' \circ \phi^{-1} \circ \pi_{T'} \circ \phi$. Then one can check that $\pi_{T'} \circ f(x) = x^{q^{2\theta}+q^\theta} + c^{q^\theta-1}x^{q^\theta+1}$ has Q-rank only 2, see Figure 2.

Thus the removal of the first equation does not change the min-Q-rank. However, because of the sparsity of the central map and the condition that $\gcd(\theta, d) = 1$, each subsequent *pair* of equations removed up to $d/2$ equations increases the Q-rank by 2 as illustrated in Figure 2. Thus the min-Q-rank is $2\lceil \frac{a+1}{2} \rceil$ if $gcd(\theta, d) = 1$.

We note that if k =gcd$(\theta, d) > 1$, then the map $\pi_{T'}$ is of lower rank, specifically of rank $d-k$, so it cannot model the removal of one equation. In this case, when $a < k$ each nonzero row of the quadratic form has a single nonzero entry and by symmetry so does each column. Thus the rank is the number of nonzero entries of the matrix which is $2\lceil \frac{a+1}{2} \rceil + 2$. When k equations are removed, however, we can use the same construction as above to obtain a rank two quadratic form. Thus, we conclude that the case of $\gcd(\theta, d) > 1$ is not advisable since an adversary can potentially remove more equations and reduce the rank.

The projection modifier cannot increase the Q-rank of the central map. Since we expect the Q-rank of the projected scheme to be the same as the original, we expect the security of a (n, d, s, a, p, t) scheme against MinRank attacks to be the same as that of a $(n, d, s, a, p, 0)$ scheme.

The effect of (ip) depends on the support dimension s. Since the support dimension of the (ip) summand is s, there is an s dimensional space in which the (ip) modifier adds extra randomness to the quadratic form $\mathbf{F}$. Under an appropriate change of basis, all of this contribution can be contained in the upper left $s \times s$ block of $\mathbf{F}$; hence, (ip) increases the Q-rank by at most s. We can apply projections on the input to try to kill the (ip) support, which once again makes the Q-rank low. Therefore, techniques similar to [14] may limit the effectiveness of the (ip) modifier in preventing a MinRank attack, a fact that once again highlights the close relationship between (ip) and (v).

Finally, the (+) modifier has no effect on min-Q-rank. While any nonzero linear combination of the (+) polynomials is likely to have a high rank, any linear combination of the public polynomials that eliminates the contribution of the (+) polynomials has a rank independent of the (+) modifier. Thus, if the Q-rank of a $(n, d, s, a, 0, t)$ scheme is sufficiently low, the (+) modifier adds no significant security; it merely increases the number of equations.

Thus, the min-Q-rank of a generic (n, d, s, a, p, t) scheme is $2\lceil \frac{a+1}{2} \rceil + s$ at most, and this is a tight bound for most realistic parameters. Thus, we can conclude that the complexity of a MinRank attack on (n, d, s, a, p, t) schemes is

$$\mathcal{O}\left(\binom{d+a+p+s+1}{1+2\lceil\frac{a+1}{2}\rceil+s}^{\omega}\right) \approx \mathcal{O}\left((d-a+p)^{(1+2\lceil\frac{a+1}{2}\rceil+s)\omega}\right).$$

5.3 Algebraic

Algebraic attacks aim to directly invert the public key using polynomial system solvers often based on Gröbner bases. The complexity of this attack relies on the degree of regularity, which is often defined as the smallest degree at which a nontrivial degree fall is generated in the Gröbner basis algorithm. We can find an estimate of the degree of regularity depending on the Q-rank of the system in [11].

As discussed in the previous section, the Q-rank of a C^* scheme with a equations removed and an (ip) support of s is at most $2\lceil\frac{a+1}{2}\rceil+s$. The projection and plus modifier cannot increase the rank of the central map. Using the analysis in [11] we compute an upper bound on the degree of regularity on a (n, d, s, a, p, t) scheme to be

$$\frac{(q-1)(2\lceil\frac{a+1}{2}\rceil+s)}{2}+2=(q-1)\left(\left\lceil\frac{a+1}{2}\right\rceil+\frac{s}{2}\right)+2$$

We note that d_{reg} monotonically decreases in p, thus we cannot add too many plus polynomials.

If we are more conservative, we may consider projection attacks such as those in [14] attempting to eliminate some of the support for the (ip) modifier. We note that with appropriately chosen parameters, this attack can be rendered no more effective than the standard algebraic attack.

Therefore, similar to the analysis done in [5], the complexity of the algebraic attack on a (n, d, s, a, p, t) scheme is estimated to be $\mathcal{O}\left(\binom{n+d_{reg}}{d_{reg}}^{\omega}\right)$, where $2 \leq \omega \leq 3$ is the linear algebra constant.

6 Performance analyses

The performance characteristics of C^* schemes under various modifiers are straightforward to derive. There are some significant differences, however, in the effects of the modifiers when employed for signatures versus encryption. We, thus, treat each case independently.

6.1 Key size

For a generic (n, d, s, a, p, t) scheme, the derivation of the public key size is straightforward. There are $d - a + p$ public equations in n variables. The public key therefore consists of $(d - a + p)\left(\binom{n+1}{2} + 1\right)$ elements from $\mathbb{F}_q$. This observation is not all there is to say about public key size, however.

The $(+)$ modifier significantly reduces the probability of the existence of a preimage for an arbitrary element in the codomain of the public key. We may therefore safely ignore this modifier for digital signature applications.

Similarly, there is a possibility of inversion failure for the public key if the balance of the projection, minus and plus modifiers is not correctly handled. In particular, if d is larger than n the chance of failure is high. Therefore, we will restrict our consideration in the case of signature schemes to $(n, n, s, a, 0, t)$ with t small and in particular $t < a$.

For encryption schemes we need the public key to be statistically injective. To accomplish this goal with modifiers, we either require some redundancy in the plaintext space or to have a larger codomain than the domain. To allow random plaintexts, we use a value of d much larger than n. Then inherently a projection of corank at least $t = d - n$ is required. Since a larger value of t precludes unique inversion, we are forced to set the parameter $t = d - n$.

6.2 Complexity of inversion

Inversion of the public key is accomplished with Algorithm 1. From the algorithm we can see that we have as many as q^{a+s} calls to $\mathrm{Inv} f_c$, plus as many as $q^{a+s}|\ker(L_t)|$ evaluations of f_{ip} and f_p, plus a couple of linear algebra operations. These numbers can vary depending on the parameters and the application.

Consider a $(n, d, s, a, 0, t)$ scheme designed for signatures. With $n = d$, any preimage of $\mathbf{y}$ has a high probability of producing a valid signature and so it is likely that very few of the q^a such preimages will need to be utilized in the inversion. In contrast, for an encryption scheme with n far smaller than d, one may have to search nearly the entire q^a preimages to find the valid plaintext. Similarly, for a signature scheme, the kernel of L_t may be large, as large as q^t, possibly, whereas for an encryption scheme it is necessary that L_t is an embedding. Thus, we split into two cases to consider inversion complexity.

First, we consider inversion for signatures. Under the above assumptions for valid performance of a (n, d, s, a, p, t) scheme for signatures, we have that $n = d$, $p = 0$, t is small and $t < a$. Since t is small and evaluation of f_{ip} and f_p are extremely efficient, the complexity is dominated by that of $\mathrm{Inv} f_c$. Since we expect to only require a few values of $\mathbf{w}$ to find a valid inverse, the complexity of this step is $\mathcal{O}(q^{s+t})$ times the complexity of $\mathrm{Inv} f_c$. Using an efficient inversion process based on linearization equations, the complexity of the latter is $\mathcal{O}(n^\omega)$. Thus the complexity for signature schemes is

$$\mathcal{O}\left(n^\omega q^{s+t}\right).$$

Table 1 Probability of decryption failure for specific parameters of a $(n, d, s, a, p, d - n)$ scheme

q	n	d	a	s	p	$\widehat{m}$	$n - \widehat{m}$	Decryption failure rate
2	14	26	4	2	2	24	−10	$2^{-9.62}$
2	14	26	4	2	3	25	−11	$2^{-10.71}$
2	14	27	4	2	2	25	−11	$2^{-10.86}$
2	14	27	4	2	3	26	−12	$2^{-12.23}$
2	14	28	4	2	2	26	−12	$2^{-11.68}$
2	14	28	4	2	3	27	−13	$2^{-13.23}$

In the case of encryption, we assume that d is much larger than n and that $t = d - n$. In this case, all of Invf_c and f_{ip} and f_p are evaluated q^{a+s} times, thus, once again, the complexity of inversion is dominated by that of Invf_c. In this case, however, it is likely that on the order of q^a preimages of $\mathbf{y}$ under T need to be searched. Thus the complexity of inversion for an encryption scheme is

$$\mathcal{O}\left(n^{\omega} q^{a+s}\right).$$

6.3 Decryption failure rate

The decryption failure rate for $(n, d, 0, a, 0, t)$ schemes created with the intention of encryption is discussed in [5]. Using conditional probabilities and Bernoulli trials, it was found that the probability, p, that y is the image of at least two distinct elements of $\mathbb{F}_q^n$, given that it is the image of at least one can be approximated by q^{n-m}, where $m = d - a$.

This probability was found under the simplifying heuristic that considered the $\mathbb{F}_q$−quadratic central maps f to be random injective maps. Considering (ip), we no longer have the injective property, though the (ip) modifier has a large codomain, so it is unlikely to have collisions between it and anything else. Therefore, (ip) should not have an effect on the decryption failure rate because we would just be adding another random summand. Adding a plus modifier will decrease the decryption failure rate because it will add extra equations that need to be satisfied.

We can model the central map without the minus modifier as a random function $G : q^n \rightarrow q^{\widehat{m}}$, where $\widehat{m} := d + p - a$. We will again use Bernoulli trials, and compute: $P(|G^{-1}(y)| \geq 2 \mid |G^{-1}(y)| \geq 1) \leq \frac{1-(1-q^{n-\widehat{m}})-q^{n-\widehat{m}}+q^{2n-2\widehat{m}}-q^{n-2\widehat{m}}}{1-(1-q^{n-d}+\frac{1}{2}(q^n(q^{n-1})q^{-2d}))}$. Following an analysis equivalent to that in [5], we see that the numerator is bounded by $q^{2n-2\widehat{m}}$ while the denominator is very close to $q^{n-\widehat{m}}$; thus, a good approximation is about $q^{n-\widehat{m}} = q^{n+a-d-p}$. We can see that the plus modifier reduces the probability of decryption failure approximately by q^p, while (ip) has no significant effect on the failure rate.

We performed a series of experiments on failure rate for (n, d, s, a, p, t) schemes designed for encryption to investigate the rate of decryption failure relative to varying parameters d and p, the degree of extension and number of plus polynomials. The results are reported in Table 1. All experiments were performed by encrypting all possible plaintexts and counting the number of plaintexts producing non-unique ciphertexts. In every experiment the failure rates follow the above analysis closely without significant variance.

6.4 Parameter spaces for encryption and signatures

From Sect. 5, we obtain the following constraints for a 128-bit secure (n, d, s, a, p, t) scheme.

$$
\begin{aligned}
&at + sp > 0\\
&(d - a + p)^{(1+2\lceil\frac{a+1}{2}\rceil+s)\omega} \geq 2^{128}\\
&n^{\left((q-1)\left(\left\lceil\frac{a+1}{2}\right\rceil+\frac{s}{2}\right)+2\right)\omega} \geq 2^{128}.
\end{aligned}
$$

These constraints assure security against differential, MinRank and algebraic attacks, respectively.

For signature schemes with suggested parameters of the form $(n, n, s, a, 0, t)$ we obtain a public key size of at least $(n - a)\left(\binom{n+1}{2} + 1\right) lg(q)$ bits and a signing time on the order of $n^{\omega}q^{s+t}$ field operations. Both of these quantities seem to be optimized by making $s = 0$, having a fairly large, and having $t = 1$. This choice of parameters, unsurprisingly, produces the PFLASH scheme. It is interesting to note that these data also suggest an optimal choice of q for such schemes of 2 or 4.

For encryption schemes of the form (n, d, s, a, p, t), the public key size has the form $(d-a+p)\left(\binom{n+1}{2} + 1\right) lg(q)$ bits while decryption time is around $n^{\omega}q^{a+s}$ and the decryption failure rate is $q^{n+a-d-p}$. Here there is a much more interesting trade-off between different strategies. The quantity $a+s$ needs to remain sufficiently large to provide security but directly impacts the decryption speed. Public key size is reduced if a is increased while s and p are reduced , however this directly and negatively impacts decryption failure rate. Thus there are an array of options offering various optimizations.

7 Conclusion

We have thoroughly examined the behavior of the C^*-based schemes, which has demonstrated that there is an entire space of feasible schemes that can be optimized for different performance criteria. Historically, multivariate schemes have struggled to create secure encryption schemes. But now we see there is a space of seemingly viable encryption schemes, including EFLASH. This analysis also suggests how new combinations of modifiers may help current schemes, including a possible (+) modification to EFLASH to decrease the decryption failure rate without significant cost.

Security against known attacks is well understood in this area. It is for future work to determine whether these schemes will be long lived or whether new advances will change the landscape.

Acknowledgements The authors would like to thank the anonymous referees for their helpful comments.

Appendix A

See Table 2.

Table 2 Resistance of C^* to attacks under certain modifiers plotted against the effect on efficiency

Modifier	Alg	Diff	Lin Eq	MinRank	$\Delta\lvert PK\rvert$	ΔT_{inv}	Δp_{fail}
(ip)	+	0	+	+	0	$\cdot q^s$	0
(−)	+	0	+	+	$-a(\binom{n+1}{2}+1)lg(q)$	$\cdot q^a, 0$	$\cdot q^a$
(+)	−	+	+	0	$+p(\binom{n+1}{2}+1)lg(q)$	0	$\cdot q^{-p}$
(p)	−	0	0	0	$0^\dagger$	$0, \cdot q^t$	0
(v)*	+	+	+	+	$0^\dagger$	NC	0
(ip+)	−	+	+	+	$+p(\binom{n+1}{2}+1)lg(q)$	$\cdot q^s$	$\cdot q^{-p}$
(+−)	+	+	+	+	$(p\text{-}a)(\binom{n+1}{2}+1)lg(q)$	$\cdot q^a$	$\cdot q^{a-p}$
(p−)	+−	1	+	+	$-a(\binom{n+1}{2}+1)lg(q)^\dagger$	$\cdot q^a, \cdot q^t$	$\cdot q^a$

The left hand side of the table can be read as probabilities of resistance to the given attack. Thus 0 means that the modifier(s) provide(s) no security in the attack model, 1 means the modifiers(s) provide(s) provable security, and + or − mean increases, respectively decreases in security. The efficiency data on the right hand side of the table are additive unless specified to be a multiplicative factor by a ·. Where applicable, data are split by a comma indicating the effect of the modifier in different modes—(encryption effect, signature effect)

* These schemes use HFEv inversion and are not C^* schemes *per se*

† The modifier affects the quantity by changing the choice of n

References

1. Berlekamp E.R.: Factoring polynomials over large finite fields. Math. Comput. **24**, 713–735 (1970).
2. Bettale L., Faugère J., Perret L.: Cryptanalysis of HFE, multi-HFE and variants for odd and even characteristic. Des. Codes Cryptogr. **69**, 1–52 (2013).
3. Billet O., Macario-Rat G.: Cryptanalysis of the square cryptosystems. In: ASIACRYPT 2009. LNCS, vol. 5912, pp. 451–486 (2009).
4. Cartor R., Smith-Tone D.: An updated security analysis of PFLASH. In: Post-quantum Cryptography-8th International Workshop, PQCrypto 2017, Utrecht, The Netherlands, June 26–28, 2017, Proceedings, pp. 241–254 (2017).
5. Cartor R., Smith-Tone D.: EFLASH: a new multivariate encryption scheme. IACR Cryptol. ePrint Arch. **2017**, 1184 (2017).
6. Casanova A., Faugere J.C., Macario-Rat G., Patarin J., Perret L., Ryckeghem J.: Gemss: a great multivariate short signature. NIST CSRC (2017). https://csrc.nist.gov/CSRC/media/Projects/Post-Quantum-Cryptography/documents/round-1/submissions/GeMSS.zip.
7. Chen M.S., Yang B.Y., Smith-Tone D.: Pflash-secure asymmetric signatures on smart cards. In: Lightweight Cryptography Workshop (2015). http://csrc.nist.gov/groups/ST/lwc-workshop2015/papers/session3-smith-tone-paper.pdf.
8. Cryptographic Technology Group: Post-quantum cryptography standardization. NIST CSRC (2018) https://csrc.nist.gov/Projects/Post-Quantum-Cryptography/Post-Quantum-Cryptography-Standardization.
9. Ding J.: A new variant of the matsumoto-imai cryptosystem through perturbation. In: PKC 2004. LNCS, vol. 2947, pp. 305–318 (2004).
10. Ding J., Gower J.: Inoculating multivariate schemes against differential attacks. In: PKC 2006. LNCS, vol. 3958, pp. 290–301 (2006).
11. Ding J., Kleinjung T.: Degree of regularity for HFE. IACR Cryptol. ePrint Arch. **2011**, 570 (2011).
12. Ding J., Dubois V., Yang B.Y., Chen C.H.O., Cheng C.M.: Could SFLASH be repaired? In: Aceto L., Damgård I., Goldberg L.A., Halldórsson M.M., Ingólfsdóttir A., Walukiewicz I. (eds.) ICALP (2), vol. 5126, pp. 691–701. Lecture Notes in Computer ScienceSpringer, Berlin (2008).
13. Ding J., Chen M.S., Petzoldt A., Schmidt D., Yang B.Y.: Gui. NIST CSRC (2017). https://csrc.nist.gov/CSRC/media/Projects/Post-Quantum-Cryptography/documents/round-1/submissions/Gui.zip.
14. Ding J., Perlner R.A., Petzoldt A., Smith-Tone D.: Improved cryptanalysis of HFEV-via projection. In: Lange T., Steinwandt R. (eds.) Post-quantum Cryptography-9th International Conference, PQCrypto 2018, Fort Lauderdale, FL, USA, April 9–11, 2018, Proceedings, vol. 10786, pp. 375–395. Lecture Notes in Computer ScienceSpringer, Berlin (2018).

15. Dubois V., Fouque P.A., Stern J.: Cryptanalysis of SFLASH with Slightly Modified Parameters. In: Naor M. (ed.) EUROCRYPT, vol. 4515, pp. 264–275. Lecture Notes in Computer ScienceSpringer, Berlin (2007).
16. Dubois V., Fouque P.A., Shamir A., Stern J.: Practical Cryptanalysis of SFLASH. In: Menezes A. (ed.) CRYPTO, vol. 4622, pp. 1–12. Lecture Notes in Computer ScienceSpringer, Berlin (2007).
17. Fouque P., Granboulan L., Stern J.: Differential cryptanalysis for multivariate schemes. In: Cramer R. (ed.) Advances in Cryptology-EUROCRYPT 2005, 24th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Aarhus, Denmark, May 22–26, 2005, Proceedings, vol. 3494, pp. 341–353. Lecture Notes in Computer ScienceSpringer, Berlin (2005).
18. Hashimoto Y.: Cryptanalysis of multi-HFE. IACR Cryptol. ePrint Arch. **2015**, 1160 (2015).
19. Kipnis A., Shamir A.: Cryptanalysis of the hfe public key cryptosystem by relinearization. Adv. Cryptol. CRYPTO **1666**, 788 (1999).
20. Matsumoto T., Imai H.: Public quadratic polynomial-tuples for efficient signature verification and message-encryption. Eurocrypt'88, pp. 419–545. Springer, Berlin (1988).
21. Moody D., Perlner R.A., Smith-Tone D.: An asymptotically optimal structural attack on the ABC multivariate encryption scheme. In: Mosca M. (ed.) Proceedings on Post-quantum Cryptography-6th International Workshop, PQCrypto 2014, Waterloo, ON, Canada, October 1–3, 2014, vol. 8772, pp. 180–196. Lecture Notes in Computer ScienceSpringer, Berlin (2014).
22. Moody D., Perlner R.A., Smith-Tone D.: Improved attacks for characteristic-2 parameters of the cubic ABC simple matrix encryption scheme. In: Post-quantum Cryptography-8th International Workshop, PQCrypto 2017, Utrecht, The Netherlands, June 26–28, 2017, Proceedings, pp. 255–271 (2017).
23. Moody D., Perlner R.A., Smith-Tone D.: Key Recovery Attack on the Cubic ABC Simple Matrix Multivariate Encryption Scheme. Springer, Berlin (2017).
24. Patarin J.: Cryptoanalysis of the Matsumoto and Imai Public Key Scheme of Eurocrypt'88. In: Coppersmith D. (ed.) CRYPTO, vol. 963, pp. 248–261. Lecture Notes in Computer ScienceSpringer, Berlin (1995).
25. Patarin J., Goubin L., Courtois N.: C^{*}_{-+} and HM: variations around two schemes of T. Matsumoto and H. Imai. In: Ohta, K., Pei, D. (eds.) Advances in Cryptology-ASIACRYPT'98, International Conference on the Theory and Applications of Cryptology and Information Security, Beijing, China, October 18–22, 1998, Proceedings. Lecture Notes in Computer Science, vol. 1514, pp. 35–49. Springer, New York (1998)
26. Perlner R.A., Smith-Tone D.: A classification of differential invariants for multivariate post-quantum cryptosystems. In: Gaborit P. (ed.) PQCrypto, vol. 7932, pp. 165–173. Lecture Notes in Computer ScienceSpringer, Berlin (2013).
27. Perlner R.A., Petzoldt A., Smith-Tone D.: In: Total Break of the SRP Encryption Scheme. Springer, In press (2017).
28. Smith-Tone D.: On the differential security of multivariate public key cryptosystems. In: Yang B.Y. (ed.) PQCrypto, vol. 7071, pp. 130–142. Lecture Notes in Computer ScienceSpringer, Berlin (2011).
29. Vates J., Smith-Tone D.: Key recovery attack for all parameters of HFE. In: Lange T., Takagi T., eds.: Post-quantum Cryptography-8th International Workshop, PQCrypto 2017, Utrecht, The Netherlands, June 26–28, 2017, Proceedings. Lecture Notes in Computer Science, vol. 10346, pp. 272–288. Springer, New York (2017).

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.