Turkish Journal of Mathematics

http://journals.tubitak.gov.tr/math/

Research Article

Turk J Math
(2019) 43: 241 – 257
© TÜBİTAK
doi:10.3906/mat-1803-92

A novel 3-pass identification scheme and signature scheme based on multivariate quadratic polynomials

Sedat AKLEYLEK[1*], **Meryem SOYSALDI**[1]
[1]Department of Computer Engineering, Faculty of Engineering, Ondokuz Mayıs University, Samsun, Turkey

Received: 19.03.2018 • **Accepted/Published Online:** 29.11.2018 • **Final Version:** 18.01.2019

Abstract: Identification schemes are used to verify identities of parties and signatures. Recently, systems based on multivariate polynomials have been preferred in identification schemes due to their resistance against quantum attacks. In this paper, we propose a quantum secure 3-pass identification scheme based on multivariate quadratic polynomials. We compare the proposed scheme with the previous ones in view of memory requirements, communication length, and computation time. We define an efficiency metric by using impersonation probability and computation time. According to the comparison results, the proposed one has the same computation time as that of Monteiro et al. and reduces impersonation probability compared to the work of Sakumoto et al. We also propose a new signature scheme constructed from the proposed identification scheme. In addition, we compare the signature scheme with the previous schemes in view of signature and key sizes. We improve the signature size compared to that given in previous work by Chen et al.

Key words: Identification schemes, signature schemes, zero-knowledge, post-quantum cryptography

1. Introduction

With the digitization of everything and the widespread use of the Internet, all systems connected to a network have become able to communicate with each other. In this environment, the security of the data between communication systems must be ensured. In order to ensure the security of the data, the policies of data security such as confidentiality, integrity, nonrepudiation, and authentication should be provided. Identification means that the communicating parties can verify each other's identity [6]. Identification schemes are very important since they are the basis of identification and signature schemes. There are two parties, the verifier and the prover, in an identification scheme. The verifier and the prover need to be sure of each other's identities.

In the literature several systems have been proposed for identification schemes. At the beginning these systems were proposed based on computationally hard problems such as factorization and discrete logarithm problems [7, 10]. However, after Shor proposed polynomial time algorithms to solve prime factorization and discrete logarithm problems in polynomial time by using quantum computers, these systems began to be called insecure in the quantum world [13]. In this regard, the studies on quantum secure schemes have increased. These are based on hard problems known to be reliable against quantum attacks. Lattice-based, code-based, hash-based, isogeny-based, and multivariate cryptosystems are known to be resistant against quantum attacks [2].

In [12], 3-pass and 5-pass zero-knowledge identification schemes using multivariate quadratic polynomial

[*]Correspondence: sedat.akleylek@bil.omu.edu.tr

2010 *AMS Mathematics Subject Classification:* 94A60, 68P25 systems over a finite field were proposed. They defined the bilinearity of the polar form of multivariate quadratic polynomial systems when they constructed a new secret key dividing technique. The secret key was first divided into two parts by using the bilinear property, and then one of them was divided into two subparts. In addition, the scheme was made interactive by requesting a constant finite element from the verifier in each round.

In [9], the number of partitions was increased to 4, while it is 3 in [12]. They first divided the secret key into two parts, and then each part was divided into two parts again. With this change, impersonation probability was improved.

In [11], Sakumoto mentioned whether or not a public key identification scheme based on multivariate polynomials of degree more than two is efficient and presented new 3- and 5-pass identification schemes based on multivariate cubic polynomials. Sakumoto stated that the 3-pass identification scheme was not productive but the 5-pass identification scheme was highly efficient. In addition, the size of the public and private key in the proposed 5-pass identification scheme was significantly reduced when compared with previous schemes.

In [3], a multivariate quadratic (MQ)-based signature scheme was proposed by applying the Fiat–Shamir transform to the 3-pass identification scheme based on the MQ polynomials given in [12]. They also applied the Fiat–Shamir transform to the 5-pass identification scheme given in [12]. A modified version of this signature scheme was submitted to NIST's First Post-Quantum Cryptography Standardization Project [4].

1.1. Motivation

Cryptosystems based on multivariate polynomial systems are attractive for the post-quantum world since they are efficient and easy to construct. In [9, 11, 12], zero-knowledge identification schemes based on multivariate quadratic or cubic polynomial systems were proposed. The comparison of identification schemes includes memory requirements, communication length, impersonation probability, and computation time for efficiency. Computation time is much more important than communication length since less computation time means fewer arithmetic operations [9]. Constructing efficient schemes is always important to improve the performance. Identification schemes based on multivariate quadratic polynomials can be transformed into signature schemes. In [3], MQ-based signature schemes were obtained by applying Fiat–Shamir transform to 5-pass identification schemes based on MQ polynomials. The differences in these identification schemes mainly depend on the number of secret key partitions. It is natural to ask what will change when we use different secret key dividing techniques.

1.2. Our contribution

In this paper, we use different secret key partitions to construct a 3-pass identification scheme by using bilinearity of the polar form. In other words, the secret key has 2 main parts. The former and latter parts have 2 and 3 parts, respectively (see Figure 1). This helps us to get an efficient scheme in terms of computation time. Moreover, the efficiency metric, a combination of impersonation probability and computation time, is better than that in [12]. In addition, we apply the Fiat–Shamir transform to the proposed 3-pass identification scheme and propose a new MQ-based signature scheme. We reduce the signature size compared to the signature scheme in [3, 4]. We also provide a comparison with the signature scheme based on [9].

1.3. Organization

The rest of this paper is organized as follows. In Section 2, we provide related definitions for the proposed idea. In Section 3, we describe our method. Then we present our identification scheme. In Section 4, we compare the proposed scheme to the previous schemes. In Section 5, we present an MQ-based signature scheme by transforming to the proposed 3-pass identification scheme based on MQ polynomials. In Section 6, the conclusion and future works are stated.

2. Preliminaries

In this section, we recall some basic definitions. Let $\mathbb{F}_q$ be a finite field with q elements where q is a prime power and $\mathbb{F}_q^n$ be an n-dimensional vector space of q elements.

Definition 1 *[2] Multivariate polynomial systems consist of a large number of polynomials. A system of MQ polynomials with n variables and m equations can be given as follows:*

$$F = \begin{cases} f^{(1)}(x_1,\ldots,x_n) = \sum_i^n \sum_j^n \alpha_{i,j}^{(1)} x_i x_j + \sum_i^n \beta_{i,j}^{(1)} x_i + \gamma^{(1)} \\ f^{(2)}(x_1,\ldots,x_n) = \sum_i^n \sum_j^n \alpha_{i,j}^{(2)} x_i x_j + \sum_i^n \beta_{i,j}^{(2)} x_i + \gamma^{(2)} \\ \vdots \\ f^{(m)}(x_1,\ldots,x_n) = \sum_i^n \sum_j^n \alpha_{i,j}^{(m)} x_i x_j + \sum_i^n \beta_{i,j}^{(m)} x_i + \gamma^{(m)}, \end{cases}$$

where the coefficients $\alpha_{i,j}^{(k)}, \beta_{i,j}^{(k)}, \gamma^{(k)}$ for $1 \leq k \leq m$ are in $\mathbb{F}_q$. In Definition 2, the MQ problem is recalled.

Definition 2 *[2] Let F be a system of MQ polynomials as given in Definition 1. The MQ problem, NP-complete, is defined as finding a solution vector $\bar{\mathbf{x}} = (x_1,\ldots,x_n)$ for the system $F(\bar{\mathbf{x}}) = 0$. It is difficult to solve MQ problems over a finite field. In other words, there is no known polynomial-time algorithm to solve this problem even for quantum computers.*

In Definition 3, an identification scheme is given.

Definition 3 *[1] An identification scheme consists of two parties: the prover and the verifier. While the prover has a public and private key pair, the verifier has only the prover's public key. An identification scheme allows the prover to verify his identity without giving the secret key. An identification scheme is a set of algorithms: $Setup$, Gen, P, and V.*

- *The $Setup$ algorithm takes a security parameter 1^λ as an input and the output of this algorithm is a system parameter $param$.*
- *Gen is a key generation algorithm and uses $param$, which is the output of the $Setup$ algorithm. Gen generates a (v,s) public and private key pair.*
- *P and V are the processes that are executed respectively by the prover and the verifier. P performs the operations by using public key v and secret key s on the prover side, while V uses public key v of the prover on the verifier side. Communication between the prover and the verifier begins with the commitment of the prover and continues with challenge and response. Since there is continuous communication between the prover and verifier, the pair of (P,V) is an interactive identification scheme.*

Multivariate quadratic polynomials are mostly preferred to construct identification schemes since one can easily obtain the polar form by using bilinearity. In Definition 4, the polar form of the function is given.

Definition 4 *[12] Recall that if $G : A\ x\ B \rightarrow \mathbb{F}_q^n$ is a bilinear function, then $G(x_1+x_2, y) = G(x_1, y)+G(x_2, y)$ where $\forall\ x_1, x_2 \in A$ and $y \in B$. Let G be a bilinear function that is the polar form of F. It is defined as follows:*

$$G(x, y) = F(x + y) - F(x) - F(y), \tag{1}$$

where $x, y \in \mathbb{F}_q^n$.

The zero-knowledge identification scheme has challenge and response states between the prover and the verifier. In zero-knowledge identification schemes, the prover can convince the verifier without giving any useful information about the scheme and the secret key. The verifier challenges the prover and tries to reject the prover during the scheme. However, the prover replies to each of the verifier's challenges. Finally, although the verifier does not know any information about the scheme, he/she convinces the prover. At the end of the scheme, the verifier convinces and accepts the prover but he/she does not prove the knowledge that he/she is convinced by another. In this case, it is first thought that the verifier has unlimited computing power. Even if the verifier has unlimited computing power, the verifier cannot learn any useful information from communication about the scheme due to the statistical zero-knowledge property. An identification scheme must have completeness (Definition 5) and soundness (Definition 6) properties.

Definition 5 ***(Completeness)*** *[14] If an identification scheme is completed with the acceptance of the verifier, then completeness is achieved.*

Definition 6 ***(Soundness)*** *[14] If a cheating prover cannot convince the verifier by impersonating someone else (except with negligible probability $\varepsilon > 0$), then soundness is achieved.*

A commitment scheme, denoted as Com, allows the sender to make a commitment to the receiver and then the receiver verifies this commitment. Com has two phases. The first phase is the commitment phase. In this phase, the sender computes the commitment values to be sent to the receiver and sends to the receiver. The second one is the verification phase. The sender sends some parameters to the receiver so that the receiver can compute the commitment values. After getting the parameters, the receiver computes and verifies the commitment values. The commitment phase is used to ensure that the identification scheme can be a zero-knowledge identification scheme. For this reason, Com must satisfy statistically hiding and computationally binding properties.

Definition 7 ***(Statistically hiding)*** *[8] After the receiver takes the commitment values, the receiver cannot distinguish commitment values from each other even if he has unlimited computing power. The commitment values must be statistically close to each other. In other words, it is not known which parameters are used to compute the commitment values generated by Com.*

Definition 8 ***(Computationally binding)*** *[8] It is computationally hard for the sender to change these values after generating the commitment values.*

Digital signatures provide authentication, nonrepudiation, and integrity. In digital signature schemes, a sender, who signs a message with his or her own private key, sends a message to a receiver. The receiver can verify the signed message by using the sender's public key. A digital signature scheme is a set of algorithms $DSS = (KeyGen, Sign, Verify)$ defined as follows [1]:

- $KeyGen$ is a probabilistic key generation algorithm. $KeyGen$ takes security parameter 1^k, where $k \in \mathbb{N}$ is the input and it outputs a public and secret key pair (p_k, s_k).
- $Sign$ is a probabilistic signature algorithm that takes a secret key s_k and a message m as input and then outputs a signature σ.
- $Verify$ is a deterministic verification algorithm that takes a public key p_k, a message m, and a signature σ and generates a decision bit 1 or 0 that means the signature is accepted or rejected, respectively.

Correctness of the signature is mentioned if a signature scheme satisfies that for all $k \in \mathbb{N}$, $(p_k, s_k) \leftarrow KeyGen(1^k)$, all messages m, and corresponding signatures $\sigma \leftarrow Sign(s_k, m)$, we have $Verify(p_k, m, \sigma) = 1$.

In Section 3, we construct a 3-pass identification scheme satisfying completeness and soundness properties. Moreover, we check statistically hiding and computationally binding properties of the proposed scheme.

3. A new 3-pass identification scheme

In this section, we propose a new 3-pass identification scheme. We divide the secret key in a different way and use the polar form. To the best of our knowledge, this partition idea has not been used before. Our idea depends on modifying the idea given in [9, 12]. Let F be a multivariate quadratic polynomial system as in Eq. (1). In [12], the secret key is divided into two parts as $s = r_0 + r_1$. Then r_0 is divided into two subparts as $r_0 = t_0 + t_1$ and $F(r_0) = e_0 + e_1$. Our contribution is to divide r_1 into three parts by using bilinearity, i.e. $r_1 = d_0 + d_1 + d_2$ and $F(r_1) = u_0 + u_1 + u_2$. The secret key partitioning is demonstrated in Figure 1.

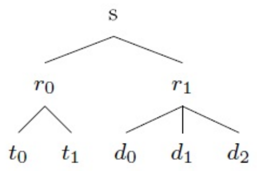
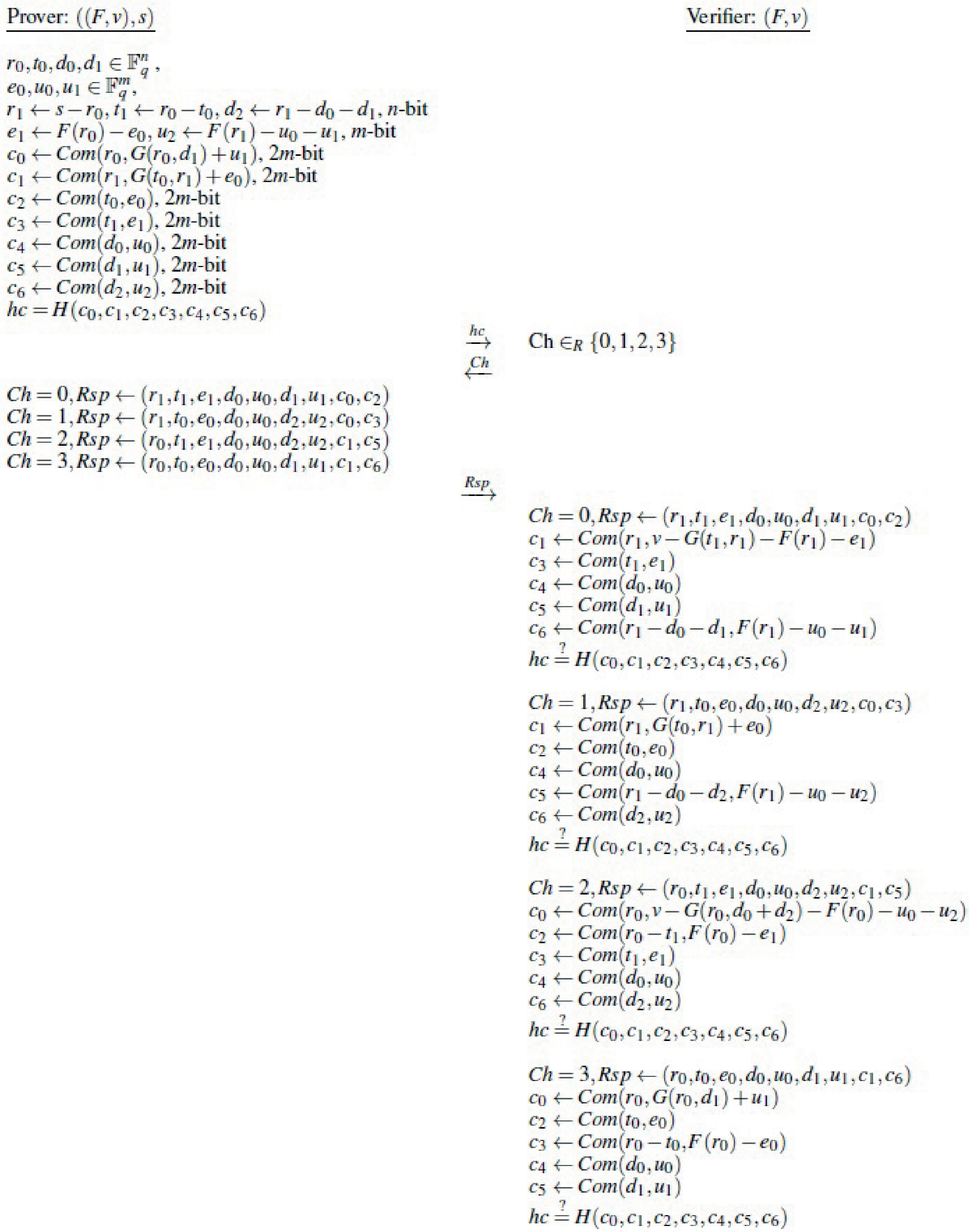


Figure 1. Dividing technique.

In Figure 2, the proposed identification scheme is given. The prover executes some algorithms before an interactive scheme begins between the prover and the verifier in [12]. By Definition 3, the identification scheme begins with the $Setup$ algorithm that generates an output as system parameter $F \in_R MQ(n, m, \mathbb{F}_q)$, where $\in_R$ stands for a randomly chosen element. Gen receives randomly generated multivariate quadratic polynomials system F and the private key s. Then Gen computes the $v = F(s)$ public key. Immediately after that, interactive communication starts. The prover randomly selects $r_0, r_1, t_0, d_0, d_1 \in \mathbb{F}_q^n$ and $e_0, u_0, u_1 \in \mathbb{F}_q^m$ for partitioning secret key s and obtains the other parts given in Figure 1. After the secret key is partitioned, the prover prepares commitment values and sends the verifier hc, the hash value of the commitment values. Then the verifier challenges the prover. The prover helps the verifier to compute the commitment values by sending a response. If the verifier gets the same hc value, he accepts the prover's claim; otherwise, he rejects it. The verifier accepts the prover's claim as long as Eqs. (2) and (3) hold.

$$G(r_0, d_1) + u_1 = v - \overbrace{(G(r_0, d_0) + G(r_0, d_2))}^{G(r_0, d_0+d_2)} - F(r_0) - u_0 - u_2 \tag{2}$$



Prover: $((F,v),s)$

Verifier: (F,v)

$r_0, t_0, d_0, d_1 \in \mathbb{F}_q^n$,
$e_0, u_0, u_1 \in \mathbb{F}_q^m$,
$r_1 \leftarrow s - r_0$, $t_1 \leftarrow r_0 - t_0$, $d_2 \leftarrow r_1 - d_0 - d_1$, n-bit
$e_1 \leftarrow F(r_0) - e_0$, $u_2 \leftarrow F(r_1) - u_0 - u_1$, m-bit
$c_0 \leftarrow Com(r_0, G(r_0, d_1) + u_1)$, $2m$-bit
$c_1 \leftarrow Com(r_1, G(t_0, r_1) + e_0)$, $2m$-bit
$c_2 \leftarrow Com(t_0, e_0)$, $2m$-bit
$c_3 \leftarrow Com(t_1, e_1)$, $2m$-bit
$c_4 \leftarrow Com(d_0, u_0)$, $2m$-bit
$c_5 \leftarrow Com(d_1, u_1)$, $2m$-bit
$c_6 \leftarrow Com(d_2, u_2)$, $2m$-bit
$hc = H(c_0, c_1, c_2, c_3, c_4, c_5, c_6)$

$\xrightarrow{hc}$ Ch $\in_R \{0,1,2,3\}$

$\xleftarrow{Ch}$

$Ch = 0, Rsp \leftarrow (r_1, t_1, e_1, d_0, u_0, d_1, u_1, c_0, c_2)$
$Ch = 1, Rsp \leftarrow (r_1, t_0, e_0, d_0, u_0, d_2, u_2, c_0, c_3)$
$Ch = 2, Rsp \leftarrow (r_0, t_1, e_1, d_0, u_0, d_2, u_2, c_1, c_5)$
$Ch = 3, Rsp \leftarrow (r_0, t_0, e_0, d_0, u_0, d_1, u_1, c_1, c_6)$

$\xrightarrow{Rsp}$

$Ch = 0, Rsp \leftarrow (r_1, t_1, e_1, d_0, u_0, d_1, u_1, c_0, c_2)$
$c_1 \leftarrow Com(r_1, v - G(t_1, r_1) - F(r_1) - e_1)$
$c_3 \leftarrow Com(t_1, e_1)$
$c_4 \leftarrow Com(d_0, u_0)$
$c_5 \leftarrow Com(d_1, u_1)$
$c_6 \leftarrow Com(r_1 - d_0 - d_1, F(r_1) - u_0 - u_1)$
$hc \overset{?}{=} H(c_0, c_1, c_2, c_3, c_4, c_5, c_6)$

$Ch = 1, Rsp \leftarrow (r_1, t_0, e_0, d_0, u_0, d_2, u_2, c_0, c_3)$
$c_1 \leftarrow Com(r_1, G(t_0, r_1) + e_0)$
$c_2 \leftarrow Com(t_0, e_0)$
$c_4 \leftarrow Com(d_0, u_0)$
$c_5 \leftarrow Com(r_1 - d_0 - d_2, F(r_1) - u_0 - u_2)$
$c_6 \leftarrow Com(d_2, u_2)$
$hc \overset{?}{=} H(c_0, c_1, c_2, c_3, c_4, c_5, c_6)$

$Ch = 2, Rsp \leftarrow (r_0, t_1, e_1, d_0, u_0, d_2, u_2, c_1, c_5)$
$c_0 \leftarrow Com(r_0, v - G(r_0, d_0 + d_2) - F(r_0) - u_0 - u_2)$
$c_2 \leftarrow Com(r_0 - t_1, F(r_0) - e_1)$
$c_3 \leftarrow Com(t_1, e_1)$
$c_4 \leftarrow Com(d_0, u_0)$
$c_6 \leftarrow Com(d_2, u_2)$
$hc \overset{?}{=} H(c_0, c_1, c_2, c_3, c_4, c_5, c_6)$

$Ch = 3, Rsp \leftarrow (r_0, t_0, e_0, d_0, u_0, d_1, u_1, c_1, c_6)$
$c_0 \leftarrow Com(r_0, G(r_0, d_1) + u_1)$
$c_2 \leftarrow Com(t_0, e_0)$
$c_3 \leftarrow Com(r_0 - t_0, F(r_0) - e_0)$
$c_4 \leftarrow Com(d_0, u_0)$
$c_5 \leftarrow Com(d_1, u_1)$
$hc \overset{?}{=} H(c_0, c_1, c_2, c_3, c_4, c_5, c_6)$

Figure 2. A new 3-pass identification scheme.

$$G(t_0, r_1) + e_0 = v - G(t_1, r_1) - F(r_1) - e_1 \tag{3}$$

The correctness of the scheme is proved by using Definition 4. Our aim is to show $v = F(s)$.

$$\begin{aligned} v &= G(r_0, d_0) + G(r_0, d_1) + G(r_0, d_2) + F(r_0) + u_0 + u_1 + u_2 \\ &= G(r_0, d_0 + d_1 + d_2) + F(r_0) + u_0 + u_1 + u_2 \\ &= G(r_0, r_1) + F(r_0) + F(r_1) \\ &= F(r_0 + r_1) = F(s) \end{aligned}$$

3.1. Statistically hiding and computationally binding properties

Now we show that the proposed scheme satisfies the statistically hiding and computationally binding properties.

Theorem 1 ***(Zero-knowledge)*** *The 3-pass identification scheme is statistically zero-knowledge when the commitment scheme Com is statistically hiding.*

Proof Let S be a simulator that will impersonate an honest prover against the cheating verifier without knowing what the secret key is. S selects a $Ch' \in_R \{0,1\}$ that the cheating verifier will not choose where $Ch' = 1 \to Ch \in_R \{0,1\}$ and $Ch' = 0 \to Ch \in_R \{2,3\}$. Then S selects the vectors $s', r'_0, t'_0, d'_0, d'_1 \in_R \mathbb{F}_q^n, e'_0, u'_0, u'_1 \in_R \mathbb{F}_q^m$ and computes $r'_1 \leftarrow s' - r'_0$, $t'_1 \leftarrow r'_0 - t'_0$, $d'_2 \leftarrow r'_1 - d'_0 - d'_1$. If $Ch' = 0$, it means that $Ch \in \{2,3\}$. Then Ch' is equal either 0 or 1. In this case, if $Ch' = 0$, S computes $e'_1 \leftarrow v - F(s') + F(r'_0) - e'_0$ and $u'_2 \leftarrow F(r_1)' - u'_0 - u'_1$; otherwise, ($Ch^{'} = 1$ case) S computes $e'_1 \leftarrow F(r'_0) - e'_0$ and $u'_2 \leftarrow v - F(s') + F(r'_1) - u'_0 - u'_1$. After all values are obtained, S computes commitment value $c_0 \leftarrow Com(r'_0, G(r'_0, d'_1) + u'_1)$, $c_1 \leftarrow Com(r'_1, G(t'_0, r'_1) + e'_0)$, $c_2 \leftarrow Com(t'_0, e'_0)$, $c_3 \leftarrow Com(t'_1, e'_1)$, $c_4 \leftarrow Com(d'_0, u'_0)$, $c_5 \leftarrow Com(d'_1, u'_1)$, $c_6 \leftarrow Com(d'_2, u'_2)$ and sends the verifier.

The verifier sends a new $Ch \in \{0,1,2,3\}$.

$$Ch = \begin{cases} 0, Rsp \leftarrow (r'_1, t'_1, e'_1, d'_0, u'_0, d'_1, u'_1, c_0, c_2) \\ 1, Rsp \leftarrow (r'_1, t'_0, e'_0, d'_0, u'_0, d'_2, u'_2, c_0, c_3) \\ 2, Rsp \leftarrow (r'_0, t'_1, e'_1, d'_0, u'_0, d'_2, u'_2, c_1, c_5) \\ 3, Rsp \leftarrow (r'_0, t'_0, e'_0, d'_0, u'_0, d'_1, u'_1, c_1, c_6) \end{cases}$$

S can impersonate the honest prover with probability $1/2$.

$$Ch' = \begin{cases} 0, Ch = 2 \text{ or } Ch = 3 \to \text{verification with } r_0 \\ 1, Ch = 0 \text{ or } Ch = 1 \to \text{verification with } r_1 \end{cases}$$

When $Ch' = 1$ and $Ch = 0$, it is obtained that $e'_1 \leftarrow v - F(s') + F(r'_0) - e'_0$ and c_1 is verified since $v - G(t_1, r_1) - F(r_1) - e_1 = G(t_0, r_1) + e_0$. In the other case, when $Ch' = 1$ and $Ch = 1$, the verification is much easier since public key v is not used. For the other case ($Ch' = 0$ and $Ch \in \{2,3\}$), the idea is similar.

Consequently, someone using S can create a copy of the scheme. However, when the commitment scheme Com is statistically hiding, one cannot get any useful information although a copy of the scheme is received. Note that only scheme transcripts are obtained. □

Theorem 2 ***(Soundness)*** *The 3-pass identification scheme is an argument of knowledge with error probability* $1/2$ *in each round when the commitment scheme* Com *is computational binding.*

Proof Let

$$\begin{aligned}
s_0 &= \{(hc_0, Ch_0, Rsp_0), (hc_1, Ch_1, Rsp_1), (hc_2, Ch_2, Rsp_2)\},\\
s_1 &= \{(hc_0, Ch_0, Rsp_0), (hc_1, Ch_1, Rsp_1), (hc_3, Ch_3, Rsp_3)\},\\
s_2 &= \{(hc_0, Ch_0, Rsp_0), (hc_2, Ch_2, Rsp_2), (hc_3, Ch_3, Rsp_3)\},\\
s_3 &= \{(hc_1, Ch_1, Rsp_1), (hc_2, Ch_2, Rsp_2), (hc_3, Ch_3, Rsp_3)\},
\end{aligned}$$

be four transcriptions of the proposed identification scheme such that $Ch_i = i$ and the decision function $Dec(F, v, hc_i, Ch_i, Rsp_i) = 1$ means that the verifier accepts the transcript with the honest prover. Let $(c_0, c_1, c_2, c_3, c_4, c_5, c_6)$ be commitment values, $hc_0 = hc_1 = hc_2 = hc_3$ be the hash values of the commitment values, and

$$\begin{aligned}
Rsp_0 &= (r_1^{(0)}, t_1^{(0)}, e_1^{(0)}, d_0^{(0)}, u_0^{(0)}, d_1^{(0)}, u_1^{(0)}, c_0, c_2),\\
Rsp_1 &= (r_1^{(1)}, t_0^{(1)}, e_0^{(1)}, d_0^{(1)}, u_0^{(1)}, d_2^{(1)}, u_2^{(1)}, c_0, c_3),\\
Rsp_2 &= (r_0^{(2)}, t_1^{(2)}, e_1^{(2)}, d_0^{(2)}, u_0^{(2)}, d_2^{(2)}, u_2^{(2)}, c_1, c_5),\\
Rsp_3 &= (r_0^{(3)}, t_0^{(3)}, e_0^{(3)}, d_0^{(3)}, u_0^{(3)}, d_1^{(3)}, u_1^{(3)}, c_1, c_6).
\end{aligned}$$

After the first set s_0 is used, then the equations are as follows:

$$\begin{aligned}
c_0 &= Com(r_0^{(2)}, v - G(r_0^{(2)}, d_0^{(2)} + d_2^{(2)}) - F(r_0^{(2)}) - u_0^{(2)} - u_2^{(2)})\\
c_1 &= Com(r_1^{(0)}, v - G(t_1^{(0)}, r_1^{(0)}) - F(r_1^{(0)}) - e_1^{(0)})\\
&= Com(r_1^{(1)}, G(t_0^{(1)}, r_1^{(1)}) + e_0^{(1)}) \qquad (4)\\
c_2 &= Com(t_0^{(1)}, e_0^{(1)}) = Com(r_0^{(2)} - t_1^{(2)}, F(r_0^{(2)}) - e_1^{(2)})\\
c_3 &= Com(t_1^{(0)}, e_1^{(0)}) = Com(t_1^{(2)}, e_1^{(2)})\\
c_4 &= Com(d_0^{(0)}, u_0^{(0)}) = Com(d_0^{(1)}, u_0^{(1)}) = Com(d_0^{(2)}, u_0^{(2)})\\
c_5 &= Com(d_1^{(0)}, u_1^{(0)}) = Com(r_1^{(1)} - d_0^{(1)} - d_2^{(1)}, F(r_1^{(1)}) - u_0^{(1)} - u_2^{(1)})\\
c_6 &= Com(r_1^{(0)} - d_0^{(0)} - d_1^{(0)}, F(r_1^{(0)}) - u_0^{(0)} - u_1^{(0)})\\
&= Com(d_2^{(1)}, u_2^{(1)}) = Com(d_2^{(2)}, u_2^{(2)}).
\end{aligned}$$

Due to the binding property of Com, we can write the following equations:

$r_1^{(0)} = r_1^{(1)}$, $t_0^{(1)} = r_0^{(2)} - t_1^{(2)}$, $e_0^{(1)} = F(r_0^{(2)}) - e_1^{(2)}$, $t_1^{(0)} = t_1^{(2)}$, $e_1^{(0)} = e_1^{(2)}$, $d_0^{(0)} = d_0^{(1)} = d_0^{(2)}$, $u_0^{(0)} = u_0^{(1)} = u_0^{(2)}$, $d_1^{(0)} = r_1^{(1)} - d_0^{(1)} - d_2^{(1)}$, $u_1^{(0)} = F(r_1^{(1)}) - u_0^{(1)} - u_2^{(1)}$, $r_1^{(0)} - d_0^{(0)} - d_1^{(0)} = d_2^{(1)} = d_2^{(2)}$, $F(r_1^{(0)}) - u_0^{(0)} - u_1^{(0)} = u_2^{(1)} = u_2^{(2)}$.

By Eq. (4),

$$v = G(t_0^{(1)}, r_1^{(1)}) + G(t_1^{(0)}, r_1^{(0)}) + F(r_1^{(0)}) + e_1^{(0)} + e_0^{(1)}. \qquad (5)$$

If we use $r_1^{(1)}$ instead of $r_1^{(0)}$, $t_1^{(2)}$ instead of $t_1^{(0)}$ and $r_0^{(2)} - t_1^{(2)}$ instead of $t_0^{(1)}$, $e_1^{(2)}$ instead of $e_1^{(0)}$, and $F(r_0^{(2)}) - e_1^{(2)}$ instead of $e_0^{(1)}$ in Eq. (5), we get $v = G(r_0^{(2)} - t_1^{(2)}, r_1^{(1)}) + G(t_1^{(2)}, r_1^{(1)}) + F(r_0^{(2)}) - e_1^{(2)} + e_1^{(2)} + F(r_1^{(1)})$, and then by using Definition 4, we can conclude that $v = G(r_0^{(2)} - t_1^{(2)} + t_1^{(2)}, r_1^{(1)}) + F(r_0^{(2)}) - e_1^{(2)} + e_1^{(2)} + F(r_1^{(1)}) = G(r_0^{(2)}, r_1^{(1)}) + F(r_0^{(2)}) + F(r_1^{(1)}) = F(r_0^{(2)} + r_1^{(1)})$, and the secret key is obtained with $r_0^{(2)} + r_1^{(1)}$.

After the second set s_1 is used, then the equations are as follows:

$$\begin{aligned}
c_0 &= Com(r_0^{(3)}, G(r_0^{(3)}, d_1^{(3)}) + u_1^{(3)}) \\
c_1 &= Com(r_1^{(0)}, v - G(t_1^{(0)}, r_1^{(0)}) - F(r_1^{(0)}) - e_1^{(0)}) \\
&= Com(r_1^{(1)}, G(t_0^{(1)}, r_1^{(1)}) + e_0^{(1)}) \\
c_2 &= Com(t_0^{(1)}, e_0^{(1)}) = Com(t_0^{(3)}, e_0^{(3)}) \\
c_3 &= Com(t_1^{(0)}, e_1^{(0)}) = Com(r_0^{(3)} - t_0^{(3)}, F(r_0^{(3)}) - e_0^{(3)}) \\
c_4 &= Com(d_0^{(0)}, u_0^{(0)}) = Com(d_0^{(1)}, u_0^{(1)}) = Com(d_0^{(3)}, u_0^{(3)}) \\
c_5 &= Com(d_1^{(0)}, u_1^{(0)}) = Com(r_1^{(1)} - d_0^{(1)} - d_2^{(1)}, F(r_1^{(1)}) - u_0^{(1)} - u_2^{(1)}) \\
&= Com(d_1^{(3)}, u_1^{(3)}) \\
c_6 &= Com(r_1^{(0)} - d_0^{(0)} - d_1^{(0)}, F(r_1^{(0)}) - u_0^{(0)} - u_1^{(0)}) = Com(d_2^{(1)}, u_2^{(1)}).
\end{aligned} \tag{6}$$

Due to the binding property of Com, we can write the following equations:

$r_1^{(0)} = r_1^{(1)}$, $t_0^{(1)} = t_0^{(3)}$, $e_0^{(1)} = e_0^{(3)}$, $t_1^{(0)} = r_0^{(3)} - t_0^{(3)}$, $e_1^{(0)} = F(r_0^{(3)}) - e_0^{(3)}$, $d_0^{(0)} = d_0^{(1)} = d_0^{(3)}$, $u_0^{(0)} = u_0^{(1)} = u_0^{(3)}$, $d_1^{(0)} = r_1^{(1)} - d_0^{(1)} - d_2^{(1)} = d_1^{(3)}$, $u_1^{(0)} = F(r_1^{(1)}) - u_0^{(1)} - u_2^{(1)} = u_1^{(3)}$, $r_1^{(0)} - d_0^{(0)} - d_1^{(0)} = d_2^{(1)}$, $F(r_1^{(0)}) - u_0^{(0)} - u_1^{(0)} = u_2^{(1)}$.

By Eq. (6),

$$v = G(t_1^{(0)}, r_1^{(0)}) + G(t_0^{(1)}, r_1^{(1)}) + F(r_1^{(0)}) + e_1^{(0)} + e_0^{(1)}. \tag{7}$$

If we use $r_1^{(0)}$ instead of $r_1^{(1)}$, $r_0^{(3)} - t_0^{(3)}$ instead of $t_1^{(0)}$, $t_0^{(3)}$ instead of $t_0^{(1)}$, $F(r_0^{(3)}) - e_0^{(3)})$ instead of $e_1^{(0)}$, and $e_0^{(3)}$ instead of $e_0^{(1)}$ in Eq. (7), we get $v = G(r_0^{(3)} - t_0^{(3)}, r_1^{(0)}) + G(t_0^{(3)}, r_1^{(0)}) + F(r_0^{(3)}) - e_0^{(3)} + e_0^{(3)} + F(r_1^{(0)})$. Then, by Definition 4, we can conclude that $v = G(r_0^{(3)} - t_0^{(3)} + t_0^{(3)}, r_1^{(0)}) + F(r_0^{(3)}) - e_0^{(3)} + e_0^{(3)} + F(r_1^{(0)}) = G(r_0^{(3)}, r_1^{(0)}) + F(r_0^{(3)}) + F(r_1^{(0)}) = F(r_0^{(3)} + r_1^{(0)})$, and for the secret key s, $r_0^{(3)} + r_1^{(0)}$ is a value.

After the third set s_2 is used, then the equations are as follows:

$$\begin{aligned}
c_0 &= Com(r_0^{(2)}, v - G(r_0^{(2)}, d_0^{(2)} + d_2^{(2)}) - F(r_0^{(2)}) - u_0^{(2)} - u_2^{(2)}) \\
&= Com(r_0^{(3)}, G(r_0^{(3)}, d_1^{(3)}) + u_1^{(3)}) \\
c_1 &= Com(r_1^{(0)}, v - G(t_1^{(0)}, r_1^{(0)}) - F(r_1^{(0)}) - e_1^{(0)}) \\
c_2 &= Com(r_0^{(2)} - t_1^{(2)}, F(r_0^{(2)}) - e_1^{(2)}) = Com(t_0^{(3)}, e_0^{(3)}) \\
c_3 &= Com(t_1^{(0)}, e_1^{(0)}) = Com(t_1^{(2)}, e_1^{(2)}) = Com(r_0^{(3)} - t_0^{(3)}, F(r_0^{(3)}) - e_0^{(3)}) \\
c_4 &= Com(d_0^{(0)}, u_0^{(0)}) = Com(d_0^{(2)}, u_0^{(2)}) = Com(d_0^{(3)}, u_0^{(3)}) \\
c_5 &= Com(d_1^{(0)}, u_1^{(0)}) = Com(d_1^{(3)}, u_1^{(3)}) \\
c_6 &= Com(r_1^{(0)} - d_0^{(0)} - d_1^{(0)}, F(r_1^{(0)}) - u_0^{(0)} - u_1^{(0)}) = Com(d_2^{(2)}, u_2^{(2)}).
\end{aligned} \tag{8}$$

Due to the binding property of Com, we can write the following equations:

$r_0^{(2)} = r_0^{(3)}$, $r_0^{(2)} - t_1^{(2)} = t_0^{(3)}$, $F(r_0^{(2)}) - e_1^{(2)} = e_0^{(3)}$, $t_1^{(0)} = t_1^{(2)} = r_0^{(3)} - t_0^{(3)}$, $e_1^{(0)} = e_1^{(2)} = F(r_0^{(3)}) - e_0^{(3)}$, $d_0^{(0)} = d_0^{(2)} = d_0^{(3)}$, $u_0^{(0)} = u_0^{(2)} = u_0^{(3)}$, $d_1^{(0)} = d_1^{(3)}$, $u_1^{(0)} = u_1^{(3)}$, $r_1^{(0)} - d_0^{(0)} - d_1^{(0)} = d_2^{(2)}$, $F(r_1^{(0)}) - u_0^{(0)} - u_1^{(0)} = u_2^{(2)}$.

By Eq. (8),

$$v = G(r_0^{(2)}, d_0^{(2)} + d_2^{(2)}) + G(r_0^{(3)}, d_1^{(3)}) + F(r_0^{(2)}) + u_0^{(2)} + u_1^{(3)} + u_2^{(2)}. \tag{9}$$

If we use $r_0^{(3)}$ instead of $r_0^{(2)}$, $d_0^{(0)}$ instead of $d_0^{(2)}$, $r_1^{(0)} - d_0^{(0)} - d_1^{(0)}$ instead of $d_2^{(2)}$, $d_1^{(0)}$ instead of $d_1^{(3)}$, $u_0^{(0)}$ instead of $u_0^{(2)}$, $u_1^{(0)}$ instead of $u_1^{(3)}$, and $F(r_1^{(0)}) - u_0^{(0)} - u_1^{(0)}$ instead of $u_2^{(2)}$ in Eq. (9), we get $v = G(r_0^{(3)}, d_0^{(0)} + r_1^{(0)} - d_0^{(0)} - d_1^{(0)}) + G(r_0^{(3)}, d_1^{(0)}) + F(r_0^{(3)}) + u_0^{(0)} + u_1^{(0)} + F(r_1^{(0)}) - u_0^{(0)} - u_1^{(0)}$. Then, by Definition 4, we can conclude that $v = G(r_0^{(3)}, r_1^{(0)} + d_0^{(0)} - d_0^{(0)} - d_1^{(0)} + d_1^{(0)}) + F(r_0^{(3)}) + u_0^{(0)} + u_1^{(0)} + F(r_1^{(0)}) - u_0^{(0)} - u_1^{(0)} = G(r_0^{(3)}, r_1^{(0)}) + F(r_0^{(3)}) + F(r_1^{(0)}) = F(r_0^{(3)} + r_1^{(0)})$, and the secret key is obtained with $r_0^{(3)} + r_1^{(0)}$.

After the last set s_3 is used, then the equations are as follows:

$$\begin{aligned}
c_0 &= Com(r_0^{(2)}, v - G(r_0^{(2)}, d_0^{(2)} + d_2^{(2)}) - F(r_0^{(2)}) - u_0^{(2)} - u_2^{(2)}) \\
&= Com(r_0^{(3)}, G(r_0^{(3)}, d_1^{(3)}) + u_1^{(3)}) \qquad (10) \\
c_1 &= Com(r_1^{(1)}, G(t_0^{(1)}, r_1^{(1)}) + e_0^{(1)}) \\
c_2 &= Com(t_0^{(1)}, e_0^{(1)}) = Com(r_0^{(2)} - t_1^{(2)}, F(r_0^{(2)}) - e_1^{(2)}) = Com(t_0^{(3)}, e_0^{(3)}) \\
c_3 &= Com(t_1^{(2)}, e_1^{(2)}) = Com(r_0^{(3)} - t_0^{(3)}, F(r_0^{(3)}) - e_0^{(3)}) \\
c_4 &= Com(d_0^{(1)}, u_0^{(1)}) = Com(d_0^{(2)}, u_0^{(2)}) = Com(d_0^{(3)}, u_0^{(3)}) \\
c_5 &= Com(r_1^{(1)} - d_0^{(1)} - d_2^{(1)}, F(r_1^{(1)}) - u_0^{(1)} - u_2^{(1)}) = Com(d_1^{(3)}, u_1^{(3)}) \\
c_6 &= Com(d_2^{(1)}, u_2^{(1)}) = Com(d_2^{(2)}, u_2^{(2)}).
\end{aligned}$$

Since Com has the binding property, we can write the following equations:

$r_0^{(2)} = r_0^{(3)}$, $t_0^{(1)} = r_0^{(2)} - t_1^{(2)} = t_0^{(3)}$, $e_0^{(1)} = F(r_0^{(2)}) - e_1^{(2)} = e_0^{(3)}$, $t_1^{(2)} = r_0^{(3)} - t_0^{(3)}$, $e_1^{(2)} = F(r_0^{(3)}) - e_0^{(3)}$, $d_0^{(1)} = d_0^{(2)} = d_0^{(3)}$, $u_0^{(1)} = u_0^{(2)} = u_0^{(3)}$, $r_1^{(1)} - d_0^{(1)} - d_2^{(1)} = d_1^{(3)}$, $F(r_1^{(1)}) - u_0^{(1)} - u_2^{(1)} = u_1^{(3)}$, $d_2^{(1)} = d_2^{(2)}$, $u_2^{(1)} = u_2^{(2)}$.

By Eq. (10),

$$v = G(r_0^{(2)}, d_0^{(2)} + d_2^{(2)}) + G(r_0^{(3)}, d_1^{(3)}) + F(r_0^{(2)}) + u_0^{(2)} + u_2^{(2)} + u_1^{(3)}. \tag{11}$$

If we use $r_0^{(3)}$ instead of $r_0^{(2)}$, $u_0^{(1)}$ instead of $u_0^{(2)}$, $F(r_1^{(1)}) - u_0^{(1)} - u_2^{(1)}$ instead of $u_1^{(3)}$, $u_2^{(1)}$ instead of $u_2^{(2)}$, $r_1^{(1)} - d_0^{(1)} - d_2^{(1)}$ instead of $d_1^{(3)}$, $d_0^{(1)}$ instead of $d_0^{(2)}$, and $d_2^{(1)}$ instead of $d_2^{(2)}$ in Eq. (11), we get $v = G(r_0^{(3)}, d_0^{(1)} + d_2^{(1)}) + G(r_0^{(3)}, r_1^{(1)} - d_0^{(1)} - d_2^{(1)}) + F(r_0^{(3)}) + u_0^{(1)} + F(r_1^{(1)}) - u_0^{(1)} - u_2^{(1)} + u_2^{(1)}$. Then, by using Definition 4, we can conclude that $v = G(r_0^{(3)}, d_0^{(1)} + d_2^{(1)} + r_1^{(1)} - d_0^{(1)} - d_2^{(1)}) + F(r_0^{(3)}) + u_0^{(1)} + F(r_1^{(1)}) - u_0^{(1)} - u_2^{(1)} + u_2^{(1)} = G(r_0^{(3)}, r_1^{(1)}) + F(r_0^{(3)}) + F(r_1^{(1)}) = F(r_0^{(3)} + r_1^{(1)})$, and the secret key is obtained with $r_0^{(3)} + r_1^{(1)}$.

The proof indicates that anyone must be able to respond to at least 3 out of 4 challenges to reach the secret key or solve the multivariate quadratic system. □

Remark 1 *Impersonation probability is closely related to Theorem 2. The error probability in Theorem 2 indicates the possibility that an attacker cannot impersonate. The error probability is the complement of the impersonation probability.*

3.2. Observations

Recall that the computation time is computed by the number of F and G functions used in Com [9]. Note that less computation time is better in view of efficient schemes. We note that the computation time remains the same when the secret key is first divided into two parts regardless of the number of subparts during the experiments. In other words, if the polar form given in Definition 4 is used and the secret key is divided into two parts, the computation time is $5/2$ due to the bilinear polar form. However, when the number of subparts of the secret key increases, the total memory also increases.

Remark 2 *The first question that comes to mind is whether the secret key can be divided into more than two parts to build more efficient scheme based on multivariate quadratic polynomials. It is not possible to construct an identification scheme with the polar form given in Definition 4 by dividing the secret key into more than two parts. Recall that the polar form given in Definition 4 has the bilinear property. For this reason, if the secret key parts differ from two, the bilinear property is not provided. Therefore, Theorem 1 and Theorem 2 cannot be achieved.*

4. Efficiency analysis and comparison

In this section, we analyze the proposed scheme and compare it with other identification schemes based on multivariate quadratic polynomials in terms of memory requirements, communication length, impersonation probability, computation time, and the efficiency metric. The results are summarized in Table 1.

Table 1. The identification schemes comparison table.

	[12]	[9]	Our scheme
Prover	$9m + 2n$	$16m + 3n$	$18m + 3n$
Verifier	$4m$	$8m$	$10m$
Communication length	$5m + 2n + 2$	$8m + 3n + 2$	$9m + 4n + 2$
Total	$18m + 4n + 2$	$32m + 6n + 2$	$37m + 7n + 2$
Impersonation probability	2/3	1/2	1/2
Efficiency metric	1.14	1	1
Computation time	4/3	5/2	5/2

Definition 9 *The efficiency metric (em) for identification schemes is computed by using impersonation probability (ip) and computation time (ct), i.e. the number of F and G functions without the repeated functions in the same challenge. Let $\Delta(t)$ be the time required for the calculation of each of $F(x)$ and $G(x,y)$ dominant parts and r be the number of rounds for the security level. Then $em = \lambda \cdot r \cdot ct \cdot \Delta(t)$, where λ is the lower bound for the required number of r rounds and is computed with $(ip)^r \leq 2^{-\lambda}$ [9].*

We provide some notations for memory requirements:

- Each subpart of the secret key is n-bit for $s \in \mathbb{F}_q^n$.
- The output of the F polynomial system is m-bit.
- The hash values and the outputs of Com are $2m$-bit.

Now we compute the memory requirement for the prover side. At the beginning of the identification scheme, the prover computes three subparts, each of which is n-bit, with the outputs of these two subparts by F that are $2m$-bit. Then the prover computes seven commitment values, each of which is $2m$-bit, and the hash value hc of the commitment values. Total memory is $18m + 3n$ bits.

Now we compute the memory requirement for the verifier side. The verifier computes five commitment values in each round, $10m$-bit. The communication length is computed as follows: it needs hc, and challenge and response values are $2m$-bit. The verifier's challenge is 2-bit and Rsp, the prover's response, is $7m + 4n$ bits.

The total communication length is $9m + 4n + 2$ bits. Total memory includes prover, verifier, and communication length. Thus, it is $37m + 7n + 2$ bits. F and G functions are the most time-consuming parts of Com. In addition, lower numbers of them mean better timing results since the computation time is obtained by dividing the number of these F and G functions into challenges [9].

Remark 3 *The more partitions that the secret key is divided into, provided that the polar form and the properties of bilinear functions are preserved, the lower the probability is that the adversary will reach the secret key. By Table 1, as the number of subparts of the secret key increases, the total memory also increases accordingly. When the total memory is computed for $n = 84$, $m = 80$ bits, there is no significant difference between these systems in view of memory requirements.*

5. A new MQ-based signature scheme

In this section, we convert the proposed 3-pass identification scheme into a signature scheme. Let $k \in \mathbb{N}$ be a security parameter (for instance, $k = 128$), while $n, m \in \mathbb{N}$ are the same as in Definition 1. Let F and $F^{len} = m \cdot \frac{n(n+1)}{2}$ [1] indicate the MQ system and size of the MQ system F as in Definition 1, respectively. We need the following functions:

- Cryptographic hash functions: $\mathcal{H} : \{0,1\}^* \rightarrow \{0,1\}^k$, $\mathcal{H}_1 = \{0,1\}^{2k} \rightarrow \{0,1,2,3\}^r$
- A string commitment function: $Com : \mathbb{F}_2^n \times \mathbb{F}_2^m \rightarrow \{0,1\}^k$
- Pseudorandom generators: $G_{S_F} = \{0,1\}^k \rightarrow \mathbb{F}_2^{F_{len}}$, $G_{S_K} = \{0,1\}^k \rightarrow \mathbb{F}_2^n$, $G_C : \{0,1\}^{2k} \rightarrow \mathbb{F}_2^{r(4n+3m)}$

Key generation (KeyGen): We first generate a key pair (p_k, s_k) and the MQ system F. We randomly select a k-bit seed that is $SK \leftarrow_R \{0,1\}^k$ for the secret key and $S_F \leftarrow_R \{0,1\}^k$ for F. By using pseudorandom generator G_{S_F} and G_{S_K} with the related seeds S_F and S_K, we construct MQ system $F = G_{S_F}(S_F)$ and $SK_{\mathbb{F}_2} = G_{S_K}(SK)$, respectively. Then we compute $PK_v = F(SK_{\mathbb{F}_2})$ to obtain part of the public key. The key generation algorithm ends with a pair of keys (p_k, s_k), where $p_k = (S_F, PK_v)$ and $s_k = (S_F, SK)$.

[1] Since we are calculating over $\mathbb{F}_2$, $x^2 = x$, we do not count quadratic terms.

Signature generation (Sign): The digital signature scheme runs the *Sign* algorithm after the key generation algorithm. The *Sign* algorithm inputs the message $m \in \{0,1\}^*$ and the secret key $s_k = (S_F, SK)$. We generate $F = G_{S_F}(S_F)$ as it is the same in *KeyGen*. We compute a message value $M = \mathcal{H}(SK||m)$, where $||$ denotes a string concatenation. Then we again compute a message digest value $D = \mathcal{H}(M||m)$. The signature has to include M to verify the same message digest in verification.

By Section 4, the prover selects the secret key parts in the beginning of the scheme. As the identification scheme, we expand (SK, D) to generate the secret key parts by using pseudorandom generator G_C for producing the values: $(\mathrm{r}_{(0,0)}, \ldots, \mathrm{r}_{(0,r)}, t_{(0,0)}, \ldots, t_{(0,r)}, e_{(0,0)}, \ldots, e_{(0,r)}, d_{(0,0)}, \ldots, d_{(0,r)}, d_{(1,0)}, \ldots, d_{(1,r)}, u_{(0,0)}, \ldots, u_{(0,r)}, u_{(1,0)}, \ldots, u_{(1,r)})$, where r indicates the number of required rounds. Then we compute the following values for each round i as in Section 4:

$\mathrm{r}_{(1,i)} = s_k - \mathrm{r}_{(0,i)}$, $t_{(1,i)} = \mathrm{r}_{(0,i)} - t_{(0,i)}$, $e_{(1,i)} = F(\mathrm{r}_{(0,i)}) - e_{(0,i)}$, $u_{(2,i)} = F(\mathrm{r}_{(1,i)}) - u_{(0,i)} - u_{(1,i)}$, $d_{(2,i)} = r_{(1,i)} - d_{(0,i)} - d_{(1,i)}$. After getting all values, we compute commitment values using the string commitment function Com. The commitment values are as follows:

$c_{(0,i)} = Com(\mathrm{r}_{(0,i)}, G(\mathrm{r}_{(0,i)}, d_{(1,i)}) + u_{(1,i)})$,

$c_{(1,i)} = Com(\mathrm{r}_{(1,i)}, G(t_{(0,i)}, r_{(1,i)}) + e_{(0,i)})$,

$c_{(2,i)} = Com(t_{(0,i)}, e_{(0,i)})$,

$c_{(3,i)} = Com(t_{(1,i)}, e_{(1,i)})$,

$c_{(4,i)} = Com(d_{(0,i)}, u_{(0,i)})$,

$c_{(5,i)} = Com(d_{(1,i)}, u_{(1,i)})$,

$c_{(6,i)} = Com(d_{(2,i)}, u_{(2,i)})$.

We compute $\sigma_0 = \mathcal{H}(c_{(0,0)}||c_{(1,0)}||c_{(2,0)}||c_{(3,0)}||c_{(4,0)}||c_{(5,0)}||c_{(6,0)}|| \ldots ||c_{(0,r-1)}||c_{(1,r-1)}||c_{(2,r-1)}||c_{(3,r-1)}||c_{(4,r-1)} ||c_{(5,r-1)}||c_{(6,r-1)})$. Then we obtain the challenge from $Ch_i = \mathcal{H}_1(D, \sigma_0)$. We concatenate all responses for each challenges ch_i and get $\sigma_1 = (\mathrm{r}_{(1,i)}||t_{(1,i)}||e_{(1,i)}||d_{(0,i)}||u_{(0,i)}||d_{(1,i)}||u_{(1,i)}||c_{(0,i)}||c_{(2,i)}||\mathrm{r}_{(1,i)}||t_{(0,i)}||e_{(0,i)}||d_{(0,i)}||u_{(0,i)}|| d_{(2,i)}||u_{(2,i)}||c_{(0,i)}||c_{(3,i)}||\mathrm{r}_{(0,i)}||t_{(1,i)}||e_{(1,i)}||d_{(0,i)}||u_{(0,i)}||d_{(2,i)}||u_{(2,i)}||c_{(1,i)}||c_{(5,i)}||\mathrm{r}_{(0,i)}||t_{(0,i)}||e_{(0,i)}||d_{(0,i)}||u_{(0,i)}||d_{(1,i)}||u_{(1,i)} ||c_{(1,i)}||c_{(6,i)})$. Note that since some $c_{(j,i)}$ commitment values cannot be computed for each Ch_i, we add them to σ_1. In other words, for $Ch_i = \{0,1,2,3\}$, σ_1 contains $\{c_{(0,i)}, c_{(2,i)}\}$, $\{c_{(0,i)}, c_{(3,i)}\}$, $\{c_{(1,i)}, c_{(5,i)}\}$, $\{c_{(1,i)}, c_{(6,i)}\}$, respectively. The signature is $\sigma = (M, \sigma_0, \sigma_1)$ and the size of the signature is $2k + r \cdot (4n + 3m + 2k)$ bits.

Signature verification (Verify): The last phase of the digital signature scheme is verification. The verification algorithm inputs message m, public key $p_k = (S_F, PK_v)$, and signature $\sigma = (M, \sigma_0, \sigma_1)$. First, the MQ system is obtained from $F = G_{S_F}(S_F)$. We use M and m to compute the message digest $D = \mathcal{H}(M||m)$. We compute the challenges $Ch_i = \mathcal{H}_1(D, \sigma_0)$ for each r rounds. Then the responses of these challenges are extracted from σ_1 for all rounds. We obtain all commitment values as follows:

The verifier computes $\sigma_0' = \mathcal{H}(c_{(0,0)}||c_{(1,0)}||c_{(2,0)}||c_{(3,0)}||c_{(4,0)}||c_{(5,0)}||c_{(6,0)}|| \ldots ||c_{(0,r-1)}||c_{(1,r-1)}||c_{(2,r-1)}||c_{(3,r-1)} ||c_{(4,r-1)}||c_{(5,r-1)}||c_{(6,r-1)})$. If $\sigma_0' = \sigma_0$, the signature is verified.

5.1. Security analysis of the proposed signature scheme

Now we give the security analysis for the proposed signature scheme.

Theorem 3 *The proposed signature scheme is existentially unforgeable under adaptively chosen-message (EU-CMA) attacks in the random oracle model if:*

$$\text{If } Ch_i = 0, \quad \begin{cases} c_{(1,i)} = Com(\mathrm{r}_{(1,i)}, PK_v - G(t_{(1,i)},\mathrm{r}_{(1,i)}) - F(\mathrm{r}_{(1,i)}) - e_{(1,i)}) \\ c_{(3,i)} = Com(t_{(1,i)}, e_{(1,i)}) \\ c_{(4,i)} = Com(d_{(0,i)}, u_{(0,i)}) \\ c_{(5,i)} = Com(d_{(1,i)}, u_{(1,i)}) \\ c_{(6,i)} = Com(\mathrm{r}_{(1,i)} - d_{(0,i)} - d_{(1,i)}, F(\mathrm{r}_{(1,i)}) - u_{(0,i)} - u_{(1,i)}) \end{cases}$$

$$\text{If } Ch_i = 1, \quad \begin{cases} c_{(1,i)} = Com(\mathrm{r}_{(1,i)}, G(t_{(0,i)},\mathrm{r}_{(1,i)}) + e_{(0,i)}) \\ c_{(2,i)} = Com(t_{(0,i)}, e_{(0,i)}) \\ c_{(4,i)} = Com(d_{(0,i)}, u_{(0,i)}) \\ c_{(5,i)} = Com(\mathrm{r}_{(1,i)} - d_{(0,i)} - d_{(2,i)}, F(\mathrm{r}_{(1,i)}) - u_{(0,i)} - u_{(2,i)}) \\ c_{(6,i)} = Com(d_{(2,i)}, u_{(2,i)}) \end{cases}$$

$$\text{If } Ch_i = 2, \quad \begin{cases} c_{(0,i)} = Com(\mathrm{r}_{(0,i)}, PK_v - G(\mathrm{r}_{(0,i)}, d_{(0,i)} + d_{(2,i)}) - F(\mathrm{r}_{(0,i)}) - u_{(0,i)} - u_{(2,i)}) \\ c_{(2,i)} = Com(\mathrm{r}_{(0,i)} - t_{(1,i)}, F(\mathrm{r}_{(0,i)}) - e_{(1,i)}) \\ c_{(3,i)} = Com(t_{(1,i)}, e_{(1,i)}) \\ c_{(4,i)} = Com(d_{(0,i)}, u_{(0,i)}) \\ c_{(6,i)} = Com(d_{(2,i)}, u_{(2,i)}) \end{cases}$$

$$\text{If } Ch_i = 3, \quad \begin{cases} c_{(0,i)} = Com(\mathrm{r}_{(0,i)}, G(\mathrm{r}_{(0,i)}, d_{(1,i)}) + u_{(1,i)}) \\ c_{(2,i)} = Com(t_{(0,i)}, e_{(0,i)}) \\ c_{(3,i)} = Com(\mathrm{r}_{(0,i)} - t_{(0,i)}, F(\mathrm{r}_{(0,i)}) - e_{(0,i)}) \\ c_{(4,i)} = Com(d_{(0,i)}, u_{(0,i)}) \\ c_{(5,i)} = Com(d_{(1,i)}, u_{(1,i)}) \end{cases}$$

- *The MQ problem that the signature scheme is based on is a problem that cannot be solved with a polynomial-time algorithm,*
- *The hash functions $\mathcal{H}, \mathcal{H}_1$ are replaced by a random oracle,*
- *The commitment function Com is statistically hiding and computationally binding and it is infeasible to construct Com to obtain a certain output,*
- *The pseudorandom generator G_{S_F} is constructed as a random oracle,*
- *The pseudorandom generators G_{S_K} and G_C generate outputs that are not distinguished from random values.*

We prove this theorem by combining the methods in [3, 5]. Our strategy is to use a divide-and-conquer approach for the proof. First, we need to recall some definitions. Then, by using these, we complete the proof. In Definition 10, the definition of Fiat–Shamir transform for the $q2$-identification scheme given in [3] is modified for the proposed identification scheme.

Definition 10 ***(Fiat–Shamir transform for identification scheme)*** *Let $k \in \mathbb{N}$ be a security parameter and $IDS = (KeyGen, Sign, Verify)$ be an identification scheme that satisfies soundness with soundness error κ. Select r that indicates the number of rounds of IDS, such that $\kappa^r = negl(k)$. For IDS^r with r rounds, the challenge is C^r and cryptographic hash function is $\mathcal{H}_1 : \{0,1\}^* \to C^r$. The signature scheme DSS transformed from IDS has three algorithms $(KeyGen, Sign, Verify)$:*

- $(pk, sk) \leftarrow KeyGen(1^k)$,
- $\sigma = (\sigma_0, \sigma_1) \leftarrow Sign(sk, m)$ *where* $\sigma_0 = Com \leftarrow P_0^r(sk), h_1 = \mathcal{H}_1(m, \sigma_0), \sigma_1 = resp \leftarrow P_1^r(sk, \sigma_0, h_1)$,
- $Verify(pk, m, \sigma)$ *algorithm extracts* $\sigma = (\sigma_0, \sigma_1)$, *derives* $h_1 = \mathcal{H}_1(m, \sigma_0)$ *and outputs* $V^r(pk, \sigma_0, h_1, \sigma_1)$.

We need to define an n-generic signature scheme and n-soundness property (given in [5]) since we prove the security of our signature scheme under chosen attacks. The proposed signature scheme by using Definition 10 is a 1-generic signature scheme according to the definition of the n-generic signature scheme given in [5]. Definition 11 is the modified version of n-generic signature scheme for the proposed scheme.

Definition 11 ***(1-generic signature scheme)*** *Let $DSS = (KeyGen, Sign, Verify)$ be a signature scheme, m be a message, and $\mathcal{H}$ be a hash function. $(\sigma_0, h_1, \sigma_1)$ is the form of the 1-generic signature where $h_1 = \mathcal{H}(m, \sigma_0)$ is the hash value. Moreover, the 1-generic signature scheme has an honest-verifier zero-knowledge property (see [5]).*

Definition 12 ***(1-Soundness)*** *Let $DSS = (KeyGen, Sign, Verify)$ be a 1-generic signature scheme. If there exists a probabilistic polynomial-time algorithm (PPT) ε, called an extractor, for any $(sk, pk) \leftarrow KeyGen(1^k)$ for any two distinct valid signatures $\sigma = (\sigma_0, h_1, \sigma_1)$ and $\sigma' = (\sigma_0, h_1', \sigma_1')$, where $h_1 \neq h_1'$, we get $sk \leftarrow \varepsilon(pk, m, \sigma, \sigma')$ with non-negligible probability.*

Now we can prove Theorem 3.

Proof First, we start with the first item in Theorem 3. In [5], it is stated that DSS is EU-CMA secure in a random oracle model if DSS is an n-generic signature scheme that achieves n-soundness with underlying hard problem. Our signature scheme is derived from our proposed identification scheme. The computational hardness of the proposed scheme is based on an MQ problem. Note that we explain that there is no polynomial-time algorithm to solve this problem in Definition 2. Considering the related theorem and its proof in [5], our 1-generic signature scheme that achieves soundness with underlying MQ problem is EU-CMA secure.

We prove the other items in Theorem 3 by defining the games. Assume that there exists an adversary A that wins the EU-CMA game with non-negligible probability. We claim that the difference between the success probabilities of A is negligible in these games. Suppose that A has an oracle machine $\mathcal{O}$ that can solve the MQ problem, disrupt the hiding and binding properties that the commitment scheme satisfies, find the distinct message corresponding to the same hash value, or distinguish the outputs of pseudorandom generators from random values.

Game 0 is the EU-CMA game for DSS.
Game 1 is Game 0 with the difference that $\mathcal{O}$ changes the outputs of G_{SK} with random values.
Game 2 is Game 1 with the difference that $\mathcal{O}$ changes the outputs of G_C with random values.
Game 3 is Game 2 with the difference that $\mathcal{O}$ generates the random coefficients for an F polynomials system like pseudorandom generator G_{S_F} that inputs S_F.

Assume that A wins Game 0 with ϵ non-negligible success probability. If the difference between Game 0 and Game 1 that A wins is non-negligible, it means that A knows the outputs of pseudorandom generators and changes them with random values. Similarly, we can say the same things for Game 1 - Game 2 and Game 2 - Game 3. For these games, A has the same success probability. Note that A wins Game 3 and Game 0 with the same success probability. Thus, Game 3 is EU-CMA for a DSS transformed identification scheme based on MQ polynomials. On the contrary, DSS is EU-CMA secure if DSS is an n-generic signature scheme satisfying n-soundness. Note that the proposed DSS, constructed by using Fiat–Shamir transform for the identification scheme given in Definition 10, is an n-generic signature scheme. Moreover, the Com commitment scheme satisfies the statistically hiding and computationally binding properties. Since the identification scheme has the zero-knowledge and soundness properties given in Theorem 1 and Theorem 2, the proposed DSS is EU-CMA secure. □

5.2. Comparison

In this section, we compare the proposed signature scheme with the previous ones in view of signature and key sizes. In Table 2, the signature schemes constructed from identification schemes are compared in terms of key and signature sizes. Note that we construct the digital signature scheme (DSS) based on [9] to give a comparison. Since we focus on our identification scheme and its effects, we omit DSS based on [9]. In Table 2, "3-pass" refers to the signature scheme that transforms from the 3-pass identification scheme in [3].

Table 2. Signature and key size of signature schemes.

		3-pass [3]	DSS based on [9]	The proposed DSS	5-pass [4]
Signature size		$2k + r(2n + m + k)$	$2k + r(3n + 2m + 2k)$	$2k + r(4n + 3m + 2k)$	$4k + r(3n\lceil log_2 q\rceil + 2k)$
Key size	Public key	$m + k$	$m + k$	$m + k$	$k + n\lceil log_2 q\rceil$
	Secret key	$2k$	$2k$	$2k$	k

Now we compute the signature and key sizes. Recall that the signature is $\sigma = (M, \sigma_0, \sigma_1)$. In the signature, M and σ_0, which are the outputs of the hash function, are k-bit. Then σ_1, the concatenation of all responses for each round, is $r(4n + 3m + 2k)$ bits. Total signature size is $2k + r(4n + 3m + 2k)$. The proposed scheme has the same key size with 3-pass signature scheme and is smaller than the 5-pass signature scheme. The signature size of the proposed scheme is smaller than the 5-pass signature scheme. The proposed scheme has lower signature and key sizes when compared to the 5-pass given in [4].

6. Conclusions and future works

We propose a novel 3-pass zero-knowledge identification scheme based on multivariate quadratic polynomials. We divide the secret key into partitions with a new partition technique by using the same polar form in [9, 12]. We present a new identification scheme having the same computation time as in [9]. We also propose a new MQ-based signature scheme by applying Fiat–Shamir transform to the proposed 3-pass identification scheme. We reduce the signature and key sizes compared to the 5-pass signature scheme in [4]. As a future work, a new digital signature scheme based on the identification scheme given in [9] can be constructed. In addition, an identification scheme and related signature scheme can be proposed by d-linear polar form.

Acknowledgments

This research was partially supported by TÜBİTAK under grant no. EEEAG-116E279. The authors would like to express their gratitude to the anonymous reviewers for their invaluable suggestions in putting the present study into its final form.

References

[1] Abdalla M, An JH, Bellare M and Namprempre C. From identification to signatures via the Fiat-Shamir transform: minimizing assumptions for security and forward-security. Lect Notes Comp Sci 2002; 2332: 418-433.

[2] Bernstein DJ. Buchmann J, Dahmen E. Post-Quantum Cryptography. Berlin, Germany: Springer, 2009.

[3] Chen MS, Hülsing A, Rijneveld J, Samardjiska S, Schwabe P. From 5-pass MQ-based identification to MQ-based signatures. In: International Conference on the Theory and Application of Cryptology and Information Security - ASIACRYPT 2016; 4–8 December 2016; Hanoi, Vietnam. Berlin, Germany: Springer, 2016. pp. 135-165.

[4] Chen MS, Hülsing A, Rijneveld J, Samardjiska S, Schwabe, P. MQDSS specifications version 1.1. In: NIST's First PQC Standardization Conference; 2018.

[5] Dagdelen O, Galindo D, Veron P, El Yousfi Alaoui SM, Cayrel PL. Extended security arguments for signature schemes. Design Code Cryptogr 2016; 78: 441-461.

[6] Feige U, Fiat A, Shamir A. Zero-knowledge proofs of identity. J Cryptol 1988; 1: 77–94.

[7] Fiat A, Shamir A. How to prove yourself: practical solutions to identification and signature problems. In: Advances in Cryptology — CRYPTO'86; 1987. pp. 186-194.

[8] Hornschuch M. Multivariate-based identification and signature schemes with additional properties. MSc, Technische Universitat Darmstadt, Germany, 2012.

[9] Monteiro FS, Goya DH, Terada R. Improved identification protocol based on the MQ problem. IEICE T Fund Electr 2015; E98-A: 1255-1265.

[10] Okamoto T. Provably secure and practical identification schemes and corresponding signature schemes. Lect Notes Comp Sci 1992; 740: 31-53.

[11] Sakumoto K. Public-key identification schemes based on multivariate cubic polynomials. Lect Notes Comp Sci 2012; 7293: 172-189.

[12] Sakumoto K, Shirai T, Hiwatari H. Public-key identification schemes based on multivariate quadratic polynomials. Lect Notes Comp Sci 2011; 6841: 706-723.

[13] Shor PW. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. SIAM J Comput 1997; 26: 1484-1509.

[14] Simari GI. A Primer on Zero Knowledge Protocols. Technical Report. Buenos Aires, Argentina: Universidad Nacional del Sur, 2002.