A New Scheme for Zero Knowledge Proof based on Multivariate Quadratic Problem and Quaternion Algebra

Mehdi Vasef
Student Member, IEEE
mehdi.vasef@ieee.org

Abstract- This paper introduces a new intractable security problem whose intractability is due to the NP completeness of multivariate quadratic problem. This novel problem uses quaternion algebra in conjunction with MQ. Starting with the simultaneous multivariate equations, we transform these equations into simultaneous quaternion based multivariate quadratic equations. A new scheme for computational zero knowledge proof based on this problem is proposed. It is proved that according to black box definition of zero knowledge proof (ZKP) system, the proposed scheme is ZKP. Our proof has two lemmas. The proof is done through two lemmas. In the first lemma it is shown that expected polynomial time machine $M_{V^*}$ halts in a polynomial time. In the second lemma, it is showed that the probability ensembles $\{M_{V^*}(x)\}_{\varepsilon \in \mathbb{E}}$ and $\{P(x), V^*(x)\}_{\varepsilon \in \mathbb{E}}$ are polynomially indistinguishable. The scheme has low computational overhead and is particularly useful in cryptographic applications such as digital signature and key agreement.

Keywords: Multivariate Quadratic, Quaternion, Zero Knowledge Proof, Authentication

1 Introduction
The zero knowledge proof (ZKP) was first introduced by Goldwasser, Micali and Rackoff [1] in 1985. In ZKP, one entity attempts to prove to another that a statement is correct, without disclosing anything other than the veracity of the statement. ZKP is an interactive method and involves two entities: Prover (P) and Verifier (V). They are probabilistic Turing machines. They interact and eventually the Verifier replies Accept or Reject as a result. ZKP satisfies the following properties:

1- Completeness: The legitimate Prover always gets acceptance.
2- Soundness: The legitimate Prover will be rejected with some probability known as "soundness error".
3- Zero-Knowledge: No cheating Verifier can learn anything other than the statement. To describe this property, suppose that every cheating Verifier has a simulator that, given only the statement to be proven, can produce an output that "looks like" an interaction between the honest Prover and cheating Verifier [3].
The ZKP can be defined in a formal mathematical way known as *black box definition* for ZKP systems [2,3].

**Definition 1:** Let \( \langle P, V \rangle \) be an interactive proof system for a language L. We say the proof system \( \langle P, V \rangle \) is zero knowledge for L, if for every expected polynomial time interactive Turing machine \( V' \), there exists an ordinary expected polynomial time machine \( M_v \) such that the probability ensembles \( \{M_{V'}(x)\}_{x \in \text{L}} \) and \( \{P(x), V'(x)\}_{x \in \text{L}} \) are polynomially indistinguishable.

This paper introduces a novel intractable problem based on the hardness of the multivariate quadratic problem in conjunction with quaternion algebra. Accordingly, it proposes a new scheme based on this problem. The scheme does not have the weaknesses of existing comparable schemes.

The remainder of paper is organized as follows. Section 2 briefs some mathematical preliminaries such as quaternion and multivariate quadratic problem. Section 3 presents the novel intractable problem, as well as the new zero knowledge proof scheme, and then compares the scheme with comparable existing ZKP schemes qualitatively. Section 4 concludes the paper.

### 2. Related Works

Fiat and Shamir [4] presented a simple identification and signature scheme that enables any user to prove his identity and the authenticity of his messages. The hardness of this task is based on RSA problem.

Micali and Shamir [5] presented an improvement to their previous scheme that reduces the verifier's complexity to less than 2 modular multiplications and leaves the prover's complexity unchanged. Although it is computationally fast, it is still based on RSA problem.

Fiege et al [6] introduced the notion of interactive proofs of assertions to interactive proofs of knowledge. This leads to the definition of unrestricted input zero-knowledge proofs of knowledge in which the Prover demonstrates possession of knowledge without revealing any computational information. Their identification scheme is provably secure if factoring is difficult and practical implementations are about 2 orders of magnitude faster than RSA-based identification schemes.

Ong-Schnorr identification and signatures [7] are variants of the Fiat-Shamir scheme with short and fast communication and signatures. This scheme uses secret keys that are square roots modulo N of the public keys, whereas Fiat-Shamir uses square roots modulo N. Its security is based on the intractability of certain discrete logarithm problems. It is also proven to be secure against passive and concurrent attacks under DLP assumption.

Guillou and Quisquater (GQ) identification scheme [8] is an extension to Fiat-Shamir scheme, which reduces the number of exchanged messages and memory requirements for secret keys. Security of GQ is based on intractability of RSA problem.
Goldwasser and Kalai [9] showed that the signature based on Fiat-Shamir (and also Fiege-Fiat-Shamir) is forgeable. Wolf [21] shows how zero know proofs can be used to solve authentication problems. Furthermore he demonstrates how the Isomorphism of Polynomials and Multivariate Quadratic equations can be combined to obtain a new and practical Zero-Knowledge scheme. Courtois [20] proposes a new Zero-knowledge scheme based on an NP-complete problem known as MinRank. It can be used to prove in Zero-knowledge a solution to any problem described by multivariate equations.

2. Mathematical Preliminaries

2.1 Quaternion

Quaternion is the extension of complex numbers in 4 dimension which was introduced by Irish mathematician, William Hamilton in 1843 [10,11,12]. A quaternion number can be represented as 
\[ q = x_1i + x_2j + x_3j + x_4k \]

or
\[ q = (x_1, x_2, x_3, x_4) \]

in which \( i^2 = j^2 = k = -1 \), \( ij = k; jk = i; ki = j \) and \( ji = -k; kj = -i; ik = j \). The conjugate of the quaternion number is shown as 
\[ q^\ast = (x_1, -x_2, -x_3, -x_4) \]

while the norm of \( q \) is represented as
\[ |q| = (N_q)^{\frac{1}{2}} = x_1^2 + x_2^2 + x_3^2 + x_4^2. \]

The addition and subtraction of two quaternion numbers is similar to those of complex numbers. The multiplication of two quaternion numbers is shown as
\[ q_3 = (z_1, z_2, z_3, z_4) \]

in which:
\[ z_1 = x_1y_1 - x_2y_2 - x_3y_3 - x_4y_4 \]
\[ z_2 = x_1y_2 + x_2y_1 + x_3y_4 - x_4y_3 \]
\[ z_3 = x_1y_3 - x_2y_4 + x_3y_1 + x_4y_2 \]
\[ z_4 = x_1y_4 + x_2y_3 - x_3y_2 + x_4y_1 \]

The transformation of a vector \( v \) under a quaternion is defined as:
\[ [0, v'] = q.[0, v].q^\ast \]

This transformation can be in matrix form \( v' = Mv \), that is:
\[ M = \begin{pmatrix}
1 - 2y^2 - 2z^2 & 2xy - 2wz & 2zx + 2wy \\
2xy + 2wz & 1 - 2x^2 - 2z^2 & 2yz - 2wx \\
2zx - 2wy & 2yz + 2wx & 1 - 2x^2 - 2y^2
\end{pmatrix} \]

2.2 Multivariate Quadratic Problem

In this section we review a multivariate quadratic problem. Let \( p_i \) and \( p_m \) denote multivariate polynomials and \( y_1, \ldots, y_m \in F \) represent field elements. The Simultaneous Multivariate quadratic Equation (SME) is defined as follows:
\[ y_1 = p_1(x_1, \ldots, x_n) \]
\[ y_2 = p_2(x_1, \ldots, x_n) \]
\[
\vdots
\]
\[ y_m = p_m(x_1, \ldots, x_n) \]

The polynomial \( p_i \) is denoted by:

\[
p_i(x_1, \ldots, x_n) = \sum_{1 \leq j \leq n} \gamma_{i,j,k} x_j x_k + \sum_{j=1}^{n} \beta_{i,j} x_j + \alpha_i
\]

Generally the SME can have \( n \) variables and \( m \) equations. Solution of SME is NP-complete and is hard on average. Based on this MQ, some public key systems have been proposed. Some of them are HEF [13], MIA [14], and UOV [15]. Patarin [16] attacked MIA, Courtois [17] broke HEF. UOV was unsecured due to the successful attack by Braeken et al [18]. We are only concerned about MQ problem as a basis for our interactive problem.

### 3. Our Proposition

#### 3.1 A New Intractable Problem

We would like to combine the quaternion algebra with MQ problem to construct a new intractable problem. The public key crypto (PKC) systems based on MQ problem have shorter key and signature lengths. The encryption and decryption time in these systems is shorter in comparison to other PKCs. Furthermore, in some cryptographic applications that require multiple keys, instead of a single key as a private key and public key, the conventional PKC systems have low computational overheads. Motivated by this, we would like to use MQ problem for such applications. We use quaternion algebra to achieve this goal. The quaternion based multivariate quadratic problem is formulated as in the following.

Starting with the SME, we reorganize these equations such that:

\[
\sum_{i=1}^{n} q_i \begin{bmatrix} 0 & v_{i,1} \end{bmatrix} q_i + \sum_{i=1}^{n} p_{i,1} q_i + q_{c,1} = q_m (q_1, q_2, \ldots, q_n)
\]

\[
\sum_{i=1}^{n} q_i \begin{bmatrix} 0 & v_{i,2} \end{bmatrix} q_i + \sum_{i=1}^{n} p_{i,2} q_i + q_{c,2} = q_m (q_1, q_2, \ldots, q_n)
\]

\[
\sum_{i=1}^{n} q_i \begin{bmatrix} 0 & v_{i,m} \end{bmatrix} q_i + \sum_{i=1}^{n} p_{i,m} q_i + q_{c,m} = q_m (q_1, q_2, \ldots, q_n)
\]

We call this set of equations Simultaneous Quaternion Multivariate Quadratic (SQMQ). Note that there is an isomorphism between SQMQ and SME. The number of variables in quaternion form is \( n \) and the number of equations is \( m \). Instead of
finding $x_i$, we would like to find $q_i = (x_{1,i}, x_{2,i}, x_{3,i}, x_{4,i})$. Note that both $x_i$ and $q_i$ have several bits. The elements of the quaternion can be regarded as a key in a cryptographic system with multiple keys. The number of variables is much greater than the number of equations.

In the next section we will propose our new ZKP scheme based on this problem.

3.2 The Proposed Scheme for Zero Knowledge Proof
The scheme for ZKP is described in this section. Any NP complete intractable security problem can be used to accomplish ZKP [19]. We use our intractable problem to propose a ZKP scheme.

Prover and Verifier play several rounds of a game that is explained step by step below. Each time they use random coin tosses.

1. At the beginning of each round, Prover generates two sets of simultaneous SQMQ and sends them to the verifier. Two sets of equations are illustrated below:

Set 1:

$$\sum_{i=1}^{n} q_i[0] v_{i,1} \sim q_i + \sum_{i=1}^{n} b_{i,1}q_i + q_{c,1} = q_{m1}(q_1, q_2, \ldots, q_n)$$

$$\sum_{i=1}^{n} q_i[0] v_{i,j} \sim q_i + \sum_{i=1}^{n} b_{i,j}q_i + q_{c,j} = q_{mj}(q_1, q_2, \ldots, q_n)$$

$$\sum_{i=1}^{n} q_i[0] v_{i,m} \sim q_i + \sum_{i=1}^{n} b_{i,m}q_i + q_{c,m} = q_{nm}(q_1, q_2, \ldots, q_n)$$

Set 2:

$$\sum_{i=1}^{n} q_i[0] v_{i,1} \sim q_i + \sum_{i=1}^{n} b_{i,1}q_i + q_{c,1} = q_{m1}(q_1, q_2, \ldots, q_n)$$

$$\sum_{i=1}^{n} q_i[0] v_{i,j} \sim q_i + \sum_{i=1}^{n} b_{i,j}q_i + q_{c,j} = q_{mj}(q_1, q_2, \ldots, q_n)$$

$$\sum_{i=1}^{n} q_i[0] v_{i,m} \sim q_i + \sum_{i=1}^{n} b_{i,m}q_i + q_{c,m} = q_{nm}(q_1, q_2, \ldots, q_n)$$

2. Verifier chooses randomly $\alpha \in \{0,1\}$, and sends it to Prover; intuitively, Verifier asks Prover to find the hash function of the solution of these two sets of SQMQ equations, that is to find $H(q_1), \ldots, H(q_i), \ldots, H(q_n)$ or $H(q_1), \ldots, H(q_i), \ldots, H(q_n)$.

3. Verifier randomly chooses one of two questions to ask Prover. He can then ask either to find a solution to first set of SQMQ equations or the second one.

4. In this step, Prover is asked to send the solution for these sets of equations. Prover does not send the exact solution of these two sets of equations.
Instead, he uses one way hash function of their solutions. He uses $H(q_1)$ instead of $q_1$. The hash function $H(q_1)$ is the function that given the quaternion $q_1$, generates quaternion $q_1$. For Set 1 we have:

$$\sum_{i=1}^{n} H(q_i) [0 \ v_{i,1}] \sim H(q_1) + \sum_{i=1}^{n} b_{i,1} H(q_i) + q_{c,1} = T_{m1}$$

$$\sum_{i=1}^{n} H(q_i) [0 \ v_{i,j}] \sim H(q_1) + \sum_{i=1}^{n} b_{i,j} H(q_i) + q_{c,j} = T_{mj}$$

$$\sum_{i=1}^{n} H(q_i) [0 \ v_{i,m}] \sim H(q_1) + \sum_{i=1}^{n} b_{i,m} H(q_i) + q_{c,m} = T_{mm}$$

If Prover is asked to find the solution of first set of equations, he sends the hash function of the solution of these equations and the $T_1, \ldots, T_m$. Verifier accepts if by inserting the hash function of the solutions in the SQMQ, the answers are $T_1, \ldots, T_m$.

5. Prover is asked to send the solution for the second set of equations. For Set 2 we have:

$$\sum_{i=1}^{n} H(q_i) [0 \ v_{i,1}] \sim H(q_1) + \sum_{i=1}^{n} b_{i,1} H(q_i) + q_{c,1} = T'_{m1}$$

$$\sum_{i=1}^{n} H(q_i) [0 \ v_{i,j}] \sim H(q_1) + \sum_{i=1}^{n} b_{i,j} H(q_i) + q_{c,j} = T'_{mj}$$

$$\sum_{i=1}^{n} H(q_i) [0 \ v_{i,m}] \sim H(q_1) + \sum_{i=1}^{n} b_{i,m} H(q_i) + q_{c,m} = T'_{mm}$$

Prover sends $H(q_1)$, $T'_{m1}, \ldots, T'_{mm}$. He sends the solution to Verifier. Verifier checks if by inserting the hash function of the solutions in the SQMQ, the answers are $T_1, \ldots, T_m$.

During each round, Prover does not know which question he will be asked until after giving him the solution to the equation. So in order to answer both questions, he should know how to solve the QMQ and Verifier becomes convinced after some number of rounds that Prover knows this information.

However, the Prover's answer does not reveal the solution of the QMQ equations. Each round, Verifier will gain only the hash function of the solution of these two sets of SQMQ equations and cannot obtain the exact solution of these equations. Therefore, the chance of fooling Verifier is $1/2^n$.

If Verifier successfully completes $m$ iterations of the above steps, he accepts.

**Theorem 1**: The proposed scheme constitutes a zero knowledge interactive proof for quaternion multivariate quadratic according to Definition 1 (given in Section 1).

**Proof**: According to Definition 1, we must show that:

1. The expected interactive Turing machine stops polynomially
2. The two probability distribution of $\{M(v')_x\}_{x \in L}$ and $\{P(x).V'(x)\}_{x \in L}$ are indistinguishable
Let $V$ be an arbitrary, fixed, expected polynomial time interactive machine. We consider that the expected polynomial time machine $M_{\nu}$. generates a probability distribution is identical to the probability distribution induced by $V$ tapes, while it interacts with Prover. The interactive machine $V$ is used in order to construct the machine $M_{\nu}$. Intuitively speaking, $M_{\nu}$ attempts to guess which question he will be asked (computing the hash function of the solution of the first set of SQMQ equations or computing the second set of SQMQ equations). $M_{\nu}$ is constructed in a lucky way, that is the cases which $M_{\nu}$ fails is discarded and nothing appears in its output. It is evident that the cases which lead to $M_{\nu}$ success and the cases that lead to its failure are identical. The probability spaces of these two cases are similar and both of them are uniform.

We would like to explain more about the simulation machine $M_{\nu'}$. When the interactive machine is invoked for several times, machine $M_{\nu'}$ will place the common input on the input tape of $V$ and a fixed sequence of randomly chosen bits on its random tape. Hence, the contents of the random tape are not changed for the rest of the simulation. We encounter two cases. In the first case, all the rounds of the $V$ on input $x$ take at most $n$ steps, which is a fixed polynomial. On the contrary, in the second case, $V$ is expected to be polynomial and there may be no bound on the number of coin tosses that may be used on a specific input $x$. In simple words, the number of iterations ($m$) in the proposed scheme can be finite or infinite. $V$ requires some random string of bits (denoted by $r$) when the simulation starts and never is changed during the simulation.

The contents of the communication-record tape are constructed in the following $m$ rounds:

**Round i:**
1. $M_{\nu'}$ chooses randomly $\beta \in \{0,1\}$ and sets its output $o$ with $H(q_1), H(q_i), H(q_n)$. These are the solution that Prover sends in round $i$. Note that $o$ is the message that $V$ sends an input $x$. This input is in fact the two sets of SQMQ and internal coin tosses $r$ after receiving the solution of the equation.

2. There are two cases:
   - **Case 1 (lucky case):**
     Machine $M_{\nu'}$ appends $H(q_1),...,H(q_i),...,H(q_n)$ to a communication tape; machine $M_{\nu'}$ asks $V$ to send the hash function of solution and validates the answer.
   - **Case 2 (unlucky case):**
     Machine $M_{\nu'}$ repeats the current round. It is noteworthy that nothing is appended to the tape, $i$ is not increased and the steps 1 and 2 are repeated.
If all the rounds are successfully completed, \( M_{V'} \) stops outputting \((o, r_i)\), where \( r_i \) is the prefix of string \( r \) scanned by \( V^* \) on input \( x \), and \( r \) is internal coin tosses.

We should prove the validity of above communication-record tape construction. First in Lemma 1, we show that \( M_{V'} \) halts in expected polynomial time. Second, in Lemma 2, we prove that the output distribution generated by \( M_{V'} \) equals the distribution over \( V^* \) tape when interacting with \( P \).

**Lemma 1**: Machine \( M_{V'} \) halts in expected polynomial time.

**Proof**: Let \( o \) show the output of machine \( M_{V'} \), and for each iteration of the \( i \)-th round, the output of this machine is either null or \( o \). Therefore, we have:

\[
\Pr(V^*(x, r) = o) = 1/2
\]

implying that the expected number of time that \( M_{V'} \) repeats each round is exactly 2.

**Lemma 2**: The probability distribution \( \{M_{V'}(x)\}_{x \in L} \) and \( \{P(x), V^*(x)\}_{x \in L} \) are identical.

**Proof**: Both of these distributions depend on \( r \), the random sequence of coin tosses for Verifier, which is uniform. We need only to show that for every value of \( r \), the conditional distributions are equal. For any fixed infinite sequence \( r \), and \( 0 \leq i \leq m \), let \( Pr_{M_{V'}}^{(x,r,i)} \) represent the probability distribution defined by the first \( i \) rounds of message exchange between Prover and interactive machine \( V^* \). Similarly, \( Pr_{M}^{(x,r,i)} \) denotes the probability distribution defined by first \( i \) rounds of output by \( M_{V'} \) on input \( x \) and regarding \( r \) as the Verifier's source of internal coin tosses. We prove the equality of \( Pr_{M_{V'}}^{(x,r,i)} = Pr_{M}^{(x,r,i)} \) by induction on \( i \) as follows:

- **Induction Base** \((i = 0)\) holds true.

- **Induction Step** \((i \rightarrow i + 1)\)
  
  By induction hypothesis:
  
  \[
  Pr_{M_{V'}}^{(x,r,i)} = Pr_{M}^{(x,r,i)}
  \]

  And set a condition on the event:

  \[
  Pr_{M}^{(x,r,i)} = \delta
  \]

  We consider the \((i + 1)\) th step in \( Pr_{M_{V'}}^{(x,r,i)} \) (and respectively \( Pr_{M}^{(x,r,i)} \)).

As noted before, both of these distributions are uniform and rely on \( r \). The \( r \) in \( M_{V'} \) is the internal coin tosses, and in the \( V^* \) is the Verifier's source of internal coin tosses. So in the \( i \)-th round, the two probability distributions become equal. Since the two lemmas have been proved, therefore, the theorem is true and the proof is done.

### 3.3 Comparison

Table 1 shows a qualitative comparison of our proposed ZKP scheme with some of existing ZKP schemes. Three factors are used for comparison: 1) hardness of the scheme, 2) security of the scheme, and 3) computational speed.
The ZKP schemes in the literature can be categorized into three groups. In the first group, the intractable problem is integer factorization. They are insecure and forgeable and have low computation speed in execution. Unlikely, in the second group, Schnorr scheme for example, the hardness of the schemes is based on discrete logarithm problem. They are fast in comparison to the first group. Up to now these schemes have remained secure. In the third group, the intractable problem is multivariate quadratic problem. They are computationally the fastest.

The hardness of our proposed scheme is based on multivariate quadratic problem. Since any zero knowledge proof system is provably secure, our scheme is also secure. Since the proposed scheme does not use exponentiation, it is faster than previous schemes. The advantage of our scheme to third group is that its secrecy is not easier than third group.

Definition:

Table 1: Comparison of our scheme with some existing schemes

<table>
<thead>
<tr>
<th>The ZKP schemes</th>
<th>Hardness</th>
<th>Security</th>
<th>Speed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fiat -Shamir</td>
<td>Integer factorization</td>
<td>No</td>
<td>Lower</td>
</tr>
<tr>
<td>Guillou-Quisquater</td>
<td>Integer factorization</td>
<td>No</td>
<td>Low</td>
</tr>
<tr>
<td>Fiege-Fiat-Shamir</td>
<td>Integer factorization</td>
<td>No</td>
<td>Low</td>
</tr>
<tr>
<td>Schnorr</td>
<td>Discrete logarithm problem</td>
<td>Yes</td>
<td>Fast</td>
</tr>
<tr>
<td>Our Scheme</td>
<td>Based on QMQ</td>
<td>Yes</td>
<td>Faster</td>
</tr>
</tbody>
</table>

4. Conclusion

We introduced a novel intractable problem bases on the hardness of the NP-completeness of multivariate quadratic. This problem uses quaternion algebra. Based on this problem, we also proposed a scheme for zero knowledge proof. We proved that the proposed scheme has the two properties mentioned in black box definition for ZKP. The use of quaternion adds some advantages to conventional MQ problems. One advantage is that one can have multiple keys. The other benefit is that the key lengths can be short just like public key systems based on MQ. The proposed ZKP scheme can be used as a method for identification and is faster than some existing ZKP schemes.

References


