Public-Key Identification Schemes Based on Multivariate Quadratic Polynomials

Koichi Sakumoto, Taizo Shirai, and Harunaga Hiwatari

Sony Corporation
5-1-12 Kitashinagawa Shinagawa-ku, Tokyo 141-0001, Japan
{Koichi.Sakumoto,Taizo.Shirai,Harunaga.Hiwatari}@jp.sony.com

Abstract. A problem of solving a system of multivariate quadratic polynomials over a finite field, which is called an MQ problem, is a promising problem in cryptography. A number of studies have been conducted on designing public-key schemes using the MQ problem, which are known as multivariate public-key cryptography (MPKC). However, the security of the existing schemes in MPKC relies *not* only on the MQ problem but *also* on an Isomorphism of Polynomials (IP) problem. In this paper, we propose public-key identification schemes based on the conjectured intractability of the MQ problem under the assumption of the existence of a non-interactive commitment scheme. Our schemes do *not* rely on the IP problem, and they consist of an identification protocol which is zero-knowledge argument of knowledge for the MQ problem. For a practical parameter choice, the efficiency of our schemes is highly comparable to that of identification schemes based on another problem including Permuted Kernels, Syndrome Decoding, Constrained Linear Equations, and Permuted Perceptrons. Furthermore, even if the protocol is repeated in parallel, our scheme can achieve the security under active attack with some additional cost.

Keywords: identification scheme, zero knowledge, MQ problem.

1 Introduction

A problem of solving a system of multivariate quadratic polynomials over a finite field, which is called an MQ problem, is a promising problem in cryptography. The associated decision problem is known to be NP-complete [24,40], and a random instance of the MQ problem is widely believed to be intractable. In contrast to factorization or a discrete logarithm problem, there is no known polynomial-time quantum algorithm to solve the MQ problem. A function consisting of multivariate quadratic polynomials, which we call an MQ function, can be used as a one-way function with short input and output. Complexity of generic attacks using Gröbner basis is known to be exponential in time and space [3,16], and the best known attack to break the MQ function over $\mathbb{F}_2$ with 84-bit input and 80-bit output requires $2^{88.7}(> 2^{80})$ bit operations [10].

P. Rogaway (Ed.): CRYPTO 2011, LNCS 6841, pp. 706–723, 2011.

A number of studies on designing primitives based on the MQ function have been conducted both in symmetric and in asymmetric cryptography. In symmetric cryptography, a stream cipher which is named QUAD is proposed by Berbain et al. [7]. The security of QUAD is provably reducible to the conjectured intractability of the MQ problem. In asymmetric cryptography, several public-key schemes have been proposed, which are known as multivariate public-key cryptography (MPKC) [30,35,39]. However, the security of the existing schemes in MPKC relies not only on the MQ problem but also on an Isomorphism of Polynomials (IP) problem. The IP problem consists of recovering a particular transformation between two sets of multivariate polynomials, and some cryptanalyses of the problem have been reported [11,17,22,41]. In fact, some schemes in MPKC have been already shown to be insecure [11,14,31,38].

In this paper, we propose public-key identification schemes based on the conjectured intractability of the MQ problem under the assumption of the existence of a non-interactive commitment scheme which is statistically-hiding and computationally-binding. We emphasize that our schemes do not rely on the IP problem. The assumption for the commitment scheme is natural, since it can be constructed from a collision resistant hash function [27]. Our identification protocols are non-trivial constructions of statistical zero-knowledge argument of knowledge for the MQ problem. Assuming the intractability of the MQ function, our identification schemes consisting of the *sequential* composition and the *parallel* composition of the protocols are secure against impersonation under *active* attack and *passive* attack, respectively. These security levels are the same as those of known identification schemes based on another problem including Permuted Kernels (PK) [46], binary Syndrome Decoding (SD) [47,49], Constrained Linear Equations (CLE) [48], Permuted Perceptrons (PP) [42,43], and q-ary SD [12].

For a practical parameter choice, the sizes of a public key, a secret key, and communication data of our schemes are comparable to those of the schemes based on PK, SD, CLE, PP, and q-ary SD. In particular, the sizes of a public key and a secret key of our 3-pass scheme are only 80 bits and 84 bits for 80-bit security, respectively. These are smaller than those of the known schemes [12,42,43,46,47,48,49]. This is due to the fact that the MQ function has short input and output. The size of communication data in our 3-pass protocol is 29,640 bits when the impersonation probability is less than 2^{-30}. This is also small compared to those of the existing 3-pass protocols [42,43,47,48,49], which are between 45,517 bits and 100,925 bits. Although the data size of system parameter of our scheme is relatively large, it can be reduced to some small seed, e.g. 128 bits, by employing a pseudo-random number generator. The technique is also used in the implementation of QUAD [2].

Furthermore, we consider the case that our scheme employs the MQ function which is substantially compressing (e.g., mapping 160 bits to 80 bits), although the sizes of the secret key and the communication data increase compared to those of the practical parameter choice. In this case, when such a function is preimage resistant, our scheme is secure under active attack even if the protocol is repeated in parallel. The proof of the security is non-trivial, since zero knowledge

is not preserved under the parallel composition. Thus we prove the security by also showing that the MQ function is *second-preimage* resistant if such a function is *preimage* resistant, although the MQ function is known not to have the collision resistance [9].

Techniques for Our Constructions. Our protocols employ the cut-and-choose approach, where a prover first divides her secret into shares and then proves the correctness of some shares depending on the choice of a verifier without revealing the secret itself. The property of group homomorphism such as a modular exponentiation $x \mapsto g^x \mod p$ and a linear function $x \mapsto Mx$ is useful for this approach, since dividing a secret $s = r_0 + r_1$ simply corresponds to dividing its image $g^s = (g^{r_0})(g^{r_1})$ and $Ms = Mr_0 + Mr_1$, respectively. However, the MQ function $(x_1, \ldots, x_n) \mapsto (y_1, \ldots, y_m)$ where $y_l = \sum_{i,j} a_{l,i,j} x_i x_j + \sum_i b_{l,i} x_i$ does not seem to have such a property.

Therefore, we introduce new dividing techniques using the bilinearity of a *polar form* of the MQ function. The polar form $\mathbf{G}$ of the MQ function $\mathbf{F}$ is a function $\mathbf{G}(\mathbf{x}_1, \mathbf{x}_2) = \mathbf{F}(\mathbf{x}_1 + \mathbf{x}_2) - \mathbf{F}(\mathbf{x}_1) - \mathbf{F}(\mathbf{x}_2)$ and is known to be bilinear. It was introduced as the differential of the quadratic system, and has been used for cryptanalysis of MPKC so far [14,15,21,22]. To our knowledge, this is the first time that it is constructively used in a context of a public-key identification scheme.

Our dividing techniques are briefly described as follows. Let $\mathbf{s}$ and $\mathbf{v} = \mathbf{F}(\mathbf{s})$ be a secret key and a public key, respectively. When the secret key is divided as $\mathbf{s} = \mathbf{r}_0 + \mathbf{r}_1$, the public key $\mathbf{v} = \mathbf{F}(\mathbf{r}_0 + \mathbf{r}_1)$ can be represented as $\mathbf{v} = \mathbf{F}(\mathbf{r}_0) + \mathbf{F}(\mathbf{r}_1) + \mathbf{G}(\mathbf{r}_0, \mathbf{r}_1)$ by using the polar form $\mathbf{G}$ of $\mathbf{F}$. However, this representation still contains the term $\mathbf{G}(\mathbf{r}_0, \mathbf{r}_1)$ which depends on both $\mathbf{r}_0$ and $\mathbf{r}_1$. Consider that $\mathbf{r}_0$ and $\mathbf{F}(\mathbf{r}_0)$ are further divided as $\mathbf{r}_0 = \mathbf{t}_0 + \mathbf{t}_1$ and $\mathbf{F}(\mathbf{r}_0) = \mathbf{e}_0 + \mathbf{e}_1$, respectively. In this case, the public key can be divided into two parts $\mathbf{v} = \big(\mathbf{G}(\mathbf{t}_0, \mathbf{r}_1) + \mathbf{e}_0\big) + \big(\mathbf{F}(\mathbf{r}_1) + \mathbf{G}(\mathbf{t}_1, \mathbf{r}_1) + \mathbf{e}_1\big)$, due to the bilinearity of $\mathbf{G}$. Each of the two parts is represented by either a tuple $(\mathbf{r}_1, \mathbf{t}_0, \mathbf{e}_0)$ or a tuple $(\mathbf{r}_1, \mathbf{t}_1, \mathbf{e}_1)$, while no information on the secret key $\mathbf{s}$ can be obtained from one out of the two tuples.

Related Work. Identification schemes based on PK [46], SD [47,49], CLE [48], PP [42,43], and q-ary SD [12] have some features similar to our schemes as follows. First, these schemes rely on the hardness of a random instance of each of the problems whose associated decision version is known to be NP-complete. Second, their protocols have perfect correctness. Finally, assuming the existence of a non-interactive commitment scheme, the sequential version and the parallel version of the schemes are secure against impersonation under active attack and passive attack, respectively. However, it is not explicitly known that the parallel versions of these schemes achieve the security under active attack.

On the other hand, lattice-based schemes [29,33,34,36] have other features. They are based on an average-case problem which is as hard as worst-case problems, and some of them [29,33,34] are secure under active attack even if repeated in parallel. Lyubashevsky's scheme [34] is stated to be more practically efficient

than the others [29,33,36]. The size of communication data of the scheme [34] for 80-bit security against impersonation is only about 65,000 bits, although the scheme has small correctness error 2^{-20}. Both of the sizes of the public key and the secret key are 16,000 bits.

In a context of post-quantum cryptography, Komano et al. proposed a signature scheme based on a section finding problem on algebraic surface [32]. Their construction, similarly to our schemes, does not rely on a property of homomorphism. However, their scheme is universally forgeable under key-only attack, and their technique turned out to be unsuccessful to realize a signature scheme [45].

Paper Organization. The remainder of this paper is organized as follows. In Section 2 we present several notions and tools that are used in our constructions. In Section 3 and Section 4, our 3-pass and 5-pass constructions are presented, respectively. In Section 5 we discuss their security and efficiency for a practical parameter choice. In Section 6 we study the security of the parallel composition of our scheme at the expense of the efficiency. In Section 7 we mention some extensions of our scheme.

2 Preliminaries

A finite field of order q is denoted by $\mathbb{F}_q$. If an element x is randomly chosen from a finite set S, it is expressed by $x \in_R S$. If A and B are sets, and $R \subset A \times B$ is a binary relation, then we define $R(x) = \{s : (x, s) \in R\}$. If $s \in R(x)$, then s is called a solution for x.

Identification Scheme. An identification scheme is a tuple of algorithms ($\mathtt{Setup}$, $\mathtt{Gen}, \mathtt{P}, \mathtt{V}$) defined as follows. $\mathtt{Setup}$ is a setup algorithm which takes a security parameter 1^λ and outputs a system parameter *param*. $\mathtt{Gen}$ is a key-generation algorithm which takes *param*, and outputs a public key and a secret key (pk, sk). A pair of a prover $\mathtt{P}$ and a verifier $\mathtt{V}$ is an interactive protocol where a common input is $(param, pk)$ and an auxiliary input of $\mathtt{P}$ is sk. After interactions, $\mathtt{V}$ outputs a bit as a verification result. The protocol $(\mathtt{P}, \mathtt{V})$ is called an identification protocol.

Security against impersonation under *passive/active* attacks considers an adversary whose goal is to impersonate the prover without the knowledge of the secret key. The adversary under *passive* attack has access to interactions between the real prover and an honest verifier. The adversary under *active* attack can interact with the prover. Requiring security against impersonation under active attack is stronger than under passive attack. The details are described in [1,18].

The definitions of zero knowledge, witness indistinguishability, and argument of knowledge are omitted. For formal definitions, refer to textbooks, e.g., [25].

String Commitment Scheme. A string commitment function is denoted by *Com*. The commitment scheme runs in two phases. In the first phase, the sender computes a commitment value $c \leftarrow Com(s; \rho)$ and sends c to the receiver, where s

is a string and ρ is a random string. In the second phase, the sender gives (s, ρ) to the receiver and the receiver verifies $c = Com(s; \rho)$. We require two security properties of *Com*, statistically hiding and computationally binding. Informally, the former means that, at the end of the first phase, no receiver can distinguish two commitment values generated from two distinct strings even if the receiver is computationally unbounded. The latter means that, no polynomial-time sender can change the committed string after the first phase. The formal definitions and a practical construction are given in [27]. Throughout this paper, we assume the existence of such a commitment scheme. The assumption is natural, since it can be constructed from a collision resistant hash function [27]. Note that such an *interactive* commitment scheme can be constructed from any one-way function including the MQ function [26].

The MQ Function. We denote by $\mathcal{MQ}(n, m, \mathbb{F}_q)$ a family of functions

$$\left\{ \mathbf{F}(\mathbf{x}) = (f_1(\mathbf{x}), \ldots, f_m(\mathbf{x})) \middle| \begin{array}{c} f_l(\mathbf{x}) = \sum_{i,j} a_{l,i,j} x_i x_j + \sum_i b_{l,i} x_i, \\ a_{l,i,j}, b_{l,i} \in \mathbb{F}_q \ \text{ for } \ l = 1, \ldots, m \end{array} \right\}$$

where $\mathbf{x} = (x_1, \ldots, x_n)$. For the simplicity, constant terms are omitted without any security loss. We call $\mathbf{F} \in \mathcal{MQ}(n, m, \mathbb{F}_q)$ an MQ function. A function $\mathbf{G}(\mathbf{x}, \mathbf{y}) = \mathbf{F}(\mathbf{x} + \mathbf{y}) - \mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{y})$ is called the polar form of $\mathbf{F}$. The function $\mathbf{G} = (g_1, \ldots, g_m)$ is bilinear, since $g_l(\mathbf{x}, \mathbf{y}) = \sum_{i,j} a_{l,i,j}(y_i x_j + x_i y_j)$ where $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbf{y} = (y_1, \ldots, y_n)$. An intractability assumption for a random instance of $\mathcal{MQ}(n, m, \mathbb{F}_q)$ is defined as follows.

Definition 1. *For polynomially bounded functions $n = n(\lambda)$, $m = m(\lambda)$, and $q = q(\lambda)$, it is said that $\mathcal{MQ}(n, m, \mathbb{F}_q)$ is intractable if there is no polynomial-time algorithm that takes $(\mathbf{F}, \mathbf{v})$ generated via $\mathbf{F} \in_R \mathcal{MQ}(n, m, \mathbb{F}_q)$, $\mathbf{s} \in_R \mathbb{F}_q^n$, and $\mathbf{v} \leftarrow \mathbf{F}(\mathbf{s})$ and finds a preimage $\mathbf{s}' \in \mathbb{F}_q^n$ such that $\mathbf{F}(\mathbf{s}') = \mathbf{v}$ with non-negligible probability $\epsilon(\lambda)$.*

All the state-of-the-art solving techniques have exponential complexity to break the intractability [8,10,16]. In particular, it is known that complexity of generic attacks using Gröbner basis is exponential in time and space [3,16]. Bouillaguet et al. stated that it would not outperform exhaustive search in the practically interesting range $m = n \leq 200$ [10]. They proposed an improved exhaustive search algorithm to break $\mathcal{MQ}(n, m, \mathbb{F}_2)$ in $2^{n+2} \cdot \log_2 n$ bit operations, which is the best known algorithm [10].

In addition, for $\mathbf{F} \in \mathcal{MQ}(n, m, \mathbb{F}_q)$, we define a binary relation $R_{\mathbf{F}} = \{(\mathbf{v}, \mathbf{x}) \in \mathbb{F}_q^m \times \mathbb{F}_q^n : \mathbf{v} = \mathbf{F}(\mathbf{x})\}$. Given an instance $\mathbf{F} \in \mathcal{MQ}(n, m, \mathbb{F}_q)$ and a vector $\mathbf{v} \in \mathbb{F}_q^m$, the MQ problem is finding a solution $\mathbf{s} \in R_{\mathbf{F}}(\mathbf{v})$.

3 A 3-pass Identification Scheme

In this section, we construct an identification scheme which consists of a 3-pass statistical zero-knowledge argument of knowledge for $R_{\mathbf{F}}$ with knowledge error 2/3, assuming the existence of a non-interactive commitment scheme *Com* which is statistically hiding and computationally binding.

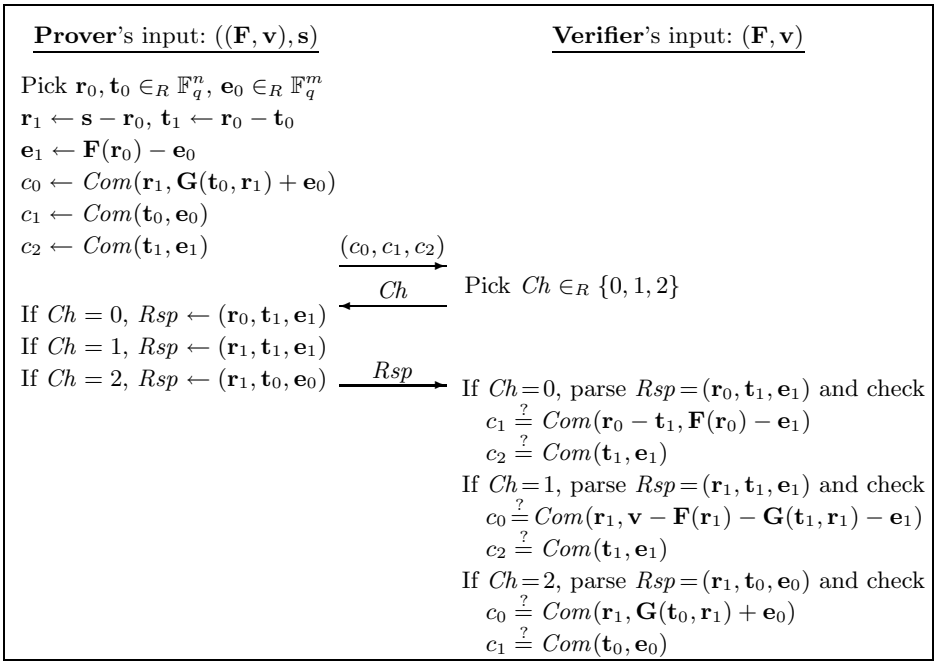

Fig. 1. Our 3-pass identification protocol

Key Generation. We begin with describing a setup algorithm and a key-generation algorithm. Let λ be a security parameter. Let $n = n(\lambda)$, $m = m(\lambda)$, and $q = q(\lambda)$ be polynomially bounded functions. The setup algorithm `Setup` takes 1^λ and outputs a system parameter $\mathbf{F} \in_R \mathcal{MQ}(n, m, \mathbb{F}_q)$ which consists of m-tuple of random multivariate quadratic polynomials. The key-generation algorithm `Gen` takes $\mathbf{F}$. After choosing a random vector $\mathbf{s} \in_R \mathbb{F}_q^n$, `Gen` computes $\mathbf{v} \leftarrow \mathbf{F}(\mathbf{s})$, then outputs $(pk, sk) = (\mathbf{v}, \mathbf{s})$.

An Identification Protocol. The basic idea for our 3-pass construction is that a prover proves that she has a tuple $(\mathbf{r}_0, \mathbf{r}_1, \mathbf{t}_0, \mathbf{t}_1, \mathbf{e}_0, \mathbf{e}_1)$ satisfying

$$\mathbf{G}(\mathbf{t}_0, \mathbf{r}_1) + \mathbf{e}_0 = \mathbf{v} - \mathbf{F}(\mathbf{r}_1) - \mathbf{G}(\mathbf{t}_1, \mathbf{r}_1) - \mathbf{e}_1 \quad (1)$$

$$\text{and} \quad (\mathbf{t}_0, \mathbf{e}_0) = (\mathbf{r}_0 - \mathbf{t}_1, \mathbf{F}(\mathbf{r}_0) - \mathbf{e}_1), \quad (2)$$

since if the tuple satisfies (1) and (2) then $\mathbf{v} = \mathbf{F}(\mathbf{r}_0 + \mathbf{r}_1)$. Note that $\mathbf{G}$ is the polar form of $\mathbf{F}$. In the concrete protocol, corresponding to a challenge $Ch \in \{0, 1, 2\}$ of a verifier, the prover reveals one out of three tuples $(\mathbf{r}_0, \mathbf{t}_1, \mathbf{e}_1)$, $(\mathbf{r}_1, \mathbf{t}_1, \mathbf{e}_1)$, and $(\mathbf{r}_1, \mathbf{t}_0, \mathbf{e}_0)$. The verifier can check each side of each equations (1) and (2) by using either of the three tuples. Such vectors $\mathbf{r}_0, \mathbf{r}_1, \mathbf{t}_0, \mathbf{t}_1, \mathbf{e}_0, \mathbf{e}_1$ are produced by using the dividing techniques described in Section 1. Thus, when $\mathbf{r}_0$, $\mathbf{t}_0$, and $\mathbf{e}_0$ are randomly chosen, the verifier can obtain no information on the secret key $\mathbf{s}$ from only one out of the three tuples.

The 3-pass identification protocol is described in Figure 1. For the simplicity, a random string ρ in *Com* is not written explicitly. The verifier finally outputs 1 if both the checks of "$\stackrel{?}{=}$" are passed, otherwise outputs 0. This is denoted by $0/1 \leftarrow Dec(\mathbf{F}, \mathbf{v}; (c_0, c_1, c_2), Ch, Rsp)$. It is easy to see that the verifier always accepts an interaction with the honest prover. Thus the 3-pass scheme has perfect correctness.

Now we show two properties of the protocol in Theorem 2 and Theorem 3 as follows.

Theorem 2. *The 3-pass protocol is statistically zero knowledge when the commitment scheme Com is statistically hiding.*

Proof sketch. Let $\mathcal{S}$ be a simulator which takes $\mathbf{F}$ and $\mathbf{v}$ without knowing $\mathbf{s}$, and interacts with a cheating verifier $\mathcal{CV}$. We show that the simulator $\mathcal{S}$ can impersonate the honest prover with probability 2/3. The simulator $\mathcal{S}$ randomly chooses a value $Ch^* \in_R \{0, 1, 2\}$ and vectors $\mathbf{s}', \mathbf{r}'_0, \mathbf{t}'_0 \in_R \mathbb{F}_q^n$, $\mathbf{e}'_0 \in_R \mathbb{F}_q^m$, where Ch^* is a prediction of what value the cheating verifier $\mathcal{CV}$ will *not* choose. Then, it computes $\mathbf{r}'_1 \leftarrow \mathbf{s}' - \mathbf{r}'_0$ and $\mathbf{t}'_1 \leftarrow \mathbf{r}'_0 - \mathbf{t}'_0$. If $Ch^* = 0$ then it computes $\mathbf{e}'_1 \leftarrow \mathbf{v} - \mathbf{F}(\mathbf{s}') + \mathbf{F}(\mathbf{r}'_0) - \mathbf{e}'_0$, else $\mathbf{e}'_1 \leftarrow \mathbf{F}(\mathbf{r}'_0) - \mathbf{e}'_0$. If $Ch^* = 2$ then it computes $c'_0 \leftarrow Com(\mathbf{r}'_1, \mathbf{v} - \mathbf{F}(\mathbf{r}'_1) - \mathbf{G}(\mathbf{t}'_1, \mathbf{r}'_1) - \mathbf{e}'_1)$, else $c'_0 \leftarrow Com(\mathbf{r}'_1, \mathbf{G}(\mathbf{t}'_0, \mathbf{r}'_1) + \mathbf{e}'_0)$. It computes $c'_1 \leftarrow Com(\mathbf{t}'_0, \mathbf{e}'_0)$ and $c'_2 \leftarrow Com(\mathbf{t}'_1, \mathbf{e}'_1)$ and sends (c'_0, c'_1, c'_2) to $\mathcal{CV}$. Due to the statistically hiding property of *Com*, a challenge *Ch* from $\mathcal{CV}$ is different from Ch^* with probability 2/3. If $Ch \neq Ch^*$ then $(\mathbf{r}'_0, \mathbf{t}'_1, \mathbf{e}'_1)$, $(\mathbf{r}'_1, \mathbf{t}'_1, \mathbf{e}'_1)$, and $(\mathbf{r}'_1, \mathbf{t}'_0, \mathbf{e}'_0)$ are accepted responses to $Ch = 0$, 1, and 2, respectively. Note that if $Ch^* = 0$ and $Ch = 1$ then it is seen that $\mathbf{v} - \mathbf{F}(\mathbf{r}'_1) - \mathbf{G}(\mathbf{t}'_1, \mathbf{r}'_1) - \mathbf{e}'_1 = \mathbf{G}(\mathbf{t}'_0, \mathbf{r}'_1) + \mathbf{e}'_0$, since $\mathbf{e}'_1 = \mathbf{v} - \mathbf{F}(\mathbf{s}') + \mathbf{F}(\mathbf{r}'_0) - \mathbf{e}'_0$, $\mathbf{F}(\mathbf{s}') = \mathbf{F}(\mathbf{r}'_0) + \mathbf{F}(\mathbf{r}'_1) + \mathbf{G}(\mathbf{r}'_0, \mathbf{r}'_1)$, and $\mathbf{r}'_0 - \mathbf{t}'_1 = \mathbf{t}'_0$.

The details of the proof are given in the full paper, where we formally construct a black-box simulator $\mathcal{S}$ which has oracle access to a cheating verifier $\mathcal{CV}$, and outputs a successful transcript with probability 2/3. Furthermore, the distribution of the output of $\mathcal{S}$ is shown to be statistically close to the distribution of the real transcript. □

Theorem 3. *The 3-pass protocol is argument of knowledge for $R_{\mathbf{F}}$ with knowledge error 2/3 when the commitment scheme Com is computationally binding.*

Proof sketch. Let $((c_0, c_1, c_2), Ch_0, Rsp_0)$, $((c_0, c_1, c_2), Ch_1, Rsp_1)$, and $((c_0, c_1, c_2), Ch_2, Rsp_2)$ be three transcripts such that $Ch_i = i$ and $Dec(\mathbf{F}, \mathbf{v}; (c_0, c_1, c_2), Ch_i, Rsp_i) = 1$ for $i \in \{0, 1, 2\}$. Then, by using the three transcripts, it is shown to be able to either break the binding property of *Com* or extract a solution for $\mathbf{v}$. Consider the situation where the responses are parsed as $Rsp_0 = (\tilde{\mathbf{r}}_0^{(0)}, \tilde{\mathbf{t}}_1^{(0)}, \tilde{\mathbf{e}}_1^{(0)})$, $Rsp_1 = (\tilde{\mathbf{r}}_1^{(1)}, \tilde{\mathbf{t}}_1^{(1)}, \tilde{\mathbf{e}}_1^{(1)})$, and $Rsp_2 = (\tilde{\mathbf{r}}_1^{(2)}, \tilde{\mathbf{t}}_0^{(2)}, \tilde{\mathbf{e}}_0^{(2)})$. Then, it is seen that

$$\begin{aligned} c_0 &= Com(\tilde{\mathbf{r}}_1^{(1)}, \mathbf{v} - \mathbf{F}(\tilde{\mathbf{r}}_1^{(1)}) - \mathbf{G}(\tilde{\mathbf{t}}_1^{(1)}, \tilde{\mathbf{r}}_1^{(1)}) - \tilde{\mathbf{e}}_1^{(1)}) \\ &= Com(\tilde{\mathbf{r}}_1^{(2)}, \mathbf{G}(\tilde{\mathbf{t}}_0^{(2)}, \tilde{\mathbf{r}}_1^{(2)}) + \tilde{\mathbf{e}}_0^{(2)}), \end{aligned} \tag{3}$$

$$c_1 = Com(\tilde{\mathbf{r}}_0^{(0)} - \tilde{\mathbf{t}}_1^{(0)}, \mathbf{F}(\tilde{\mathbf{r}}_0^{(0)}) - \tilde{\mathbf{e}}_1^{(0)}) = Com(\tilde{\mathbf{t}}_0^{(2)}, \tilde{\mathbf{e}}_0^{(2)}), \quad \text{and} \tag{4}$$

$$c_2 = Com(\tilde{\mathbf{t}}_1^{(0)}, \tilde{\mathbf{e}}_1^{(0)}) = Com(\tilde{\mathbf{t}}_1^{(1)}, \tilde{\mathbf{e}}_1^{(1)}). \tag{5}$$

If the two pairs of the arguments of *Com* are distinct on any one of the above equations, the binding property of *Com* is broken. Otherwise, the equation (3) yields $\mathbf{v} = \mathbf{F}(\tilde{\mathbf{r}}_1^{(2)}) + \mathbf{G}(\tilde{\mathbf{t}}_1^{(1)} + \tilde{\mathbf{t}}_0^{(2)}, \tilde{\mathbf{r}}_1^{(2)}) + \tilde{\mathbf{e}}_0^{(2)} + \tilde{\mathbf{e}}_1^{(1)}$. Combining it with the equations (4) and (5), it is seen that $\mathbf{v} = \mathbf{F}(\tilde{\mathbf{r}}_1^{(2)}) + \mathbf{G}(\tilde{\mathbf{r}}_0^{(0)}, \tilde{\mathbf{r}}_1^{(2)}) + \mathbf{F}(\tilde{\mathbf{r}}_0^{(0)}) = \mathbf{F}(\tilde{\mathbf{r}}_0^{(0)} + \tilde{\mathbf{r}}_1^{(2)})$. It means that a solution $\tilde{\mathbf{r}}_0^{(0)} + \tilde{\mathbf{r}}_1^{(2)}$ for $\mathbf{v}$ is extracted.

The details of the proof are given in the full paper, where we formally construct a knowledge extractor which has oracle access to a message specification function $P_{\mathbf{F},\mathbf{v},\mathbf{s},r}$, and either breaks the binding property of *Com* or outputs a solution for $\mathbf{v}$. □

Extension. The trick mentioned in [49] for saving one hash value can be applied to our 3-pass identification protocol as follows. In the first pass, by using a collision resistant hash function H, one hash value $c = H(c_0, c_1, c_2)$ instead of three commitments (c_0, c_1, c_2) is sent. In the third pass, for a challenge Ch of a verifier, a prover sends $c_i|_{i=Ch}$ in addition to Rsp. Consequently, a verifier computes $c_i|_{i\neq Ch}$ by using Rsp and checks $c = H(c_0, c_1, c_2)$. The modified version of 3-pass protocol is also shown to be zero-knowledge argument of knowledge with knowledge error 2/3.

4 A 5-pass Identification Scheme

In this section, we construct a 5-pass identification protocol which is statistical zero-knowledge argument of knowledge for $R_{\mathbf{F}}$ with knowledge error $1/2 + 1/2q$, assuming the existence of a non-interactive commitment scheme *Com* which is statistically hiding and computationally binding. The knowledge error of the 5-pass protocol is smaller than that of the 3-pass protocol when $q \geq 4$. The setup algorithm and the key-generation algorithm of the 5-pass scheme are identical to those of the 3-pass scheme.

In the 5-pass protocol, a prover also divides the secret key $\mathbf{s}$ and the public key $\mathbf{F}(\mathbf{s})$ as $\mathbf{s} = \mathbf{r}_0 + \mathbf{r}_1$ and $\mathbf{F}(\mathbf{s}) = \mathbf{F}(\mathbf{r}_0 + \mathbf{r}_1) = \mathbf{F}(\mathbf{r}_0) + \mathbf{F}(\mathbf{r}_1) + \mathbf{G}(\mathbf{r}_0, \mathbf{r}_1)$, respectively. The difference from the 3-pass protocol is that $\mathbf{r}_0$ and $\mathbf{F}(\mathbf{r}_0)$ are divided as $\alpha\mathbf{r}_0 = \mathbf{t}_0 + \mathbf{t}_1$ and $\alpha\mathbf{F}(\mathbf{r}_0) = \mathbf{e}_0 + \mathbf{e}_1$ where $\alpha \in \mathbb{F}_q$ is a choice of a verifier. After sending $(\mathbf{t}_1, \mathbf{e}_1)$ to the verifier, corresponding to a challenge $Ch \in \{0, 1\}$ of the verifier, the prover reveals one out of two vectors $\mathbf{r}_0$ and $\mathbf{r}_1$. When $\mathbf{r}_0$, $\mathbf{t}_0$, and $\mathbf{e}_0$ are randomly chosen, the verifier can obtain no information on the secret key $\mathbf{s}$ from only one out of the two vectors $\mathbf{r}_0$ and $\mathbf{r}_1$. On the other hand, the argument-of-knowledge property comes from that, for more than one choice of $\alpha \in \mathbb{F}_q$, an impersonator cannot response both of verifier's challenges $Ch = 0$ and $Ch = 1$ unless the impersonator has a solution $\mathbf{s}$ for $\mathbf{v}$.

The 5-pass identification protocol is described in Figure 2 where $\mathbf{G}$ is the polar form of $\mathbf{F}$. For the simplicity, a random string ρ in *Com* is not written

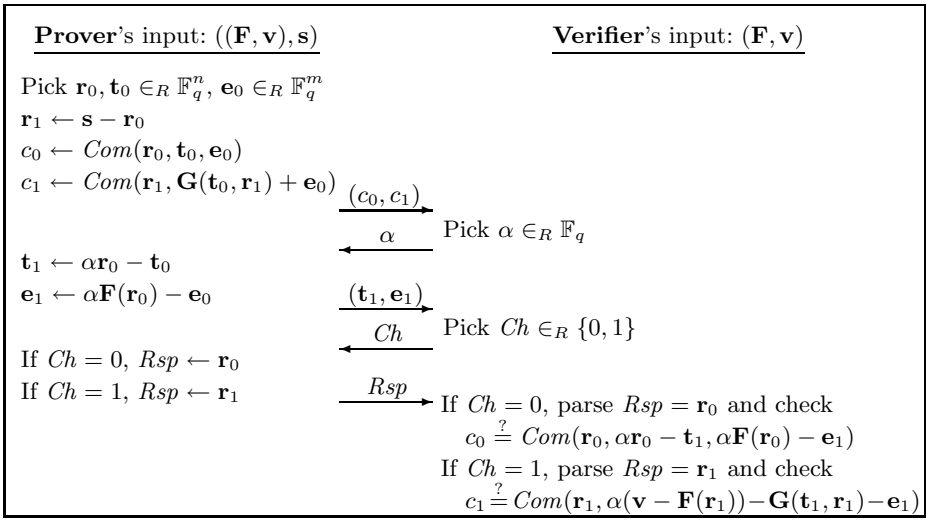

Fig. 2. Our 5-pass identification protocol

explicitly. The verifier finally outputs 1 if the check of "$\stackrel{?}{=}$" is passed, otherwise outputs 0. This is denoted by $0/1 \leftarrow Dec(\mathbf{F},\mathbf{v};(c_0,c_1),\alpha,(\mathbf{t}_1,\mathbf{e}_1),Ch,Rsp)$. It is easy to see that the verifier always accepts an interaction with the honest prover. Thus the 5-pass scheme has perfect correctness.

Now we show two properties of the protocol in Theorem 4 and Theorem 5 as follows.

Theorem 4. *The 5-pass protocol is statistically zero knowledge when the commitment scheme Com is statistically hiding.*

Proof sketch. Let $\mathcal{S}$ be a simulator which takes $\mathbf{F}$ and $\mathbf{v}$ without knowing $\mathbf{s}$, and interacts with a cheating verifier $\mathcal{CV}$. We show that the simulator $\mathcal{S}$ can impersonate the honest prover with probability 1/2. The simulator $\mathcal{S}$ randomly chooses a value $Ch^* \in_R \{0,1\}$ and vectors $\mathbf{s}',\mathbf{r}'_0,\mathbf{t}'_0 \in_R \mathbb{F}_q^n$, $\mathbf{e}'_0 \in_R \mathbb{F}_q^m$, where Ch^* is a prediction of what value the cheating verifier $\mathcal{CV}$ *will* choose. Then, it computes $\mathbf{r}'_1 \leftarrow \mathbf{s}'-\mathbf{r}'_0$, $c'_0 \leftarrow Com(\mathbf{r}'_0,\mathbf{t}'_0,\mathbf{e}'_0)$, and $c'_1 \leftarrow Com(\mathbf{r}'_1,\mathbf{G}(\mathbf{t}'_0,\mathbf{r}'_1)+\mathbf{e}'_0)$. It sends (c'_0,c'_1) to $\mathcal{CV}$. Receiving a challenge α from $\mathcal{CV}$, it computes $\mathbf{t}'_1 \leftarrow \alpha\mathbf{r}'_0-\mathbf{t}'_0$. If $Ch^*=0$ then it computes $\mathbf{e}'_1 \leftarrow \alpha\mathbf{F}(\mathbf{r}'_0)-\mathbf{e}'_0$, else $\mathbf{e}'_1 \leftarrow \alpha(\mathbf{v}-\mathbf{F}(\mathbf{s}')+\mathbf{F}(\mathbf{r}'_0))-\mathbf{e}'_0$. It sends $(\mathbf{t}'_1,\mathbf{e}'_1)$ to $\mathcal{CV}$. Due to the statistically hiding property of Com, a challenge Ch from $\mathcal{CV}$ is equal to Ch^* with probability 1/2. If $Ch=Ch^*$ then $\mathbf{r}'_0$ and $\mathbf{r}'_1$ are accepted responses to $Ch=0$ and 1, respectively. Note that the case of $\alpha=0$ does not spoil the zero-knowledge property. The details of the proof are given in the full paper, where we formally construct a black-box simulator $\mathcal{S}$ which outputs a successful transcript with probability $1/2+1/2q$. □

Theorem 5. *The 5-pass protocol is argument of knowledge for $R_{\mathbf{F}}$ with knowledge error $1/2 + 1/2q$ when the commitment scheme Com is computationally binding.*

Proof sketch. Let $((c_0, c_1), \alpha_i, (\tilde{\mathbf{t}}_1^{(i)}, \tilde{\mathbf{e}}_1^{(i)}), Ch_j, Rsp^{(i,j)})$ be four transcripts for $i, j \in \{0, 1\}$ such that $Dec(\mathbf{F}, \mathbf{v}; (c_0, c_1), \alpha_i, (\tilde{\mathbf{t}}_1^{(i)}, \tilde{\mathbf{e}}_1^{(i)}), Ch_j, Rsp^{(i,j)}) = 1$, $\alpha_0 \neq \alpha_1$, and $Ch_j = j$. Then, by using the four transcripts, it is shown to be able to either break the binding property of *Com* or extract a solution for $\mathbf{v}$. Consider that the responses are parsed as $Rsp^{(0,0)} = \tilde{\mathbf{r}}_0^{(0)}$, $Rsp^{(0,1)} = \tilde{\mathbf{r}}_1^{(0)}$, $Rsp^{(1,0)} = \tilde{\mathbf{r}}_0^{(1)}$, and $Rsp^{(1,1)} = \tilde{\mathbf{r}}_1^{(1)}$. Then, it is seen that

$$\begin{aligned} c_0 &= Com(\tilde{\mathbf{r}}_0^{(0)}, \alpha_0 \tilde{\mathbf{r}}_0^{(0)} - \tilde{\mathbf{t}}_1^{(0)}, \alpha_0 \mathbf{F}(\tilde{\mathbf{r}}_0^{(0)}) - \tilde{\mathbf{e}}_1^{(0)}) \\ &= Com(\tilde{\mathbf{r}}_0^{(1)}, \alpha_1 \tilde{\mathbf{r}}_0^{(1)} - \tilde{\mathbf{t}}_1^{(1)}, \alpha_1 \mathbf{F}(\tilde{\mathbf{r}}_0^{(1)}) - \tilde{\mathbf{e}}_1^{(1)}) \quad \text{and} \qquad (6) \\ c_1 &= Com(\tilde{\mathbf{r}}_1^{(0)}, \alpha_0(\mathbf{v} - \mathbf{F}(\tilde{\mathbf{r}}_1^{(0)})) - \mathbf{G}(\tilde{\mathbf{t}}_1^{(0)}, \tilde{\mathbf{r}}_1^{(0)}) - \tilde{\mathbf{e}}_1^{(0)}) \\ &= Com(\tilde{\mathbf{r}}_1^{(1)}, \alpha_1(\mathbf{v} - \mathbf{F}(\tilde{\mathbf{r}}_1^{(1)})) - \mathbf{G}(\tilde{\mathbf{t}}_1^{(1)}, \tilde{\mathbf{r}}_1^{(1)}) - \tilde{\mathbf{e}}_1^{(1)}). \qquad (7) \end{aligned}$$

If the two tuples of the arguments of *Com* are distinct on either of the above equations, the binding property of *Com* is broken. Otherwise, it is seen that $(\alpha_0 - \alpha_1)(\mathbf{v} - \mathbf{F}(\tilde{\mathbf{r}}_1^{(0)})) = \mathbf{G}(\tilde{\mathbf{t}}_1^{(0)} - \tilde{\mathbf{t}}_1^{(1)}, \tilde{\mathbf{r}}_1^{(0)}) + \tilde{\mathbf{e}}_1^{(0)} - \tilde{\mathbf{e}}_1^{(1)}$ from the equation (7). Combining it with the equation (6) yields $(\alpha_0 - \alpha_1)(\mathbf{v} - \mathbf{F}(\tilde{\mathbf{r}}_1^{(0)})) = \mathbf{G}((\alpha_0 - \alpha_1)\tilde{\mathbf{r}}_0^{(0)}, \tilde{\mathbf{r}}_1^{(0)}) + (\alpha_0 - \alpha_1)\mathbf{F}(\tilde{\mathbf{r}}_0^{(0)})$. Thus, $\mathbf{v} = \mathbf{F}(\tilde{\mathbf{r}}_1^{(0)}) + \mathbf{G}(\tilde{\mathbf{r}}_0^{(0)}, \tilde{\mathbf{r}}_1^{(0)}) + \mathbf{F}(\tilde{\mathbf{r}}_0^{(0)}) = \mathbf{F}(\tilde{\mathbf{r}}_1^{(0)} + \tilde{\mathbf{r}}_0^{(0)})$ is obtained, since $\alpha_0 \neq \alpha_1$. It means that a solution $\tilde{\mathbf{r}}_1^{(0)} + \tilde{\mathbf{r}}_0^{(0)}$ for $\mathbf{v}$ is extracted. The details of the proof are given in the full paper, where a knowledge extractor is formally constructed. □

5 Security and Efficiency

In this section, we summarize the security which is easily derived from the properties of zero-knowledge argument of knowledge, and give a practical parameter choice for each of the 3-pass scheme and the 5-pass scheme.

5.1 Security of the Identification Schemes

Here we briefly mention the security of each of the sequential and the parallel compositions when $\mathcal{MQ}(n, m, \mathbb{F}_q)$ is intractable and the commitment scheme *Com* is statistically hiding and computationally binding. Let $(\mathtt{P}, \mathtt{V})$ be an identification protocol described in Section 3 or Section 4. Then identification protocols which consist of repeating $(\mathtt{P}, \mathtt{V})$ N-times in sequential and in parallel are denoted by $(\mathtt{P}_N^{(s)}, \mathtt{V}_N^{(s)})$ and $(\mathtt{P}_N^{(p)}, \mathtt{V}_N^{(p)})$, respectively. The security of our identification schemes $(\mathtt{Setup}, \mathtt{Gen}, \mathtt{P}_N^{(s)}, \mathtt{V}_N^{(s)})$ and $(\mathtt{Setup}, \mathtt{Gen}, \mathtt{P}_N^{(p)}, \mathtt{V}_N^{(p)})$ is evaluated as follows.

First, we consider $(\mathtt{Setup}, \mathtt{Gen}, \mathtt{P}_N^{(s)}, \mathtt{V}_N^{(s)})$. From Theorem 2 and the sequential composition lemma [25], $(\mathtt{P}_N^{(s)}, \mathtt{V}_N^{(s)})$ is statistically zero knowledge. Furthermore,

it is directly shown that the sequential repetition reduces the knowledge error at an optimal rate in the same way as [48,49]. We note that Bellare and Goldreich showed the theorem for a general reduction of a knowledge error by the sequential repetition [4]. Therefore, the identification scheme $(\mathtt{Setup}, \mathtt{Gen}, \mathtt{P}_N^{(s)}, \mathtt{V}_N^{(s)})$ is secure against impersonation under *active* attack where $N = \omega(\log \lambda)$.

Second, consider $(\mathtt{Setup}, \mathtt{Gen}, \mathtt{P}_N^{(p)}, \mathtt{V}_N^{(p)})$. It is easy to see that the parallel repetition of $(\mathtt{P}, \mathtt{V})$ reserves zero-knowledge with respect to an *honest verifier*. Because if the simulator $\mathcal{S}$ knows a challenge Ch which $\mathcal{CV}$ will choose, then $\mathcal{S}$ can *always* output a successful transcript in both case of the 3-pass protocol and the 5-pass protocol. Furthermore, Pass and Venkitasubramaniam mentioned that the parallel repetition reduces a knowledge error in a constant-round public-coin argument of knowledge [37]. In particular, the error rate drops exponentially with the number of repetitions N. Therefore, the identification scheme $(\mathtt{Setup}, \mathtt{Gen}, \mathtt{P}_N^{(p)}, \mathtt{V}_N^{(p)})$ is secure against impersonation under *passive* attack where $N = \omega(\log \lambda)$. In addition, for a certain parameter choice, the parallel version of our scheme is also secure under *active* attack as shown in Section 6.

5.2 Efficiency

We estimate practical sizes of system parameters, a public key, a secret key, and a transcript of our schemes. The numbers of arithmetic operations, computing permutations, and computing hash functions are also estimated as computational cost. Almost all arithmetic operations are done in evaluations of $\mathbf{F}$ and $\mathbf{G}$. The efficiency is compared with that of the identification schemes based on binary SD, q-ary SD, CLE, PP, and PK. The key lengths of these schemes for around 80-bit security is estimated in [12,23]. In our evaluation, the key lengths given in [12] are used, where lengths of a hash value and a random seed are 160 bits and 128 bits, respectively.

First, we consider the 3-pass identification scheme employing $\mathcal{MQ}(84, 80, \mathbb{F}_2)$. Following the same way as [7,3], the time complexity of the F_5 algorithm to break $\mathcal{MQ}(84, 80, \mathbb{F}_2)$ is estimated to be more than 2^{80}. Furthermore, the complexity of the improved exhaustive search algorithm to break $\mathcal{MQ}(84, 80, \mathbb{F}_2)$ [10], which is stated as the best known algorithm, is $2^{88.7}$ and thus also more than 2^{80}. Table 1 shows comparison of the sequential version of our scheme and the 3-pass schemes based on binary SD, CLE, and PP when each protocol is repeated until impersonation probability is less than 2^{-30}. In the SD-based scheme, the CLE-based scheme, and ours, we consider the case that $H(c_0, c_1, c_2)$ is sent in the first pass instead of (c_0, c_1, c_2) as mentioned at the end of Section 3. In the PP-based scheme, we consider the efficient version using hash tree [43]. The sizes of public/secret keys and communication of our scheme are smaller than those of the others. Although the size of system parameter of our scheme is relatively large, it can be reduced to some small seed, e.g. 128 bits, if a pseudo-random number generator is used as the implementation of QUAD [2]. Although the cost of arithmetic operations of our scheme is relatively high, it is still reasonable. In particular, our scheme does not require random permutations.

Second, consider the 5-pass identification scheme. As the order q of a field becomes larger, the knowledge error of the 5-pass protocol $1/2 + 1/2q$ is smaller. Here we use $\mathcal{MQ}(45, 30, \mathbb{F}_{2^4})$ which is one of the minimal recommended parameters given in [8] for 80-bit security. Table 2 shows efficiency of the sequential

Table 1. Comparison of 3-pass schemes on 80-bit security against key-recovery attack when the impersonation probability is less than 2^{-30}

	SD [47,49]	CLE [48]	PP [43]	Our
round	52	52	73	52
system parameter (bit)	122,500[*1]	4,608[*1]	28,497[*1]	285,600[*1]
public key (bit)	350	288[*2]	245	80
secret key (bit)	700	192	177	84
communication (bit)	59,800[*6]	45,517[*3 *4 *6]	100,925[*6]	29,640
arithmetic ops. (times/field)	2^{24} / $\mathbb{F}_2$	2^{16} / $\mathbb{F}_{257}$	2^{22} / $\mathbb{F}_{127}$	2^{26} / $\mathbb{F}_2$
permutations[*5] (times/size)	$2/S_{700}$	$4/S_{24}$	$2/S_{161}$,S_{177}	NO
hash function (times)	4	4	8	4
best known key-recovery attack	2^{87}	2^{84}	$> 2^{74}$	2^{80}

Table 2. Comparison of 5-pass schemes on 80-bit security against key-recovery attack when the impersonation probability is less than 2^{-30}

	SD [47,49]	SD [12]	PK [46]	CLE [48]	PP [42,43]	Our
round	31	31	31	31	52	33
system parameter (bit)	122,500[*1]	32,768[*1]	4,608[*1]	4,608[*1]	28,497[*1]	259,200[*1]
public key (bit)	2450	512	384	288[*2]	245	120
secret key (bit)	4900	1024	203[*7]	192	177	180
communication (bit)	120,652[*6]	61,783[*6]	27,234[*6]	27,528[*3 *6]	105,060[*6]	26,565
arithmetic ops. (times/field)	$2^{23}/\mathbb{F}_2$	$2^{18}/\mathbb{F}_{256}$	$2^{15}/\mathbb{F}_{251}$	$2^{15}/\mathbb{F}_{257}$	$2^{21}/\mathbb{F}_{127}$	$2^{22}/\mathbb{F}_{2^4}$
permutations[*5] (times/size)	$8/S_{700}$	$2/S_{128}$	$3/S_{48}$	$4/S_{24}$	$2/S_{161}$,S_{177}	NO
hash function (times)	2	2	2	2	5	2
best known key-recovery attack	2^{87}	2^{87}	2^{85}	2^{84}	$> 2^{74}$	2^{83}

[*1] These values can be reduced to 128 bit if a pseudo-random number generator is used.

[*2] For the verification, only one vector P is required for the public key whose size is 96 bits. However, as mentioned in [48,12], zero-knowledge property of the scheme can only be stated if two quantities ($S\sigma$ and $T\tau$) are public in addition to the vector P.

[*3] It is estimated for the case where elements in $\mathbb{F}_{257}$ are regarded as 8 bits.

[*4] In the original paper [48], a prover sends $(U\sigma, V\tau, (U+S)\sigma, (V-T)\tau)$ in the third pass. However, if the prover sends $(U\sigma, V\tau, S\sigma, T\tau)$ instead of $(U\sigma, V\tau, (U+S)\sigma, (V-T)\tau)$, then the communication cost is reduced. Our estimation employs the efficient version.

[*5] This shows the number of times of computing permutations and the size of the permutation, where S_n means a permutation over $\{1, \ldots, n\}$.

[*6] By following [46,48], the data size of S_n is regarded as $\lceil \log_2(n!) \rceil$ bits. Furthermore, in the same way as [48,49,12], the data size of a random permutation or a random vector is estimated at the length of random seed as 128 bits if it is over 128 bits.

[*7] We follow the original paper [46] and estimate the length of the secret key as $\lceil \log_2(n!) \rceil$ bits, although it is regarded as the length of the random seed in [12].

version of our scheme and the 5-pass schemes based on binary SD, q-ary SD, CLE, PK, and PP when each protocol is repeated until impersonation probability is less than 2^{-30}. This table tells us some advantages of our 5-pass scheme, which are similar to those of our 3-pass scheme.

6 On the Security against Active Attack in Parallel Repetition

In this section, we focus on the case that the underlying MQ function is substantially compressing, in particular, mapping $\mathbb{F}_q^n$ to $\mathbb{F}_q^m$ where $n = m + k$ and $k = \omega(\log \lambda)$. For example, the MQ function $\mathbf{F} \in \mathcal{MQ}(2m, m, \mathbb{F}_q)$ satisfies the requirement where $m = \omega(\log \lambda)$. In this case, the parallel version ($\mathtt{Setup}, \mathtt{Gen}$, $\mathtt{P}_N^{(p)}, \mathtt{V}_N^{(p)}$) of our 3-pass scheme is shown to be secure against impersonation under *active* attack, although the sizes of the secret key and the communication data increase at most double compared to those of Section 5.2. This argument can also be applied to our 5-pass scheme.

First, we define the preimage resistance and the second-preimage resistance of the MQ function as follows. The preimage resistance is slightly different from the intractability assumption of Definition 1 in the distribution of the challenge $\mathbf{v}$, but is also widely believed.

Definition 6. *For polynomially bounded functions $n = n(\lambda)$, $m = m(\lambda)$, and $q = q(\lambda)$, it is said that $\mathcal{MQ}(n, m, \mathbb{F}_q)$ is preimage resistant if there is no polynomial-time algorithm that takes $(\mathbf{F}, \mathbf{v})$ generated via $\mathbf{F} \in_R \mathcal{MQ}(n, m, \mathbb{F}_q)$ and $\mathbf{v} \in_R \mathbb{F}_q^m$ and finds a preimage $\mathbf{s} \in \mathbb{F}_q^n$ such that $\mathbf{F}(\mathbf{s}) = \mathbf{v}$ with non-negligible probability $\epsilon(\lambda)$. On the other hand, it is said that $\mathcal{MQ}(n, m, \mathbb{F}_q)$ is second-preimage resistant if there is no polynomial-time algorithm that takes $(\mathbf{F}, \mathbf{x})$ generated via $\mathbf{F} \in_R \mathcal{MQ}(n, m, \mathbb{F}_q)$ and $\mathbf{x} \in_R \mathbb{F}_q^n$ and finds a second preimage $\mathbf{x}' \in \mathbb{F}_q^n$ such that $\mathbf{F}(\mathbf{x}') = \mathbf{F}(\mathbf{x})$ and $\mathbf{x}' \neq \mathbf{x}$ with non-negligible probability $\epsilon(\lambda)$.*

When a second-preimage resistant hash function is substantially compressing, it is known to be preimage resistant [44]. Conversely, with respect to the MQ function, the following lemma is also shown.

Lemma 7. *If $\mathcal{MQ}(n, m, \mathbb{F}_q)$ is preimage resistant, then $\mathcal{MQ}(n + 1, m, \mathbb{F}_q)$ is second-preimage resistant.*

Proof sketch. Given $\mathbf{F} = (f_1, \ldots, f_m) \in_R \mathcal{MQ}(n, m, \mathbb{F}_q)$ and $\mathbf{v} = (v_1, \ldots, v_m) \in_R \mathbb{F}_q^m$, we show that a preimage $\mathbf{x}$ satisfying $\mathbf{v} = \mathbf{F}(\mathbf{x})$ can be found by using an algorithm $\mathcal{A}$ that breaks the second-preimage resistance of $\mathcal{MQ}(n + 1, m, \mathbb{F}_q)$, where $f_l(x_1, \ldots, x_n) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{l,i,j} x_i x_j + \sum_{i=1}^{n} b_{l,i} x_i$. For the simplicity, suppose that the algorithm $\mathcal{A}$ takes $\tilde{\mathbf{F}} = (\tilde{f}_1, \ldots, \tilde{f}_m) \in \mathcal{MQ}(n + 1, m, \mathbb{F}_q)$ and $\mathbf{t} = (t_1, \ldots, t_{n+1}) \in \mathbb{F}_q^{n+1}$ and outputs a second preimage $\mathbf{t}+\Delta$ such that $\tilde{\mathbf{F}}(\mathbf{t}+\Delta) = \tilde{\mathbf{F}}(\mathbf{t})$ and $\Delta = (d_1, \ldots, d_n, 1)$, where $\tilde{f}_l(x_1, \ldots, x_{n+1}) = \sum_{i=1}^{n+1} \sum_{j=1}^{n+1} \tilde{a}_{l,i,j} x_i x_j$

$+\sum_{i=1}^{n+1}\tilde{b}_{l,i}x_i$. In this case, the equation $\tilde{\mathbf{F}}(\mathbf{t}+\Delta)-\tilde{\mathbf{F}}(\mathbf{t})=\mathbf{0}$ is expanded as follows:

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\tilde{a}_{l,i,j}d_id_j+\sum_{i=1}^{n}(\sum_{j=1}^{n+1}(\tilde{a}_{l,i,j}+\tilde{a}_{l,j,i})t_j+\tilde{b}_{l,i}+(\tilde{a}_{l,i,n+1}+\tilde{a}_{l,n+1,i}))d_i + \sum_{j=1}^{n+1}(\tilde{a}_{l,n+1,j}+\tilde{a}_{l,j,n+1})t_j+\tilde{b}_{l,n+1}+\tilde{a}_{l,n+1,n+1}=0$$

for $l=1,\ldots,m$. From the above equation, we can see that the output $\mathbf{t}+\Delta$ of $\mathcal{A}$ satisfies $\mathbf{v}=\mathbf{F}(d_1,\ldots,d_n)$ if the input $(\tilde{\mathbf{F}},\mathbf{t})$ of $\mathcal{A}$ is produced as follows.

- The vector $\mathbf{t}$ is generated via $\mathbf{t}\in_R\mathbb{F}_q^{n+1}$.
- For $1\le i\le n$ and $1\le j\le n$ do $\tilde{a}_{l,i,j}\leftarrow a_{l,i,j}$, otherwise $\tilde{a}_{l,i,j}\in_R\mathbb{F}_q$.
- For $1\le i\le n$ do $\tilde{b}_{l,i}\leftarrow b_{l,i}-(\tilde{a}_{l,n+1,i}+\tilde{a}_{l,i,n+1})-\sum_{j=1}^{n+1}(\tilde{a}_{l,i,j}+\tilde{a}_{l,j,i})t_j$, otherwise $\tilde{b}_{l,n+1}\leftarrow -v_l-\tilde{a}_{l,n+1,n+1}-\sum_{j=1}^{n+1}(\tilde{a}_{l,n+1,j}+\tilde{a}_{l,j,n+1})t_j$.

The details of the proof of Lemma 7 are described in the full paper. □

Moreover, the following lemma is also shown.

Lemma 8. *Let $n=m+k$, $k=\omega(\log\lambda)$, and $N=\omega(\log\lambda)$. Suppose that $\mathcal{MQ}(n,m,\mathbb{F}_q)$ is second-preimage resistant. Then, $(\mathtt{P}_N^{(p)},\mathtt{V}_N^{(p)})$ achieves the security against impersonation under* active *attack when Com is statistically hiding and computationally binding.*

Proof sketch. The proof of the lemma follows standard techniques used in [19,29,33]. We construct an algorithm $\mathcal{B}$ breaking the second-preimage resistance of $\mathcal{MQ}(n,m,\mathbb{F}_q)$ by using an impersonator $\mathcal{I}=(\mathcal{CP},\mathcal{CV})$ which succeeds impersonation under active attack. Given $(\mathbf{F},\mathbf{x})$, the algorithm $\mathcal{B}$ runs the cheating verifier $\mathcal{CV}$ on input $(\mathbf{F},\mathbf{v})$ where $\mathbf{v}=\mathbf{F}(\mathbf{x})$. Using the secret key $\mathbf{x}$, $\mathcal{B}$ can simulate the prover oracle perfectly. After obtaining a state for $\mathcal{CP}$ from $\mathcal{CV}$, $\mathcal{B}$ feeds the state to $\mathcal{CP}$ and acts as the legitimate verifier. By using standard rewinding techniques, $\mathcal{B}$ either breaks the binding property of *Com* or obtains $\mathbf{x}'$ satisfying $\mathbf{v}=\mathbf{F}(\mathbf{x}')$, in the same way as the proof of Theorem 3. Furthermore, the event $\mathbf{x}'\neq\mathbf{x}$ occurs with non-negligible probability, because of the following (1) and (2): (1) $(\mathtt{P}_N^{(p)},\mathtt{V}_N^{(p)})$ is statistically witness indistinguishable when *Com* is statistically hiding, due to Theorem 2. (2) The probability that there is not another $\mathbf{x}'\in\mathbb{F}_q^n\setminus\{\mathbf{x}\}$ such that $\mathbf{F}(\mathbf{x})=\mathbf{F}(\mathbf{x}')$ is at most q^{-k} which is negligible, since $k=\omega(\log(\lambda))$. In the case of $\mathbf{x}'\neq\mathbf{x}$, $\mathcal{B}$ finds a second preimage $\mathbf{x}'$. The details of the proof of Lemma 8 are described in the full paper. □

We note that the above proof can be extended into that of the security under *concurrent* attack [6] as in the proof of Kawachi et al. [29].

Finally, combining Lemma 7 and Lemma 8 yields the following theorem.

Theorem 9. *Let $n=m+k$, $k=\omega(\log\lambda)$, and $N=\omega(\log\lambda)$. Suppose that $\mathcal{MQ}(n-1,m,\mathbb{F}_q)$ is preimage resistant. Then, $(\mathtt{P}_N^{(p)},\mathtt{V}_N^{(p)})$ achieves the security against impersonation under* active *attack when Com is statistically hiding and computationally binding.*

7 Extensions of Our Scheme

In this section we mention the following two extensions of our scheme.

Slightly Efficient Parallelization. The trick mentioned in the end of Section 3 can also be applied into the parallel version of our 3-pass scheme ($\mathtt{Setup}, \mathtt{Gen}$, $\mathtt{P}_N^{(p)}, \mathtt{V}_N^{(p)}$) without losing the security. After that, a hash value of $3N$-tuple of commitments $c = H((c_{0,1}, c_{1,1}, c_{2,1}), \ldots, (c_{0,N}, c_{1,N}, c_{2,N}))$ is sent by a prover in the first pass, where $c_{i,j}$ is a commitment and H is a collision resistant hash function. The sizes of a public key, a secret key, and communication data of the modified scheme are only 80 bits, 84 bits, and $160 + 410N$ bits, respectively.

A Signature Scheme. The Fiat-Shamir method is a generic technique which transforms an identification scheme into a signature scheme [20]. The signature scheme is secure against chosen-message attack in the random oracle model if the underlying identification scheme is secure against impersonation under passive attack [1]. Thus the transform yields a signature scheme based on the conjectured intractability of the MQ problem from the parallel version of our 3-pass identification scheme. Using the signature scheme, our identification/signature scheme can also be extended to an identity-based one in a natural way [5].

8 Conclusion

We introduced the dividing techniques using bilinearity of the polar form of the MQ function and proposed public-key identification schemes consisting of a non-trivial construction of zero-knowledge argument of knowledge for the MQ problem, assuming the existence of a non-interactive commitment scheme. For a practical parameter choice, the efficiency of our schemes is highly comparable to identification schemes based on another problem including PK, SD, CLE, and PP. Furthermore, even if the protocol is repeated in parallel, our scheme can achieve the security under active attack with some additional cost.

References

1. Abdalla, M., An, J.H., Bellare, M., Namprempre, C.: From Identification to Signatures via the Fiat-Shamir Transform: Minimizing Assumptions for Security and Forward-Security. In: Knudsen, L. (ed.) EUROCRYPT 2002. LNCS, vol. 2332, pp. 418–433. Springer, Heidelberg (2002)
2. Arditti, D., Berbain, C., Billet, O., Gilbert, H.: Compact FPGA Implementations of QUAD. In: Bao, F., Miller, S. (eds.) ASIACCS, pp. 347–349. ACM, New York (2007)
3. Bardet, M., Faugère, J.-C., Salvy, B.: Complexity of Gröbner Basis Computation for Semi-regular Overdetermined Sequences over F_2 with Solutions in F_2. Research Report RR-5049, INRIA (2003)
4. Bellare, M., Goldreich, O.: On Defining Proofs of Knowledge. In: Brickell, E.F. (ed.) CRYPTO 1992. LNCS, vol. 740, pp. 390–420. Springer, Heidelberg (1993)

5. Bellare, M., Namprempre, C., Neven, G.: Security Proofs for Identity-Based Identification and Signature Schemes. J. Cryptology 22(1), 1–61 (2009)
6. Bellare, M., Palacio, A.: GQ and Schnorr Identification Schemes: Proofs of Security against Impersonation under Active and Concurrent Attacks. In: Yung, M. (ed.) CRYPTO 2002. LNCS, vol. 2442, pp. 162–177. Springer, Heidelberg (2002)
7. Berbain, C., Gilbert, H., Patarin, J.: A Practical Stream Cipher with Provable Security. In: Vaudenay [50], pp. 109–128
8. Bettale, L., Faugère, J.-C., Perret, L.: Hybrid Approach for Solving Multivariate Systems over Finite Fields. Journal of Mathematical Cryptology 3(3), 177–197 (2009)
9. Billet, O., Robshaw, M.J.B., Peyrin, T.: On Building Hash Functions from Multivariate Quadratic Equations. In: Pieprzyk, J., Ghodosi, H., Dawson, E. (eds.) ACISP 2007. LNCS, vol. 4586, pp. 82–95. Springer, Heidelberg (2007)
10. Bouillaguet, C., Chen, H.-C., Cheng, C.-M., Chou, T., Niederhagen, R., Shamir, A., Yang, B.-Y.: Fast Exhaustive Search for Polynomial Systems in F_2. In: Mangard, S., Standaert, F.-X. (eds.) CHES 2010. LNCS, vol. 6225, pp. 203–218. Springer, Heidelberg (2010)
11. Bouillaguet, C., Faugère, J.-C., Fouque, P.-A., Perret, L.: Practical Cryptanalysis of the Identification Scheme Based on the Isomorphism of Polynomial with One Secret Problem. Cryptology ePrint Archive, Report 2010/504 (2010)
12. Cayrel, P.-L., Véron, P., El Yousfi Alaoui, S.M.: A Zero-Knowledge Identification Scheme Based on the q-ary Syndrome Decoding Problem. In: Biryukov, A., Gong, G., Stinson, D.R. (eds.) SAC 2010. LNCS, vol. 6544, pp. 171–186. Springer, Heidelberg (2011)
13. Cramer, R. (ed.): EUROCRYPT 2005. LNCS, vol. 3494. Springer, Heidelberg (2005)
14. Dubois, V., Fouque, P.-A., Shamir, A., Stern, J.: Practical Cryptanalysis of SFLASH. In: Menezes, A. (ed.) CRYPTO 2007. LNCS, vol. 4622, pp. 1–12. Springer, Heidelberg (2007)
15. Dubois, V., Fouque, P.-A., Stern, J.: Cryptanalysis of SFLASH with Slightly Modified Parameters. In: Naor, M. (ed.) EUROCRYPT 2007. LNCS, vol. 4515, pp. 264–275. Springer, Heidelberg (2007)
16. Faugère, J.C.: A New Efficient Algorithm for Computing Gröbner Bases without Reduction to Zero (F5). In: Proceedings of the 2002 International Symposium on Symbolic and Algebraic Computation, ISSAC 2002, pp. 75–83. ACM, New York (2002)
17. Faugère, J.-C., Perret, L.: Polynomial Equivalence Problems: Algorithmic and Theoretical Aspects. In: Vaudenay [50], pp. 30–47
18. Feige, U., Fiat, A., Shamir, A.: Zero-Knowledge Proofs of Identity. J. Cryptology 1(2), 77–94 (1988)
19. Feige, U., Shamir, A.: Witness Indistinguishable and Witness Hiding Protocols. In: STOC, pp. 416–426. ACM, New Orleans (1990)
20. Fiat, A., Shamir, A.: How to Prove Yourself: Practical Solutions to Identification and Signature Problems. In: Odlyzko, A.M. (ed.) CRYPTO 1986. LNCS, vol. 263, pp. 186–194. Springer, Heidelberg (1987)
21. Fouque, P.-A., Granboulan, L., Stern, J.: Differential Cryptanalysis for Multivariate Schemes. In: Cramer [13], pp. 341–353
22. Fouque, P.-A., Macario-Rat, G., Stern, J.: Key Recovery on Hidden Monomial Multivariate Schemes. In: Smart, N.P. (ed.) EUROCRYPT 2008. LNCS, vol. 4965, pp. 19–30. Springer, Heidelberg (2008)

23. Gaborit, P., Girault, M.: Lightweight Code-Based Identification and Signature. In: IEEE International Symposium on Information Theory, ISIT, pp. 191–195 (2007)
24. Garey, M.R., Johnson, D.S.: Computers and Intractability; A Guide to the Theory of NP-Completeness. W. H. Freeman & Co., New York (1979)
25. Goldreich, O.: Foundations of Cryptography: Volume I. Basic Tools. Cambridge University Press, Cambridge (2001)
26. Haitner, I., Reingold, O.: Statistically-Hiding Commitment from Any One-Way Function. In: Johnson, Feige [28], pp. 1–10
27. Halevi, S., Micali, S.: Practical and Provably-Secure Commitment Schemes from Collision-Free Hashing. In: Koblitz, N. (ed.) CRYPTO 1996. LNCS, vol. 1109, pp. 201–215. Springer, Heidelberg (1996)
28. Johnson, D.S., Feige, U. (eds.): Proceedings of the 39th Annual ACM Symposium on Theory of Computing, San Diego, California, USA, June 11-13. ACM, New York (2007)
29. Kawachi, A., Tanaka, K., Xagawa, K.: Concurrently Secure Identification Schemes Based on the Worst-Case Hardness of Lattice Problems. In: Pieprzyk, J. (ed.) ASIACRYPT 2008. LNCS, vol. 5350, pp. 372–389. Springer, Heidelberg (2008)
30. Kipnis, A., Patarin, J., Goubin, L.: Unbalanced Oil and Vinegar Signature Schemes. In: Stern, J. (ed.) EUROCRYPT 1999. LNCS, vol. 1592, pp. 206–222. Springer, Heidelberg (1999)
31. Kipnis, A., Shamir, A.: Cryptanalysis of the Oil & Vinegar Signature Scheme. In: Krawczyk, H. (ed.) CRYPTO 1998. LNCS, vol. 1462, pp. 257–266. Springer, Heidelberg (1998)
32. Komano, Y., Akiyama, K., Hanatani, Y., Miyake, H.: ASS-CC: Provably Secure Algebraic Surface Signature Scheme. In: The 2010 Symposium on Cryptography and Information Security 4A2-4 (2010)
33. Lyubashevsky, V.: Lattice-Based Identification Schemes Secure Under Active Attacks. In: Cramer, R. (ed.) PKC 2008. LNCS, vol. 4939, pp. 162–179. Springer, Heidelberg (2008)
34. Lyubashevsky, V.: Fiat-Shamir with Aborts: Applications to Lattice and Factoring-Based Signatures. In: Matsui, M. (ed.) ASIACRYPT 2009. LNCS, vol. 5912, pp. 598–616. Springer, Heidelberg (2009)
35. Matsumoto, T., Imai, H.: Public Quadratic Polynominal-Tuples for Efficient Signature-Verification and Message-Encryption. In: Gunther, C.G. (ed.) EUROCRYPT 1988. LNCS, vol. 330, pp. 419–453. Springer, Heidelberg (1988)
36. Micciancio, D., Vadhan, S.P.: Statistical Zero-Knowledge Proofs with Efficient Provers: Lattice Problems and More. In: Boneh, D. (ed.) CRYPTO 2003. LNCS, vol. 2729, pp. 282–298. Springer, Heidelberg (2003)
37. Pass, R., Venkitasubramaniam, M.: An Efficient Parallel Repetition Theorem for Arthur-Merlin Games. In: Johnson, Feige [28], pp. 420–429
38. Patarin, J.: Cryptanalysis of the Matsumoto and Imai Public Key Scheme of Eurocrypt '88. In: Coppersmith, D. (ed.) CRYPTO 1995. LNCS, vol. 963, pp. 248–261. Springer, Heidelberg (1995)
39. Patarin, J.: Hidden Fields Equations (HFE) and Isomorphisms of Polynomials (IP): Two New Families of Asymmetric Algorithms. In: Maurer, U. (ed.) EUROCRYPT 1996. LNCS, vol. 1070, pp. 33–48. Springer, Heidelberg (1996)
40. Patarin, J., Goubin, L.: Trapdoor One-Way Permutations and Multivariate Polynominals. In: Han, Y., Okamoto, T., Qing, S. (eds.) ICICS 1997. LNCS, vol. 1334, pp. 356–368. Springer, Heidelberg (1997)
41. Perret, L.: A Fast Cryptanalysis of the Isomorphism of Polynomials with One Secret Problem. In: Cramer [13], pp. 354–370

42. Pointcheval, D.: A New Identification Scheme Based on the Perceptrons Problem. In: Santis, A.D. (ed.) EUROCRYPT 1995. LNCS, vol. 950, pp. 319–328. Springer-Verlag, Heidelberg (1995)
43. Pointcheval, D., Poupard, G.: A New NP-Complete Problem and Public-key Identification. Des. Codes Cryptography 28(1), 5–31 (2003)
44. Rogaway, P., Shrimpton, T.: Cryptographic Hash-Function Basics: Definitions, Implications, and Separations for Preimage Resistance, Second-Preimage Resistance, and Collision Resistance. In: Roy, B., Meier, W. (eds.) FSE 2004. LNCS, vol. 3017, pp. 371–388. Springer, Heidelberg (2004)
45. Sakumoto, K., Shirai, T., Hiwatari, H.: On the Security of the Algebraic Surface Signature Scheme. IEICE Technical Report ISEC2010-39 (2010-9) (2010)
46. Shamir, A.: An Efficient Identification Scheme Based on Permuted Kernels (Extended Abstract). In: Brassard, G. (ed.) CRYPTO 1989. LNCS, vol. 435, pp. 606–609. Springer, Heidelberg (1990)
47. Stern, J.: A New Identification Scheme Based on Syndrome Decoding. In: Stinson, D.R. (ed.) CRYPTO 1993. LNCS, vol. 773, pp. 13–21. Springer, Heidelberg (1994)
48. Stern, J.: Designing Identification Schemes with Keys of Short Size. In: Desmedt, Y.G. (ed.) CRYPTO 1994. LNCS, vol. 839, pp. 164–173. Springer, Heidelberg (1994)
49. Stern, J.: A New Paradigm for Public Key Identification. IEEE Transactions on Information Theory, 13–21 (1996)
50. Vaudenay, S. (ed.): EUROCRYPT 2006. LNCS, vol. 4004. Springer, Heidelberg (2006)