Private Minutia-Based Fingerprint Matching

Neyire Deniz Sarier

Private Minutia-based Fingerprint Matching

Neyire Deniz Sarier

MEF University,
Department of Computer Engineering
Istanbul, Turkey
sarierd@mef.edu.tr

Abstract. In this paper, we propose an efficient biometric authentication protocol for fingerprints particularly suited for the minutia-based representation. The novelty of the protocol is that we integrate the most efficient (linear complexity) private set intersection cardinality protocol of Cristofaro et al. and a suitable helper data system for biometrics in order to improve the accuracy of the system. We analyze the security of our scheme in the standard model based on well-exploited assumptions, considering malicious parties, which is essential to eliminate specific attacks on biometric authentication schemes designed for semi-honest adversaries only. Finally, the complexity is compared to the existing provably secure schemes for fingerprint matching, which shows that the new proposal outperforms them both in semi-honest and malicious security models.
Keywords: Secure Remote Authentication, Biometrics, Set difference, Private Set Intersection, Standard model

1 Introduction

Over the last decade, it has been shown that biometrics have some advantages in authentication systems compared to password-based systems, as passwords can be easily lost, forgotten or compromised using various attacks.

However, biometrics is sensitive data, thus, biometric data, either stored on a central database or on a tamper-proof smartcard, should be protected using cryptographic techniques. For instance, biometric cryptosystems such as fuzzy extractors, fuzzy vault and bipartite biotokens are used for biometric key generation, key binding and key release, respectively. Juels and Wattenberg [20] introduce the fuzzy commitment scheme as a cryptographic primitive, which is is applicable for biometrics that can be represented as an ordered set of features. However, biometrics can be affected from two types of noise, i.e. white noise that represents the slight perturbation of each feature and the replacement noise caused by the replacement of some features. Thus, Juels and Sudan have developed the *fuzzy vault* [19], which assumes that biometrics consists of an unordered set of features and is designed for the set difference metric. Specifically, fuzzy vault [19] is a key binding system that hides an encoded secret among some chaff points, where the secret key is encoded as the coefficients of a polynomial that is evaluated at the biometric feature locations such as fingerprint minutia coordinates. Implementation of fuzzy vault for fingerprints are

given in [8] and [33, 32], where the latter two include helper data constructed from the high curvature points of the fingerprint minutia, which does not leak any information about the minutia locations and used for easing the alignment of the query fingerprint to the original template.

However, the implementation of biometric cryptosystems come along with various attacks that question the security of them [28, 27]. In fact, the first paper that considers provable security in biometric remote authentication is the work of Bringer et al. [6] that proposed a hybrid protocol distributing the server side functionality in order to detach the biometric data storage from the authentication server. The common point of this work and the following papers designed for security against semi-honest adversaries -where security is guaranteed if each party follows the protocol- is that they are all implemented for biometric data represented as a binary string such as Iris. Hence, they depend on the hamming distance metric for the matching operation of the verification protocol. For this particular metric, an efficient face-identification protocol between a client C and server S are described in [22] that is based on Secure Function Evaluation (SFE) -a special case of Secure Multiparty Computation-. Within the same framework, biometric identification [3, 2] and authentication [29] protocols are described for iris and fingerprint (in particular fingercode), all of which are based on euclidean distance metric.

Finally, one should note that the most popular and widely used techniques in fingerprint identification extract information about minutiae from a fingerprint and store that information as a set of points in the two-dimensional plane as in fuzzy vault. Fingerprint matching can also be performed using a different type of information extracted from fingerprint image, i.e. FingerCode, that uses texture information from a fingerprint scan to form fingerprint representation. Although FingerCodes are not as distinctive as minutiae-based representations, [3, 2] describe privacy-preserving protocols for FingerCodes due to the efficient implementation within the euclidean distance.

2 Related Work

It is quite surprising that despite the various papers on minutia-based biometric cryptosystems [19, 8, 33, 32, 30, 31] designed for the set difference metric, the only paper that describes a private minutia-based fingerprint authentication protocol based on SFE and set difference metric is [12]. In particular, the authors of [12] design an efficient minutia-based biometric authentication scheme for a client server architecture based on the Private Set Intersection (PSI) protocol of [13] that is secure against semi-honest parties in the standard model and malicious adversaries in the random oracle model (ROM). This PSI protocol is based on homomorphic encryption and polynomial interpolation and its computation complexity is quadratic, although the number of modular exponantiations can be reduced to $O(n \log \log m)$. Here, m denotes the size of the client set and n denotes the size of the server set with $m \approx n$ in the authentication mode. Besides, [3, 29] describe private minutia-based fingerprint matching using homomorphic

encryption for euclidean distance, the former considering semi-honest adversaries only in a system based on garbled circuit evaluation. The latter is also based on polynomial interpolation idea of [13] but it is much more complex compared to the original scheme as it can be deduced from the computation complexity that is $O(nmwh)$ for the semi-honest case, where w and h denote the pixel sizes of the fingerprint image.

As one can notice, current minutia-based biometric authentication schemes, whose security is proven against semi-honest attackers are based on PSI, in particular the combination of homomorphic encryption and polynomial interpolation. A natural question is whether there exists more efficient constructions of PSI that is applicable to input sets that can be represented as an unorderded set of elements such as fingerprint minutia. To answer this, we need to inverstigate several techniques that realize PSI protocols such as Public-Key-Based PSI, Circuit-Based PSI, OT-Based PSI and Third Party-Based PSI as summerized in [23]. Specifically, the first PSI protocol based on the Diffie-Hellman (DH) key agreement scheme was presented in [16] without any security analysis. This protocol is based on the commutative properties of the DH function and was used for private preference matching, which allows two parties to verify if their preferences match to some degree. The Diffie-Hellman-based protocol of [16], which was the first PSI protocol, is actually the most efficient w.r.t. communication (when implemented using elliptic-curve crypto) [23]. Therefore it is suitable for settings with distant parties which have limited connectivity. Lastly, it is possible to incorporate a relatively efficient zero-knowledge proof and authenticated inputs that each party is following the protocol honestly, so that active cheating by either party will be detected. In this context, [10] extends the protocol of [16] for malicious server and semi-honest client by incorporating zero-knowlege proofs and two additional communication rounds and provides a simulation based proof in ROM in order to build a Private Set Intersection Cardinality (PSI-CA) protocol. Similarly, [18] also extends the protocol of [16] so that security is guaranteed for malicious parties (both C and S) in ROM. The protocols in [10, 18] provide linear complexity in the sizes of the two input sets, however the PSI protocol in [18] cannot be converted to a PSI-CA scheme due to its ROM based security proof that reveals the common elements of the intersection set to one of the parties (C or S).

2.1 Motivation and Contributions

When confronted with the PSI problem, most novices come up with a solution where both parties apply a cryptographic hash function to their inputs and then compare the resulting hashes. Although this protocol is very efficient, it is insecure if the input domain is not large or does not have high entropy, since one party could easily run a brute force attack that applies the hash function to all items that are likely to be in the input set and compare the results to the received hashes. This is exactly the case for minutia based fingerprint data. To avoid this attack, our solution is to incorporate a malicious-secure PSI to biometric authentication.

First of all, when designing a secure biometric authentication protocol, one should consider three major points: The matching should be performed privately for both sides, namely, for the two parties, a client C and a server S who jointly compute a function of their private inputs, the parties should only learn the output of the matching and nothing else. Secondly, the protocol should consider both honest-but curious adversaries and malicious adversaries. This is required for a secure biometric system in order to protect against the attack of [1], which regenarates the enrolled biometric image from a random template with a hill climbing attack, that depends on the matching score. However, a recent publication [15] shows that with malicious behaviour against the cryptographic identification protocol SciFI [22] designed for the semi-honest adversaries, one can reconstruct a full face image with the help of computer vision techniques although SciFI does not output any matching score. The attack relies on the fact that a dishonest adversary is able to input vectors of any form, not just vectors that are properly formatted [15]. The attack learns the client's face code bit-by bit through the output of $'match'$ or $'nomatch'$ decision. Thus, the new protocol should be designed in the malicious security model so that neither learning the matching score nor the accept/reject decision could help a malicious party to learn additional information about the private data of the other party including the common elements of the intersection set as in PSI schemes. Finally, the protocol should be practical and efficiently implementable with linear complexity (in terms of computation and communication cost) and it should depend on widely adopted representations of biometric data.

With these goals in mind, we present a new minutia-based fingerprint authentication protocol for set difference metric between a client and a server based on PSI techniques. In particular, the only work within this framework is the work of [12], that depends on the PSI scheme of [13].

Specifically, our protocol is inspired by the PSI-CA scheme of [10] although our scheme is defined on an elliptic curve group that simplifies the PSI-CA protocol of [10] slightly by removing the last step of the protocol (i.e. hashing), but more importantly, the need for a random oracle which questions the security of the systems when the ROM is replaced by a real hash function. In fact, certain artificial signature and encryption schemes are known which are proven secure in the ROM, but which are trivially insecure when any real function is substituted for the random oracle [7]. This way, we also reduce the communication complexity since the communication overhead of [10] amounts to $2(m+1)$ $|p|$-bit values with $|p| = 1024$ or $|p| = 2048$, whereas our protocol requires $2(m+1)$ $|q|$-bit values with $|q| = 160$ or $|q| = 224$. Thus, our scheme is a scalable and efficient protocol with linear complexity and its security relies on well exploited cryptographic assumptions (DDH and l-DDHI) in the standard model. Besides, our protocol reveals neither the server nor the client the elements of the intersection set S, but only the size of the intersection set $d = |S|$ is learned by a single party (C or S). Similar to the scheme of [12], the computation complexity of [29] is also quadratic, i.e. $O(nmwh)$ for the semi-honest case, where w and h denote the pixel sizes of the fingerprint image. Thus, our proposal is more

efficient compared to the current private fingerprint matching schemes [12, 29] that are based on Oblivious Polynomial Evaluation (OPE) of [13].

Furthermore, we discuss the security of our scheme in malicious model in order to prevent the attacks presented in [1, 15]. Unfortunately, the PSI-CA scheme of [10] can only achieve one-sided simulatibility in ROM, i.e. the scheme only provides privacy of the server against a semi-honest client. Thus, we extend the security of our protocol so that both parties can be corrupted by a malicious adversary in standard model.

To the best of our knowledge, the proposed scheme is the first private minutia-based fingerprint authentication protocol for set difference metric that achieves complexities linear in the size of input sets, i.e. set of user's minutia that is secure in the standard model both for semi-honest and malicious adversaries.

3 Building Blocks

3.1 Fingerprint data

The approach that forms the basis for the biometric data representation of our scheme is the Minutiae Fuzzy Vault Implementation of Uludag et al. [33, 32]. Our system operates on the fingerprint minutiae that are generally represented as (x_i, y_i, θ_i) triplets, denoting their row indices (x_i), column indices (y_i) and angle of the associated ridge, respectively. Next, we concatenate x_i and y_i coordinates of a minutia as $[x_i|y_i]$ to arrive at the data unit b_i for $i \in [1, m]$. To account for slight variations in minutiae data (due to fingerprint distortions), raw minutiae data are first quantized. We require an alignment step where the query minutiae templates are aligned to the registered template based on using auxiliary alignment data aux, i.e. helper data derived from the orientation field of fingerprints. Naturally, it is required that the helper data does not leak any information about the minutiae-based fingerprint template. Another approach could be the use of alignment-free features, i.e. features that do not depend on the finger's rotation or displacement. The reader is referred to [33, 32] for the details of this representation.

3.2 Cryptographic tools

Since our system works in set difference metric, we need to compare/match aligned query template to the registered template in a private manner. In particular, our protocol is inspired by the (reversed) PSI-CA scheme of [10] that enables two parties, i.e. a client C which has a set $B' = (b'_1, ..., b'_m)$ of size m and a server S which has a set $B = (b_1, ..., b_n)$ of size n to compute the size of the intersection of their respective sets without disclosing anything about their inputs including the common elements of the intersection set. After the computation the server has obtained the size of the intersection $d = |B \cap B'|$ and the client has learnt nothing other than the accept/reject notification based on the system threshold t.

In short, PSI and PSI-CA can be achieved using OPE [13], Oblivious Pseudo-Random Functions (OPRF) [17], Bloom filters [11] and blind signatures [10], where the latter is the primitive we require in our protocol to achieve linear complexity. As different from the scheme of PSI-CA of [10] we eliminate the last step of the protocol, namely hashing the result of the verification and computing the size of the intersection on these hashes. Besides, we swap the roles of the server and the client in [10], thus, the biometric server obtains a signature on its input without disclosing it. This simplification is caused by describing our protocol on a suitably chosen elliptic curve group where DDH (and l-DDHI) assumption holds, whereas PSI-CA of [10] works on groups where DDH (and One-More-Gap-DH) assumption holds. Thus, the client performs $2(m+1)$ exponentiations and server computes $(m+n)$ modular exponentiations modulo p-bit prime with $p = 1024$ or $p = 2048$, whereas in our scheme the same operations are performed modulo q-bit prime with $q = 160$ or $q = 224$. In [10], communication overhead amounts to $2(m+1)$ p-bit values and n κ-bit values, where κ is a security parameter of H':$\{0.1\}^* \rightarrow \{0.1\}^\kappa$. Since, we eliminate H' and work on an elliptic curve group, the communication complexity is reduced from p-bit values to q-bit values. To provide client and server privacy against malicious adversaries, we employ standard techniques of cryptography such as zero knowledge proof of knowledge (PoK).

3.3 Security Model

We provide efficient biometric authentication protocols with security in the presence of both semi-honest and malicious adversaries. Here, the term adversary refers to insiders, i.e., protocol participants. Outside adversaries are not considered, since their actions can be mitigated via standard network security techniques. Informally, we have the following goals for our protocols.

Client Privacy: No information is leaked about client C biometrics, except an upper bound on its size m and the matching score, i.e. the number of common elements between the biometric template registered at the server and the client's fresh template.

Server Privacy: C learns no information beyond an upper bound on the size of his registered feature set n at the server and the accept/reject notification.

Unlinkability: Neither party can determine if any two instances of the protocol are related, i.e., executed on the same input by client or server, unless this can be inferred from the actual protocol output [10].

Our first protocols for authentication are presented in the semi-honest model, i.e. adversaries that are honest-but-curious, who follow the protocols and try to gain more information than they should on the other parties' inputs. An honest-but-curious party is a party that follows the instructions of the protocol, but may record the communications it receives and try to infer extra information using such recordings. In this case, the traditional real-versus-ideal definition is applied in the security proof. Basically, the protocol privately computes a function for an honest-but-curious Client C (resp. Server S) if there exists a PPT algorithm SIM that is able to simulate the view of C (resp. S), given only

Client's (resp. Server's) (private and public) input and output. The random variable representing the view of Client (resp. Server) during an execution of the protocol with Client's private input $B' = \{b'_i\}$, Server's private input $B = \{b_i\}$ is denoted here by $View_S(B, B', P)$ (resp. $View_C(B, B', P)$).

Definition 1. *(Privacy against Honest-but-curious Adversaries).*
Let $View_S(B, B')$ be a random variable representing server's view during execution of PSI-CA with inputs B, B', P. There exists a PPT algorithm SIM that is able to simulate the view of Server (resp. Client), given only Server's (resp. Client's) respective (private and public) input and output; i.e., $\forall(B, B', P)$:

$$View_S(B, B', P) \stackrel{c}{\equiv} SIM_S(B, P, |B \cap B'|))$$
$$(resp.\ View_C(B, B', P) \stackrel{c}{\equiv} SIM_C(B', P))$$

The security of our protocols relies on the following assumptions.

Definition 2. *Decisional Diffie-Hellman (DDH). Let $x, y, z \stackrel{R}{\leftarrow} \mathbb{Z}_q^*$ and $g \in \mathbb{G}$ be a random generator of the prime order group $\mathbb{G}$. Given (g, g^x, g^y) distinguishing between the distributions (g, g^x, g^y, g^{xy}) and (g, g^x, g^y, g^z) is hard.*

Definition 3. *l-Diffie-Hellman inversion problem (l-DHI). Let $l \in \mathbb{Z}$, $z \stackrel{R}{\leftarrow} \mathbb{Z}_q^*$ and $g \in \mathbb{G}$ as above. Given $(g, g^z, g^{z^2}, ..., g^{z^l})$ computing $g^{\frac{1}{z}}$ is hard.*

Definition 4. *l-Decisional Diffie-Hellman inversion problem (l-DDHI). Let $l \in \mathbb{Z}$, $z \stackrel{R}{\leftarrow} \mathbb{Z}_q^*$, $g \in \mathbb{G}$. Given $(g, g^z, g^{z^2}, ..., g^{z^l}, v)$ deciding whether $v = g^{\frac{1}{z}}$ is hard.*

In section 7, we present our last protocol for authentication in malicious model, where a malicious adversary uses any kind of strategy to learn information. A malicious party is a part that does not necessarily follow the instructions of the protocol. Finally, the number of minutiae used in the protocol, namely n and m, are considered to be public. If privacy of the number of minutiae is required, C and S can simply agree on a size (or two sizes) beforehand and then adjust the number of minutiae they use as input by either omitting a number of minutiae or adding a number of chaff minutiae to their set.

4 The new Protocol

As a warm up, this section presents our first construction in authentication mode, secure in the presence of semi-honest adversaries in the ROM. An overview of the scheme is given in Fig. 1. Although our scheme integrates the PSI-CA of [10], its security is based on a different assumption. Besides, we work on a group $\mathbb{G}$ implemented using a group of points on a certain elliptic curve with generator g of prime order q and require a MaptoPoint hash function (modeled as a random oracle) H: $\{0.1\}^* \rightarrow \mathbb{G}$ together with two random permutations $\mathbf{P}$ and $\mathbf{P}'$.

The client C registers his biometric features b_i for $i \in [1, n]$ at the server S as described in section 3.1 and stores the helper data aux publicly. For verification,

C presents his fresh biometrics, aligns it with the help of aux, and obtains $\{b'_i\}$ for $i \in [1, m]$. Next, C makes an authentication request and the server S replies by masking the hashed biometric feature set items corresponding to the client C with a random exponent $k \in \mathbb{Z}_q$ and sends resulting w'_is to C, which blindly exponentiates them with its own random value $\alpha \in \mathbb{Z}_q$. Next, C shuffles these v'_is and sends to S the resulting u'_is together with the exponentiations of client's items $\mathrm{H}(\underline{b}'_j)'$s to randomness $\alpha \in \mathbb{Z}_q$ as x'_js. Finally, S tries to match these x'_j values received from C with the shuffled u_i values, stripped of the initial randomness $k \in \mathbb{Z}_q$. S learns the set intersection cardinality (and nothing else) by counting the number of such matches and notifies C based on the system threshold t with an accept/reject decision.

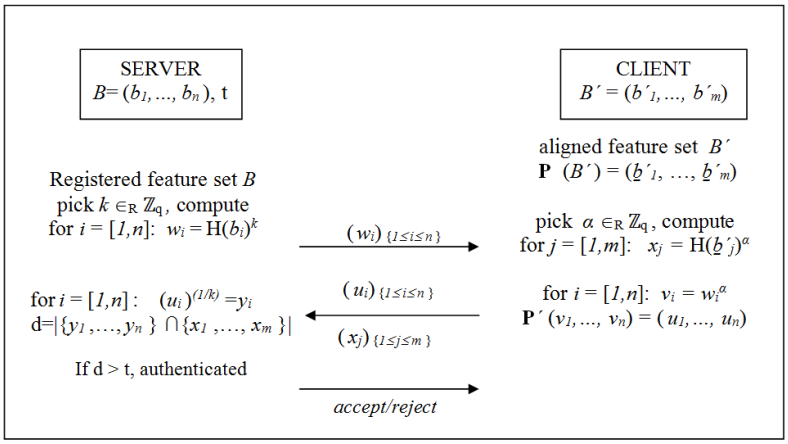

Fig. 1. Protocol in ROM: $m \approx n$

Lemma 1. *The proposed scheme achieves client privacy against a semi-honest server based on the l-DDHI assumption in the random oracle model.*

Lemma 2. *The proposed scheme achieves server privacy against a semi-honest client based on the DDH assumption in the random oracle model.*

Due to page limitations, the proofs will appear in the full version of the paper.

By designing the protocol for an elliptic curve group $\mathbb{G}$, we do not require a second hash function H', hence our scheme is less complex compared to [10], since the elements of $\mathbb{G}$ are already 160 or 224-bits instead of 1024 or 2048-bit as in [10]. Hence, the comparison performed over the H' values as in [10], can be performed on x'_js and y'_is directly. Since the protocol is designed for semi-honest adversaries, the attack of [1] does not work since the parties are passive attackers and do follow the protocol specifications. However, the distance/matching score

or accept/reject notification could be useful for a malicious server for a brute force attack against the privacy of the client or the opposite, namely, a malicious client trying to impersonate a user. In other words, this information is only helpful as in the case of malicious behaviour by one of the parties. However, to prevent malicious behaviour as presented in [1, 15], where the latter attack is able to break the secure face identification scheme SciFI even if no matching score or distance information is output by the protocol, one should extend the security of the new scheme for malicious adversaries.

5 Security in Standard model

As described above, our protocol requires one hash function that is assumed as a random oracle. However, by slightly modifying the protocol, we are able to prove the security of our scheme in the standard model. In particular, instead of extracting the input set of each party via the random oracle queries as in [10], we use the Proof of Knowledge (PoK) to extract the randomness k used by each party and determine the input set as in [17, 18]. Hence, we use the input set of the semi-honest (resp. malicious) party directly in the simulation due to the extraction of sender's inputs given this randomness that is obtained by running the extractor algorithm for PoK with the semi-honest party to extract k, such that it satisfies the commitment g^k sent by that semi-honest party. As an example application, we can replace the hash function with the MapToPoint hash function of [14, 4], we are able to prove the security in the standard model.

For instance, [14] relies on a variant of Dodis-Yampolskiy's Pseudo-Random Function (PRF) based on the Boneh-Boyen unpredictable function [17]. The Boneh-Boyen function is $f_y(x) = g^{1/(y+x)}$ where $g \in \mathbb{G}$ generates a group $\mathbb{G}$ of prime order q, and y is a random element in $\mathbb{Z}_q^*$. This function is unpredictable under the computational l-DHI assumption on $\mathbb{G}$ [17]. Thus, the decisional l-DHI assumption on group $\mathbb{G}$ implies that the Boneh-Boyen function is a PRF. Besides, the OPRF construction of [17] is also based on the Boneh-Boyen PRF with the sole modification being a substitution of a prime-order group $\mathbb{G}$ with a group whose order is a safe RSA modulus.

Lemma 3. *The proposed scheme achieves client privacy against a semi-honest server in the standard model.*

Proof. We show that server's view can be efficiently simulated by a probabilistic polynomial time algorithm SIM_S. The server's view includes his inputs B, randomnesses he uses, and messages he receives. The server has inputs of the registered feature set $B = \{b_i\}$ and randomness $k \in \mathbb{Z}_q$. We follow a similar proof technique that is presented in [17]. The simulator is constructed as follows:

1. Upon receiving g^k, π_1 and $w_1, ..., w_n$ from the server, if the server succeeds in the proof π_1, then SIM_S runs the extractor algorithm for π_1 with the server to extract k. Then when getting the randomness k from S, SIM_S tries every possible input in the range of the hash function -which is identical to

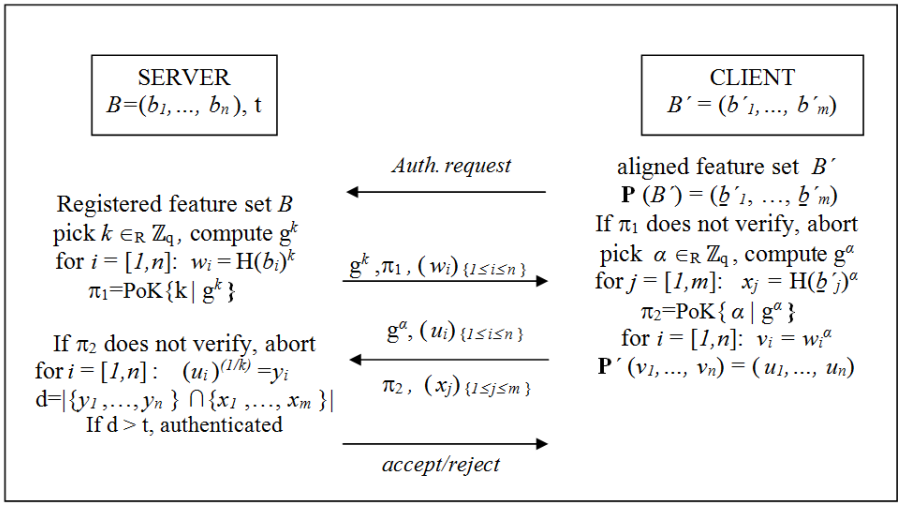

Fig. 2. Protocol in standard model: $m \approx n$

the Boneh-Boyen PRF- to reconstruct the set B as in OPRF proof of [17]. This can be performed due to the fact that the domain of this hash/PRF is polynomially-sized [17].

2. SIM_S picks at random $\alpha \leftarrow \mathbb{Z}_q$, computes g^α, computes π_2 and adds distinct pairs $(\mathrm{H}(b_i), x_i)=(h_i, x_i)$, where x_i $=\mathrm{H}(b_i)^\alpha$ and b_is (i.e. the set B) are computed as in the previous step. SIM_S computes $v_i = w_i^\alpha$ and sends $\mathbf{P'}(v_1, ..., v_n) = (u_1, ..., u_n)$ and $(x_1, ..., x_m)$ to the server. Here, $(x_1, ..., x_d)$ denotes the intersection of the client and server's input set constructed by selecting a random subset of x_i $=\mathrm{H}(b_i)^\alpha$ values with size $|d|$. For the remaining $m - d$ elements, the simulator padds the set with random values, i.e. c_i^α for $i \in [d+1, m]$.

Server learns nothing either interacting with the real world client or interacting with SIM_S, therefore, the environment (distinguisher) D's views in the real world and ideal world are indistinguishable. Now we show that this SIM_S does a successful simulation. Consider the following series of games:

1. In the first game, the public parameters are generated as in the definition of the protocol, and then the adversary A interacts with the real world party as defined above.
2. In the second game, the parameters are generated the same way, but now A interacts with a $\overline{SIM}$ which behaves as the real protocol for step 1, but then behaves as SIM_S for step 2. The only difference then is that this simulator padds the set with random values, i.e. c_i^α for $i \in [d+1, m]$ for the remaining $m - d$ elements. This differs from the first game only in that the elements not common with the set B and the simulated set B' are randomly chosen in

order to simulate the fresh biometric reading of the client biometrics which cannot be equal to the registered biometric set B totally due to the nature of biometrics. Thus, this is indistinguishable from the first game by the randomness of these padded elements chosen from the underlying group.

3. In the last game, the public parameters are generated the same way, and then adversary A interacts with SIM_S. This differs from the second game only in that SIM_S extracts k from the proof, and uses this k to form the registered biometric set of the authenticating client at the server. Note that if the proof is sound, then this set will be identical to that used in the previous game. Thus this is indistinguishable from the previous game by the extraction property of the ZK proof system.

Since the first game is indistinguishable from the third, the probability that the adversary A can detect the simulation in each game can differ only negligibly. Thus, the simulation is successful.

Lemma 4. *The proposed scheme achieves server privacy against a semi-honest client in the standard model.*

Due to page limitations, the proof will appear in the full version of the paper.

6 Use of multi-modal biometrics for high-entropy inputs

One factor limiting the security of biometric cryptosystems is the entropy of the biometric feature data. To increase the entropy of biometric data and to achieve higher privacy levels in biometric cryptosystems, one combines the information of several biometric traits (e.g. fingerprints with finger vein, or face with iris) or several instances of the same biometric trait, denoted as multi-biometrics systems. Compared to traditional (uni)biometric authentication, multibiometric systems offer several advantages such as better recognition accuracy, increased population coverage, greater security, flexibility and user convenience. For these systems, different fusion approaches exist, and in [21], fusion at the feature level is performed for both multi-modal and multi-instances that the key entropy in the biometric cryptosystem is increased to sufficient levels required in security applications. In [26, 25, 24], another fusion at the feature level is described in the context of biometric IBE in order to avoid the collusion attacks inherent in fuzzy IBE systems. Considering our biometric matching system, one can follow a similar strategy as described in [28]. Specifically, 2048 bits Iriscode b has inherent entropy of 249 bits. If we implement the Iris fuzzy commitment scheme of [5], we can see this Iris code as $z = b \oplus c$, where c is a codeword that is stored in form of $H(c)$ as a helper data together with z. If we concataneted to each biometric feature (for instance fingerprint minutia value) this c, each of the biometric data has enough input entropy for the hash function. To further increase the input-entropy, a client password can be concatanated to the biometric inputs, where a randomly generated 8-character password can have 52-bit entropy [21].

7 Security in Malicious model

Consider a malicious client (or an adversary trying to impersonate a user) that implements one of the attacks presented in [1, 15] against the biometric authentication system. To prevent this, the security should be guaranteed considering malicious behaviour of both parties. We note that the PSI-CA protocol of [10] provides security against semi-honest server and malicious client, when the roles of server and client are swapped, namely the protocol provides one-sided simulatibility in ROM.

To upgrade our scheme presented in Fig. 2 to malicious parties in the standard model, we add one additional zero-knowledge proof π_3 as in [10], where π_3= PoK $\{\alpha|(\prod_{i=1}^{m} w_i)^{\alpha} = \prod_{i=1}^{m} u_i\}$ since a proof of logical *and* of n separate statements $w_i^{\alpha} = u_i$ would reveal the relationship between each index i of w_i and corresponding index j of u_j with $w_i^{\alpha} = u_j$ after permutation $\mathbf{P'}$ allowing the server to determine which elements belong to the intersection, rather than just how many [10]. We note that considering our protocol in a group equipped with a bilinear map does not solve the problem since the server can check $\hat{e}(w_i, g^{\alpha}) = \hat{e}(u_j, g)$ for each u_j until he determines all the common elements instead of just their cardinality.

The commitments g^k, g^{α} together with the proofs of knowledge allows the simulator to extract the malicious party's input and may help to ensure that the inputs are consistent and that the same values are used along the protocol. However, since any logical *and* of n separate PoK as in the above sense would reveal the common elements themselves instead of just their total number, a challenge/response mechanism similar to the one in [10] is needed to guarantee that the same α is used on each w_i. An overview of the protocol is presented in Fig. 3.

Lemma 5. *The proposed scheme achieves client privacy against a malicious server in the standard model.*

Sketch of the Proof. A malicious server against a honest client can behave arbitrarly as in the following ways.

Case1: A malicious server can pick a random set of inputs instead of the registered user information B or does not apply the same random exponent k that is committed in $w_i = \mathrm{H}(b_i)^k$ and g^k. To avoid this, one can include a zero knowledge proof in order to prove the honest client that the malicious server knows the underlying registered biometric feature hashes and another zero knowledge proof to prove that the committed value in g^k is consistently used in all w_is. However, as it is proven in [18], the server (i.e. the receiver of the PSI scheme of [18]) cannot change its input set B after sending the w_is since the server's input set is committed in the first and only message he sends regarding the biometric data. With this behaviour, the server does not gain any advantage since the honest client can detect the malicious behaviour from the authentication result (i.e. a reject decision for a honest client that should be accepted) as the malicious server cannot compute the matching score and

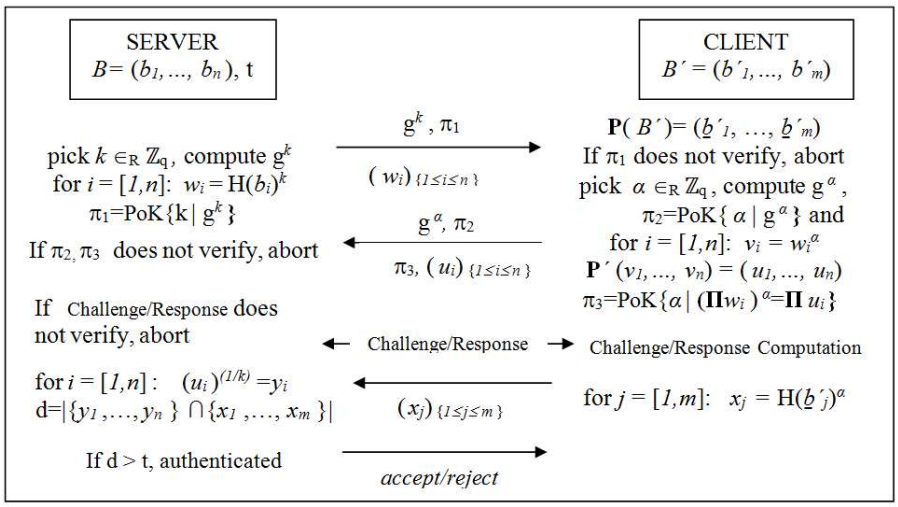

Fig. 3. Protocol in malicious model: $m \approx n$

returns a random accept/reject notification or aborts the protocol without any notification. We note that an accept decision of that server for a honest client that should be authenticated remains undetected. Hence, to prove that the committed input set of the server belongs to the particular client that tries to authenticate to the system, authorization of server input must be enforced. This can be achieved via the signature of the sensor on the inputs of the server during the registration phase of each client to the server, since the sensor, which captures the biometric data of each client is fully trusted in any biometric authentication system [6]. An example application in a different context is presented in the Authorized PSI-CA scheme of [10], which we can integrate into our construction with the sole modification of substitution the prime-order group $\mathbb{G}$ with a group whose order is a safe RSA modulus N. It is shown that prime-order groups also imply that the Boneh-Boyen function in a composite-order group N remains a PRF under the l-DDHI assumption on such groups (and hardness of factoring) and the same generic-group argument which motivated trust in the l-DDHI assumption on the prime-order groups carries to composite-order groups as well [17]. Hence, if we use the MapToPoint hash function of [14] that is identical to the Boneh-Boyen PRF, we can integrate authorization of server inputs via the signatures of the trusted sensor at the registration.

Case2: Hence, the only misbehaviour left for the malicious server is to abort without sending the final decision although it computed the (correct) matching score. This can be eliminated by providing fairness via integrating an optimistic fairness protocol, i.e. a semi-trusted offline third party arbiter. Fairness is out of the scope of this paper.

Lemma 6. *The proposed scheme achieves server privacy against a malicious client in the standard model.*

Due to page limitations, the proofs will appear in the full version of the paper.

8 Comparison

As it is noted in [23], the Diffie-Hellman-based private matching protocol of [16], which was the first PSI protocol, is actually the most efficient w.r.t. communication (when implemented using elliptic-curve crypto). Besides, the PSI scheme of [18], PSI-CA scheme of [10] and our scheme are based on small variations of the protocol in [16], this protocol is suitable for settings with distant parties which have limited connectivity. To the best of our knowledge, the only schemes that provide private fingerprint matching protocols with a concrete security analysis based on fingerprint minutia representation are described in [3], [29],[12], where the latter considers set difference metric, whereas the others implement the protocols for euclidean distance. All three of the protocols provide security against semi-honest adversaries, although the scheme of [29] includes an extention of his semi-honest protocol for malicious adversaries without any security analysis. Thus, the comparison is based on the protocols for semi-honest adversaries for consistency and we assume $m \approx n$ for the authentication mode since the total number of minutia m registered at the server and captured at the client side n will be close to each other as opposed to the identification mode as in [3].

Table 1. Comparison of time complexity

	Complexity Estimate, i.e. Number of exponentiations	Underlying Method
Blanton et al.* [3]	quadratic in m + m OT protocols	Homomorphic encryption and Garbled Circuits
Shahandashti et al. [29]	quadratic in m	OPE
Feng et al.† [12]	quadratic in m	OPE
Our Construction‡	linear in m	PSI-CA

$*$:in authentication mode;
†: [13] reduces the number of exponentiations to $O(n \log\log m)$ using Horner's rule and hashing for bucket allocation; ‡ $m \approx n$ with $20 < m < 40$;

Therefore, our construction is the most efficient authentication protocol for minutia-based fingerprint authentication based on PSI techniques, in particular the OPE of [13]. In addition, our protocol is more efficient compared to the garbled circuit-based construction of [3], as it is shown in [9], the PSI and PSI-CA constructions of [10] are more efficient compared to garbled-circuit based constructions. Finally, the only scheme that considers malicious parties is [29] (without any security analysis). Similar to the comparison in the semi honest

model, our scheme outperforms [29] also in malicious model due to the additional PoKs at each step of the protocol which is already complex enough for semi-honest model.

9 Conclusion

In this paper, we design an efficient biometric authentication protocol for a client-server architecture based on one of the most efficient PSI-CA technique. Our scheme is suitable for any type of biometrics that can be represented as an unordered set of features similar to the constructions of fuzzy vault. We provide the security in standard model based on the well-exploited assumptions and consider malicious parties, which is essential to eliminate specific attacks on biometric schemes. A future work could be integration of fairness protocol to prevent a malicious abort of the server.

References

1. A. Adler. Vulnerabilities in biometric encryption systems. In *AVBPA'05*, pages 1100–1109, 2005.
2. M. Barni, T. Bianchi, D. Catalano, M. Di Raimondo, R. Donida Labati, P. Failla, D. Fiore, R. Lazzeretti, V. Piuri, F. Scotti, and A. Piva. Privacy-preserving fingercode authentication. In *MMSec'10*, pages 231–240. ACM, 2010.
3. M. Blanton and P. Gasti. Secure and efficient protocols for iris and fingerprint identification. In *ESORICS'11*, volume 6879 of *LNCS*, pages 190–209. Springer, 2011.
4. D. Boneh and M. Franklin. Identity-based encryption from the weil pairing. In *CRYPTO'01*, volume 2139 of *LNCS*, pages 213–229. Springer, 2001.
5. J. Bringer, H. Chabanne, G. Cohen, B. Kindarji, and G. Zemor. Optimal iris fuzzy sketches. In *BTAS'07*, pages 1–6. IEEE, 2007.
6. J. Bringer, H. Chabanne, M. Izabachène, D. Pointcheval, Q. Tang, and S. Zimmer. An application of the goldwasser-micali cryptosystem to biometric authentication. In *ACISP'07*, volume 4586 of *LNCS*, pages 96–106. Springer, 2007.
7. R. Canetti, O. Goldreich, and S. Halevi. The random oracle methodology, revisited. *J. ACM*, 51(4):557–594, 2004.
8. T. C. Clancy, N. Kiyavash, and D. J. Lin. Secure smartcard based fingerprint authentication. In *WBMA'03*, pages 45–52. ACM, 2003.
9. E. De Cristofaro and G. Tsudik. Experimenting with fast private set intersection. In *Trust and Trustworthy Computing*, volume 7344 of *LNCS*, pages 55–73. Springer, 2012.
10. E.D. Cristofaro, P. Gasti, and G. Tsudik. Fast and private computation of cardinality of set intersection and union. In *CANS'12*, volume 7712 of *LNCS*, pages 218–231. Springer, 2012.
11. C. Dong, L. Chen, and Z. Wen. When private set intersection meets big data: an efficient and scalable protocol. In *ACMCCS'13*, pages 789–800. ACM, 2013.
12. Q. Feng, F. Su, and A. Cai. Privacy-preserving authentication using fingerprint. *IJICIC*, 8(11):8001–8018, 2012.

13. M. J. Freedman, K. Nissim, and B. Pinkas. Efficient private matching and set intersection. In *EUROCRYPT'04*, volume 3027 of *LNCS*, pages 1–19. Springer, 2004.
14. V. Goyal, A. O'Neill, and V. Rao. Correlated-input secure hash functions. In *TCC'11*, volume 6597 of *LNCS*, pages 182–200. Springer, 2011.
15. K. Grauman, M. Gerbush, A. Luong, and B. Waters. Reconstructing a fragmented face from a cryptographic identification protocol. In *WACV'13*, pages 238–245. IEEE, 2013.
16. B. A. Huberman, M. Franklin, and T. Hogg. Enhancing privacy and trust in electronic communities. In *Proceedings of the 1st ACM Conference on Electronic Commerce*, EC '99, pages 78–86. ACM, 1999.
17. S. Jarecki and X. Liu. Efficient oblivious pseudorandom function with applications to adaptive ot and secure computation of set intersection. In *TCC'09*, volume 5444 of *LNCS*, pages 577–594. Springer, 2009.
18. S. Jarecki and X. Liu. Fast secure computation of set intersection. In *SCN'10*, volume 6280 of *LNCS*, pages 418–435. Springer, 2010.
19. A. Juels and M. Sudan. A fuzzy vault scheme. *Des. Codes Cryptography*, 38(2):237–257, 2006.
20. A. Juels and M. Wattenberg. A fuzzy commitment scheme. In *ACM CCS'99*, pages 28–36, 1999.
21. S. Kanade, D. Petrovska-Delacre'taz, and B. Dorizzi. Multi-biometrics based cryptographic key regeneration scheme. In *Biometrics: Theory, Applications, and Systems, 2009. BTAS '09. IEEE 3rd International Conference on*, pages 1–7, 2009.
22. M. Osadchy, B. Pinkas, A. Jarrous, and B. Moskovich. Scifi - a system for secure face identification. In *IEEE Symposium on Security and Privacy*, pages 239–254, 2010.
23. B. Pinkas, T. Schneider, and M. Zohner. Faster private set intersection based on OT extension. In *Usenix'04*, pages 797–812. USENIX Association, 2014.
24. N. D. Sarier. A New Biometric Identity Based Encryption Scheme. In *International Symposium on Trusted Computing - TrustCom'08*, pages 2061–2066. IEEE, 2008.
25. N. D. Sarier. A new Biometric Identity Based Encryption Scheme secure against DoS attacks. *Security and Communication Networks*, 3(1):268–274, 2010.
26. N. D. Sarier. Generic Constructions of Biometric Identity Based Encryption Systems. In *WISTP'10*, volume 6033 of *LNCS*, pages 90–105. Springer, 2010.
27. N. D. Sarier. Security Notions of Biometric Remote Authentication Revisited. In *STM'11*, volume 7170 of *LNCS*, pages 72–89. Springer, 2011.
28. N. D. Sarier. *Biometric Cryptosystems: Authentication, Encryption and Signature for Biometric Identities*. PhD thesis, Bonn University, Germany, 2013.
29. S. F. Shahandashti, R. Safavi-Naini, and P. Ogunbona. Private fingerprint matching. In *ACISP'12*, volume 7372 of *LNCS*, pages 426–433. Springer, 2012.
30. B. Tams. Absolute fingerprint pre-alignment in minutiae-based cryptosystems. In *BIOSIG'13*, pages 1–12. IEEE, 2013.
31. B. Tams. Attacks and countermeasures in fingerprint based biometric cryptosystems. *CoRR*, abs/1304.7386, 2013.
32. U. Uludag and A. Jain. Securing fingerprint template: Fuzzy vault with helper data. In *CVPRW'06*. IEEE, 2006.
33. U. Uludag, S. Pankanti, and A. K. Jain. Fuzzy vault for fingerprints. In *AVBPA'05*, volume 3546 of *LNCS*, pages 310–319. Springer, 2005.