

PassBio: Privacy-Preserving User-Centric Biometric Authentication

Kai Zhou¹, *Student Member, IEEE*, and Jian Ren, *Senior Member, IEEE*

Abstract—The proliferation of online biometric authentication has necessitated security requirements of biometric templates. The existing secure biometric authentication schemes feature a *server-centric* model, where a service provider maintains a biometric database and is fully responsible for the security of the templates. The end-users have to fully trust the server in storing, processing, and managing their private templates. As a result, the end-users' templates could be compromised by outside attackers or even the service provider itself. In this paper, we propose a *user-centric* biometric authentication scheme (PassBio) that enables end-users to encrypt their own templates with our proposed light-weighted encryption scheme. During authentication, all the templates remain encrypted such that the server will never see them directly. However, the server is able to determine whether the distance of two encrypted templates is within a pre-defined threshold. Our security analysis shows that no critical information of the templates can be revealed under both passive and active attacks. PassBio follows a “compute-then-compare” computational model over encrypted data. More specifically, our proposed threshold predicate encryption (TPE) scheme can encrypt two vectors x and y in such a manner that the inner product of x and y can be evaluated and compared to a pre-defined threshold. TPE guarantees that only the comparison result is revealed and no key information about x and y can be learned. Furthermore, we show that TPE can be utilized as a flexible building block to evaluate different distance metrics, such as Hamming distance and Euclidean distance over encrypted data. Such a compute-then-compare computational model, enabled by TPE, can be widely applied in many interesting applications, such as searching over encrypted data while ensuring data security and privacy.

Index Terms—Biometric authentication, data security and privacy, computation over encrypted data, predicate encryption, inner product encryption.

I. INTRODUCTION

BIOMETRIC authentication has been incredibly useful in services such as access control to authenticate individuals based on their biometric traits. Unlike passwords or identity documents used in conventional authentication systems, biometric traits, such as fingerprint, iris and behavioral characteristics are physically linked to an individual that

cannot be easily manipulated. Also due to such a strong connection, security and privacy of the biometric templates used in the authentication process is a critical issue [1]–[3].

Existing biometric authentication systems generally employ a two-phase mechanism [3]. In a registration phase, an end-user submits her biometric template to the service provider who will store the template along with the end-user's ID in a central database. In a query phase, the end-user requesting access to certain services will submit a fresh template to the service provider for authentication. Based on the end-user's ID, the service provider will retrieve the enrolled template for comparison. Only if the two templates are close enough under certain distance metric, the end-user is successfully authenticated.

The above biometric authentication model can be regarded as *server-centric*. That is, the service provider will receive end-users' biometric templates in plaintext and is fully responsible for the security of the templates. Such an approach has several inherent deficiencies. First, the end-users have to fully trust the service provider to properly handle their templates; otherwise the security and privacy of the templates are at risk. For example, different service providers may cross-check their databases to discover possible duplications, meaning that the same end-user may get enrolled in different services. As a consequence, the privacy of the end-user is violated. Second, unlike password, biometric templates are inherently noisy. As a result, the fresh template to be authenticated is not necessarily the same as the registered template. Such a property prevents the service provider from keeping the templates encrypted during the whole authentication process. At some point, the templates have to be recovered in plaintext for distance computation and comparison. This renders the adversaries with the opportunity to spy the registered or freshly submitted templates.

To address the above issues, we propose a *user-centric* model for biometric authentication. In terms of security, such a user-centric model has several unique features, compared to the server-centric model. First, biometric templates are encrypted at user side and then transmitted to the server. The service provider is only able to see encrypted versions of the registered templates and query templates. Second, the secret keys and the templates are generated and processed locally thus never leaving the local environment. Third, computations involved in authentication are all carried out on ciphertext, meaning that no templates are exposed in plaintext during the authentication. These features can effectively reduce the possibility for the server as well as outside adversaries to learn any key information of the biometric templates.

Manuscript received November 7, 2017; revised March 14, 2018 and April 26, 2018; accepted May 13, 2018. Date of publication May 21, 2018; date of current version June 5, 2018. This work was supported in part by the National Science Foundation under Grant CNS: 1524520 and in part by the Semiconductor Research Corporation under Grant 2015-TS-2639. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Aris Gkoulalas Divanis. (*Corresponding author: Jian Ren.*)

The authors are with the Department of Electrical and Computer Engineering, Michigan State University, East Lansing, MI 48824-1226 USA (e-mail: zhoukai@msu.edu; renjian@msu.edu).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TIFS.2018.2838540

To meet the demands of the proposed user-centric model, the underlying encryption scheme should be efficient and expose as little information as possible. Since the key management and encryption are carried out at the user side, the encryption scheme should be computationally efficient. Some existing encryption schemes relying on heavy cryptographic operations such as Predicate Encryption (PE) [8], [9], Inner Product Encryption (IPE) [4]–[7] and Homomorphic Encryption (HE) [10], [11] may not be practical in such a scenario. Also, the encryption scheme should support certain kinds of computation on encrypted data. For example, given two encrypted vector, the server should be able to decide whether the two vectors are close enough (e.g., within a certain threshold) under some distance metric. The encryption scheme should expose as little template information as possible for security and privacy. Although some distance preserving transformation schemes [12] have been proposed for private nearest neighbor search on encrypted data, these schemes will inevitably expose the distance information between the registered and query template, which makes them vulnerable to security attacks [12].

In this paper, we propose a new primitive named Threshold Predicate Encryption (TPE). TPE encrypts two vectors \mathbf{x} and \mathbf{y} respectively as C_x and C_y . Unlike traditional cryptosystems, the decryption of TPE will only reveal whether the inner product of \mathbf{x} and \mathbf{y} is within a threshold θ or not, instead of the plaintext. Therefore, no more information about the vectors and the inner product are exposed. TPE is fundamentally different from the previous schemes such as IPE [4] and PE [8]. IPE reveals the inner product of \mathbf{x} and \mathbf{y} thus the distance between the registered template and the query template, which makes the scheme vulnerable to security attacks [12]. PE can only reveal whether the inner product equals to a threshold or not. It is not flexible enough for biometric authentication since generally we want to know whether the distance between the two templates is within a threshold. In comparison, our proposed TPE provides an excellent trade-off between information leakage and flexibility, which makes it uniquely suitable for biometric authentication.

TPE enables a compute-then-compare computational model over encrypted data. In this model, given ciphertexts, any party is able to **compute** the distance between the underlying plaintexts and then **compare** the distance with a threshold. The output is an indicator showing whether the distance is within the threshold or not. We show that such a computational model captures the essence of various applications such as privacy-preserving biometric identification and searching over encrypted data. TPE based schemes are able to fulfill the requirements of such applications while ensuring the security and privacy of the data.

The main contributions of this paper are summarized as follows:

- We propose a user-centric biometric authentication scheme enabling end-users to utilize their biometric templates for authentication while preserving template privacy.
- We propose a new primitive named TPE that can encrypt two vector \mathbf{x} and \mathbf{y} in such a manner that the decryption

result only reveals whether the inner product of \mathbf{x} and \mathbf{y} is within a threshold or not.

- The proposed TPE enables a compute-then-compare computational model over encrypted data. We show that such a computational model can be applied to many privacy-preserving applications such as biometric identification and searching over encrypted data.

The rest of this paper is organized as follows. In Section III, we introduce the system model as well as the threat model where different attacks are identified. We then illustrate the design of TPE and give a detailed implementation in Section IV. Based on TPE, we propose the user-centric biometric authentication scheme in Section V, where different similarity measurements are considered. We give detailed security analysis of TPE in Section VI. In Section VII, we introduce some applications of TPE such as outsourced biometric identification and searching over encrypted data. We analyze the complexity of TPE and evaluate the performance of TPE through some simulations in Section VIII. We conclude in Section IX.

II. RELATED WORK

The proposed TPE scheme can be regarded as an instance of functional encryption. That is, given the decryption key, the decryption process actually produces a function of the underlying plaintext, instead of the plaintext itself. In this section, we will review some related works on functional encryption and secure biometric computation.

A. Functional Encryption and Controlled Disclosure

In conventional cryptosystem, the decryption process will eventually recover the underlying plaintext m . As a result, all information of m is disclosed. Many applications, however, require only *partially disclosure* of the information of m . For example, a financial organization wants to filter out those customers whose transactions exceed certain amount. For privacy concern, all the transactions of the customers are encrypted. In this case, instead of decrypting the transactions, a more desirable approach is to determine whether a transaction exceeds certain amount without disclosing the transaction. Such application scenarios motivate the research of functional encryption [15], [16], [23]. In a functional encryption scheme, a decryption key S_f is associated with a function f . Given the ciphertext C , the decryption process will evaluate the function $f(m)$, where m is the underlying plaintext. Note that in this process, the plaintext m cannot be recovered. Thus, by issuing different decryption keys S_{f_i} , functional encryption can actually implement *controlled disclosure* of the plaintext m .

Much research effort has been devoted to designing various functions f_i for functional encryption schemes. Representative works are Predicate Encryption (PE) [8], [9] and Inner Product Encryption (IPE) [4]–[7]. In PE, a message is modeled as a vector \mathbf{x} and a decryption key is associated with a vector \mathbf{y} . The decryption result is meaningful (otherwise, a random number) if and only if the inner product of \mathbf{x} and \mathbf{y} is equal to 0. Based on this basic implementation, different predicates are

realized such as exact threshold, polynomial evaluation and set comparison. In contrast, IPE schemes will recover the value of inner product of \mathbf{x} and \mathbf{y} , without revealing neither \mathbf{x} nor \mathbf{y} . In the context of controlled disclosure, IPE discloses more information of the plaintext than PE. This is because with PE, one can only decide whether the inner product of \mathbf{x} and \mathbf{y} is equal to a certain value or not while with IPE, one can know the value of the inner product. In comparison, with TPE, what we seek is to control the amount of information to be disclosed between those of PE and IPE. As a result, TPE can efficiently fulfill the task of biometric authentication while exposing less information about the templates.

B. Secure Biometric Computation

The applications of biometric traits can largely be divided into two categories [39]: biometric authentication and biometric identification. The existing research mainly focused on privacy-preserving biometric identification [40]–[42], in which a server database is assumed to contain the templates of the registered users and their corresponding identities. A client submits a template to the server and the server will compare the template with all the templates in the database. The main goal of privacy-preserving biometric identification is to ensure that only the client is able to know the identification result meaning that even the server knows nothing about it. In this way, the privacy of the client is well protected. The implementations were based on either homomorphic encryption ([42]) and secure two-party computation ([40] and [41]). In biometric authentication, the client submits a freshly generated template together with a claimed identity. The server verifies the claim by comparing the submitted template with the registered template. Several research papers (e.g., [43], [44]) considered protecting the privacy of the biometric templates by dividing the authentication server into three functionally independent components. However, they assumed that these components would not collude with each other which may not be quite realistic.

In general, the existing research on secure biometric authentication and identification can be regarded as “server-centric” in that it assumes that the servers will properly store and process the clients’ biometric templates. In comparison, in our unique user-centric computational model, the end-user will encrypt their templates at the local side and register them to the authentication server. For authentication, the end-user submits the encrypted templates to the server. The authentication at the server side is based on the encrypted templates.

Our proposed TPE scheme utilizes similar techniques as that in [14]. However, the computational models as well as the security requirements of our model are fundamentally different. In the biometric identification scheme of [14], given a query template, the server tries to identify the closest template in the database and return it to the end-user. After decryption of the returned template, the end-user calculates the distance and determines whether the distance is within a threshold. We note that such a computational model cannot be easily applied to biometric authentication. This is because in biometric authentication, while the server needs to compare

the distance with a threshold, it is not allowed to decrypt the templates thus calculating the distance. Moreover, secure k -nn based approaches will inherently expose more information than needed, such as the relative distances between a query template and all the templates in the database. Such information is superfluous since for biometric authentication and identification, the server only needs to know whether the distance exceeds a pre-defined threshold.

III. PROBLEM STATEMENT

A. System Model

We consider an online biometric authentication system consisting of two parties: an online service provider and a set of end-users. The service provider provides certain online services such as storage to its authenticated end-users. We assume that every end-user possesses a device such as a mobile phone that is able to collect the her biometric traits and transform the traits to biometric templates at the local side. Without loss of generality, we assume that each biometric template is represented by an n -dimensional vector $\mathbf{T} = (t_1, t_2, \dots, t_n)$ of real numbers.

The biometric authentication process consists of two phases. In the registration phase, an end-user U_i will register with her biometric template $\mathbf{T}_i = (t_{i1}, t_{i2}, \dots, t_{in})$ along with a unique identifier \mathcal{ID}_i . We note that the template \mathbf{T}_i is sent to the service provider in encrypted form denoted as $\text{Enc}(\mathbf{T}_i)$ and \mathcal{ID}_i can be any pseudorandom string that uniquely identifies U_i within the system. The tuple $\langle \text{Enc}(\mathbf{T}_i), \mathcal{ID}_i \rangle$ for the end-user U_i is then stored at the server side by the service provider. In the query phase, when the end-user U_i desires to authenticate herself to the service provider, U_i will locally generate a fresh biometric template \mathbf{T}'_i and send the tuple $\langle \text{Enc}(\mathbf{T}'_i), \mathcal{ID}_i \rangle$ to the service provider, where $\text{Enc}(\mathbf{T}'_i)$ is the encrypted form of \mathbf{T}'_i . On receiving the query, the service provider will retrieve the record $\langle \text{Enc}(\mathbf{T}_i), \mathcal{ID}_i \rangle$ through searching \mathcal{ID}_i in the server. Then distance between \mathbf{T}_i and \mathbf{T}'_i are computed based on $\text{Enc}(\mathbf{T}_i)$ and $\text{Enc}(\mathbf{T}'_i)$. If the distance is within a certain threshold θ , then the service provider will view the end-user U_i as a valid user. We also note that during the query phase, the service provider is only able to derive whether the distance between \mathbf{T}_i and \mathbf{T}'_i is within the threshold θ , instead of the exact distance between them.

B. Threat Model

We assume the end-users are fully trusted in the registration phase. That is, they will honestly generate their own biometric templates and register at the service provider using the encrypted templates. In the query phase, we assume the encryption and decryption algorithms are publicly known. However, the secret keys are generated and kept secret at the local side throughout the whole authentication process. We do allow the adversaries to submit their own biometric templates through the local device. In this case, the local device acts as an oracle to encrypt templates and submit the encrypted templates to the service provider. The service provider can be honest-but-curious or malicious. In the former case, the service provider will honestly follow the protocol but will try to

obtain any useful information of end-users' biometric templates based only on the encrypted templates. In the latter case, the adversaries may collude with the service provider such as sharing with the service provider the invalid templates that are submitted through the local devices. In summary, depending on the different capabilities of the service provider and the adversaries, we propose two attack models as follows.

- 1) **Passive Attack:** the service provider is able to know the registered record $\langle \text{Enc}(\mathbf{T}_i), \mathcal{ID}_i \rangle$ for end-user U_i and observe a series of m submitted queries $\text{Enc}(\mathbf{T}_i^j)$, $j = 1, 2, \dots, m$. However, the service provider does not know the underlying templates \mathbf{T}_i^j in plaintext. Such an attack model is also known as the Ciphertext-Only-Attack in cryptography.
- 2) **Active Attack:** besides the registered record $\langle \text{Enc}(\mathbf{T}_i), \mathcal{ID}_i \rangle$ for end-user U_i , the service provider is able to observe a series of m submitted queries $\text{Enc}(\mathbf{T}_i^j)$ as well as the corresponding plaintext \mathbf{T}_i^j , $j = 1, 2, \dots, m$. Such an attack model corresponds to the Chosen-Plaintext-Attack in cryptography. In practice, an adversary may submit her own templates through the local device. The service provider can then collude with the adversary to obtain the queries in plaintext as well as the encrypted queries.

Informally, the security requirement of biometric authentication is that the service provider is unable to learn any information about the templates than allowed through the authentication process. In particular, it should be possible for the service provider to determine whether the distance between two templates is within a threshold or not; but infeasible to derive any key information about the registered template as well as the query templates. We will formally define the security against both attacks in Section VI.

IV. PROPOSED THRESHOLD PREDICATE ENCRYPTION SCHEME

A user-centric privacy-preserving biometric authentication scheme requires that an end-user is able to encrypt her registered biometric template as well as the freshly generated query templates. For the service provider, given two encrypted templates, it should be able to determine the distance between the two templates and compare the distance with a threshold. In this section, we introduce Threshold Predicate Encryption (TPE) that can fulfill the functionalities required by such a biometric authentication system.

A. Framework

Our proposed privacy-preserving biometric authentication scheme is based on the new primitive named Threshold Predicate Encryption (TPE). Generally speaking, TPE can be regarded as an instance of functional encryption [15], [16], where decryption will output a function of the plaintext instead of the plaintext itself. The framework of functional encryption can be briefly summarized as follows. A plaintext vector \mathbf{x} is encrypted as C_x and a secret key associated with a vector \mathbf{y} is generated as S_y . Given C_x and S_y , the decryption will give

the value of $f(\mathbf{x}, \mathbf{y})$, where f is a pre-defined function. Two notable instances of functional encryption are Inner Product Encryption (IPE) [4] and Predicate Encryption (PE) [8]. The function f in IPE is the inner product. That is, the decryption of IPE will give the inner product of \mathbf{x} and \mathbf{y} . In comparison, PE will produce a meaningful decryption result (e.g., a flag number 0) if and only if the inner product of \mathbf{x} and \mathbf{y} is 0. Otherwise, the decryption result is just some random number. An important predicate is that the inner product of \mathbf{x} and \mathbf{y} equals 0. Based on this, an extension of PE can implement exact threshold predicate encryption, meaning that the decryption result is meaningful only if the inner product of \mathbf{x} and \mathbf{y} is equal to a pre-defined threshold θ determined by the security requirements.

At the high-level view, functional encryption aims at revealing only limited information about the plaintext. As introduced above, IPE reveals the inner product of the plaintext and a vector. PE reveals whether the inner product is equal to 0 (or a threshold) or not. In application scenarios like biometric authentication, the amount of information revealed by IPE and PE are both inappropriate. As shown in our latter analysis, the inner product of \mathbf{x} and \mathbf{y} can be modeled as the distance between the registered template and the query template. As a result, IPE will give the exact distance between the two templates, which exposes too much information. With PE, one can decide whether the distance of the two templates is equal to a certain threshold, which is not sufficient for authentication purpose. What we need is an functional encryption scheme that can determine whether the distance between the two templates is within a threshold or not. Specifically, a TPE is composed of five algorithms:

- **TPE.Setup()** $\rightarrow param$: the set up algorithm generates system parameters $param$.
- **TPE.KeyGen(λ)** $\rightarrow sk$: on input of a security parameter λ , the key generation algorithm will generate a secret key sk .
- **TPE.Enc(sk, \mathbf{x})** $\rightarrow C_x$: given a vector \mathbf{x} and the secret key sk , the encryption algorithm will encrypt \mathbf{x} as ciphertext C_x .
- **TPE.TokenGen(sk, \mathbf{y})** $\rightarrow T_y$: given a vector \mathbf{y} and the secret key sk , the token generation algorithm will generate a token T_y for \mathbf{y} .
- **TPE.Dec(C_x, T_y)** $\rightarrow \Lambda = \{0, 1\}$: given the ciphertext C_x and the token T_y , the decryption algorithm will output a result Λ satisfying

$$\Lambda = \begin{cases} 1, & \mathbf{x} \circ \mathbf{y} \leq \theta \\ 0, & \text{otherwise,} \end{cases}$$

where $\mathbf{x} \circ \mathbf{y}$ is the inner product of \mathbf{x} and \mathbf{y} .

B. Design of TPE

While our proposed TPE scheme utilizes some similar techniques as the biometric identification scheme in [14], the settings of biometric authentication are fundamentally different. In particular, our proposed TPE is designed to address the following challenges.

Challenge 1: The system and threat model of outsourced biometric identification and biometric authentication are different. In biometric identification, the database owner possesses the encryption and decryption keys. The aim of the server is to identify the template closest to the query template. Then the database owner will retrieve the template, decrypt it and compare the distance to a threshold. However, in our scenario, the server does not possess the decryption key thus is unable to decrypt the encrypted template and calculate the distance. What we need is an encryption scheme that can directly determine whether the distance between the query template and the registered template is within the threshold based only on ciphertexts.

Challenge 2: The computation involved in biometric identification and authentication are different. In biometric identification, the sever needs to compute and compare the distances between a query template and all the templates in the database. However, in biometric authentication, we need to compute the distance and compare it with a threshold.

Challenge 3: The decryption process in [14] will output a randomized distance between a query template and registered template. From this randomized distance, it is not easy to directly compare it with a threshold without first recovering the actual distance.

To address the above challenges, we first embed the threshold into the registered templates. To enhance security, we pad the templates with one-time randomness in a special manner and make random permutation to both the query template and registered template. After all these transformations, the decryption process can derive $\text{dist}(\mathbf{x}, \mathbf{y}) - \theta$, where $\text{dist}(\mathbf{x}, \mathbf{y})$ denotes the distance between a registered template \mathbf{x} and a query template \mathbf{y} . However, if we output this value directly, it is inevitable that the exact value of $\text{dist}(\mathbf{x}, \mathbf{y})$ will be exposed. Therefore, we introduce more one-time randomness into the encrypted templates. As a result, the decryption result becomes $\alpha\beta(\text{dist}(\mathbf{x}, \mathbf{y}) - \theta)$, where α and β are positive one-time random numbers associated with \mathbf{x} and \mathbf{y} , respectively. This design reveals only adequate information to determine whether the distance between is within the threshold and at the same time conceals the exact value of the distance.

C. Construction of TPE

Follow the aforementioned design of our threshold predicate encryption scheme, we give a detailed implementation in Protocol 1.

Now, we prove the correctness of the proposed TPE scheme. For a square matrix Y , the trace $\text{Tr}(Y)$ is defined as the sum of the diagonal entries of Y . Given an invertible matrix M_1 of the same size, the transformation $M_1 Y M_1^{-1}$ is called similarity transformation of Y . We have the following lemma from linear algebra ([45, Ch. 3.3]).

Lemma 1: The trace of a square matrix remains unchanged under similarity transformation. That is, $\text{Tr}(Y) = \text{Tr}(M_1 Y M_1^{-1})$.

Based on Lemma 1, we have the following theorem.

Theorem 1: For the proposed TPE scheme in Protocol 1, $\Lambda \leftarrow \text{TPE.Dec}(C_x, T_y)$ equals 1 if and only if $\mathbf{x} \circ \mathbf{y} \leq \theta$, where $\mathbf{x} \circ \mathbf{y}$ denotes the inner product of \mathbf{x} and \mathbf{y} .

Protocol 1 Threshold Predicate Encryption (TPE) Scheme

Input: $\mathbf{x} = \{x_1, \dots, x_n\}, \mathbf{y} = \{y_1, \dots, y_n\}, \theta$.

Output: $\Lambda = \{0, 1\}$.

TPE.Setup() $\rightarrow param$:

1: set $param = \{n, \theta\}$.

TPE.KeyGen(λ) $\rightarrow sk$:

- 1: Randomly generate two non-singular $(n+3) \times (n+3)$ matrices M_1 and M_2 and calculate their inversions M_1^{-1} and M_2^{-1} .
- 2: Choose a random permutation $\pi: \mathbb{R}^{n+3} \rightarrow \mathbb{R}^{n+3}$
- 3: Set $sk = \{M_1, M_2, M_1^{-1}, M_2^{-1}, \pi\}$.

TPE.Enc(sk, \mathbf{x}) $\rightarrow C_x$:

- 1: Generate two random real numbers β and r_x , where β is positive.
- 2: Extend the vector \mathbf{x} to an $(n+3)$ -dimensional vector $\mathbf{x}' = (\beta x_1, \beta x_2, \dots, \beta x_n, -\beta\theta, r_x, 0)$.
- 3: Permute \mathbf{x}' to obtain $\mathbf{x}'' = \pi(\mathbf{x}')$.
- 4: Transform \mathbf{x}'' to a diagonal matrices X with $\text{diag}(X) = \mathbf{x}''$.
- 5: Generate a random $(n+3) \times (n+3)$ lower triangular matrices S_x with the diagonal entries fixed as 1.
- 6: Compute $C_x = M_1 S_x X M_2$.

TPE.TokenGen(sk, \mathbf{y}) $\rightarrow T_y$:

- 1: Generate two random real numbers α and r_y , where α is positive.
- 2: Extend \mathbf{y} to an $(n+3)$ -dimensional vector $\mathbf{y}' = (\alpha y_1, \alpha y_2, \dots, \alpha y_n, \alpha, 0, r_y)$.
- 3: Permute \mathbf{y}' to obtain $\mathbf{y}'' = \pi(\mathbf{y}')$.
- 4: Transform \mathbf{y}'' to a diagonal matrix Y with \mathbf{y}'' being the diagonal.
- 5: Generate a random $(n+3) \times (n+3)$ lower triangular matrix S_y with the diagonal entries fixed as 1.
- 6: Compute $T_y = M_2^{-1} Y S_y M_1^{-1}$.

TPE.Dec(C_x, T_y) $\rightarrow \Lambda = \{0, 1\}$:

- 1: Compute $I = \text{Tr}(C_x T_y)$, where $\text{Tr}(\cdot)$ denotes the trace of a matrix.
 - 2: Set $\Lambda = 1$ if $I \leq 0$; otherwise set $\Lambda = 0$.
-

Proof: Following the notations in Protocol 1, it is equivalent to prove that $I \leq 0$ if and only if $\mathbf{x} \circ \mathbf{y} \leq \theta$. In the following, we will show that $I = \alpha\beta(\mathbf{x} \circ \mathbf{y} - \theta)$.

In Protocol 1, the vector \mathbf{x} is first extended to \mathbf{x}' . Through a random permutation π , \mathbf{x}' is transformed to \mathbf{x}'' , which is further transformed to a diagonal matrix X with \mathbf{x}'' being the diagonal. Then the matrix X is transformed to $C_x = M_1 S_x X M_2$. In a similar manner, the vector \mathbf{y} is transformed to $T_y = M_2^{-1} Y S_y M_1^{-1}$ as detailed in Protocol 1.

First, we have $C_x T_y = M_1 S_x X Y S_y M_1^{-1}$. From Lemma 1, $I = \text{Tr}(C_x T_y) = \text{Tr}(S_x X Y S_y)$. Second, since X is a diagonal matrix and S_x is selected as a non-singular lower triangular matrix, where all the diagonal entries are set to 1, the product $S_x X$ is also a lower triangular matrix whose diagonal is \mathbf{x}'' (since \mathbf{x}'' is the diagonal of X). Similarly, $Y S_y$ is a lower triangular matrix whose diagonal is \mathbf{y}'' . Third, from the law of matrix multiplication, it is obvious that the diagonal of $S_x X Y S_y$ is the same as that of $X Y$, since both $S_x X$

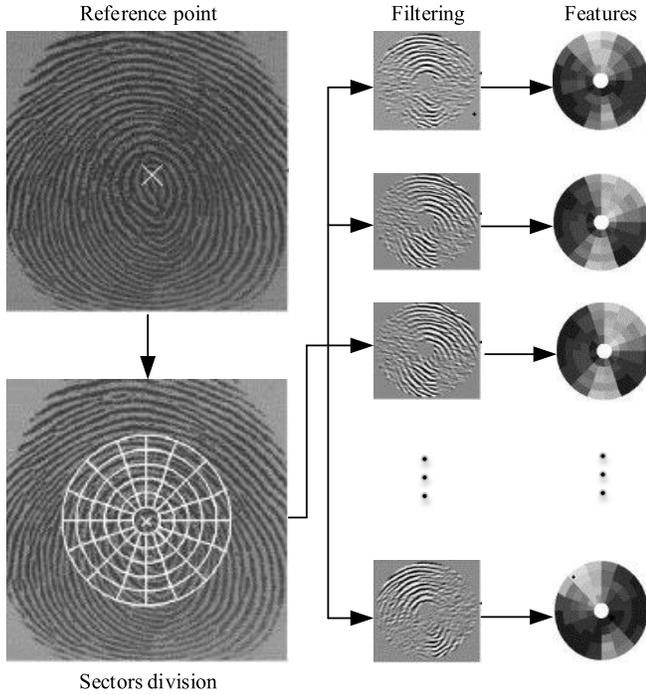


Fig. 1. Feature extraction of fingerprints: (i) Identify reference point; (ii) Divide region of interest into sectors around reference point; (iii) Filter region of interest; (iv) Extract features.

and $Y S_y$ are lower triangular matrices. As a result, we have $I = \text{Tr}(S_x X Y S_y) = \text{Tr}(X Y) = \mathbf{x}'' \circ \mathbf{y}'' = \mathbf{x}' \circ \mathbf{y}'$. The last equality holds because \mathbf{x}'' and \mathbf{y}'' are obtained through the same permutation π . From the detailed construction of \mathbf{x}' and \mathbf{y}' as in Protocol 1, it is obvious that $I = \mathbf{x}' \circ \mathbf{y}' = \alpha\beta(\mathbf{x} \circ \mathbf{y} - \theta)$.

Since α and β are selected as positive random numbers, we have $\Lambda \leftarrow \text{TPE.Dec}(C_x, T_y)$ equals 1 if and only if $\mathbf{x} \circ \mathbf{y} \leq \theta$. ■

V. BIOMETRIC AUTHENTICATION UNDER DIFFERENT DISTANCE METRICS

In this section, we will first introduce some necessary background on biometric authentication. Then, we show how to construct privacy-preserving biometric authentication systems utilizing our proposed TPE scheme under different distance metrics.

A. Backgrounds

The first critical step in biometric authentication is to efficiently transform biometric traits into templates that are easy for computation. Such a process is often called *feature extraction*. The extracted features are often represented as *feature vectors*. Depending on the biometric traits, the process as well as the result of feature extraction could differ. For example, a fingerprint can be transformed to a FingerCode [24]–[26] that is a vector of integers with dimension 640. An Iris image is often represented as a binary string of 2048 bits. In the following, we briefly review the feature extraction process of fingerprints. The details can be found in [25] and [26].

As illustrated in Fig. 1,¹ given an image of a fingerprint, the first step is to identify a reference point. Then the region

of interest around the reference point is divided into 5 bands and 16 sectors. Those sectors are further normalized and filtered by 8 different Gabor filters. At last, the features are extracted from each filtered image. The final result is a 640-dimensional vector (FingerCode) representing each fingerprint image, where each entry in the vector is an 8-bit integer. An important feature of the FingerCode is that it is translation invariant, meaning that translation of the fingerprint image would not result in much difference in the FingerCode. However, FingerCode is not rotation invariant. As a result, rotation of images will often cause different FingerCodes. To resolve this issue, a user is often associated with several (for example, 5) FingerCodes captured from rotated images in the database. In the following discussion, we assume that at the local side, there exists a sensor that can capture the end-user's biometric trait and transform it to a multi-dimensional vector.

B. Overview

A user-centric biometric authentication protocol is composed of four phases: setup, registration, query and authentication.

- 1) **Setup**: an end-user will generate his own secret key sk that will be stored at the local side and will never be exposed to any party. The end-user will also share some public parameters with the server.
- 2) **Registration**: the end-user will encrypt his biometric template T_1 using his secret key sk and then register the encrypted template $\text{Enc}(sk, T_1)$ to the server together with his identity ID .
- 3) **Query**: the end-user will generate a fresh biometric template T_2 , encrypt T_2 with the same secret key sk and send the encrypted template $\text{Enc}(sk, T_2)$ together with his identity ID to the server.
- 4) **Authentication**: the server will retrieve the encrypted template $\text{Enc}(sk, T_1)$ based on the received ID . From $\text{Enc}(sk, T_1)$ and $\text{Enc}(sk, T_2)$, the server will output an authentication result.

C. Authentication Based on Euclidean Distance

Different biometric templates often rely on different similarity measurements. For example, in Iris recognition, the templates are represented by binary vectors and the similarity is generally measured by Hamming distance. For fingerprint, the Euclidean distance is normally utilized to measure the similarity. Our proposed TPE scheme is highly flexible in that it can be applied to measuring similarity based on different distance metrics. As a result, TPE can be utilized as the critical component to build different privacy-preserving biometric authentication systems. In this section, we will illustrate how to utilize TPE to construct a biometric authentication scheme based on Euclidean distance.

Euclidean distance is often used to measure the similarity between vectors of non-binary entries. A FingerCode representing a fingerprint is an n -dimensional vector, where each entry is an l -bit integer. Typically, $n = 640$ and $l = 8$. We denote a registered FingerCode as $\mathbf{x} = (x_1, x_2, \dots, x_n)$

¹This figure is partially obtained from [26].

and a query FingerCode as $\mathbf{y} = (y_1, y_2, \dots, y_n)$. Let $d_E(\mathbf{x}, \mathbf{y})$ be the Euclidean distance between \mathbf{x} and \mathbf{y} . Then we have

$$d_E^2(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^n x_i^2 + \sum_{i=1}^n y_i^2 - 2\mathbf{x} \circ \mathbf{y},$$

where $\mathbf{x} \circ \mathbf{y}$ is the inner product of \mathbf{x} and \mathbf{y} . Let θ be a pre-defined threshold. Our goal is to extend \mathbf{x} and \mathbf{y} to vectors \mathbf{x}' and \mathbf{y}' respectively such that the relation $d_E^2(\mathbf{x}, \mathbf{y}) < \theta^2$ can be determined through computing $\mathbf{x}' \circ \mathbf{y}'$. In light of this, we let $\mathbf{x}' = (2\beta x_1, 2\beta x_2, \dots, 2\beta x_n, -\beta \sum_{i=1}^n x_i^2, \beta, \beta\theta^2, r_x, 0)$ and $\mathbf{y}' = (\alpha y_1, \alpha y_2, \dots, \alpha y_n, \alpha, -\alpha \sum_{i=1}^n y_i^2, \alpha, 0, r_y)$, where α, β, r_x, r_y are one-time random numbers. Then we have

$$\begin{aligned} \mathbf{x}' \circ \mathbf{y}' &= \alpha\beta(2\mathbf{x} \circ \mathbf{y} + \theta^2 - \sum_{i=1}^n x_i^2 - \sum_{i=1}^n y_i^2) \\ &= \alpha\beta(\theta^2 - d_E^2(\mathbf{x}, \mathbf{y})). \end{aligned}$$

To secure the biometric templates, we further add different randomnesses (i.e., α, β, r_x and r_y) to the extended vectors as shown in Protocol 2. The rest of the encryption procedures is then the same as those in TPE.Enc and TPE.TokenGen.

As presented in Protocol 2, during the registration phase, an end-user encrypts his template \mathbf{x} as C_x and registers C_x along with her identity at the service provider. During the query phase, the end-user encrypts a freshly generated template \mathbf{y} as T_y and sends T_y to the service provider. Then the service provider runs TPE.Dec with inputs C_x and T_y and outputs an authentication result. The correctness of this scheme is guaranteed by Theorem 1. Specifically, we have $I = \mathbf{x}' \circ \mathbf{y}' = \alpha\beta(\theta^2 - d_E^2(\mathbf{x}, \mathbf{y}))$. As a result, $\Lambda = \text{Authenticated}$ if and only if $d_E(\mathbf{x}, \mathbf{y}) \leq \theta$.

Remark 1: For any random pair \mathbf{x} and \mathbf{y} , the probability for $\text{dist}_E(\mathbf{x}, \mathbf{y}) \leq \theta^2$ is $(\theta/nR)^n$, where R is the upper bound of x_i and y_i , $i = 1, \dots, n$, and n is the dimension of the \mathbf{x} and \mathbf{y} .

This remark shows the probability of false positive rate. As an example, in iris authentication, in general $\theta/n \leq 0.3$ and $n = 1024$. In this case, for any pair of randomly selected templates, the likelihood for their distance to be less than θ^2 is $(\theta/nR)^n \leq 0.3^{1024} \approx 3.7 \times 10^{-536}$. In fact, even for $\theta/n = 0.5$ and $n = 400$, the $(\theta/nR)^n \leq 0.5^{400} \approx 3.8 \times 10^{-121}$, which is negligible.

D. Distance in Hamming Space

From the construction of Euclidean distance, we know that the critical part in computing the distance through inner product lies in proper design of the extended vectors. Thus, in the following, we will focus on how to design the vectors in order to compute different distances.

Hamming distance is a popular metric to measure the similarity of binary template such as Iris. Now, we assume the registered template and query template are $\mathbf{x} = (x_1, x_2, \dots, x_n)$ and $\mathbf{y} = (y_1, y_2, \dots, y_n)$ respectively, where x_i and y_i are 0 or 1. To calculate the Hamming distance $d_H(\mathbf{x}, \mathbf{y})$ between \mathbf{x} and \mathbf{y} , we first map the 0's in \mathbf{x} and \mathbf{y} to -1 and map 1's

Protocol 2 Privacy Preserving Biometric Authentication

Input: $\mathbf{x} = \{x_1, \dots, x_n\}, \mathbf{y} = \{y_1, \dots, y_n\}, \theta$.

Output: $\Lambda = \{\text{Denied}, \text{Authenticated}\}$.

Setup (End-user U):

- 1: Set the public parameters as $param = \{n, \theta\}$.
- 2: Randomly generate two matrices M_1 and M_2 with dimension $(n+5) \times (n+5)$ and a permutation $\pi : \mathbb{R}^{n+5} \rightarrow \mathbb{R}^{n+5}$.
- 3: Set secret key $sk = \{M_1, M_2, M_1^{-1}, M_2^{-1}, \pi\}$.

Registration (End-user U):

- 1: Generate random numbers β and r_x . Extend \mathbf{x} to an $(n+5)$ -dimensional vector $\mathbf{x}' = (2\beta x_1, 2\beta x_2, \dots, 2\beta x_n, -\beta \sum_{i=1}^n x_i^2, \beta, \beta\theta^2, r_x, 0)$.
- 2: Permute \mathbf{x}' to obtain $\mathbf{x}'' = \pi(\mathbf{x}')$.
- 3: Transform \mathbf{x}'' to a diagonal matrices X with \mathbf{x}'' being the diagonal.
- 4: Generate a random $(n+5) \times (n+5)$ lower triangular matrix S_x with the diagonal entries fixed as 1. Compute $C_x = M_1 S_x X M_2$.
- 5: Register the record $\langle ID_U, C_x \rangle$ to the service provider SP , where ID_U is the identity of end-user U .

Query (End-user U):

- 1: Generate random numbers α and r_y .
- 2: Extend \mathbf{y} to an $(n+5)$ -dimensional vector $\mathbf{y}' = (\alpha y_1, \alpha y_2, \dots, \alpha y_n, \alpha, -\alpha \sum_{i=1}^n y_i^2, \alpha, 0, r_y)$.
- 3: Permute \mathbf{y}' to obtain $\mathbf{y}'' = \pi(\mathbf{y}')$.
- 4: Transform \mathbf{y}'' to a diagonal matrix Y with diagonal being \mathbf{y}'' .
- 5: Generate a random $(n+5) \times (n+5)$ lower triangular matrix S_y with the diagonal entries fixed as 1. Compute $T_y = M_2^{-1} Y S_y M_1^{-1}$.
- 6: Send the query $\langle ID_U, T_y \rangle$ to SP .

Authentication (Service Provider SP):

- 1: On receiving a query from the end-user U , retrieve the registered record according to ID_U .
 - 2: Compute $I = \text{Tr}(C_x T_y)$.
 - 3: Set $\Lambda = \text{Authenticated}$ if $I \geq 0$; otherwise set $\Lambda = \text{Denied}$.
-

to 1. Then we have

$$2d_H(\mathbf{x}, \mathbf{y}) = n - \mathbf{x} \circ \mathbf{y}.$$

The condition $d_H(\mathbf{x}, \mathbf{y}) - \theta \leq 0$ is equivalent to $\mathbf{x} \circ \mathbf{y} + 2\theta - n \geq 0$. Thus, we need to design vectors \mathbf{x}' and \mathbf{y}' such that $\mathbf{x} \circ \mathbf{y} + 2\theta - n$ can be represented as $\mathbf{x}' \circ \mathbf{y}'$. In light of this, we let $\mathbf{x}' = (\beta x_1, \beta x_2, \dots, \beta x_n, \beta(2\theta - n), r_x, 0)$ and $\mathbf{y}' = (\alpha y_1, \alpha y_2, \dots, \alpha y_n, \alpha, 0, r_y)$. Then the rest of the authentication process is similarly as in Protocol 2. The correctness of this scheme is guaranteed by Theorem 1. Specifically, we have $I = \mathbf{x}' \circ \mathbf{y}' = 2\alpha\beta(\theta - d_H(\mathbf{x}, \mathbf{y}))$. As a result, $\Lambda = \text{Authenticated}$ if and only if $d_H(\mathbf{x}, \mathbf{y}) \leq \theta$.

In fact, the Hamming distance between two binary vectors is just one specific distance metric. There are many other different metrics such as Minkowski distance, Sokal & Michener similarity and Sokal & Sneath-II [31] introduced for different

applications. Using our proposed TPE scheme, we are able to evaluate such metrics and compare them to a pre-defined threshold. The critical part is to properly design the vectors \mathbf{x}' and \mathbf{y}' given two binary vectors \mathbf{x} and \mathbf{y} .

VI. SECURITY ANALYSIS

In this section, we analyze the security of PassBio under both passive attack and active attack as defined in Section III. PassBio is designed so that the service provider is unable to learn any critical information about the registered and query templates other than what is already revealed by the decryption process, given an encrypted registered template and a sequence of encrypted query templates.

Since PassBio is based on our proposed TPE, we will focus on the security analysis of TPE in the following discussion. An important difference between TPE and some traditional symmetric encryption schemes is that it is the service provider (could be malicious) that carries out the decryption process. The decryption process will reveal whether the inner product is within a threshold or not. Therefore, in the security analysis of TPE, it is necessary to analyze the security of both the encryption and decryption process, which will be discussed separately in the following sections.

A. Encryption Security

We first give a sketch of encryption security analysis. We will first utilize two experiments to model the ability of the adversary in passive attack and active attack, respectively. Then, we define the security of TPE under both passive and active attacks. At last, we prove the security of TPE under active attack since it implies the security under passive attack.

1) *Security Against Passive Attack*: In our scenario, the passive attack corresponds to the ciphertext-only-attack [17], where an adversary \mathcal{A} observes a sequence of ciphertext. We define an experiment $\text{Passive}_{\mathcal{A},\text{TPE}}^{\text{mult}}(\lambda)$ to simulate passive attacks, where the superscript **mult** denotes that the adversary \mathcal{A} is able to submit multiply messages instead of one single message.

Experiment 1 Passive Attack Experiment $\text{Passive}_{\mathcal{A},\text{TPE}}^{\text{mult}}(\lambda)$:

- 1: Given a security parameter λ , the adversary \mathcal{A} outputs two sequences of messages $M_0 = (m_{0,1}, m_{0,2}, \dots, m_{0,t})$ and $M_1 = (m_{1,1}, m_{1,2}, \dots, m_{1,t})$, where the length of each message $|m_{0,i}| = |m_{1,i}|, i = 1, 2, \dots, t$.
 - 2: The challenger \mathcal{C} runs $\text{TPE.KeyGen}(\lambda)$ to generate the secret key sk .
 - 3: \mathcal{C} chooses a uniform bit $b \in \{0, 1\}$ and computes the ciphertext $c_i = \text{TPE.TokenGen}(m_{b,i}, sk)$. The sequence $C = (c_1, c_2, \dots, c_t)$ is returned to \mathcal{A} .
 - 4: The adversary \mathcal{A} outputs a bit b' .
 - 5: The output of the experiment is 1 if $b = b'$, and 0 otherwise.
-

Based on $\text{Passive}_{\mathcal{A},\text{TPE}}^{\text{mult}}(\lambda)$, we now define the security of TPE under passive attack.

Definition 1: The proposed TPE scheme is secure against passive attack if for all polynomial-time adversary \mathcal{A} , there is

a negligible function negl such that the probability

$$\left| \Pr(\text{Passive}_{\mathcal{A},\text{TPE}}^{\text{mult}}(\lambda) = 1) - \frac{1}{2} \right| \leq \text{negl}(\lambda).$$

Remark 2: In the above security definition, we only use the token generation function TPE.TokenGen as a representative. This is because the operations involved in TPE.Enc and TPE.TokenGen are almost the same. The security analysis for TPE.Token applies for TPE.Enc . However, in our security proof, we will show that both TPE.Enc and TPE.TokenGen meet the security requirement.

Based on Definition 1, we have the following theorem.

Theorem 2: The proposed TPE scheme is secure against passive attack.

Remark 3: We will skip the proof of Theorem 2. Instead, we will focus on the proof of Theorem 3, which states that TPE is secure against active attack. This is because an adversary in an active attack has more information than an adversary in a passive attack. In particular, an active attacker has continuous oracle access to $\text{TPE.TokenGen}(sk, \cdot)$ as defined in the active attack experiment. As a result, the proof for Theorem 3 is also applicable for Theorem 2.

2) *Security Against Active Attack*: Under the active attack, the service provider is able to observe a sequence of pairs of query templates as well as their encrypted version. This can happen when, for example, some adversaries submit their templates and collude with the service provider. This attack scenario corresponds to the Chosen-Plaintext-Attack (CPA) in cryptography. Accordingly, an encryption scheme has CPA-security if it is secure against CPA. To prove that TPE has CPA-security, we model the active attack using and experiment $\text{Active}_{\mathcal{A},\text{TPE}}(\lambda)$. We define CPA-security for TPE as follows.

Experiment 2 Active Attack Experiment $\text{Active}_{\mathcal{A},\text{TPE}}(\lambda)$:

- 1: The function $\text{TPE.KeyGen}(\lambda)$ generates a secret key sk .
 - 2: The adversary \mathcal{A} is given oracle access to the function $\text{TPE.TokenGen}(sk, \cdot)$ and outputs two messages m_0 and m_1 of the same length to the challenger \mathcal{C} .
 - 3: The challenger \mathcal{C} chooses a uniform bit $b \in \{0, 1\}$, then computes $c = \text{TPE.TokenGen}(sk, m_b)$ and returns to \mathcal{A} .
 - 4: \mathcal{A} continues to have oracle access to $\text{TPE.TokenGen}(sk, \cdot)$ and outputs a bit b' . Note however, \mathcal{A} cannot use $\text{TPE.TokenGen}(sk, \cdot)$ to generate tokens for messages somehow related to m_0 and m_1 .
 - 5: The output of the experiment is 1 if $b = b'$, and 0 otherwise.
-

Definition 2: The proposed TPE is secure against active attack if for all polynomial-time adversary \mathcal{A} , there is a negligible function negl such that the probability

$$\left| \Pr(\text{Active}_{\mathcal{A},\text{TPE}}(\lambda) = 1) - \frac{1}{2} \right| \leq \text{negl}(\lambda).$$

Remark 4: Different from the passive attack experiment, the adversary will continually have oracle access to the token generation function. This models the situation where the adversary is able to observe multiple pairs of messages and their ciphertexts.

As a result, in the active attack, what an adversary can observe through decryption is a series of results $I_i = \alpha_i \beta(\mathbf{x} \circ \mathbf{y}_i - \theta)$. Since α_i are selected independently, the final results I_i only reveals whether $\alpha_i \beta(\mathbf{x} \circ \mathbf{y} - \theta)$ is positive or not. No more key information can be derive from I_i .

C. The Effect of Randomness on Security

Besides the randomly generated long-time keys (i.e., M_1 , M_2 and π), we also introduce different one-time randomness in the encryption scheme. At the high-level view, the one-time randomness provides TPE with CPA-security similar to that of the one-time pad. From a cryptographic point of view, the one-time pad encryption scheme provides perfect security. However, it is not practical since the one-time secret key has the same length as the message itself. The most notable difference between TPE and the traditional encryption schemes is that TPE actually does not decrypt the message. Instead, TPE evaluate a function of the ciphertext in order to obtain the function value of the plaintext. As a result, TPE does not require the one-time randomness in the decryption process. In this sense, TPE can achieve the security comparable to the one-time pad while avoiding the impractical key management requirement.

It is important to understand the effect of different randomness on security. We briefly categorize the one-time randomness utilized by TPE into three types.

- Type I: *result-disguising randomness*. When extending the vectors in both TPE.Enc and TPE.TokenGen, we use random β and α respectively to multiply with each entry of \mathbf{x} and \mathbf{y} . Since α and β will remain in the decryption result, we name it as result-disguising randomness.
- Type II: *vector-extension randomness*. In both TPE.Enc and TPE.TokenGen, we extend the vector and pad it with a random r .
- Type III: *matrix-multiplication randomness*. In both TPE.Enc and TPE.TokenGen, we multiply the extended matrices (X and Y) with random matrices (S_x and S_y).

These one-time randomnesses together ensure the CPA-security of the encryption process of TPE as analyzed in Section VI-A. The main function of decryption is to evaluate the trace of the matrix. We note that the trace function will cancel Type II and Type III randomness. However, Type I randomness will remain in the decryption result. This is important since it will only reveal partial information of the plaintext, which is just adequate for the purpose of biometric authentication. We will further demonstrate the effect of Type I randomness in Section VII-A.

VII. OTHER APPLICATIONS OF TPE

Our proposed threshold predicate encryption scheme enables a compute-then-compare computational model over encryption data. That is, given two encrypted vector \mathbf{x} and \mathbf{y} , an untrusted party is able to determine whether the inner product of \mathbf{x} and \mathbf{y} is greater than or within a threshold θ . No other key information about the value of \mathbf{x} , \mathbf{y} or $\mathbf{x} \circ \mathbf{y}$ is exposed. Previously, we also showed that utilizing the inner product of \mathbf{x} and \mathbf{y} , we are able to compute many distance

and similarity metrics. Such properties of TPE are critical for many applications that require data security and privacy.

A. Improved Security for Outsourced Biometric Identification

Outsourcing of different computational problems to the cloud while preserving the security and privacy of the outsourced problem has becoming a new trend. Many previous works have considered secure outsourcing of different problems [13], [14], [18]–[21]. In [14], a secure outsourcing scheme is proposed for biometric identification. The system models of outsourced biometric identification and biometric authentication are fundamentally different. In outsourced biometric identification, a data owner possesses a database of users' biometric templates. The goal of biometric identification is that given a query template, the data owner needs to identify a user to whom the query template belongs to.

To protect the security and privacy of biometric templates, [14] proposed an outsourcing scheme where the database owner will first encrypt the templates and then outsource the encrypted data to the cloud. Specifically, the data owner encrypts a biometric template \mathbf{x} as C_x using a symmetric key sk . For a given query template \mathbf{z} , it is also encrypted as C_z using the same key sk . The scheme is designed in such a manner that given two encrypted templates C_x and C_y and a query template C_z , the cloud is able to determine which template (\mathbf{x} or \mathbf{y}) is closer to \mathbf{z} , without learning any key information about \mathbf{x} , \mathbf{z} and \mathbf{y} . By repeating this process, the cloud is able to identify the template \mathbf{x} that is closest to \mathbf{z} . Then the encrypted version C_x is returned to the data owner, who can decrypt C_x to obtain \mathbf{x} and calculate the actual distance between \mathbf{x} and \mathbf{z} . Thus, the data owner can finally decide whether \mathbf{x} and \mathbf{y} are close enough such that they belong to the same person.

There are mainly two security and privacy issues regarding the above scheme. First, the registration phase is vulnerable to the registration attack [22], since an adversary (i.e., the cloud) is able to inject known templates into the database. During decryption, the cloud is able to derive the following equation (i.e., [22, eq. 3]):

$$b_{ci} = \frac{(\text{Tr}(Y'_i B'_c) - \text{Tr}(X'_i B'_c)) - (y_{i(n+1)} - x_{i(n+1)})}{y_{ii} - x_{ii}},$$

where b_{ci} is the i -th entry in a submitted query template \mathbf{b}_c . Since $\text{Tr}(Y'_i B'_c)$ and $\text{Tr}(X'_i B'_c)$ are computable and \mathbf{x} and \mathbf{y} are selected by the cloud, the cloud is able to recover b_{ci} . Repeating such attack will finally recover the whole query template \mathbf{b}_c as demonstrated in [22].

Second, from the decryption result, the cloud is able to learn more information than needed. In particular, the cloud is able to determine which one of any two encrypted templates is closer to the query template. By repeating this process, the cloud can actually rank all the templates by their distances to the query template. This unnecessarily reveals more information than what is needed in biometric identification.

We now show that our proposed TPE scheme can address these two issues. The security vulnerability of the scheme in [14] was caused due to lacking of Type I randomness as

defined in Section VI-C. The trace function $\text{Tr}(\cdot)$ will cancel the Type III randomness, resulting in [22, eq. 3].

Our proposed TPE scheme can be directly utilized in outsourced biometric identification. In the encryption part, each registered template \mathbf{x} is encrypted with TPE.Enc . A query template \mathbf{z} is encrypted with TPE.TokenGen . The decryption process will give $\alpha_z \beta_x (\text{dist}^2(\mathbf{x}, \mathbf{z}) - \theta^2)$, where α_z and β_x are one-time random numbers associated with \mathbf{z} and \mathbf{x} respectively. As a result, [22, eq. 3] is replaced by

$$b_{ci} = \frac{(\text{Tr}(PB'_c) - \text{Tr}(QB'_c)) - (p_{n+1} - q_{n+1})}{\alpha_c \beta_x (p_n - q_n)}.$$

Note that α_c is a one-time random number associated with a query b_c and β_x is a one-time random number associated with \mathbf{x} . Thus, although the adversary is able to insert known templates into the database, it cannot derive b_{ci} due to the one-time randomness. In other words, the outsourced biometric identification scheme based on TPE is able to defend against registration attack.

For the second privacy issue, the decryption result $\alpha_z \beta_x (\text{dist}^2(\mathbf{x}, \mathbf{z}) - \theta)$ will only reveal whether the distance between the query \mathbf{z} and the registered template \mathbf{x} is within a threshold or not. Since β_x is a one-time randomness associated with each registered template \mathbf{x} , the relative distance information is concealed. As a result, the cloud is not able to rank all the registered templates according to the distance to the query template.

B. Searching Over Encrypted Data

With the development of cloud computing and storage, there is a clear motivation for searching over encrypted data [27]–[30]. For example, a medical institution may store its medical data in the cloud. To ensure security of the data, the institution chooses to encrypt all the data before outsourcing. Meanwhile, the institution wishes to maintain the searching ability over the encrypted data in order to retrieve the desired data files. The proposed TPE is a promising solution for searching over encrypted data. In the following, we discuss how to utilize TPE to implement different searching functionalities.

1) *Set Intersection*: We assume that a file F_i is indexed by a set of keywords S_i . The files and their associated keyword sets are encrypted and outsourced to the cloud. A search query consists of a set of keywords S_j . Given the search query, the cloud will return the file F_i if the overlap of keyword sets S_i and S_j exceeds a certain threshold θ . That is $|S_i \cap S_j| > \theta$.

The above set intersection search function can be implemented through TPE as follows. Suppose the universe of keywords is the set S with size n . Fix the order of the keywords within S . Then, an index S_i can be formulated as an n -dimensional binary vector \mathbf{x}_i , where $x_i^t = 1$ means that the t -th keyword in S appears in S_i . The vector \mathbf{x}_i for file F_i is encrypted using TPE.Enc . Each file is then encrypted using standard symmetric encryption schemes such as AES. The encrypted files and index are outsourced to the cloud. For a search query S_j , a vector \mathbf{x}_j can be formulated in a similar manner. Then a search token can be generated

using TPE.TokenGen . With this formulation, it is obvious that $|S_i \cap S_j| = \mathbf{x}_i \circ \mathbf{x}_j$, where $\mathbf{x}_i \circ \mathbf{x}_j$ denotes the inner product of \mathbf{x}_i and \mathbf{x}_j . With TPE, the cloud is able to identify the files whose associated indices satisfy $\mathbf{x}_i \circ \mathbf{x}_j > \theta$ while not learning any useful information about the indices.

2) *Weighted Sum Evaluation*: For many numeric data, it is significant to evaluate the weighted sum of the data record with different weights. For example, the grades of each subject for a student form a vector G_i . An evaluator wants to evaluate the performance of the students via some criteria. Each criterion can be formulated as the weighed sum of the grades. The different weights reflects different emphasis on the subjects.

We assume that an administrator possess the grades for all the students. For privacy issues, all the grades are encrypted using TPE.Enc and stored in an external server. An evaluator desires to identify those students whose performance meets certain standard. In this scenario, the evaluator can submit a vector of weights W_j to the administrator, who will then generate a search token for the evaluator through TPE.TokenGen . The evaluator can submit the token generated for W_j to the sever and search over the encrypted grades. The server is then able to identify the students whose grades satisfy $G_i \circ W_j > \theta$.

VIII. PERFORMANCE EVALUATION

In this section, we evaluate the performance of PassBio. First, we give detailed analysis of both computational and communication complexity. Then, some numeric results are presented for the proposed TPE through simulation.

A. Complexity Analysis

As shown in Protocol 2, at local side an end-user needs to run the TPE.KeyGen , TPE.Enc and TPE.TokenGen algorithms. The service provider needs to run the TPE.Dec algorithm for every query. It is obvious that the computational bottleneck of these algorithms lies in matrix multiplication or matrix inversion. Thus, in the following analysis, we will focus on matrix multiplication and inversion. Without loss of generality, we assume that the matrices involved in the computation all have the same dimension $n \times n$.

For the function TPE.KeyGen , two random matrices are generated and two matrix inversions need to be calculated. Note that the setup phase is generally a one-time process. That is, TPE.KeyGen needs to be executed by the end-user only once. The function TPE.Enc and TPE.TokenGen will both take 3 matrix multiplications. As a result, they have a complexity of $\mathcal{O}(n^3)$, without optimization for matrix multiplication.

In the function TPE.Dec , the trace of $C_x T_y$ needs to be computed. There is no need to calculate the matrix multiplication before evaluating the trace. Only computing of the diagonal entries is needed. Thus, TPE.Dec has a complexity of $\mathcal{O}(n^2)$.

In terms of communication overhead, assume all the matrix or vector has the same size l . In the registration phase, the end-user needs to submit the encrypted template C_x to the service provider. Thus the communication overhead for registration is $n^2 l$. Similarly, the communication overhead for the query phase is also $n^2 l$.

B. Efficiency Improvement

The above complexity analysis shows that the computational bottleneck of both TPE.Enc and TPE.TokenGen lie in matrix multiplication. For resource-constrained devices such as mobile phones, the computation of matrix multiplication with high dimensions is still expensive, if not impossible. In the following, we will introduce two typical techniques that can reduce the computational overhead for mobile devices.

1) *Dimension Reduction*: The complexity of normal matrix multiplication is $\mathcal{O}(n^3)$, where n is the dimension of the matrices. Thus, a straight forward way to reduce the complexity is to reduce the dimension of the matrices. For applications such as biometric authentication and identification, it is critical to preserve the identification accuracy while reducing the dimension. Several works [32]–[34] have been devoted to reducing the sizes of biometric templates. In [32], two techniques are introduced to decimate the FingerCode representation. The *tesselation reduction* approach reduces the dimension of FingerCode from the feature generation phase, which is illustrated in Section V-A. Specifically, given a fingerprint image, this approach will reduce the number of sectors of the tessellation. The other approach is to directly apply some general dimension reduction methods such as PCA to the obtained FingerCodes. In this way, the most compact representation of FingerCode is found for a specific dataset.

We note that the above two approaches will both degrade the identification accuracy, however, to a satisfying level. In the experiments [32], the length of FingerCode vary from 640 to 8 in the tessellation approach. For PCA approach, the dimension of FingerCode varies from 64 to 4. Generally speaking, the shorter the FingerCode is, the worse the accuracy would be. However, the experimental result demonstrated that FingerCode of dimensions 96 (from tessellation reduction) and 8 (from PCA) can achieve a satisfactory accuracy compared to that of the original 640. We also note that the approaches in [32] quantized each entry in FingerCode resulting a reduced accuracy. However, our proposed TPE scheme can be directly utilized to real numbers. Thus, TPE is applicable to the non-quantized case in [32], which has a higher accuracy.

2) *Online/Offline Computation*: The idea of online/offline computation [35]–[37] is to divide a computational expensive process into an online phase and an offline phase. During the offline phase, some pre-computation is done without given the input. During the online phase, given the input, it is relatively easy to padding the offline computation result in order to generate the final result. Typically, the offline computation is carried out when the mobile devices are idle or getting charged. Thus, such an approach can reduce the overall responding time and battery consumption.

Our proposed TPE scheme can utilize such approach to reduce the online computational overhead. For example, in the query phase, an end-user needs to compute $M_2^{-1}YS_yM_1^{-1}$ given a transformed template Y . Then during the offline phase, the end-user can generate the random matrix S_y and compute $S_yM_1^{-1}$. The computation results can be stored for later usage. When a fresh template Y is generated, the end-user can compute $M_2^{-1}YS_yM_1^{-1}$ during the online phase.

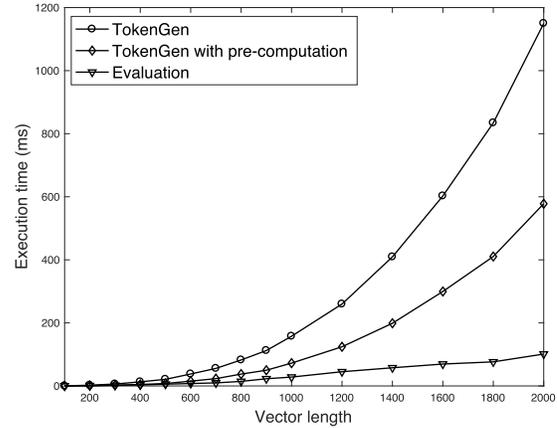


Fig. 2. Performance of token generation and evaluation simulated on laptop (with vs. without pre-computation).

This approach can reduce half of the computational overhead, which is critical for resource-constrained devices.

C. Numeric Results

In this section, we measure the performance of our proposed TPE scheme through simulation. Since the functions TPE.Setup and TPE.KeyGen are both one-time processes during the registration phase, we mainly focus on the execution time of TPE.TokenGen.

Since PassBio is a user-centric biometric authentication scheme, we measure the performance on both mobile phone and personal laptop. In the simulation, we utilize a mobile phone with Android 6.0 operating system, 2.5 GHz Cortex-A72 CPU and 4 GB RAM. We also utilize a personal laptop with macOS 10, 1.6 GHz Intel Core i5 and 4 GB RAM. The java library UJMP [38] and C++ library Armadillo are utilized for the simulation in the mobile phone and personal computer, respectively. We note that the performance relies on the selection of software packages. Our selection does not guarantee the best performance. Through complexity analysis, we know that the most important parameter affecting the performance is the dimension n of the vector. For the simulation on the mobile phone and laptop, we let n vary from 10 to 300 and from 100 to 2000, respectively. Due to the dimension reduction techniques introduced in Section VIII-B.1, the dimension $n = 300$ is sufficient for most of the biometric templates. We also utilize the online/offline computation mechanism introduced in Section VIII-B.2 to reduce to online computational overhead.

The numeric result on the laptop is shown in Fig. 2. The token generation time for moderate size template (n is around 200) is just around one millisecond with pre-computation. For high-dimensional template with $n = 2000$, the token generation time is less than 1 second with pre-computation. The numeric result on the mobile phone is shown in Fig. 3. The simulation results show that it is efficient to generate tokens for templates with moderate size. For example, when $n = 100$, the generation time is approximately 50 ms. When $n = 300$, the generation time is around 900 ms. It can be observed in both figures that the online/offline mechanism can effectively reduce the online computational overhead.

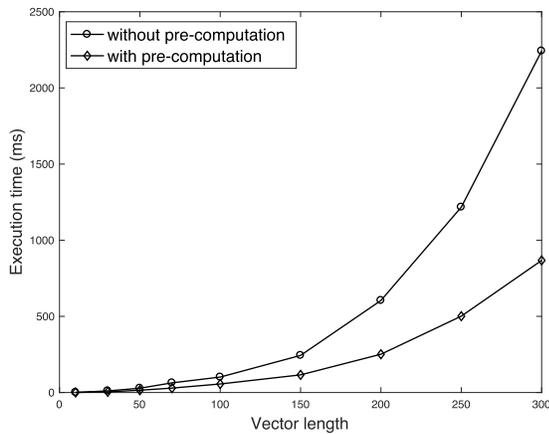


Fig. 3. Performance of token generation and evaluation on mobile phone (with vs. without pre-computation).

By pre-computation during the offline phase, the online computation time is reduced to about half of the whole processing time.

IX. CONCLUSION

In this paper, we proposed a Threshold Predicate Encryption (TPE) scheme. TPE is able to encrypt a vector \mathbf{x} and generate a token for another vector \mathbf{y} . Given the two encrypted vectors, any party is able to determine whether the inner product of \mathbf{x} and \mathbf{y} is within a pre-defined threshold or not. Our security analysis shows that no sensitive information about the vectors can be learned by the untrusted party under both passive and active attacks. Based on TPE, we proposed PassBio, a privacy preserving user-centric biometric authentication scheme. One key feature of PassBio is that end-users can encrypt their own biometric template and register it to the service provider. Then the end-user is able to encrypt their freshly generated template and submit them to the service provider for authentication usage. We show that the TPE is suitable for a compute-then-compare computational model on encrypted data. Such a computational model can be widely used in many applications requiring computations on encrypted data while preserving the data security and privacy. In particular, we presented two additional applications of TPE, searching over encrypted data and outsourced biometric identification. Our simulation results demonstrated that the proposed TPE can be efficiently implemented on both mobile phones and personal laptops.

REFERENCES

- [1] A. K. Jain, K. Nandakumar, and A. Ross, "50 years of biometric research: Accomplishments, challenges, and opportunities," *Pattern Recognit. Lett.*, vol. 79, pp. 80–105, Aug. 2016.
- [2] A. K. Jain and K. Nandakumar, "Biometric authentication: System security and user privacy," *Computer*, vol. 45, no. 11, pp. 87–92, Nov. 2012.
- [3] S. Rane, Y. Wang, S. C. Draper, and P. Ishwar, "Secure biometrics: Concepts, authentication architectures, and challenges," *IEEE Signal Process. Mag.*, vol. 30, no. 5, pp. 51–64, Sep. 2013.
- [4] S. Kim, K. Lewi, A. Mandal, H. W. Montgomery, A. Roy, and D. Wu, "Function-hiding inner product encryption is practical," IACR Cryptol. ePrint Archive, NV, USA, Tech. Rep. 20160507:001353, 2016, pp. 440–456.
- [5] M. Abdalla, F. Bourse, A. De Caro, and D. Pointcheval, "Simple functional encryption schemes for inner products," in *Proc. IACR Int. Workshop Public Key Cryptogr.* NV, USA: Springer, 2015, pp. 733–751.
- [6] A. Bishop, A. Jain, and L. Kowalczyk, "Function-hiding inner product encryption," in *Proc. Int. Conf. Theory Appl. Cryptol. Inf. Secur.* Berlin, Germany: Springer, 2015, pp. 470–491.
- [7] P. Datta, R. Dutta, and S. Mukhopadhyay, "Functional encryption for inner product with full function privacy," in *Public-Key Cryptography—PKC*. Berlin, Germany: Springer, 2016, pp. 164–195.
- [8] J. Katz, A. Sahai, and B. Waters, "Predicate encryption supporting disjunctions, polynomial equations, and inner products," in *Advances in Cryptology—EUROCRYPT*. Berlin, Germany: Springer, 2018, pp. 146–162.
- [9] E. Shen, E. Shi, and B. Waters, "Predicate privacy in encryption systems," in *Proc. TCC*. Berlin, Germany: Springer, vol. 5444, 2009, pp. 457–473.
- [10] N. P. Smart and F. Vercauteren, "Fully homomorphic encryption with relatively small key and ciphertext sizes," in *Public-Key Cryptography—PKC*, vol. 6056. Berlin, Germany: Springer, 2010, pp. 420–443.
- [11] M. Naehrig, K. Lauter, and V. Vaikuntanathan, "Can homomorphic encryption be practical?" in *Proc. 3rd ACM Workshop Cloud Comput. Secur. Workshop*, 2011, pp. 113–124.
- [12] W. K. Wong, D. W.-L. Cheung, B. Kao, and N. Mamoulis, "Secure KNN computation on encrypted databases," in *Proc. ACM SIGMOD Int. Conf. Manage. Data*, 2009, pp. 139–152.
- [13] J. Yuan and S. Yu, "Efficient privacy-preserving biometric identification in cloud computing," in *Proc. IEEE INFOCOM*, Apr. 2013, pp. 2652–2660.
- [14] Q. Wang, S. Hu, K. Ren, M. He, M. Du, and Z. Wang, "CloudBI: Practical privacy-preserving outsourcing of biometric identification in the cloud," in *Proc. Eur. Symp. Res. Comput. Secur.* Berlin, Germany: Springer, 2015, pp. 186–205.
- [15] D. Boneh, A. Sahai, and B. Waters, "Functional encryption: Definitions and challenges," in *Theory of Cryptography*. Berlin, Germany: Springer, 2011, pp. 253–273.
- [16] A. Lewko, T. Okamoto, A. Sahai, K. Takashima, and B. Waters, "Fully secure functional encryption: Attribute-based encryption and (hierarchical) inner product encryption," in *Advances in Cryptology—EUROCRYPT*, vol. 6110. Berlin, Germany: Springer, 2010, pp. 62–91.
- [17] J. Katz and Y. Lindell, *Introduction to Modern Cryptography*. Boca Raton, FL, USA: CRC Press, 2014.
- [18] K. Zhou and J. Ren, "Secure outsourcing of scalar multiplication on elliptic curves," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2016, pp. 1–5.
- [19] K. Zhou and J. Ren, "LinSOS: Secure outsourcing of linear computations based on affine mapping," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2016, pp. 1–5.
- [20] K. Zhou, M. H. Afifi, and J. Ren, "ExpSOS: Secure and verifiable outsourcing of exponentiation operations for mobile cloud computing," *IEEE Trans. Inf. Forensics Security*, vol. 12, no. 11, pp. 2518–2531, Nov. 2017.
- [21] K. Zhou and J. Ren, "Secure fine-grained access control of mobile user data through untrusted cloud," in *Proc. 25th Int. Conf. Comput. Commun. Netw. (ICCCN)*, Aug. 2016, pp. 1–9.
- [22] C. Hahn and J. Hur, "POSTER: Towards privacy-preserving biometric identification in cloud computing," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, 2016, pp. 1826–1828.
- [23] S. Goldwasser *et al.*, "Multi-input functional encryption," in *Proc. Annu. Int. Conf. Theory Appl. Cryptogr. Techn.* Berlin, Germany: Springer, 2014, pp. 578–602.
- [24] A. K. Jain, S. Prabhakar, L. Hong, and S. Pankanti, "FingerCode: A filterbank for fingerprint representation and matching," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2, Jun. 1999, pp. 187–193.
- [25] A. K. Jain, S. Prabhakar, L. Hong, and S. Pankanti, "Filterbank-based fingerprint matching," *IEEE Trans. Image Process.*, vol. 9, no. 5, pp. 846–859, May 2000.
- [26] A. K. Jain, S. Prabhakar, and L. Hong, "A multichannel approach to fingerprint classification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 4, pp. 348–359, Apr. 1999.
- [27] D. X. Song, D. Wagner, and A. Perrig, "Practical techniques for searches on encrypted data," in *Proc. IEEE Symp. Secur. Privacy*, May 2000, pp. 44–55.

- [28] E. Shi, J. Bethencourt, T.-H. H. Chan, D. Song, and A. Perrig, "Multi-dimensional range query over encrypted data," in *Proc. IEEE Symp. Secur. Privacy*, May 2007, pp. 350–364.
- [29] D. Boneh and B. Waters, "Conjunctive, subset, and range queries on encrypted data," in *Theory of Cryptography*. Berlin, Germany: Springer, 2007, pp. 535–554.
- [30] S. Tu, M. F. Kaashoek, S. Madden, and N. Zeldovich, "Processing analytical queries over encrypted data," *VLDB Endowment*, vol. 6, no. 5, pp. 289–300, 2013.
- [31] S.-S. Choi, S.-H. Cha, and C. C. Tappert, "A survey of binary similarity and distance measures," *J. Syst., Cybern. Inform.*, vol. 8, no. 1, pp. 43–48, 2010.
- [32] T. Bianchi, S. Turchi, A. Piva, R. D. Labati, V. Piuri, and F. Scotti, "Implementing fingerprint-based identity matching in the encrypted domain," in *Proc. IEEE Workshop Biometric Meas. Syst. Secur. Med. Appl. (BIOMS)*, Sep. 2010, pp. 15–21.
- [33] N. Kambhatla and T. K. Leen, "Dimension reduction by local principal component analysis," *Neural Comput.*, vol. 9, no. 7, pp. 1493–1516, Jul. 2006.
- [34] T. Mandal, Q. M. J. Wu, and Y. Yuan, "Curvelet based face recognition via dimension reduction," *Signal Process.*, vol. 89, no. 12, pp. 2345–2353, Dec. 2009.
- [35] S. S. M. Chow, J. K. Liu, and J. Zhou, "Identity-based online/offline key encapsulation and encryption," in *Proc. 6th ACM Symp. Inf., Comput. Commun. Secur.*, Mar. 2011, pp. 52–60.
- [36] S. Hohenberger and B. Waters, "Online/offline attribute-based encryption," in *Proc. Int. Workshop Public Key Cryptogr.* Berlin, Germany: Springer, 2014, pp. 293–310.
- [37] J. K. Liu and J. Zhou, "An efficient identity-based online/offline encryption scheme," in *Proc. ACNS*, vol. 5536. Berlin, Germany: Springer, 2009, pp. 156–167.
- [38] *Universal Java Matrix Package*. Accessed: Oct. 15, 2017. [Online]. Available: <https://ujmp.org>
- [39] A. Jain, A. A. Ross, and K. Nandakumar, *Introduction to Biometrics*. Berlin, Germany: Springer, 2011.
- [40] M. Blanton and P. Gasti, "Secure and efficient protocols for iris and fingerprint identification," in *Proc. Eur. Symp. Res. Comput. Secur.* Berlin, Germany: Springer, 2011, pp. 190–209.
- [41] F. Bayatbabolghani, M. Blanton, M. Aliasgari, and M. Goodrich. (2017). "Secure fingerprint alignment and matching protocols." [Online]. Available: <https://arxiv.org/abs/1702.03379>
- [42] M. Barni *et al.*, "Privacy-preserving fingerprint authentication," in *Proc. 12th ACM Workshop Multimedia Secur.*, 2010, pp. 231–240.
- [43] J. Bringer, H. Chabanne, M. Izabachène, D. Pointcheval, Q. Tang, and S. Zimmer, "An application of the Goldwasser–Micali cryptosystem to biometric authentication," in *Proc. Australas. Conf. Inf. Secur. Privacy*. Berlin, Germany: Springer, 2007, pp. 96–106.
- [44] A. Abidin, K. Matsuura, and A. Mitrokovtsa, "Security of a privacy-preserving biometric authentication protocol revisited," in *Proc. Int. Conf. Cryptol. Netw. Secur.* Berlin, Germany: Springer, 2014, pp. 290–304.
- [45] S. Lang, *Linear Algebra*. New York, NY, USA: Springer-Verlag, 1987.



Kai Zhou (S'16) received the B.S. degree in electrical engineering from Shanghai Jiao Tong University, China, in 2013. He is currently pursuing the Ph.D. degree in electrical and computer engineering with Michigan State University. His research interests include applied cryptography, cloud security and privacy, coding theory, and secure communication.



Jian Ren (SM'09) received the B.S. and M.S. degrees in mathematics from Shaanxi Normal University, and the Ph.D. degree in electrical engineering from Xidian University, China. He is currently an Associate Professor with the Department of ECE, Michigan State University. His current research interests include network security, cloud computing security, privacy-preserving communications, distributed network storage, and Internet of Things. He was a recipient of the U.S. National Science Foundation Faculty Early Career Development (CAREER) Award in 2009. He is the TPC Chair of the IEEE ICNC'17 and a General Chair of ICNC'18.