PolyCosGraph: a Privacy-Preserving Cancelable EEG Biometric System

Min Wang, *Member, IEEE,* Song Wang, Jiankun Hu, *Senior Member, IEEE*

Abstract—Recent findings confirm that biometric templates derived from electroencephalography (EEG) signals contain sensitive information about registered users, such as age, gender, cognitive ability, mental status and health information. Existing privacy-preserving methods such as hash function and fuzzy commitment are not cancelable, where raw biometric features are vulnerable to hill-climbing attacks. To address this issue, we propose the PolyCosGraph, a system based on **Poly**nomial transformation embedding **Cos**ine functions with **Graph** features of EEG signals, which is a privacy-preserving and cancelable template design that protects EEG features and system security against multiple attacks. In addition, a template corrupting process is designed to further enhance the security of the system, and a corresponding matching algorithm is developed. Even when the transformed template is compromised, attackers cannot retrieve raw EEG features and the compromised template can be revoked. The proposed system achieves the authentication performance of 1.49% EER with a resting state protocol, 0.68% EER with a motor imagery task, and 0.46% EER under a watching movie condition, which is equivalent to that in the non-encrypted domain. Security analysis demonstrates that our system is resistant to attacks via record multiplicity, preimage attacks, hill-climbing attacks, second attacks and brute force attacks.

Index Terms—EEG biometrics, authentication, cancelable template, privacy-preserving.

1 Introduction

Brain biometrics based on electroencephalography (EEG) has attracted increasing attention from both academia and industry [1]. Compared with traditional biometric techniques based on fingerprint, face or iris, EEG biometrics offers additional advantages in terms of robustness against circumvention and intrinsic liveness detection [2]. First, the biosignals used for EEG biometrics are results of cerebral activities, which are internal traits not exposed to the public as face and fingerprint. Meanwhile, since many features of EEG signals are non-volitional (i.e., beyond control or conscious apprehensions of the user), the user cannot deliberately divulge their identifier, thus protecting the biometric system [3]. Furthermore, as EEG biometrics involves conscious engagement of the user, with current sensing technologies, it is highly unlikely to capture EEG signals covertly or remotely without the user's awareness. Being difficult to steal or forge makes EEG biometrics less prone to sensor spoofing attacks than exposed biometrics [4]. In addition, due to the nature of brain signals, EEG biometrics inherently supports liveness detection, which is an important aspect in enhancing the security of biometric systems against sensor spoofing [2]. Finally, the lack of brain activity is a clinical indicator of physical death. A person has to be alive in order to present EEG signals to the sensor at the time of capture, which protects users and reduces the possibility of presentation attacks using spoofing artifacts or lifeless body parts [3].

The typical architecture of an EEG biometric system consists of a signal acquisition module to collect data under specified signal induction protocols, a feature extraction module to compute discriminative features from raw data, and a template matching or classification module for decision making, as illustrated in Fig. 1. Template matching-based systems store a template (e.g., feature vector) for each user and make a decision to accept or reject a query by comparing the query template with the stored template of the claimed user [3], [5], [6], [7], [8]. In contrast, classifier-based systems train and store a classification model for each user during registration, and use this trained model to predict whether a query sample comes from the claimed user or not [2], [9], [10]. These two types of systems respectively require user templates or models to be stored in the authentication system. This question arises: *is it secure to directly store templates or models this way?* Relevant research indicates that the answer is *no*. Assuming that an attacker manages to break into the database and successfully steals user templates or models, this would pose a huge threat to user privacy.

EEG signals contain sensitive information about the user's age, gender [11], cognitive abilities with regard to learning and memory [12], mental states on cognitive workload [13] and emotion [14], as well as health condition, especially brain disorder [15]. A recent study further examined EEG templates (features) used in biometric applications and confirmed that personal characteristics regarding age and gender, as well as information related to medication intake and neurological disorders, can be inferred from the templates [16]. These findings highlight the need to apply privacy-preserving mechanisms to protect user templates when deploying EEG biometric systems [16]. The same conclusion applies to classifier-based EEG biometric sys-

- *Min Wang and Jiankun Hu are with the School of Engineering and Information Technology, University of New South Wales, Canberra, ACT 2612, Australia (e-mail: maggie.wang1@adfa.edu.au; j.hu@adfa.edu.au)*
- *Song Wang is with the School of Engineering and Mathematical Sciences, La Trobe University, VIC 3086, Australia (e-mail: song.wang@latrobe.edu.au)*

This work was supported by the Australian Research Council through the discovery grant DP200103207. (Corresponding author: Jiankun Hu)
Manuscript received xx xx, 2022; revised xx xx, 2022.

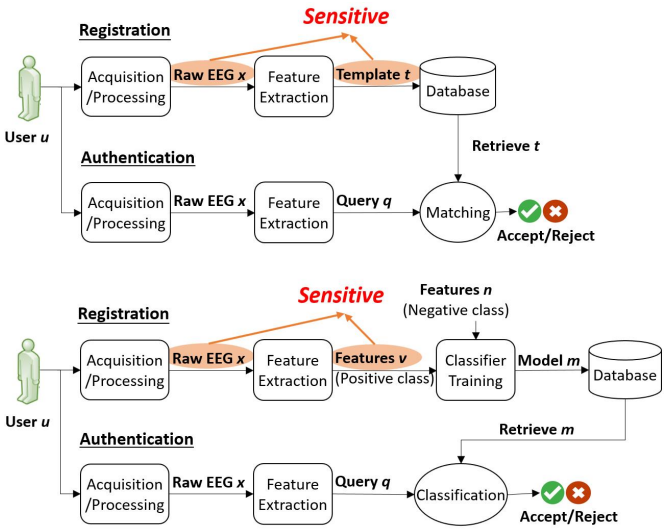

Fig. 1: Traditional EEG biometric systems, template matching-based systems (top) and classifier-based systems (bottom).

tems. Although classifier-based systems do not store user templates other than classification models, they can still reveal user-sensitive information. This is because a classifier is essentially a decision maker that takes a query (e.g., EEG signals or features) as input and outputs the probability of the query belonging to the genuine user (positive class). An adversary can run an evolutionary algorithm and use the probabilities produced by the classifier to generate a synthetic input that enables false acceptance [17]. Since the obtained synthetic input is an approximation of the raw EEG signals or features, the system is factually compromised and user information leaked. Such attacks most likely happen in a remote authentication environment, e.g., a wireless network, where the sensor has been bypassed.

So far, most studies on EEG biometrics have focused on the optimization of registration and authentication, such as signal acquisition protocols, feature extraction methods and classification algorithms. Meager efforts have been made to protect EEG biometric systems from privacy leakage and security breaches. In this work, we propose the PolyCos Transform, short for **Poly**nomial transformation embedding **Cos**ine functions, a privacy-preserving and cancelable biometric design that protects EEG templates and supports secure biometric applications. The proposed PolyCos Transform is used to generate cancelable EEG graph templates, denoted as PolyCosGraph. This paper is a pioneering study addressing the security concerns of EEG biometrics and provides insights for future research in this direction. Specifically, we contribute to the existing studies on EEG biometrics in the following aspects:

- Existing privacy-preserving methods such as hash function and fuzzy commitment are not cancelable, where the raw biometric features are vulnerable to hill-climbing attacks. In this paper, a new privacy-preserving and cancelable EEG biometric system is designed, which consists of a non-invertible transformation, a template corrupting process, and a filter-embedded matching algorithm. Even when the transformed template is compromised, attackers cannot retrieve the raw EEG features and the compromised template can be revoked.
- An innovative non-linear, non-invertible transformation is proposed based on a system of multivariate polynomial equations embedding trigonometric functions. To the best of our knowledge, there exists no systematic method to solve such equations.
- A template corrupting mechanism is designed to create 'corrupted' equations in the system, which can mislead attackers in the solution finding process.
- Considering the template corrupting operation, we propose a filter-embedded matching algorithm to match queries with corrupted templates.
- A comprehensive evaluation and security analysis is carried out that verifies the capacity of the proposed system against attacks via record multiplicity (ARM), preimage attacks, hill-climbing attacks, second attacks, and brute force attacks. To date, few published studies have investigated EEG biometric systems against these attacks.

The rest of this paper is organized as follows. Section 2 presents a brief review of state-of-the-art research on EEG biometrics and template protection mechanisms. Section 3 elaborates on the proposed method, PolyCosGraph, followed by experimental and analytical results in Section 5 and security analysis in Section 6. Section 7 summarizes the study and indicates future directions.

2 Related Work

2.1 EEG Biometrics

Existing studies on EEG biometrics mainly focuses on signal acquisition protocols, feature extraction methods, and decision-making algorithms, with the aim of improving recognition accuracy and inter-session stability. For signal acquisition, different protocols are proposed, including the resting state protocol [18], protocols based on internal and volitional tasks such as pass-thoughts and motor imagery [5], and event-related potential protocols using external stimulation [3]. Among these, the resting state protocol provides convenient data collection and has been shown effective and robust for EEG biometric applications [7], [18].

Feature extraction is another critical element in EEG biometrics in that the discriminative power of the extracted features directly affects the recognition accuracy. Important features for EEG biometric applications include those based on autoregressive (AR) models [6], entropy estimation [19], Fourier transform [6], [7], and wavelet packet decomposition [9]. These features capture the temporal dependency and complexity of the EEG time series in the time domain, and the spectral characteristics in the frequency domain, respectively. Moreover, recent studies investigating the performance of bivariate features based on EEG functional connectivity in user identification and authentication have shown that these features are more robust to changes in user state and provide higher inter-session stability than univariate features [8], [20].

The decision-making methods for EEG biometrics can be divided into template matching-based and classifier-based.

For user authentication, the template matching method compares the similarity between the query and the stored template using a predefined threshold to decide whether the query is accepted or not. Existing works defined the similarity according to different distance measures, such as the Euclidean [21], Mahalanobis [7], and Manhattan distances [6], as well as cosine similarity [6] and cross-correlation [3]. Another group of works applied machine learning algorithms, such as deep learning models, to classification for EEG biometrics. Widely used classification algorithms in EEG biometrics include the linear discriminant analysis [9], support vector machines and neural networks [2], [10]. However, as discussed in the section of Introduction, for either template matching-based or classifier-based EEG biometric systems, corresponding privacy protection mechanisms are in demand in order to address user privacy and data security concerns.

2.2 Template Protection and cancelable Mechanisms

To protect the templates in EEG biometric systems, existing studies applied hash functions [22] and cryptographic schemes [23]. Specifically, He *et al.* [22] hashed EEG autoregressive features using the fast Johnson-Lindenstrauss algorithm, and applied a naive Bayes probabilistic model to classify the hash vectors. Bidgoly *et al.* [24] used a neural network model to generate EEG templates and analogized this feature extraction process as a hashing process that can hide users' private information. Damaševičius *et al.* [23] proposed a cryptographic scheme based on fuzzy commitment and error-correcting codes for EEG-based authentication, where the statistical features derived from the covariance matrix of EEG data were hidden through a fuzzy commitment construct. The turbo codes and modulation constellations were also used for protecting EEG biometric templates [25]. The system derives a codeword by turbo coding and modulating a randomly generated binary key, and then binds the EEG features with the codeword to obtain a helper data template through an operator whose outputs reveal no information about its arguments. Hence, the helper data template can be made publicly available, together with a hashed version of the binary key. While these methods protect user-specific sensitive information contained in the EEG template, they do not support cancelability to revoke compromised templates, which makes the system vulnerable to hill-climbing attacks and second attacks.

EEG biometrics are sometimes referred to as cancelable biometrics, since they can be elicited by numerous distinct brain systems through different acquisition protocols [3]. For example, different brain responses can be elicited with sophisticated visual stimuli. Therefore, it is possible to reset and change brain biometrics when the current biometric credential is divulged [11]. The Neurokey [26], a key generation method, was proposed based on this concept of cancelable EEG biometrics. Specifically, to replace a user's Neurokey, the system changes the signal acquisition protocol and uses the data collected under the new protocol to generate a new key. However, such 'cancelable' schemes protect neither EEG features/templates nor user privacy. Moreover, alternative options for signal elicitation are limited, and using different protocols can impact on the authentication performance [27]. It is worth noting that cancelability defined on signal acquisition protocols is different from the one defined on non-invertible transformations.

Cancelable template design based on non-invertible transformations offers data privacy protection and template revocability. It performs a one-way transformation on the raw biometric template to derive a transformed template such that an adversary is unable to obtain the raw template even if both the stored (transformed) template and the transformation method are compromised [28]. However, little research has been done on EEG biometrics in this area. While many non-invertible transformations were proposed for other biometric modalities (e.g., fingerprint), most of them have drawbacks. For example, transformations relying on underdertermined systems of linear equations are subject to ARM attacks [29]. A recent study developed a non-invertible transformation based on multivariate polynomial equations, improving the resistance to the ARM attack [30]. However, it is still possible to find analytical solutions to the system [31].

3 Methodology

This section describes the proposed privacy-preserving and cancelable EEG biometric system that protects data privacy and renders revocability at the same time. This is mainly achieved by the designed non-invertible transformation that converts EEG features into encrypted templates, the template corrupting process, and the corresponding matching algorithm, as illustrated in Fig. 2.

3.1 Feature Extraction

A resting state signal acquisition protocol is adopted for data collection, which asks the user to stay relaxed with eyes open during data collection. After data pre-processing, we extract the β band (13-30 Hz) signals with a bandpass filter since EEG in the β band shows higher correlation with human distinctiveness [2], [8]. Then, we estimate the functional connectivity between every two channels of the N_{ch}-channel signals using the ρ index, a general synchronization index based on the Shannon entropy [32]. After functional connectivity estimation, a fully-collected network of dimension $N_{ch} \times N_{ch}$ is constructed, where each node represents an EEG channel and each edge reflects the phase synchronization degree of signals of the two corresponding channels. Then the following graph features, as summarized in Table 1, are extracted from the established ρ-index functional connectivity networks. These features have been shown effective in capturing individuals' unique EEG patterns and are therefore suitable for authentication applications [8]. The resultant feature vector is of length $N_{ch} + 6$, with $N_{ch} = 64$ in a standard setup.

3.2 Feature Transformation

Motivated by the idea of multivariate polynomial transformation [30], we propose a non-linear system of multivariate polynomial equations embedding trigonometric functions.

Let $\mathbf{v}$ denote the feature vector extracted from EEG data, $\mathbf{v} = \{v_1, v_2, ..., v_N\} \in \mathbb{R}^{1\times N}$ where N is the number of

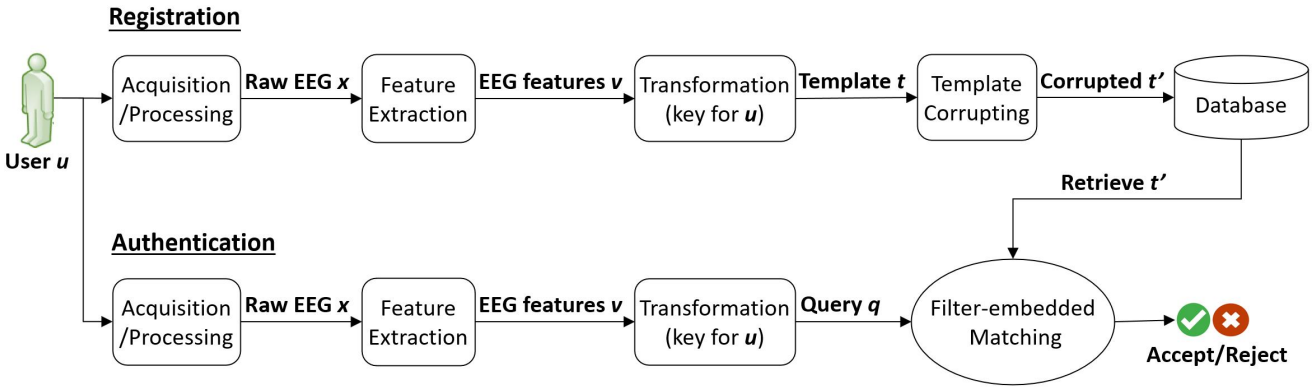

Fig. 2: Framework of the proposed method.

TABLE 1: Graph features extracted from the functional connectivity networks.

Nodal features	*Descriptions*
Pagerank centrality	Time spent at each node during a random walk
Global features	*Descriptions*
Transitivity	Interconnection degree of adjacent nodes
Modularity	Strength of division of a network into modules
Average path length	Information transport efficiency
Global efficiency	Information exchanges efficiency
Radius and diameter	Eccentricity of the network

features in the vector and each feature is a real number. A multivariate polynomial function of input $\mathbf{v}$ can be written in the following form:

$$\sum_{1}^{p}\prod_{1}^{\mathbf{Q}}\mathbf{v}^{\mathbf{D}}=t \tag{1}$$

where p is the number of monomials in the function, $\mathbf{Q}$ denotes the number of variables in the monomials, and $\mathbf{D}$ denotes the power of variables in the monomials. We have $\mathbf{Q}=\{q_1,q_2,...,q_p\}$ and $\mathbf{D}=\{d_1,d_2,...,d_{\sum Q}\}$. Evaluating the function at input $\mathbf{v}$ results in a transformed value t at the right side of the equation.

Repeating the above process N times, we can establish a well-defined multivariate polynomial system of equations as follows:

$$\begin{cases}\sum^{p_1}\prod^{\mathbf{Q}_1}\mathbf{v}^{\mathbf{D}_1}=t_1\\ \sum^{p_2}\prod^{\mathbf{Q}_2}\mathbf{v}^{\mathbf{D}_2}=t_2\\ ...\\ \sum^{p_N}\prod^{\mathbf{Q}_N}\mathbf{v}^{\mathbf{D}_N}=t_N\end{cases} \tag{2}$$

where $\mathbf{t}=\{t_1,t_2,...,t_N\}$ denotes the encrypted feature vector after transformation.

Assume that an attacker is able to obtain $\mathbf{t}$ and the corresponding parameters p, $\mathbf{Q}$, and $\mathbf{D}$, recovering the user's EEG biometric template $\mathbf{v}$ requires to solve the polynomial system (2). It is known that solving large systems of quadratic multivariate polynomial equations is an NP-hard problem [31]. For a well-defined system, where the number of equations is the same as the number of unknown variables as in our case, the most efficient methods known to date are exhaustive search for a small field and the Gröbner basic algorithm for a large field. However, with a large exponential complexity, these algorithms are unable to handle systems with $\geqslant 15$ unknown variables. In the proposed method, we establish a higher-order multivariate polynomial system of N ($N=70$) unknown variables, which is an NP-hard problem infeasible to be solved in practice.

To further increase the complexity of solving the system, we generate trigonometric terms $cos(v_n)$ and insert them into (2), as follows:

$$\begin{cases}\sum^{p_1}\prod^{\mathbf{Q}_1}\mathbf{v}^{\mathbf{D}_1}\cdot E[cos(v_n)]=t_1\\ \sum^{p_2}\prod^{\mathbf{Q}_2}\mathbf{v}^{\mathbf{D}_2}\cdot E[cos(v_n)]=t_2\\ ...\\ \sum^{p_N}\prod^{\mathbf{Q}_N}\mathbf{v}^{\mathbf{D}_N}\cdot E[cos(v_n)]=t_N\end{cases} \tag{3}$$

where $E[\cdot]$ denotes the rule that determines the existence of the trigonometric term in the monomial based on certain conditions of the monomial. Transforming $\mathbf{v}$ into $\mathbf{t}$ protects the raw EEG features and data privacy since there is no systematic way to solve the equations in (3).

In addition, we offer certain flexibility in customizing the transformation and adjusting the complexity of the system. Specifically, we allow the system operator to have a different setup for the number of monomials in a polynomial equation (N_m), the maximum number of variables in each monomial (M_v), and the maximum power of variables (M_p). If unspecified, the default values are 3, 10, and 3, respectively. The details of the PolyCos transform are described in Algorithm 1, and the functions used therein are explained in Table 2. During the registration phase, the system generates and stores a key k for each user. This key is used as the seed to initialize the pseudorandom number generators: $rng(k)$. Next, two matrices, $\mathbf{Q}$ and $\mathbf{D}$, are initialized. The $\mathbf{Q}$ is a two-dimensional matrix with each entry Q_{zx} indicating the number of variables in monomial x of equation z. The $\mathbf{D}$ is a three-dimensional matrix with each entry D_{xyz} indicating the power of variable y in monomial x of equation z. Lines 5-11 in Algorithm 1 set up the matrix $\mathbf{D}$: for each monomial in each equation, it computes a variable index vector idx and the corresponding powers of these indexed variables pw. Looping over all the monomials in all equations yields

the final matrix $\mathbf{D}$, which contains all parameters of the multivariate polynomial transformation. Then Lines 12-25 transform $\mathbf{v}$ to $\mathbf{t}$: for each monomial in each equation, it retrieves the variable powers pw and the corresponding variable indices idx, and applies the multivariate polynomial transformation. In particular, if idx satisfies the condition $mod(numel(idx), 2) = 1$, a $consine$ function term is inserted into the monomial. The resultant $\mathbf{t}$ is the output of the PolyCos transform. Revoking a template simply requires to replace the user key k.

Algorithm 1: PolyCos Transform

Setup : N_m = 3; M_v = 10; M_p = 3
Input : feature vector $\mathbf{v}$; user key k
Output: template $\mathbf{t}$

1 $N = numel(\mathbf{v})$
2 set seed for random number generators: $rng(k)$
3 $\mathbf{Q} = randi([1, M_v], N, N_m)$
4 $\mathbf{D} = zeros(N_m, N, N)$
5 **for** $n = 1$ **to** N **do**
6 **for** $m = 1$ **to** N_m **do**
7 $idx = randperm(N, \mathbf{Q}(n, m))$
8 $pw = randi([1, M_p], 1, \mathbf{Q}(n, m))$
9 $\mathbf{D}(m, idx, n) = pw$
10 **end**
11 **end**
12 $\mathbf{t} = zeros(1, N)$
13 **for** $n = 1$ **to** N **do**
14 $z = 0$
15 **for** $m = 1$ **to** N_m **do**
16 $pw = \mathbf{D}(m, :, n)$
17 $idx = find(pw)$
18 **if** $mod(numel(idx), 2) = 1$ **then**
19 $z = z + \prod_{idx} \mathbf{v}(idx)^{pw(idx)} \cdot cos(\mathbf{v}(idx(1)))$
20 **else**
21 $z = z + \prod_{idx} \mathbf{v}(idx)^{pw(idx)}$
22 **end**
23 **end**
24 $\mathbf{t}(n) = z$
25 **end**

3.3 Template Corrupting Process

At the end of registration, after deriving $\mathbf{t}$ through the proposed transformation, we randomly replace a few elements in $\mathbf{t}$ with dummy values to get $\mathbf{t}'$, a corrupted version of $\mathbf{t}$ to be stored in the system. We refer to this operation as template corrupting, and N_r denotes the number of elements being replaced. Details of the template corrupting process are summarized in Algorithm 2. Specifically, it generates an index vector idx to randomly select N_r elements from N elements in $\mathbf{t}$: $idx = randi([1, N], 1, N_r)$. Then, the selected elements are replaced with dummy values: $\mathbf{t}'(idx) = min(\mathbf{t}) + (max(\mathbf{t}) - min(\mathbf{t})) \cdot \mathbf{c}$, where $\mathbf{c} = rand(1, N_r)$ represents the coefficients of the dummy values. The above process ensures that the resulting dummy values are in the same range of the original values, so it is impossible for an attacker to distinguish them. The default value of N_r is 4, but different settings are allowed. The corresponding analysis is in Section 5.1.2. Note that our proposed template corrupting process is not restricted to a specific transformation algorithm. It can be applied to other methods as an additional security layer since it adds extra complexity to finding a solution. An attacker would have to filter out the dummy equations, which is a combination problem, before solving the transformation system. In addition, this process helps the system resist hill-climbing attacks as the dummy values can misguide the optimization algorithm to an invalid solution.

TABLE 2: Descriptions of functions used in the algorithms.

Func.	Descriptions
$rng(k)$	sets random number generators with seed k
$randi([a, b], m, n)$	returns $m \times n$ pseudorandom integers in $[a, b]$ (uniform)
$rand(m, n)$	returns $m \times n$ pseudorandom values in $(0, 1)$ (uniform)
$randperm(N, k)$	returns k unique integers selected randomly from $[1, N]$
$find(X)$	returns indices of nonzero elements in X
$numel(X)$	returns the number of elements in X

Algorithm 2: Template Corrupting

Setup : N_r = 4 (default)
Input : template $\mathbf{t}$; user key k
Output: corrupted template $\mathbf{t}'$

1 $N = numel(\mathbf{t})$
2 set seed for random number generators: $rng(k)$
3 randomly select: $idx = randi([1, N], 1, N_r)$
4 dummy value coefficients: $\mathbf{c} = rand(1, N_r)$
5 initialize: $\mathbf{t}' = \mathbf{t}$
6 replace: $\mathbf{t}'(idx) = min(\mathbf{t}) + (max(\mathbf{t}) - min(\mathbf{t})) \cdot \mathbf{c}$
7 output $\mathbf{t}'$

3.4 Filter-embedded Matching in the Encrypted Domain

During authentication, the system generates a query $\mathbf{q}$ following the same signal acquisition, feature extraction and transformation procedures as in the registration phase, and then computes a matching score between $\mathbf{q}$ and $\mathbf{t}'$, the stored template of the claimed user. Since $\mathbf{t}'$ is a corrupted version of the transformed template $\mathbf{t}$ with N_r elements replaced, a genuine query will have a high probability of generating N_r transformed elements that are different from the stored template. To eliminate the effects of corrupted elements on matching scores, we embed the matching algorithm with a filtering mechanism. Specifically, it first calculates the element-wise absolute distances between $\mathbf{q}$ and $\mathbf{t}'$: $\mathbf{d} = abs(\mathbf{q} - \mathbf{t}')$, then sorts the distances in descending order and removes the N_r largest elements: $\mathbf{d} \leftarrow sort(\mathbf{d}, \text{'descend'})$ and $\mathbf{d} \leftarrow \mathbf{d}(:, 1 : N_r) = 0$. Finally, the inverse of the sum of the element-wise distances, $s = 1/sum(\mathbf{d})$, is employed as the matching score, which is then compared with the operating threshold θ to output the final decision $\hat{o}$. Algorithm 3 illustrates the matching process. In our experiments, we increment the threshold until reaching the equal error rate (EER) point, that is, when the false acceptance rate (FAR) equals the false rejection rate (FRR). The FAR gives the percentage of queries in

which impostors are incorrectly accepted, whereas the FRR expresses the percentage of queries in which genuine users are incorrectly rejected.

Algorithm 3: Filter-embedded Matching

Setup : N_r = 4 (default); operating threshold θ
Input : query $\mathbf{q}$; template $\mathbf{t}'$
Output: decision $\hat{o}$

$\mathbf{d} = abs(\mathbf{q} - \mathbf{t}')$
$\mathbf{d} \leftarrow sort(\mathbf{d}, \text{'descend'})$
$\mathbf{d} \leftarrow \mathbf{d}(:, 1 : N_r) = 0$
$s = 1/sum(\mathbf{d})$
if $s \geqslant \theta$ **then**
  $\hat{o} =$ accept
else
  $\hat{o} =$ reject
end

4 Evaluation Procedure

4.1 Databases and Pre-processing

The proposed method is evaluated over two publicly available databases, which are the EEG Motor Movement/Imagery Database (MMIDB) [33] and SEEDiv database [34]. The MMIDB provides EEG signals of 109 healthy subjects under resting states and motor imagery tasks, including opening/closing and imagining opening/closing fists or feet. We refer to these tasks as resting with eyes open (EO), resting with eyes closed (EC), motor movement (MM), and motor imagery (MI). More detailed descriptions are available on the webpage of the database [35]. MMIDB has been widely used in EEG biometric studies due to its relatively large number of subjects and multiple recording conditions [2], [7], [8], [9], [20]. A 64-electrode BCI2000 system [36] was used for signal acquisition. The sampling rate was 160 Hz, and the recorded EEG was referenced to the earlobes. The SEEDiv database contains EEG recordings of 15 subjects watching movie clips. This database was originally collected for EEG-based emotion recognition, where the movie clips were used as visual stimuli to induce happiness, sadness, fear and neutral emotions from the subjects. We selected recordings under the neutral emotion setting for this study. Details of the two databases are summarized in Table 3. In terms of data pre-processing, we first removed the DC offset and extracted signal within the frequency range [0.5 42] Hz, which is the canonical EEG frequency range. Then EEG artifacts induced by eye and muscle movement and loose contact of electrodes were removed using independent component analysis and the Multiple Artifact Rejection Algorithm (MARA) [37]. A non-overlapping sliding window was applied to signal segmentation, and each frame has 2-second EEG data, i.e., 64×320 for MMIDB and 62×400 for SEEDiv.

TABLE 3: Databases

Databases	#Subj.	#Ch.	Samp. rate	Devices	Protocols
MMIDB	109	64	160 Hz	BCI2000	EO EC MM MI
SEEDiv	15	62	200 Hz	ESI NeuroScan	Movie clips

4.2 Comparison Methods

The proposed method is first compared with the baseline approaches, where the raw templates are directly used for comparison in the non-encrypted domain without transformation. Three popular feature types are considered, which are the reflection coefficients of autoregressive models [6], [18], band power features [6], [7], and fuzzy entropy features [19], denoted as ARr, PSD, and FuzzEn, respectively. To be specific, the ARr features are obtained through a 5th-order AR model using the Burg method [18]. The PSD features are derived from the EEG power spectrum estimated by the fast Fourier Transform [7]. The ARr, PSD, and FuzzEn features have been shown effective for EEG biometrics. In addition, we evaluate the combination of these three types of features as well as graph features defined on the EEG functional connectivity networks [8].

Furthermore, we compare the proposed method with four state-of-the-art privacy-preserving methods for EEG biometrics [22], [23], [24], [25]. These four methods are reviewed in Section 2.2. Note that research on privacy and security issues of EEG biometrics is still in the early stage. Currently, there is no published paper about cancelable EEG templates and therefore this study is a pioneering work on this topic. The four comparison methods, which are privacy-preserving but not cancelable, are the most closely related works in the literature.

5 Result

This section reports the experimental results of the authentication performance of the proposed method in the lost-key scenario and the analytical results in terms of decidability, revocability, diversity and unlinkability.

5.1 Performance in the Lost-key Scenario

In the lost-key scenario, we assume the user key used in the transformation is exposed to the attacker so that the attacker can take advantage of this to penetrate the authentication system, which is the worst case for a cancelable biometric system. In our experiment, we use a fixed parameter key k for feature transformation during registration and authentication for all users to obtain performance under the lost-key scenario.

5.1.1 Performance Comparison

Table 4 reports the EER results of the proposed method and the baseline approaches under different signal acquisition protocols in the lost-key scenario. Authentication systems based on transformed templates typically sacrifice some performance compared to their original versions using raw biometric templates without transformation. This is because the irreversible transformation often requires reordering or repositioning the feature set, which impairs the discriminative power of the feature set and introduces additional variations within the user [38]. From the EER results of PolyCosGraph and Graph in Table 4, we can see that our design exhibits equivalent authentication performance to the raw biometric feature templates, achieving 1.49%, 5.85%, 0.68%, 1.15%, and 0.46% EER with $Fe = 10$ and $Ft = 5$

under the EO, EC, MM, MI, and watching movie conditions, respectively. Furthermore, comparing PolyGraph and PolyCosGraph, we observe that integrating the trigonometric components into the multivariate polynomial transformation further improves authentication performance while increasing the complexity of solution finding.

Fig. 3 shows the detection error trade-off (DET) curves of PolyCosGraph under different signal acquisition protocols. Table 5 compares the EER results of the proposed PolyCosGraph and four state-of-the-art privacy-preserving methods for EEG biometrics (i.e., [22], [23], [24], [25]). Clearly, PolyCosGraph outperforms these comparison methods in terms of both authentication accuracy and security.

TABLE 4: Authentication performance (EER) of the proposed and comparison methods under different signal acquisition protocols.

MMIDB database - EO

Methods	Fe=10, Ft=1	Fe=10, Ft=5	Domain
ARr	19.41%	8.97%	Non-encrypted
FuzzEn	24.45%	14.66%	Non-encrypted
PSD	28.64%	21.69%	Non-encrypted
ARr+PSD+FuzzEn	17.14%	8.09%	Non-encrypted
Graph	5.18%	1.1%	Non-encrypted
PolyGraph (this study)	7.9%	1.67%	Encrypted
PolyCosGraph (this study)	7.34%	1.49%	Encrypted

MMIDB database - EC

Methods	Fe=10, Ft=1	Fe=10, Ft=5	Domain
ARr	16.72%	9.56%	Non-encrypted
FuzzEn	24.43%	18.17%	Non-encrypted
PSD	30.28%	23.21%	Non-encrypted
ARr+PSD+FuzzEn	15.17%	8.59%	Non-encrypted
Graph	11.13%	5.03%	Non-encrypted
PolyGraph (this study)	14.35%	6.15%	Encrypted
PolyCosGraph (this study)	13.81%	5.85%	Encrypted

MMIDB database - MM

Methods	Fe=10, Ft=1	Fe=10, Ft=5	Domain
ARr	19.17%	9.4%	Non-encrypted
FuzzEn	22.64%	13.75%	Non-encrypted
PSD	27.66%	22.02%	Non-encrypted
ARr+PSD+FuzzEn	14.96%	7.12%	Non-encrypted
Graph	4.02%	0.4%	Non-encrypted
PolyGraph (this study)	6.88%	0.82%	Encrypted
PolyCosGraph (this study)	6.1%	0.68%	Encrypted

MMIDB database - MI

Methods	Fe=10, Ft=1	Fe=10, Ft=5	Domain
ARr	18.43%	10.78%	Non-encrypted
FuzzEn	23.09%	15.42%	Non-encrypted
PSD	26.25%	20.63%	Non-encrypted
ARr+PSD+FuzzEn	14.12%	7.89%	Non-encrypted
Graph	4.82%	0.98%	Non-encrypted
PolyGraph (this study)	7.61%	1.26%	Encrypted
PolyCosGraph (this study)	7.02%	1.15%	Encrypted

SEEDiv database - Watching movie clips

Methods	Fe=10, Ft=1	Fe=10, Ft=5	Domain
ARr	15.63%	9.79%	Non-encrypted
FuzzEn	16.48%	11.14%	Non-encrypted
PSD	17.72%	6.09%	Non-encrypted
ARr+PSD+FuzzEn	9.78%	3.07%	Non-encrypted
Graph	2.39%	0.15%	Non-encrypted
PolyGraph (this study)	4.66%	0.76%	Encrypted
PolyCosGraph (this study)	4.2%	0.46%	Encrypted

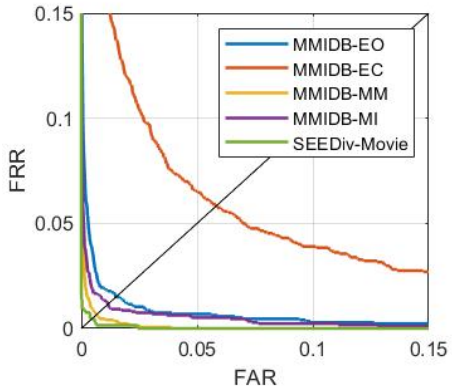

Fig. 3: DET curves of the proposed PolyCosGraph ($F_e = 10$, $F_t = 5$, $N_r = 4$) under different signal acquisition protocols.

Signal acquisition protocol. An important aspect of EEG biometrics is the signal elicitation protocol as it is the prerequisite for obtaining distinctive neural responses from individuals. From the results, we can observe that the signal acquisition protocol has an impact on the biometric performance. Specifically, the use of cognitive tasks (e.g., motor imagery) and external sensory stimulation (e.g., visual stimuli) provides better authentication performance than the resting states [39]. This is because the internal and external stimulation can elicit corresponding brain responses associated with cognitive processing or evoke activity in particular brain functional areas, which is considered distinctive for humans [3]. On the other hand, the resting state with EO offers a simple and convenient signal acquisition protocol for EEG biometrics as it does not involve sensory stimulation or complex instructions [7]. The proposed method does not rely on specific signal acquisition protocols, and the results validate its effectiveness under different types of protocols, including resting states (spontaneous brain activity), volitional tasks, and external visual stimulation. Both the resting state and motor imagery protocols represent volitional tasks, meaning that subjects are aware and in control of the responses. This should protect users of brain biometric systems against social assessment threats: users are able to intentionally or unintentionally invalidate brain biometrics when coerced [40].

Feature analysis. We evaluate the importance of each graph feature using recursive feature elimination. Specifically, we implement a support vector machine-based recursive feature elimination algorithm with correlation bias reduction [41]. This algorithm has been demonstrated to be effective for feature selection in bioinformatics. There are a total of 70 and 68 features for MMIDB and SEEDiv databases, respectively. We first generate a ranking list for these features using recursive feature elimination, then compute the EER performance as the number of top-ranked features increases. Fig. 4 presents the EER results of the top-ranked graph features. It can be observed that the EER decreases as the number of features increases, and this trend is consistent in all states for both databases, especially in the resting state EO. The results indicate that all the graph features employed in the proposed system are important for user authentication. Although the contribution of different

TABLE 5: Authentication performance (EER) comparison between the proposed PolyCosGraph and state-of-the-art privacy-preserving methods for EEG biometrics.

Method	Authentication performance (EER)					Security	
	MMIDB-EO	MMIDB-EC	MMIDB-MM	MMIDB-MI	SEEDiv-Movie	Encrypted	Cancelable
ref [25]	10.01%	9.38%	6.52%	8.21%	37.82%	Yes	No
ref [23]	44.05%	41.33%	33.59%	32.37%	38.66%	Yes	No
ref [22]	30.38%	27.5%	42.5%	26.42%	35.07%	Yes	No
ref [24]*	10.2%	12.5%	5.64%	8.2%	3.14%	Yes	No
PolyCosGraph	1.49%	5.85%	0.68%	1.15%	0.46%	Yes	Yes

*This method needs to train a neural network model with approximately 80% of the data from the databases.

features varies, there are no redundant ones as all of them help improve performance. We further visualize the importance (ranks) of the nodal graph features (Pagerank centrality of each node/channel) on authentication performance in scalp topological maps in Fig. 5, and report the importance (ranks) of global graph features in Table 6. The observation is that top features vary in different states, suggesting that the same features would contribute differently in different states and conditions. The reason for these variations is that different signal acquisition protocols and states elicit different neural responses and functional brain activities, resulting in various EEG characteristics from the scalp. Hence, the unique identity-bearing features may vary accordingly. It is necessary to retain some redundancy in the feature set for a reliable system in different states.

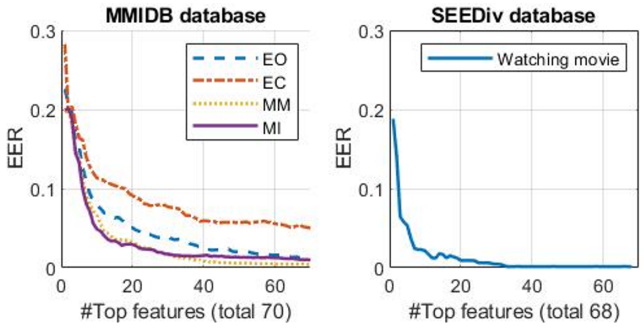

Fig. 4: EER results of top-ranked graph features.

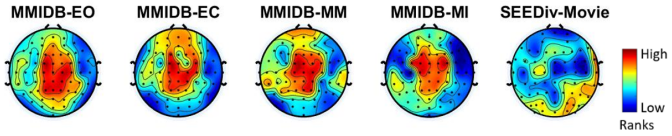

Fig. 5: Ranks of nodal features. The color indicates the ranks of the Pagerank centrality features of nodes over the scalp, with importance decreasing from red to blue.

TABLE 6: Ranks of global graph features.

Global graph features	MMIDB				SEEDiv
	EO	EC	MM	MI	Watching movie
Transitivity	1	6	2	2	40
Modularity	59	58	20	32	29
Average path length	3	4	3	3	44
Efficiency	2	1	1	1	30
Radius	4	4	4	4	47
Diameter	5	5	5	5	45

The proposed transformation is not confined to specific system configurations. In the following analysis, we evaluate the impact of system configurations on the authentication performance of the proposed method. This includes the effects of the number of EEG frames used to generate templates and queries (i.e., F_e and F_t), the random replacement parameter N_r, and electrode configurations. The EO resting state is used for the following analysis.

5.1.2 Impact of System Configuration

Number of EEG frames. EEG is a continuous data source, so it is natural and practical to use multiple frames of data rather than a single signal segment to generate templates. In this analysis, we examine how the authentication performance is impacted by F_e, the number of frames used in templates during registration, and F_t, the number of frames for generating queries during authentication. The database provides 30 frames EEG for each of the 109 subjects (60 seconds of EO recording per subject and 2 seconds per frame), leading to 3161 genuine tests at $F_e = 1$ and $F_t = 1$, and 436 genuine tests at $F_e = 10$ and $F_t = 5$. For impostor testing, we use the first F_t frame(s) of all subjects other than the user to generate query samples, leading to 11772 impostor tests.

The EER results are shown in Fig. 6. With $F_t = 1$, i.e., a single EEG frame for a query, a decreasing trend in the EER is observed as F_e increases, achieving EER = 14.69% at $F_e = 1$, EER = 7.34% at $F_e = 10$, and EER = 6.88% at $F_e = 20$. The performance improvement is significant, especially when F_e increases from 1 to 10. The interpretation of this result is related to the nature of brain signals. EEG signals contain transient components, presenting a momentary variation in the recorded data. In addition, EEG signals contain nonstationary ingredients, so their statistical characteristics change with time. The basic source of the observed nonstationarity in the EEG is not due to the casual influences of the external stimuli on the brain, but rather a reflection of switching the inherent metastable states of neural assemblies during brain functioning [42]. The quasi-stationary state has a short duration, and therefore, a longer segment or multiple short segments are usually used in practical applications. Likewise, when fixing $F_e = 10$ and increasing F_t, we can observe that the EER further decreases from 7.34% at $F_t = 1$ to 1.49% at $F_t = 5$ and 0.67% at $F_t = 10$. Our result shows that the use of multiple EEG frames enhances the stability of the template and query, which substantially improves authentication performance. This is consistent with the evidence from neuroscience research.

In the practical deployment of the proposed system, parameters F_e and F_t can be adjusted according to application scenarios and requirements. Proper settings require a balance between authentication performance and system usability (in terms of data acquisition time and convenience), as larger values of F_e and F_t enhance accuracy but increase data acquisition time required for registration and authentication. Therefore, we set $F_e = 10$ and $F_t = 5$ in all other analyses.

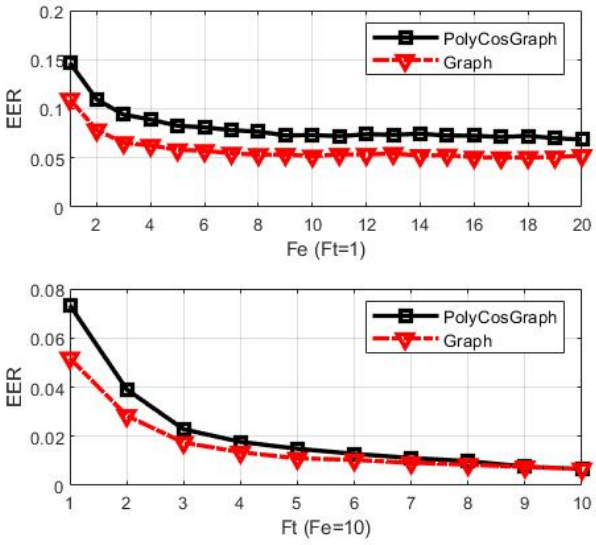

Fig. 6: Authentication performance (EER) of the proposed PolyCosGraph and the baseline Graph approach versus numbers of EEG frames during enrollment and authentication, i.e., F_e and F_t.

Number of corrupted elements. The template corrupting process in registration replaces N_r elements in $\mathbf{t}$ with dummy values randomly produced considering the distribution of values in $\mathbf{t}$. A corresponding matching protocol is established to account for the bias introduced by these dummy elements. In this analysis, we evaluate the effect of different values of N_r on the authentication performance. Fig. 7 presents the EER of the system at $N_r = \{0, 1, \cdots, 10\}$. Although there are small fluctuations in the range of 1.49% to 1.84%, the results indicate a relatively small impact of N_r on the EER. Note that when $N_r = 0$, the matching protocol degrades to the traditional case, where no random replacement takes place. Based on the results, $N_r = 4$ is selected as it provides the lowest EER among all the tested settings.

Electrode configuration. We evaluate the proposed method using four electrode configurations, the standard 64-electrode setup of the 10-20 international system and the setup of three widely used commercial EEG devices, namely Quick30 and Quick20 from Cognionics and Emotiv Epoc+. Fig. 8 illustrates the placement of electrodes in the aforementioned four configurations, with corresponding authentication performance (EER) summarized in Fig. 9. We can see that as the electrode density decreases, the EER increases from 1.49% with All64 to 4.7% with Quick30, 5.28% with Quick20, and 6.18% with Emotiv. The same

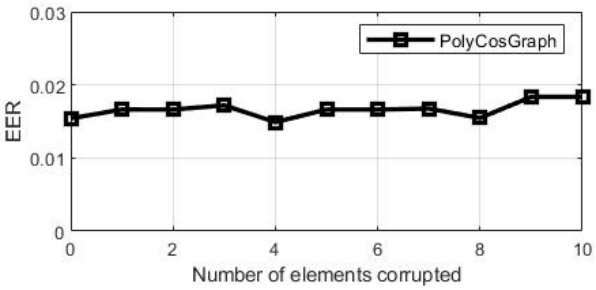

Fig. 7: Authentication performance (EER) of the proposed method versus the number of elements randomly replaced in $\mathbf{t}$ during enrollment, i.e., N_r. Other system settings: $F_e = 10$ and $F_t = 5$.

trend is observed from the results of the Graph method (i.e., without transformation). The reason for this phenomenon is simple. As the number of electrodes decreases, fewer resources are available for extracting unique features from subjects, resulting in less discriminative feature sets and thus adversely affecting authentication performance.

The proposed method is based on the functional connectivity of EEG signals, which usually requires a sufficient number of channels for a reliable estimation and feature extraction. To address performance degradation due to insufficient electrodes, we can use the channel density augmentation method proposed in a previous study [43], which has been demonstrated effective in addressing this issue. Another way to enhance channel density is to use a pre-trained machine learning model, where the relationship between channels was encoded during the training stage, to generate data for missing channels [44].

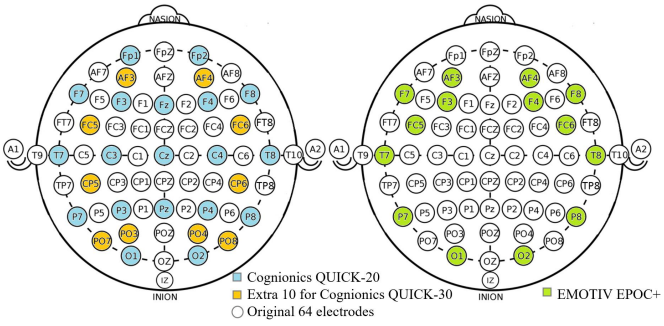

Fig. 8: Electrode configurations of four commercial EEG acquisition devices, denoted as All64, Quick30, Quick20, and Emotiv, equipped with 64, 29, 19 and 14 electrodes, respectively.

The designed system can be employed for access control in application scenarios that require high security levels, or continuous authentication for human-machine interaction systems and brain-computer interaction systems. The proposed method itself is not confined to specific EEG scanning systems or channel configurations. It is flexible to select a proper EEG scanning system (e.g., BCI 2000 system with 64 channels or Emotiv Epoc+ with 14 channels) for signal acquisition in accordance with application requirements. The signal acquisition time depends on the setting of F_e and F_t parameters. For example, with $F_e = 10$ and $F_t = 5$,

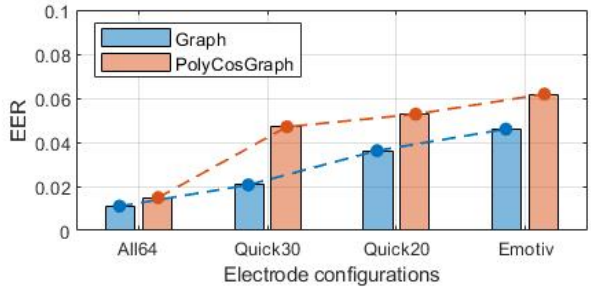

Fig. 9: Authentication performance (EER) of the proposed method with four different electrode configurations. Other system settings: $F_e = 10$, $F_t = 5$, and $N_r = 4$.

it takes 20 seconds and 10 seconds to acquire data during enrollment and authentication, respectively. Changing to $F_e = 5$ and $F_t = 1$ will reduce the time to 10 seconds and 2 seconds, respectively.

5.2 Decidability Analysis

Biometric authentication can be considered a classification task that distinguishes the user from impostors. To achieve good matching performance, a feature or template set with strong discriminative power is required. The decidability index d' [45] is used to measure the discriminative ability of the templates generated by our method. The index d' is widely used in the decidability analysis of biometric systems, defined as:

$$d' = (\mu_{intra} - \mu_{inter})/\sqrt{(\delta_{intra}^2 + \delta_{inter}^2)/2}, \qquad (4)$$

where μ_{intra} and δ_{intra} denote the mean and standard deviation of genuine scores, respectively, and μ_{inter} and δ_{inter} denote the mean and standard deviation of impostor scores, respectively. The genuine score is computed by matching two samples from the same user, and the impostor score is computed by matching the user sample against the sample of other subjects, yielding 435 genuine scores and 97200 impostor scores for each user. The score distributions of the proposed method, PolyCosGraph, under the two signal elicitation protocols, EO and MM, are shown in Fig. 10. We can see that the PolyCos transform enhances the decidability of the template: from 1.34 to 3.88 under the resting state and from 1.53 to 4.86 under motor movement tasks. The same conclusion can be drawn from the reduction in overlap between the two score distributions.

5.3 Unlinkability

Unlinkability is defined as 'a property of two or more biometric references that cannot be linked to each other or to the subject(s) from which they were derived' [46]. We follow the framework proposed by Gomez-Barrero *et al.* [47] to evaluate the unlinkability of the proposed PolyCosGraph design. This framework defines two types of scores: the mated score is computed between two templates from the same user, and the non-mated score is computed between two templates from two different users. On top of the mated and non-mated score distributions, the score-wise linkability $D_{\leftrightarrow}(s)$ and the system overall linkability $D_{\leftrightarrow}^{sys}$ are defined,

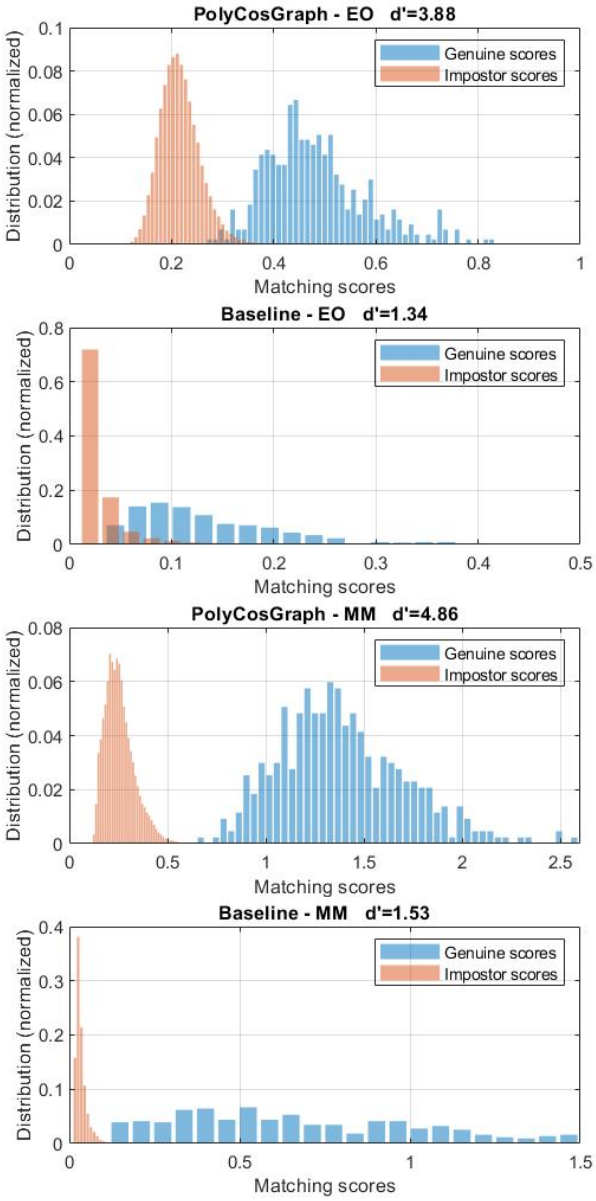

Fig. 10: Genuine and impostor score distributions and the corresponding decidability index d' in the decidability analysis, demonstrated for Users 5 and 69 under the EO and MM signal acquisition protocols.

which are local and global measures, respectively [47]. The value range of both measures is 0 to 1, where 0 indicates fully unlinkable. For the unlinkability test, we generate six different transformed templates from every sample of each user using six different keys [30] and calculate mated and non-mated scores according to their definitions. Fig. 11 reports the results of the unlinkability analysis of the proposed method under two signal acquisition protocols, EO and MM. It shows that the mated score distribution with different keys (cross-matching) is largely overlapped with the non-mated score distribution, which means that templates derived from the same user using different keys are as disparate as templates of different users. In addition, the global linkability indices are $D_{\leftrightarrow}^{sys} = 0.02$ and $D_{\leftrightarrow}^{sys} = 0.01$ under EO and MM, respectively, indicating the high unlinkability

of the proposed method.

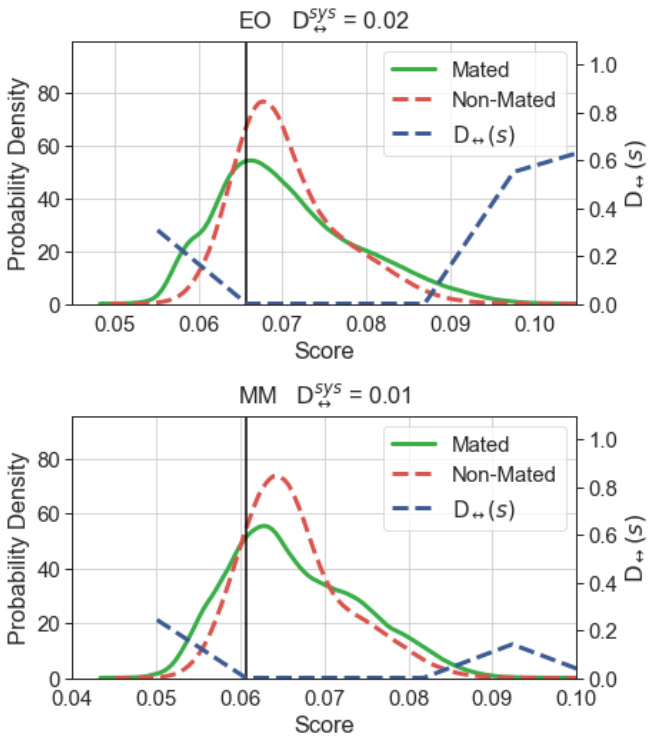

Fig. 11: Distributions of mated and non-mated scores and the corresponding local measure $D_{\leftrightarrow}$(s) and global measure $D^{sys}_{\leftrightarrow}$ in the unlinkability analysis under the EO and MM signal acquisition protocols.

5.4 Diversity

Diversity means that different templates can be generated using the same biometric data, and these templates should be unrelated so that it is impossible to match them. We compute the pseudo-impostor score [30] to evaluate whether the proposed PolyCosGraph design meets the requirement of diversity. Specifically, we apply 50 different keys $(k_1, k_2, \cdots, k_{50})$ and generate the corresponding pseudo-impostor templates from the first sample of each user. The pseudo-impostor score is then computed by matching the original user templates (generated with k_0) against the pseudo-impostor templates (generated with $k_1, k_2, \cdots, k_{50}$) of the same user. Fig.12 shows the distributions of genuine and pseudo-impostor scores for the proposed PolyCosGraph method. No overlap is observed between the two distributions, indicating that adversaries are unlikely to match across applications or break into the system using compromised templates. The same finding can be reached by the decidability indices, 5.65 and 5.76 under the EO and MM signal acquisition protocols, respectively, suggesting that templates generated from the same user with different keys are not related.

6 Security Analysis

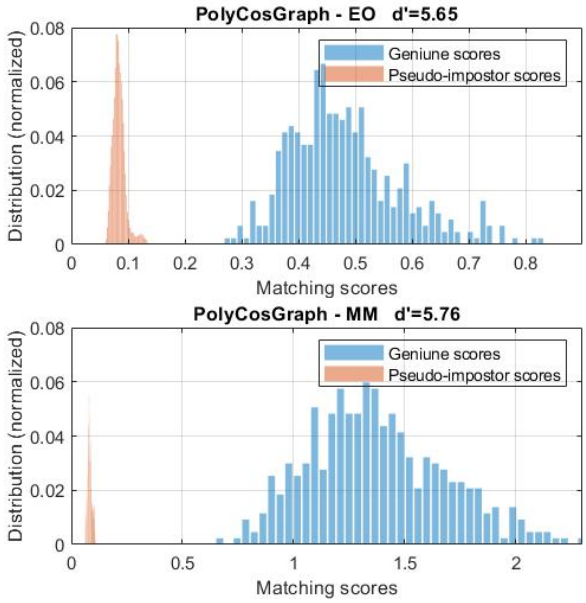

Fig. 12: Distributions of genuine scores (same user with the same key) and pseudo-impostor scores (same user with different keys) in the revocability and diversity analysis under the EO and MM signal acquisition protocols.

6.1 Attacks via Record Multiplicity (ARM)

With the principle of diversity, a cancelable biometric template design supports the generation of different transformed templates $\mathbf{t}$ from the same raw biometrics $\mathbf{x}$ by changing the transformation key k. Assume that the attacker is able to obtain multiple transformed templates $\{\mathbf{t}_1, \mathbf{t}_2, \cdots, \mathbf{t}_n\}$ from one or multiple applications and knows the transformation and user keys $\{k_1, k_2, \cdots, k_n\}$. The system is then exposed to the ARM attack, which exploits these transformed templates to recover the biometric features $\mathbf{v}$ [29]. For example, cancelable design based on classical linear random projections is vulnerable to ARM attacks, since combining multiple transformed templates will result in a well-defined system of linear equations, from which a unique solution (i.e., raw biometric features) can be determined. In other words, this type of algorithm only provides one-time-pad security.

The proposed PolyCosGraph algorithm is resistant to ARM attacks, due to three main components of the algorithm, namely the multivariate polynomial system, the embedding of trigonometric components in the system, and the template corrupting process. As analyzed by Courtois *et al.* in their study [31], solving large systems of quadratic multivariate polynomial equations is an NP-hard problem in any field. For well-defined systems, the most efficient methods known to date are exhaustive search and the Gröbner basic algorithm, for small and large fields, respectively. However, the Gröbner basic algorithm has a prohibitively high exponential complexity, and it is computationally infeasible to apply such algorithms to systems with $\geqslant 15$ unknown variables in practice. To successfully launch an ARM attack to the proposed method, a polynomial system with 70 unknown variables needs to be solved. This is considered NP-hard and infeasible to solve in practice. We further in-

crease the complexity of solving such a system of equations by applying higher-degree multivariate polynomials and embedding trigonometric functions in it. Even with over-defined systems, there is no systematic way to solve it. In addition, the random replacement procedure at the end of the registration renders extra complexity to find a solution. Attackers have to filter out dummy equations, which is a combination problem, before solving a system of higher-order multivariate polynomial equations embedded with trigonometric functions.

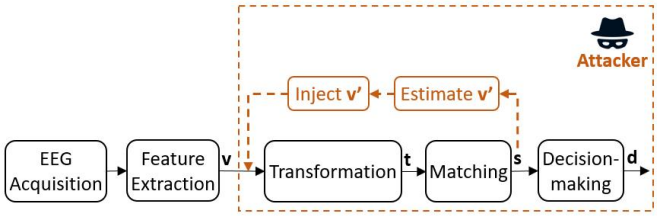

Fig. 13: Illustration of hill-climbing attacks on the authentication system.

6.2 Preimage Attacks

Taking into account the properties of cancelable biometrics, a recent study [48] extended the preimage attack, which was defined for cryptographic hash functions, in the context of cancelable biometric templates: given a transformed template $\mathbf{y}$, it should be difficult to find the true solution $\mathbf{x} = \mathbf{x_0}$ such that $\mathbf{y} = f(\mathbf{x_0}, k)$, where $f(\cdot)$ is the transformation function with key k and $\mathbf{x}_0$ the raw biometric template. That is to say, collision resistance is a property of cryptographic hash functions, but it is not necessarily required by the non-invertible transformation in a cancelable template design. This is because the cancelable template design allows a compromised template to be revoked and a new one to be generated using a different key k'. Solution $\mathbf{x} \neq \mathbf{x_0}$ would then become invalid when the key is changed and the compromised template is revoked, i.e., $f(\mathbf{x}, k') \neq f(\mathbf{x_0}, k')$.

In the ARM attack analysis, we have discussed that it is not computationally feasible to solve the system of equations in a systematic way to obtain the raw biometric template. Assuming that an attacker is able to submit queries to the system and get the corresponding matching scores, hill-climbing algorithms can be used to launch a preimage attack. This type of attack is specifically referred to as a hill-climbing attack.

6.3 Hill-Climbing Attacks and Second Attacks

In hill-climbing attacks, adversaries iteratively submit synthetic representations of a user's biometric and exploit the corresponding matching scores to guide the iteration process until a false acceptance is attained [17]. As illustrated in Fig. 13, an attacker runs an algorithm to iteratively generate $\mathbf{v}'$ and inject it to the system, then uses the corresponding matching score to guide its estimation direction until $\mathbf{v}'$ is accepted by the system. Hill-climbing attacks are a big threat to traditional biometric systems because once the attacker obtains a synthetic feature vector accepted by the system, raw biometric features are considered exposed forever. In the following experiment, we verify that the proposed cancelable template design can effectively protect raw biometric features and that hill-climbing attacks do not pose a major threat to the system.

In our experiment, we implement the Nelder-Mead algorithm, a downhill simplex method for derivative-free optimization, to perform the hill-climbing attack [17]. The iteration ends when the matching score between the template generated from the submitted input and the reference template is $\geqslant \theta$, or when the maximum number of submissions (20,000) is reached. The results of the hill-climbing attack are presented in terms of success rate (SR) and efficiency N_{att}, which are defined as the percentage of user accounts that are compromised and the average number of submissions (attempts) used to break a user account. Fig. 14 reports the SR and N_{att} results of the hill-climbing attack on the proposed system. At the EER operating point, which is $\theta = 0.33$ for EO and $\theta = 0.34$ for MM, the SRs are around 0.67 and 0.33, respectively. As the threshold θ increases, the SR of finding a synthetic input $\mathbf{v}'$ to enter the user account drops dramatically, along with the efficiency. A similar trend is observed under the two signal acquisition protocols.

In the following analysis, we demonstrate that the synthetic input $\mathbf{v}'$ obtained through hill-climbing attacks is unlikely to reflect the true biometric feature $\mathbf{v}$; therefore it becomes invalid once the compromised user template is revoked. Let $\mathbf{t}_0$ denote the user reference originally stored in the system and $\mathbf{v}'$ the synthetic feature vector estimated by the hill-climbing attack. This $\mathbf{v}'$ is tentatively accepted by the system since it produces a $\mathbf{t}'_0$ that matches $\mathbf{t}_0$. To defend, the system would revoke the compromised template $\mathbf{t}_0$ and replace it with a new one $\mathbf{t}_1$ which has no relation with $\mathbf{t}_0$. Let $\mathbf{t}'_1$ denote the template transformed from the obtained synthetic feature vector $\mathbf{v}'$ using the same new key. We demonstrate that $\mathbf{t}'_1$ is far from $\mathbf{t}_1$ so that the attacker will not be able to break in the system again. Re-entering user accounts using previously estimated feature vectors after template revocation is referred to as second attacks [48].

The success rate of launching a second attack on the proposed method is reported in Fig. 15. At each θ, we first performed a hill-climbing attack on each user, then for users whose templates were compromised, we revoked their templates and launched a second attack on each of them. For example, at $\theta = 0.33$ (the EER operating point under the EO), the success rate of hill-climbing attacks is 0.67, that is, 73 out of 109 users are successfully cracked. Then for each of the 73 users, we carry out the second attack using the estimated features obtained in the hill-climbing attack. The results indicate that even with a threshold smaller than the EER operating point, the system is unlikely to be compromised by second attacks. In addition, matching scores between real biometric features and estimated features obtained through hill-climbing attacks show that estimated features do not reflect true biometric features.

6.4 Brute Force Attacks

Brute force attacks aim to obtain raw biometric features through exhaustive search. It is important to ensure that the search space is relatively large so that the probability of successfully finding the secret is low. In our method, the

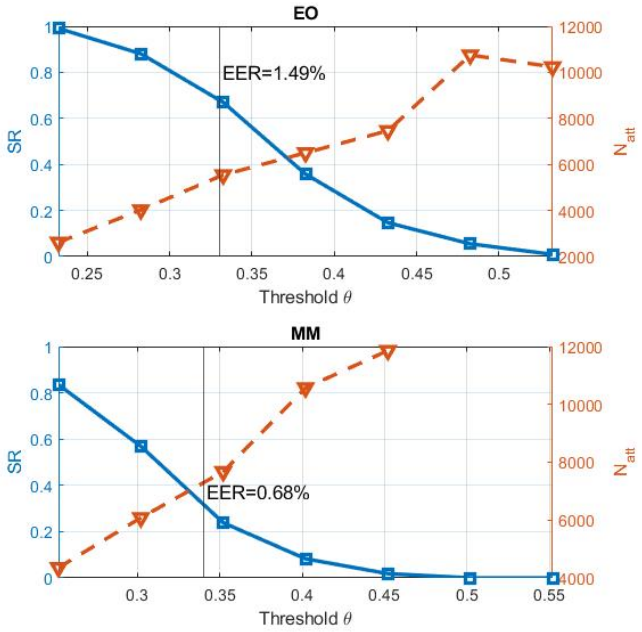

Fig. 14: Success rate (SR) and efficiency (N_{att}) of hill-climbing attacks on the system.

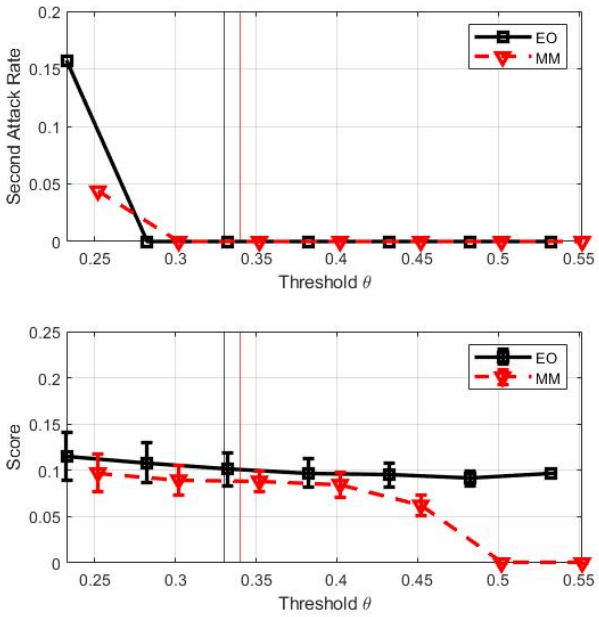

Fig. 15: Success rate of second attacks at different thresholds (top), and matching scores between real biometric features and estimated features obtained through hill-climbing attacks at different thresholds (bottom). The vertical lines indicate the EER operating point.

EEG feature vector $\mathbf{v}$ has m elements (m=70), where each element is a real number (double precision floating number). Hence, the number of trials to traverse all possible guesses in the search space is 2^{4480}, which is enormous. In the actual deployment of the system, m is related to the number of electrodes of the EEG acquisition device, and it can be adjusted according to application scenarios and requirements. Having a larger value of m can improve security strength, but at the same time it means less efficient data collection and more computational costs. Hence, a proper value for m should be set in order to balance security and efficiency.

7 Conclusion

This paper addresses two security concerns of EEG biometric systems: 1) the stored raw EEG templates leak users' private or personal information; 2) the systems are vulnerable to attacks such as ARM and hill-climbing attacks. A privacy-preserving and cancelable EEG biometric system was designed, in which we proposed a non-invertible transformation based on multivariate polynomial equations embedding trigonometric functions, a template-corrupting process and a corresponding matching algorithm. The proposed method not only protects the privacy of raw EEG features and users' sensitive information that can be inferred from raw EEG features, but also provides revocability that allows the replacement of compromised templates. The proposed system achieved the authentication performance of 1.49% EER with a resting state protocol, 0.68% EER with a motor imagery task, and 0.46% EEG under a watching movie condition, in the encrypted domain, which is comparable to the performance of EEG biometric systems in the non-encrypted domain. A comprehensive security analysis shows that the proposed method can effectively defend against ARM attacks, preimage attacks, hill-climbing attacks, second attacks and brute force attacks.

Research on the security of EEG biometric systems has just begun, so we will continue this line of study and develop cryptographic methods (e.g., the Zero-knowledge proof) in building secure bio-cryptographic EEG systems. In addition, this study targeted the template matching-based systems, however, how to tackle the security issues of classifier-based systems is still an open question. Since classifier-based authentication systems store a classification model for each user rather than a template, cancelable template design is not applicable in this case. Therefore, appropriate protection methods need to be designed. Our future work will also investigate this problem.

References

[1] P. Campisi and D. La Rocca, "Brain waves for automatic biometric-based user recognition," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 5, pp. 782–800, 2014.

[2] M. Wang, H. El-Fiqi, J. Hu, and H. A. Abbass, "Convolutional neural networks using dynamic functional connectivity for EEG-based person identification in diverse human states," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 12, pp. 3259–3272, 2019.

[3] M. V. Ruiz-Blondet, Z. Jin, and S. Laszlo, "CEREBRE: A novel method for very high accuracy event-related potential biometric identification," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 7, pp. 1618–1629, 2016.

[4] Q. Gui, M. V. Ruiz-Blondet, S. Laszlo, and Z. Jin, "A survey on brain biometrics," *ACM Computing Surveys (CSUR)*, vol. 51, no. 6, pp. 1–38, 2019.

[5] J. Chuang, H. Nguyen, C. Wang, and B. Johnson, "I think, therefore I am: Usability and security of authentication using brainwaves," in *International conference on financial cryptography and data security*. Springer, 2013, pp. 1–16.

[6] E. Maiorana, D. La Rocca, and P. Campisi, "On the permanence of EEG signals for biometric recognition," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 1, pp. 163–175, 2015.

[7] D. La Rocca, P. Campisi, B. Vegso, P. Cserti, G. Kozmann, F. Babiloni, and F. D. V. Fallani, "Human brain distinctiveness based on EEG spectral coherence connectivity," *IEEE transactions on Biomedical Engineering*, vol. 61, no. 9, pp. 2406–2412, 2014.

[8] M. Wang, J. Hu, and H. A. Abbass, "BrainPrint: EEG biometric identification based on analyzing brain connectivity graphs," *Pattern Recognition*, vol. 105, p. 107381, 2020.

[9] S. Yang, F. Deravi, and S. Hoque, "Task sensitivity in EEG biometric recognition," *Pattern Analysis and Applications*, vol. 21, no. 1, pp. 105–117, 2018.

[10] E. Debie, N. Moustafa, and A. Vasilakos, "Session invariant EEG signatures using elicitation protocol fusion and convolutional neural network," *IEEE Transactions on Dependable and Secure Computing*, 2021.

[11] F. Lin, K. W. Cho, C. Song, W. Xu, and Z. Jin, "Brain password: A secure and truly cancelable brain biometrics for smart headwear," in *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services*, 2018, pp. 296–309.

[12] W. Klimesch, "EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis," *Brain research reviews*, vol. 29, no. 2-3, pp. 169–195, 1999.

[13] S. I. Dimitriadis, Y. Sun, K. Kwok, N. A. Laskaris, N. Thakor, and A. Bezerianos, "Cognitive workload assessment based on the tensorial treatment of EEG estimates of cross-frequency phase interactions," *Annals of biomedical engineering*, vol. 43, no. 4, pp. 977–989, 2015.

[14] W.-L. Zheng and B.-L. Lu, "Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks," *IEEE Transactions on Autonomous Mental Development*, vol. 7, no. 3, pp. 162–175, 2015.

[15] O. Landau, R. Puzis, and N. Nissim, "Mind your mind: EEG-based brain-computer interfaces and their security in cyber space," *ACM Computing Surveys (CSUR)*, vol. 53, no. 1, pp. 1–38, 2020.

[16] Y. Höller and A. Uhl, "Do EEG-biometric templates threaten user privacy?" in *Proceedings of the 6th ACM Workshop on Information Hiding and Multimedia Security*, 2018, pp. 31–42.

[17] E. Maiorana, G. E. Hine, and P. Campisi, "Hill-climbing attacks on multibiometrics recognition systems," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 5, pp. 900–915, 2014.

[18] P. Campisi, G. Scarano, F. Babiloni, F. D. Fallani, S. Colonnese, E. Maiorana, and L. Forastiere, "Brain waves based user recognition using the "eyes closed resting conditions" protocol," in *2011 IEEE International Workshop on Information Forensics and Security*. IEEE, 2011, pp. 1–6.

[19] Z. Cao and C.-T. Lin, "Inherent fuzzy entropy for the improvement of EEG complexity evaluation," *IEEE Transactions on Fuzzy Systems*, vol. 26, no. 2, pp. 1032–1035, 2017.

[20] M. Fraschini, S. M. Pani, L. Didaci, and G. L. Marcialis, "Robustness of functional connectivity metrics for EEG-based personal identification over task-induced intra-class and inter-class variations," *Pattern Recognition Letters*, 2019.

[21] M. Fraschini, A. Hillebrand, M. Demuru, L. Didaci, and G. L. Marcialis, "An EEG-based biometric system using eigenvector centrality in resting state brain networks," *IEEE Signal Processing Letters*, vol. 22, no. 6, pp. 666–670, 2014.

[22] C. He, X. Lv, and Z. J. Wang, "Hashing the mAR coefficients from EEG data for person authentication," in *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2009, pp. 1445–1448.

[23] R. Damaševičius, R. Maskeliūnas, E. Kazanavičius, and M. Woźniak, "Combining cryptography with EEG biometrics," *Computational Intelligence and Neuroscience*, vol. 2018, 2018.

[24] A. J. Bidgoly, H. J. Bidgoly, and Z. Arezoumand, "Towards a universal and privacy preserving EEG-based authentication system," *Scientific Reports*, vol. 12, no. 1, pp. 1–12, 2022.

[25] E. Maiorana, D. La Rocca, and P. Campisi, "Cognitive biometric cryptosystems a case study on EEG," in *2015 International Conference on Systems, Signals and Image Processing (IWSSIP)*. IEEE, 2015, pp. 125–128.

[26] G. Bajwa and R. Dantu, "Neurokey: Towards a new paradigm of cancelable biometrics-based key generation using electroencephalograms," *Computers & Security*, vol. 62, pp. 95–113, 2016.

[27] S. Yang and F. Deravi, "On the usability of electroencephalographic signals for biometric recognition: A survey," *IEEE Transactions on Human-Machine Systems*, vol. 47, no. 6, pp. 958–969, 2017.

[28] N. Kumar *et al.*, "Cancelable biometrics: a comprehensive survey," *Artificial Intelligence Review*, vol. 53, no. 5, pp. 3403–3446, 2020.

[29] C. Li and J. Hu, "Attacks via record multiplicity on cancelable biometrics templates," *Concurrency and Computation: Practice and Experience*, vol. 26, no. 8, pp. 1593–1605, 2014.

[30] Q. N. Tran and J. Hu, "A multi-filter fingerprint matching framework for cancelable template design," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 2926–2940, 2021.

[31] N. Courtois, A. Klimov, J. Patarin, and A. Shamir, "Efficient algorithms for solving overdefined systems of multivariate polynomial equations," in *International Conference on the Theory and Applications of Cryptographic Techniques*. Springer, 2000, pp. 392–407.

[32] A. Wilmer, M. De Lussanet, and M. Lappe, "A method for the estimation of functional brain connectivity from time-series data," *Cognitive neurodynamics*, vol. 4, no. 2, pp. 133–149, 2010.

[33] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "Physiobank, physiotoolkit, and physionet," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[34] W.-L. Zheng, W. Liu, Y. Lu, B.-L. Lu, and A. Cichocki, "Emotionmeter: A multimodal framework for recognizing human emotions," *IEEE transactions on cybernetics*, vol. 49, no. 3, pp. 1110–1122, 2018.

[35] (2009) EEG motor movement/imagery dataset. [Online]. Available: https://physionet.org/content/eegmmidb/1.0.0/

[36] G. Schalk, D. J. McFarland, T. Hinterberger, N. Birbaumer, and J. R. Wolpaw, "BCI2000: a general-purpose brain-computer interface (BCI) system," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 6, pp. 1034–1043, 2004.

[37] I. Winkler, S. Brandl, F. Horn, E. Waldburger, C. Allefeld, and M. Tangermann, "Robust artifactual independent component classification for BCI practitioners," *Journal of Neural Engineering*, vol. 11, no. 3, p. 035013, 2014.

[38] A. B. Teoh, Y. W. Kuan, and S. Lee, "Cancellable biometrics and annotations on biohash," *Pattern recognition*, vol. 41, no. 6, pp. 2034–2044, 2008.

[39] M. Wang, X. Yin, Y. Zhu, and J. Hu, "Representation learning and pattern recognition in cognitive biometrics: A survey," *Sensors*, vol. 22, no. 14, p. 5111, 2022.

[40] J. F. Cavanagh and J. J. Allen, "Multiple aspects of the stress response under social evaluative threat: An electrophysiological investigation," *Psychoneuroendocrinology*, vol. 33, no. 1, pp. 41–53, 2008.

[41] K. Yan and D. Zhang, "Feature selection and analysis on correlated gas sensor data with recursive feature elimination," *Sensors and Actuators B: Chemical*, vol. 212, pp. 353–363, 2015.

[42] W. Klonowski, "Everything you wanted to ask about EEG but were afraid to get the right answer," *Nonlinear biomedical physics*, vol. 3, no. 1, pp. 1–5, 2009.

[43] M. Wang, K. Kasmarik, A. Bezerianos, K. C. Tan, and H. Abbass, "On the channel density of EEG signals for reliable biometric recognition," *Pattern Recognition Letters*, vol. 147, pp. 134–141, 2021.

[44] H. El-Fiqi, M. Wang, K. Kasmarik, A. Bezerianos, K. C. Tan, and H. A. Abbass, "Weighted gate layer autoencoders," *IEEE Transactions on Cybernetics*, 2021.

[45] G. O. Williams, "The use of d' as a "decidability" index," in *1996 30th Annual International Carnahan Conference on Security Technology*. IEEE, 1996, pp. 65–71.

[46] "Information technology-security techniques-biometric information protection," *ISO/IEC 24745:2011*, 2011.

[47] M. Gomez-Barrero, J. Galbally, C. Rathgeb, and C. Busch, "General framework to evaluate unlinkability in biometric template protection systems," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 6, pp. 1406–1420, 2017.

[48] M. Wang, S. Wang, and J. Hu, "Cancellable template design for privacy-preserving EEG biometric authentication systems," *arXiv preprint arXiv:2203.16730*, 2022.

Min Wang (Member, IEEE) is a postdoctoral research fellow with the School of Engineering and Information Technology, University of New South Wales, Canberra, Australia. She received her Ph.D. degree in computer science from the University of New South Wales in 2020. Her research interests include biometrics, privacy and security, pattern recognition, machine learning, and bio-cryptography.

Song Wang received the B.Eng. (Electrical Engineering) from Xi'an Jiaotong University, China in 1991 and PhD (Control Theory) from the University of Melbourne, Australia in 2001. She worked as a design engineer at NEC Australia in 2000-2004. Since January 2005 she has been with the Department of Engineering, La Trobe University, Australia. Her main research interest is biometric security. She has published many high-quality papers in highly ranked journals, such as PATTERN RECOGNITION, IEEE COMMUNICATIONS MAGAZINE, IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS and IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY.

Jiankun Hu (Senior Member, IEEE) is a full professor of cyber security, School of Engineering and Information Technology, University of New South Wales, Canberra, Australia. His main research interest is in the field of cyber security, including biometrics security, where he has publications at top venues including the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE. He has served on the editorial boards of up to seven international journals including the IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY.